JGR Atmospheres

RESEARCH ARTICLE
10.1029/2018JD029637

Effect of Meteorological Variability on Fine Particulate Matter Simulations Over the Contiguous United States

Rajesh Kumar[1], Jared A. Lee[1], Luca Delle Monache[1,2], and Stefano Alessandrini[1]

[1]National Center for Atmospheric Research, Boulder, CO, USA, [2]Now at Scripps Institution of Oceanography, University of California, San Diego, CA, USA

Key Points:
- A hierarchical cluster analysis (HCA) technique is applied to down-select meteorological drivers of CMAQ
- Meteorology-induced changes in CMAQ-simulated $PM_{2.5}$ are dominated by emissions, vertical diffusion, and aerosol processes
- Meteorology-induced variability in $PM_{2.5}$ is estimated to be 0.08–24 $\mu g/m^3$ over the CONUS with larger variability over the eastern United States

Supporting Information:
- Supporting Information S1

Correspondence to:
R. Kumar,
rkumar@ucar.edu

Citation:
Kumar, R., Lee, J. A., Delle Monache, L., & Alessandrini, S. (2019). Effect of meteorological variability on fine particulate matter simulations over the contiguous United States. *Journal of Geophysical Research: Atmospheres*, *124*, 5669–5694. https://doi.org/10.1029/2018JD029637

Received 13 SEP 2018
Accepted 26 APR 2019
Accepted article online 3 MAY 2019
Published online 31 MAY 2019

Author Contributions:
Conceptualization: Rajesh Kumar, Luca Delle Monache, Stefano Alessandrini
Data curation: Rajesh Kumar
Formal analysis: Rajesh Kumar
Funding acquisition: Rajesh Kumar, Luca Delle Monache, Stefano Alessandrini
Investigation: Rajesh Kumar
Methodology: Rajesh Kumar, Jared A. Lee, Luca Delle Monache, Stefano Alessandrini
Project administration: Rajesh Kumar, Luca Delle Monache
Resources: Rajesh Kumar
Software: Rajesh Kumar, Jared A. Lee
Supervision: Luca Delle Monache, Stefano Alessandrini
(continued)

Abstract This study quantifies the impact of meteorological variability on the Community Multiscale Air Quality (CMAQ) model-simulated particulate matter of aerodynamic diameter 2.5 μm or smaller (particulate matter 2.5 [$PM_{2.5}$]) over the contiguous United States (CONUS). The meteorological variability is represented using the Short-Range Ensemble Forecast (SREF) produced operationally by the National Oceanic and Atmospheric Administration. A hierarchical cluster analysis technique is applied to down-select a subset of the SREF members that objectively accounts for the overall meteorological forecast variability of SREF. Three SREF members are selected to drive off-line CMAQ simulations during January, April, July, and October 2016. Changes in emissions, vertical diffusion, and aerosol processes due to meteorological variability dominate changes in aerosol mass concentrations over 55-73% of the domain except in July when dry deposition dominates emissions and aerosol processes. Weather Research and Forecasting-Advanced Research WRF (WRF-ARW) simulations reproduced the variability of surface temperature very well but overestimated the 10-m wind speed, precipitation, and at some sites the planetary boundary layer height. Averaged over CONUS, CMAQ simulations driven by all three meteorological configurations capture the observed daytime low and nighttime high $PM_{2.5}$ mass concentrations but underestimated the observed concentrations likely due to faster advection and higher wet deposition in the model. $PM_{2.5}$ levels across the three simulations agreed well during daytime but showed larger variability during nighttime due to dominance of aerosol, clouds, and advection processes in nighttime. The meteorology-induced variability in $PM_{2.5}$ is estimated to be 0.08–24 $\mu g/m^3$ over the CONUS with larger variability over the eastern United States.

1. Introduction

Particulate matter of aerodynamic diameter 2.5 μm or smaller (particulate matter 2.5 [$PM_{2.5}$]), also known as fine particulate matter, is a criteria pollutant because it adversely affects human health. Exposure to elevated levels of $PM_{2.5}$ can cause premature deaths through respiratory or cardiovascular diseases (Burnett et al., 2014; Fann et al., 2012) and tremendous social and economic costs. For instance, air pollution is reported to have caused 160,000 premature deaths in 2010 in the United States with a total economic loss of about $175 billion (Im et al., 2018). Another study estimated that from 2002 to 2011 the total cost of air pollution exposure from energy production in the United States is on the order of $130–170 billion annually (Jaramillo & Muller, 2016). The exposure to air pollution levels even below the U.S. Environmental Protection Agency (EPA) defined National Ambient Air Quality Standards (NAAQS) is not safe and reported to adversely affect health of men, underrepresented groups, and people with Medicaid eligibility (Di et al., 2017).

To mitigate these effects of air pollution, air quality managers across the United States analyze air quality and weather observations, numerical weather prediction (NWP) model output, and $PM_{2.5}$ guidance from the National Air Quality Forecasting Capability (NAQFC) to warn the public of anticipated air pollution episodes. This information allows the public to take actions (e.g., reduce outdoor activities) to limit their exposure to harmful pollutants. However, the lack of knowledge of uncertainties associated with forecasted $PM_{2.5}$ values makes it difficult for the air quality managers to assess the value of NAQFC products in their decision-making process.

The NAQFC predictions are based on an off-line chemistry transport model called the Community Multiscale Air Quality (CMAQ) model (Byun & Schere, 2006; Lee et al., 2017), which simulates the spatial and temporal distribution of air pollutants using advanced numerical techniques, sophisticated algorithms to process emission inventories, and by parameterizing a variety of atmospheric physical and chemical

Validation: Rajesh Kumar, Jared A. Lee
Visualization: Rajesh Kumar
Writing - original draft: Rajesh Kumar
Writing – review & editing: Rajesh Kumar, Jared A. Lee, Luca Delle Monache, Stefano Alessandrini

processes. Several factors contribute to uncertainties in air quality simulations, including errors in initial and boundary conditions, inadequate understanding of some of the air quality-related processes, inaccurate emission inventories, meteorological biases, and numerical approximations. Improving initialization of CMAQ via assimilation of satellite aerosol optical depth retrievals is already shown to significantly improve the accuracy of the $PM_{2.5}$ forecasts (Kumar et al., 2019; Tang et al., 2017). Here we focus on understanding the uncertainties in NAQFC $PM_{2.5}$ forecasts due to meteorological errors and biases.

Daily variation in meteorology is reported to explain up to 50% of the $PM_{2.5}$ variability over the contiguous United States (CONUS; Tai et al., 2010). Increase in $PM_{2.5}$ over the United States during 1994–2012 due to changes in weather parameters was shown to cause additional 14,700 premature deaths (Jhun et al., 2015). Dawson et al. (2007) reported that $PM_{2.5}$ mass concentrations are sensitive to perturbations in meteorological parameters (temperature, wind speed, absolute humidity, mixing height, cloud cover, and precipitation) and changed considerably (20–170 ng/m^3) over the eastern United States in response to perturbations in meteorological parameters.

In this study, the meteorological errors and biases are represented using the Short-Range Ensemble Forecast (SREF; Du et al., 2015) produced by the National Oceanic and Atmospheric Administration's (NOAA's) National Centers for Environmental Prediction (NCEP). SREF is based on two widely used NWP models, namely, the Nonhydrostatic Multiscale Model on the B grid (NMMB) and the Weather Research and Forecasting (WRF) model (Skamarock et al., 2008). SREF is used to assure that meteorological configurations employed here can be easily replicated in the NAQFC operations. More importantly, unlike other NCEP ensemble products that either do not extend to the NAQFC forecast lead time of 48 hr or do not have any existing preprocessor for coupling their meteorological fields with CMAQ, SREF is the only NCEP meteorological ensemble product at the time this study was conducted that was technically suitable to use.

The paper is organized as follows. Section 2 describes the SREF and a hierarchical cluster analysis (HCA) technique that is employed to down-select the SREF members that capture the overall variability embodied in the SREF. This down-selection is important to reduce the computational cost of our experiments. The CMAQ modeling configuration and the observations used to evaluate the performance of CMAQ are described in section 3. The meteorological variability and its impact on $PM_{2.5}$ mass concentrations as well as different processes controlling $PM_{2.5}$ are assessed in section 4, and the results are summarized in section 5.

2. Methodology

2.1. Description of NCEP SREF

The NCEP SREF system, version 7.0, contains 26 ensemble members, evenly divided between the Advanced Research WRF (WRF-ARW) and the NOAA Environmental Modeling System (NEMS) NMMB models. For each set of 13 NMB and 13 ARW members, a control member is associated with 12 perturbed members (Tables S1 and S2 in the supporting information). The SREF is a multianalysis, multiphysics ensemble that obtains the initial conditions (ICs) from one of three sources: the North American Model (NAM) Data Assimilation System (NDAS; *Rogers* et al., 2009; Kleist et al., 2009), the NCEP Global Forecast System (GFS; Environmental Modeling Center, 2003), and the Rapid Refresh model (RAP; Benjamin et al., 2016). IC perturbations are generated via blending of the SREF and Global Ensemble Forecast System (GEFS; Wei et al., 2006), with perturbations of opposite sign for the “positive” (“p”) and “negative” (“n”) members.

The lateral boundary conditions (LBCs) are provided by the GFS for the control members and by a GEFS member for all other members. While there are some random perturbations that affect each SREF member owing to the GEFS LBCs and IC perturbations, all SREF ensemble members are statistically distinguishable from one another because each member has a unique suite of physics parameterization schemes (Fraley et al., 2010). This distinguishability allows us to meaningfully apply statistical postprocessing and verification techniques that assess the performance of specific ensemble members over a long period of time.

Unfortunately, we could not use all the SREF members to drive CMAQ in this study for two reasons. First, the Meteorology-Chemistry Interface Processor (MCIP; Otte & Pleim, 2010), which is used to prepare CMAQ meteorological inputs, can process only ARW output, and thus, we had to exclude NMMB-based SREF members from our analysis. Second, the Ferrier (Rogers et al., 2001) and Ferrier-Aligo (Aligo et al., 2014)

Table 1
Clustering Results on the Nine ARW Members of SREF With Physics Schemes Compatible With MCIP/CMAQ

Time period	Cluster members (by proximity to centroid)
January 2016	1: **arw_p5**, arw_n3, arw_n6
	2: **arw_ctl**, arw_n4, arw_p2
	3: **arw_p3**, arw_p1, arw_p6
April 2016	1: **arw_p2**, arw_n4, arw_n6
	2: **arw_ctl**, arw_p5, arw_p3
	3: **arw_p1**, arw_n3, arw_p6
July 2016	1: **arw_p2**, arw_n6, arw_n4
	2: **arw_ctl**, arw_p5, arw_p3
	3: **arw_p1**, arw_n3, arw_p6
October 2016	1: **arw_n4**, arw_p2, arw_n6
	2: **arw_ctl**, arw_p5, arw_p3
	3: **arw_n3**, arw_p6, arw_p1
All months	1: **arw_p2**, arw_n4, arw_n6
	2: **arw_ctl**, arw_p5, arw_p3
	3: **arw_p1**, arw_n3, arw_p6

Note. ARW, Advanced Research WRF; CMAQ, Community Multiscale Air Quality; MCIP, Meteorology-Chemistry Interface Processor; SREF, Short-Range Ensemble Forecast; WRF, Weather Research and Forecasting.

microphysics schemes do not output enough microphysical species to be compatible with CMAQ, so we exclude arw_n1, arw_n2, arw_p4, and arw_n5 from our analysis as well (Table S2). Therefore, we perform our analysis on the remaining nine ARW members of SREF (arw_ctl, arw_p1, arw_p2, arw_n3, arw_p3, arw_n4, arw_p5, arw_n6, and arw_p6).

2.2. Hierarchical Cluster Analysis (HCA)

We apply an inexpensive statistical postprocessing technique called HCA to the nine SREF ARW members mentioned above to reduce the computational burden and to objectively determine a subset of the SREF ARW members, which objectively account for the meteorological forecast variability. HCA has been used previously to group similar NWP (Alhamed et al., 2002; Johnson et al., 2011; Lee, 2012; Lee et al., 2016; Yussouf et al., 2004) and air quality ensemble (Solazzo et al., 2013) members together. We follow Lee et al. (2016) for ensemble down-selection.

HCA is an iterative algorithm where each member starts as a singleton cluster. For each iteration, two closest clusters are determined using a distance metric and are then merged. Ward's minimum variance method is used as distance metric (Wilks, 2006) here, which combines the two clusters (say r and s) with the smallest sum of squares of the distances between each point in the cluster and the cluster centroid. This distance metric $d(r, s)$ is defined as

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r + n_s)}} \|\overline{x_r} - \overline{x_s}\|_2 \quad (1)$$

where $\|\|_2$ is the Euclidean distance, $\overline{x_r}$ and $\overline{x_s}$ are the respective cluster centroids, and n_r and n_s are the numbers of elements in the respective clusters. The algorithm iterates until all the data vectors have been merged to form one overarching cluster.

We define each data vector as the normalized forecast errors (forecast minus observation) of a single ensemble member. Normalizing the forecast errors enables us to cluster on multiple variables together. The forecast variables we considered are 2-m temperature, 2-m dewpoint, 10-m wind speed, and 10-m wind direction, at forecast lead times of 9, 21, 33, and 45 hr (corresponding to valid times of 0000 and 1200 UTC). The forecast errors for each SREF member are calculated using the *ensemble-stat* tool of the Model Evaluation Tools (MET) (*Brown* et al., 2016; www.dtcenter.org/met/users) toolkit. METAR observations from the Meteorological Assimilation Data Ingest System (MADIS; https://madis.ncep.noaa.gov) are used as ground truth.

We perform the HCA both on each month of SREF data individually (January, April, July, and October 2016) and on all 4 months combined. After running the HCA algorithm, we compute the centroid of each cluster and order all the members of each cluster according to their proximity to the cluster centroid. We identify the member closest to its cluster centroid as the most representative member of that cluster.

For each time period, Table 1 lists this ordered membership of each cluster, and the dendrogram for the all-months' time period is seen in Figure 1 (dendrograms for the single-month periods are not shown for brevity). Dendrograms provide visual presentation of clustering of data vectors. The lower the height on the dendrogram at which two subclusters merge, the more similar they are and vice versa. There are consistently three distinct clusters in this nine-member subset of SREF, and from Table 1, the membership of each cluster is the same for every time period except January only. Based on this, we conclude that seasonal variability in the clustering of these members is not of primary importance (at least for 2016), and thus, we focus on the HCA results from the all-month time period. Therefore, the SREF members that we select to drive our own ARW/CMAQ miniensemble are arw_ctl, arw_p1, and arw_p2.

Interestingly, the control member (arw_ctl) is closest to its cluster centroid in every time period. This makes sense statistically, as one would expect that a control member would be closer to the center of a forecast

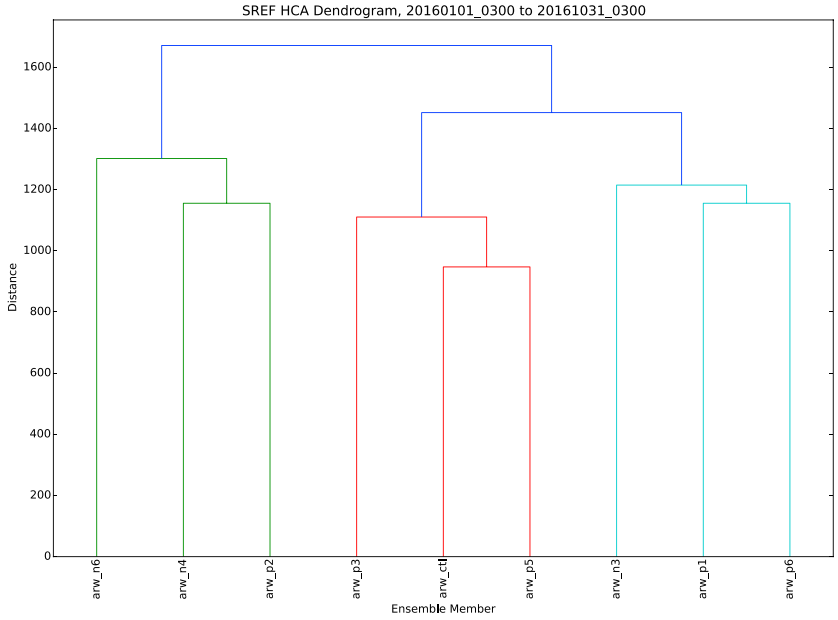

Figure 1. Dendrogram for all months when clustering on the 9 Advanced Research Weather Research and Forecasting (ARW) members of Short-Range Ensemble Forecast (SREF) with physics compatible with Meteorology-Chemistry Interface Processor/Community Multiscale Air Quality (MCIP/CMAQ).

distribution on average, when the other members have perturbations added to and subtracted from the control member.

To examine how much of the 26-member SREF variability was captured, we use the standard method of binned spread-skill plots to examine the statistical consistency of our miniensemble. Ideally, the binned spread-skill should fall along the 1:1 line, indicating that the ensemble spread can predict the ensemble error (or skill). To evaluate the statistical consistency of our miniensemble, 48-hr forecasts for the 15-day periods in January and July 2016 were used. Binned spread-skill plots are generated for 2-m temperature (Figure S1 in the supporting information), 2-m dew point temperature (Figure S2), 10-m wind speed (Figure S3), and 10-m wind direction (Figure S4), for this "3mem" ensemble, the "all_sref (26 members)," and "arw_only (13 members)" direct SREF output. These results show that the statistical consistency of the 3mem ensemble is comparable to both the "all_sref" and the "arw_only" members for all these meteorological variables. For 10-m wind speed, however (Figure S3), the 3mem ensemble is even more reliable (i.e., closer to the 1:1 line) than the other ensembles for the January test days. Similar results are obtained for this analysis with full month output for January, April, July, and October 2016. These results demonstrate that the 3mem ensemble is an adequate subset of the full SREF ensemble in terms of statistical consistency and is generally representative of the full SREF ensemble reliability.

3. Model Configuration Details

3.1. WRF-ARW

The native SREF output stored on the NCEP servers does not include a large number of variables required by MCIP/CMAQ, so we need to run our own "SREF-like" WRF-ARW simulations to generate the necessary meteorological parameters. Thus, we perform three WRF-ARW simulations with version 3.8.1 corresponding to our miniensemble. Our 481 × 369 domain covers the CONUS at a grid spacing of 12 km × 12 km with 43 vertical levels, a model top of 50 hPa, and a time step of 20 s.

The physics configuration for all three ARW members correspond to the physics schemes from the respective SREF members as summarized in Table 2. The ICs and LBCs come from SREF members arw_ctl (RAP/GFS), arw_p1 (RAP/GEFS 13), and arw_p2 (RAP/GEFS 15). We slightly modified some other physics options from SREF, including reducing the radiation scheme call frequency from 60 to 10 min, turning on

Table 2
Proposed Miniensemble If We Cluster on Only the Nine ARW Members of SREF With Physics Schemes Compatible With MCIP/CMAQ

Mem.	IC/LBC	Conv.	PBL	Sfc. Lay.	MP	LW Rad.	SW Rad.	LSM
1	arw_ctl	KF	YSU	MM5	WSM6	RRTMG	RRTMG	Noah
2	arw_p1	Grell	MYNN	MYNN	Thompson	RRTM	Goddard	Noah
3	arw_p2	BMJ	MYJ	MYJ	Thompson	RRTMG	RRTMG	Noah

Note. ARW, Advanced Research WRF; CMAQ, Community Multiscale Air Quality; IC, initial condition; LBC, lateral boundary condition; LSM, land surface model; LW, longwave; MCIP, Meteorology-Chemistry Interface Processor; MYJ, Mellor-Yamada-Janjic; MYNN, Moller-Yamada-Nakanishi-Niino; PBL, planetary boundary layer; RRTMG, Rapid Radiative Transfer Model for General Circulation Models; SREF, Short-Range Ensemble Forecast; SW, shortwave; WRF, Weather Research and Forecasting; WSM6, WRF Single-Moment 6-Class; YSU, Yonsei University.

the interpolation of shortwave radiation between radiation calls based on solar zenith angle (*swint_opt=1*), turning on a topographic correction for surface winds (*topo_wind=1*; Jiménez & Dudhia, 2012] for the Yonsei University (YSU) planetary boundary layer scheme (Hong et al., 2006), and turning on sixth-order horizontal diffusion (HDIF). For this study, we analyze and produce 48 hr of model forecasts initialized once daily at 0300 UTC in January, April, July, and October 2016, in order to get a fair representation of performance through all four seasons.

3.2. CMAQ

The CMAQ domain replicates the NAQFC domain settings and is defined on a Lambert conformal projection centered at (40°N, 97°W) with 442 × 265 grid points, 42 vertical levels from the surface to ~20 km, and a horizontal grid spacing of 12 km × 12 km. The meteorological fields from the WRF-ARW simulation are mapped to the CMAQ domain using the MCIP. The Carbon Bond mechanism-2005 (CB-05) with an updated toluene chemistry (Whitten et al., 2010) is selected to represent the gas-phase chemistry. CMAQ uses three lognormal modes, namely, Aitken, accumulation, and coarse modes to represent aerosols (Binkowski & Roselle, 2003). The AERO6 aerosol module of CMAQ that includes specification of trace metals (Appel et al., 2013; Reff et al., 2009) and source-specific ratios of organic mass to organic carbon (Simon & Bhave, 2012) is employed here. The ISORROPIA thermodynamic equilibrium module (version II) is used to calculate thermodynamic equilibrium of inorganic aerosols in the Aitken and accumulation modes (Fountoukis & Nenes, 2007). The gas-particle partitioning between the gas phase and coarse mode particles is treated dynamically (Kelly et al., 2010), and the secondary organic aerosol (SOA) formation is estimated using Carlton et al. (2010). Lateral chemical boundary conditions for CMAQ are obtained from monthly median concentrations simulated by GEOS-Chem (Tang et al., 2009). The vertical diffusion (VDIF) in CMAQ follows the Asymmetric Convective Model version 2 (ACM2; Pleim, 2007), while the advection and diffusion scheme follows Byun (1999).

The anthropogenic emissions of trace gases and aerosols from the EPA National Emission Inventory representative of the year 2011 are mapped to the CMAQ domain using the Sparse Matrix Operator Kernel Emissions (SMOKE). SMOKE is run with the three meteorological simulations to generate meteorology-dependent emissions for each configuration. The Biogenic Emissions Inventory System (BEIS) version 3.13 is used to estimate the emissions of volatile organic compounds from vegetation and nitric oxide (NO) emissions from the soil. The U.S. Bluesky Modeling Framework (https://www.airfire.org/bluesky/), which uses the NOAA Hazard Mapping System to geographically locate and estimate the strength of wildfires, is used to represent biomass burning emissions of aerosols and trace gases within the model domain. Anthropogenic and biomass burning emissions are processed via SMOKE, while biogenic emissions are calculated online within CMAQ.

In addition to the standard model output, we also track the contribution of direct emissions (EMIS), VDIF, HDIF, aerosol processes (AERO), cloud processes (CLDS), horizontal advection (HADV), vertical advection (ZADV), and dry deposition (DDEP) to hourly Aitken (AMASSI), accumulation (AMASSJ), and coarse (AMASSK) mode aerosol mass concentrations in each of the simulations using the Integrate Process Rate (IPR) analysis capability of CMAQ. AMASSI, AMASSJ, and AMASSK represent the family of aerosols belonging to the three aerosol modes used in CMAQ and are defined to reduce the computational storage requirements as follows:

$$\mathrm{AMASSI} = \mathrm{ASO4I} + \mathrm{ANO3I} + \mathrm{ANH4I} + \mathrm{ANAI} + \mathrm{ACLI} + \mathrm{AOTHRI} + \mathrm{APOCI} + \mathrm{AECI} + \mathrm{APNCOMI} \quad (2)$$

$$\begin{aligned}\mathrm{AMASSJ} = {} & \mathrm{AALK1J} + \mathrm{AALK2J} + \mathrm{AXYL1J} + \mathrm{AXYL2J} + \mathrm{AXYL3J} + \mathrm{ATOL1J} + \mathrm{ATOL2J} + \mathrm{ATOL3J} \\ & + \mathrm{ABNZ1J} + \mathrm{ABNZ2J} + \mathrm{ABNZ3J} + \mathrm{ATRP1J} + \mathrm{ATRP2J} + \mathrm{AISO1J} + \mathrm{AISO2J} + \mathrm{AISO3J} \\ & + \mathrm{ASQTJ} + \mathrm{APAH1J} + \mathrm{APAH2J} + \mathrm{APAH3J} + \mathrm{AOLGAJ} + \mathrm{AOLGBJ} + \mathrm{AORGCJ} + \mathrm{ASO4J} \\ & + \mathrm{ANO3J} + \mathrm{ANH4J} + \mathrm{ANAJ} + \mathrm{ACLJ} + \mathrm{AMGJ} + \mathrm{AKJ} + \mathrm{ACAJ} + \mathrm{AFEJ} + \mathrm{AALJ} + \mathrm{ASIJ} \\ & + \mathrm{ATIJ} + \mathrm{AMNJ} + \mathrm{AOTHRJ} + \mathrm{APOCJ} + \mathrm{AECJ} + \mathrm{APNCOMJ} \end{aligned} \quad (3)$$

$$\mathrm{AMASSK} = \mathrm{ASO4K} + \mathrm{ANO3K} + \mathrm{ANH4K} + \mathrm{ACLK} + \mathrm{ASEACAT} + \mathrm{ASOIL} + \mathrm{ACORS} \quad (4)$$

The geometric mean diameter for mass is 0.03, 0.3, and 6 μm for AMASSI, AMASSJ, and AMASSK, respectively. The CMAQ forecast on the first day of each month used idealized ICs for all the chemical species. The maximum and minimum synchronization time steps were set to 720 and 60 s, respectively. The ICs for all other CMAQ forecasts were based on the previous day's CMAQ run. The first 5 days of both WRF and CMAQ output are discarded in each month as spin-up. Hourly CMAQ output was saved for analysis.

3.3. Evaluation Data Sets

Surface $PM_{2.5}$ mass concentrations are obtained from the EPA's Air Quality System (AQS) Data Mart (https://www3.epa.gov/ttn/amtic/quality.html) for evaluation of CMAQ-simulated $PM_{2.5}$. All the $PM_{2.5}$ measurements collected under the national ambient air monitoring program are available for download at the AQS Data Mart. These measurements are performed by different tribal, state, and local agencies and undergo several quality control tests before their archival on the AQS Data Mart. For direct comparison of model output with the observations, the CMAQ utility program *combine* is used to convert CMAQ aerosol chemical composition into $PM_{2.5}$ concentrations. The resulting model output is then spatially and temporally collocated with the observed values using another CMAQ utility program called *sitecmp*. The sharp-cut $PM_{2.5}$ inlet method (Jiang et al., 2006) is used to estimate the contribution of each mode (Aitken, accumulation, and coarse) to $PM_{2.5}$ concentrations in CMAQ. While the EPA measured $PM_{2.5}$ at more than 1,600 sites across the United States during the 4 months of 2016 studied here, we include only those measurements in our analysis that had at least 50% data availability, that is, 360 hourly measurements during each month. The number of sites used in the analysis are 660, 669, 657, and 681 in January, April, July, and October 2016, respectively.

In addition, we also used in situ observations of 2-m temperature, 10-m wind speed, and planetary boundary layer height (PBLH), along with satellite retrievals of precipitation to evaluate the WRF meteorological simulations. The temperature and wind speed data are obtained from the EPA (https://aqs.epa.gov/aqsweb/airdata/download_files.html#Raw), which compiles the meteorological observations reported by the National Weather Service and uses those to account for interannual meteorological variability in air quality trends (https://www3.epa.gov/scram001/metobsdata.htm). The Tropical Rainfall Measuring Mission (TRMM) precipitation retrievals at the spatial resolution of 0.25° × 0.25° and temporal resolution of 3 hr corresponding to the algorithm 3B42 version 7 are used (https://trmm.jpl.nasa.gov/). For the PBLH, we have used the Integrated Global Radiosonde Archive (IGRA) derived estimates (https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive).

4. Results

4.1. Meteorological Evaluation and Variability

The model-simulated values are calculated at the observation sites via bilinear interpolation from the four grid points surrounding the site and paired data are used for the verification. For evaluation against the TRMM data, we mapped the WRF-simulated precipitation fields at the TRMM resolution (0.25° × 0.25°). Figure 2 compares 48-hr WRF forecasts of 2-m temperature for all the three members, that is, arw_ctl, arw_p1, and arw_p2 against the observations averaged over 837 sites in the model domain (see Figure 3 for site locations). The diurnal cycle of the 2-m temperature is reproduced very well by WRF in all the three configurations. However, we notice a cold bias in all the members during January except for a few hours in the second day of the arw_p1 forecast. Arw_p1 shows the best agreement with the observations in January, while arw_ctl and arw_p2 show similar performance. During April, July, and October, all the members show

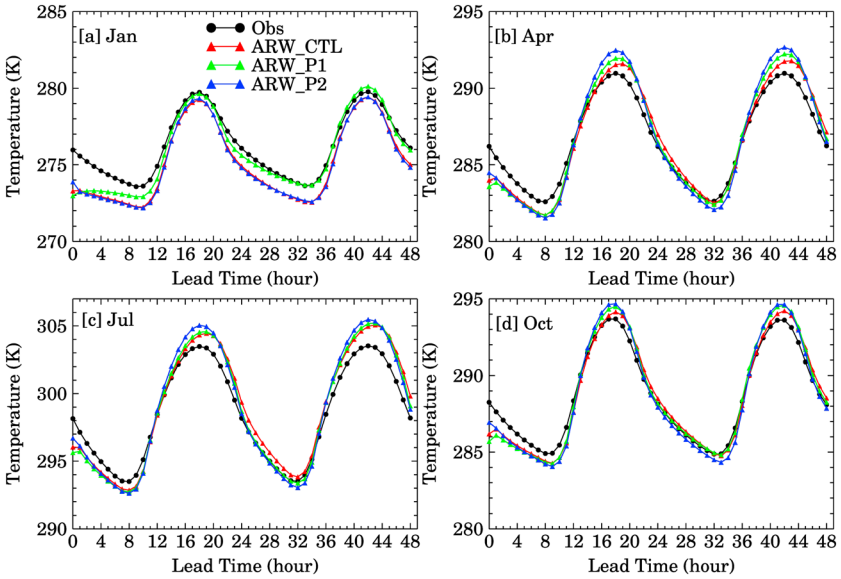

Figure 2. Evaluation of 48-hr Weather Research and Forecasting (WRF) forecasts of 2-m temperature against observations averaged over all the sites in our model domain during (a) January, (b) April, (c) July, and (d) October 2016 as a function of the lead time. All lead times are in UTC.

a warm bias during the afternoon hours and the magnitude of the nighttime cold bias is smaller than that in January. Arw_ctl even shows a slight warm bias on the second night (lead times 24-32) in July.

Figures 3 shows the maps of correlation coefficient (top panels), mean bias (middle panels), and the root-mean-square error (RMSE; bottom panel) in 2-m temperature for the three simulations in January 2016. Similar maps for April, July, and October are shown in Figures S5-S7. The correlation coefficient is much higher in the eastern United States than the western United States for all three configurations in January and October, similar between the eastern and western United States in April, and higher in the western United States in July. The correlation coefficient is greater than 0.75 for all the members at most of the sites, indicating that all WRF configurations capture the variability in 2-m temperature very well. The mean bias is within ±2K at most of the sites for all the simulations. The mean bias is smaller in arw_p1 compared to

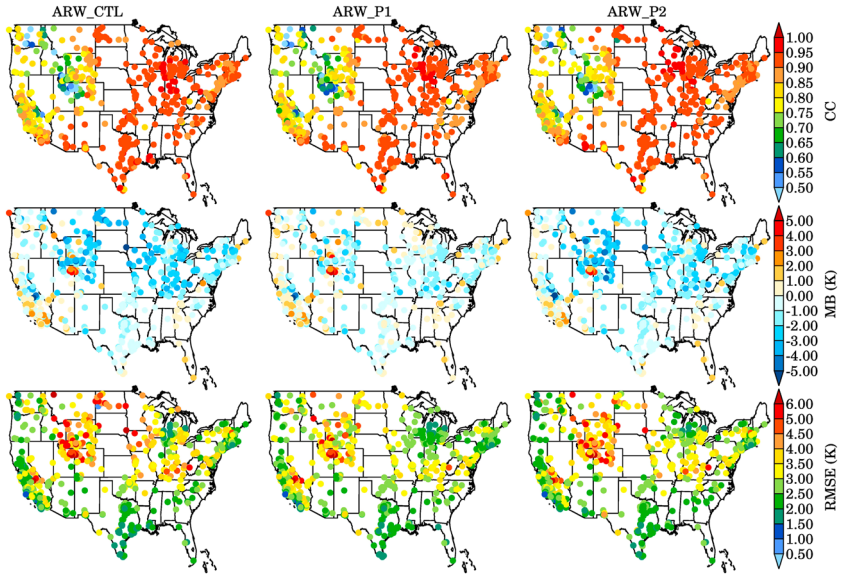

Figure 3. Spatial distributions of correlation coefficient (CC; top panel), mean bias (MB; middle panel), and root-mean-square error (RMSE; bottom panel) in 2-m temperature simulated by arw_ctl, arw_p1, and arw_p2 in January 2016.

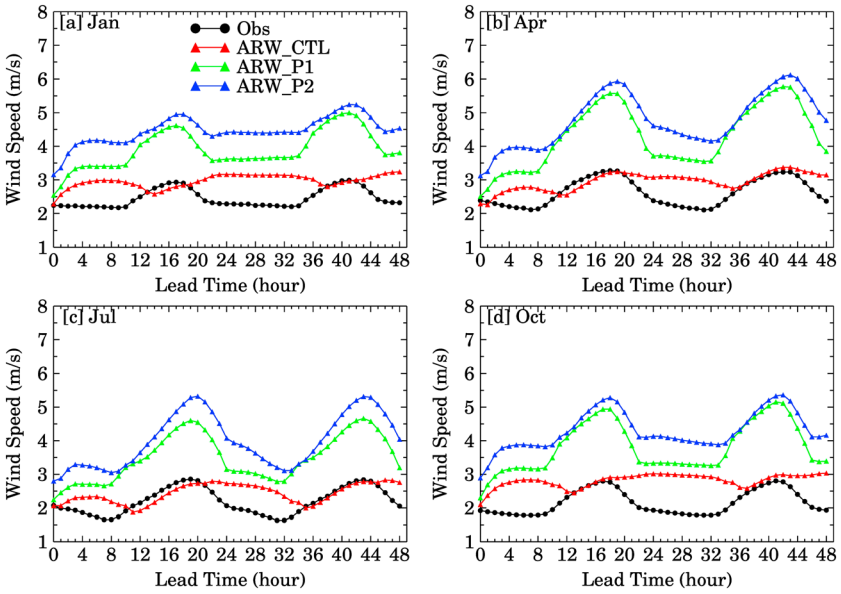

Figure 4. Evaluation of 48-hr Weather Research and Forecasting (WRF) forecasts of 10-m wind speed against observations averaged over all the sites in our model domain during (a) January, (b) April, (c) July, and (d) October 2016 as a function of the lead time. All lead times are in UTC.

arw_ctl and arw_p2 in January. All the members show similar mean bias values in other months. The RMSE values are less than 4 K in all the members during all the months except at some sites in Colorado, Wyoming, Utah, North Dakota, and parts of north-eastern United States, particularly in January and October. Similar to the mean bias, RMSE is also lower in arw_p1 in January.

Figure 4 compares 48-hr WRF forecasts of 10-m wind speed for the three WRF simulations against the observations averaged over 760 sites in the model domain (see Figure 5 for site locations). In contrast to 2-m temperature, we find large differences among the three members for 10-m wind speed. During all the months, all the members capture the diurnal patterns but overestimate the 10-m wind speed, except for the arw_ctl simulation during the afternoon hours. While arw_p1 and arw_p2 overestimate the wind speed much more than arw_ctl, they better capture the diurnal cycle compared to arw_ctl. The highest overestimation of 10-m

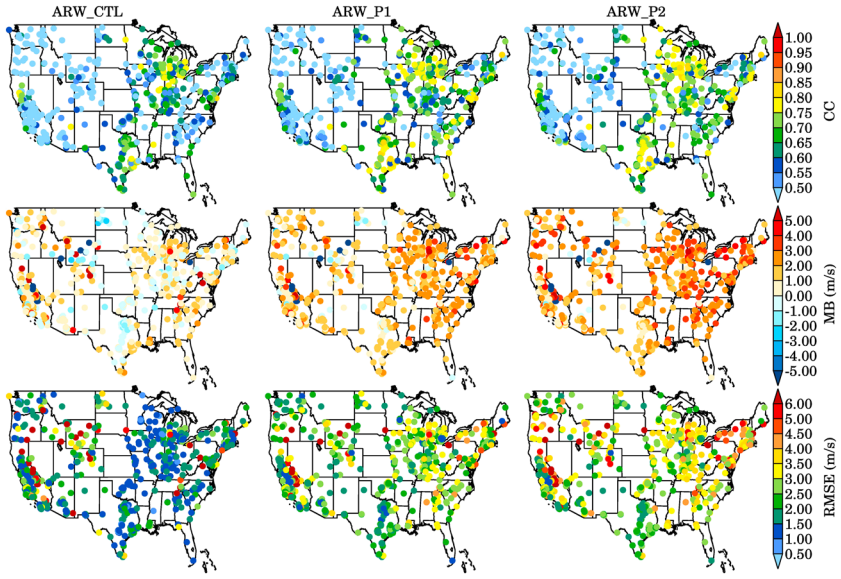

Figure 5. Spatial distribution of correlation coefficient (CC; top panel), mean bias (MB; middle panel), and root-mean-square error (RMSE; bottom panel) in 10-m wind speed simulated by arw_ctl, arw_p1, and arw_p2 in January 2016.

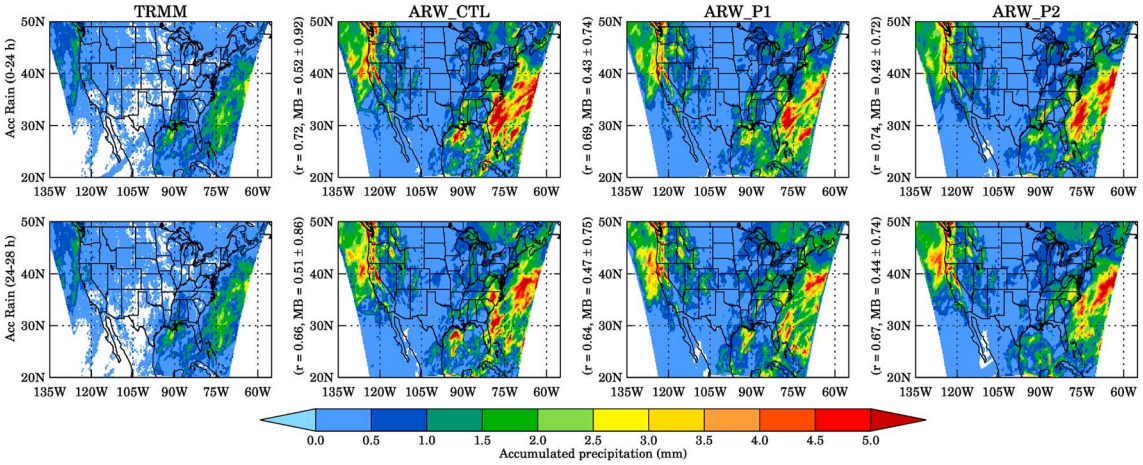

Figure 6. Spatial distribution of Tropical Rainfall Measuring Mission (TRMM)-retrieved and Weather Research and Forecasting (WRF)-simulated accumulation precipitation for the first (top panel) and second (bottom panel) days of the forecasts in January 2016. The spatial correlation coefficient and mean bias in WRF simulated precipitation for each member are also shown along the abscissa of arw_ctl, arw_p1, and arw_p2 panels.

wind speed is seen in arw_p2 simulations. Figures 5 show the maps of correlation coefficient (top panels), mean bias (middle panels), and the RMSE (bottom panel) in 10-m wind speed simulated by arw_ctl, arw_p1, and arw_p2 in January 2016. Similar maps for April, July, and October 2016 are shown in Figures S8-S10. The correlation coefficients for 10-m wind speed are much lower compared to 2-m temperature in all the months and WRF simulations, but the correlation coefficient is higher over the eastern United States, especially in January, April, and October. The WRF simulations overestimate the wind speeds at most of the sites throughout the United States in the arw_p1 and arw_p2 simulations, but there are many sites in the western United States where arw_ctl underestimates the wind speed. The spatial distribution of RMSE is similar to the MB with the arw_ctl simulations showing the lowest RMSE values and arw_p2 showing the highest RMSE values.

The spatial distributions of the TRMM retrieved 24-hr accumulated precipitation are compared against WRF-simulated accumulated precipitation for the first and second days of forecasts for all three configurations in January (Figure 6), April (Figure S11), July (Figure 7), and October (Figure S12) 2016, respectively. First- and second- day forecasts are evaluated separately to understand changes in model performance with time. During all the months, all the WRF simulations show moderate skill (0.54 < correlation coefficient < 0.74) in reproducing the spatial structure and the seasonal cycle of precipitation. However, all the members significantly overestimate the amount of precipitation. The model skill deteriorates slightly from

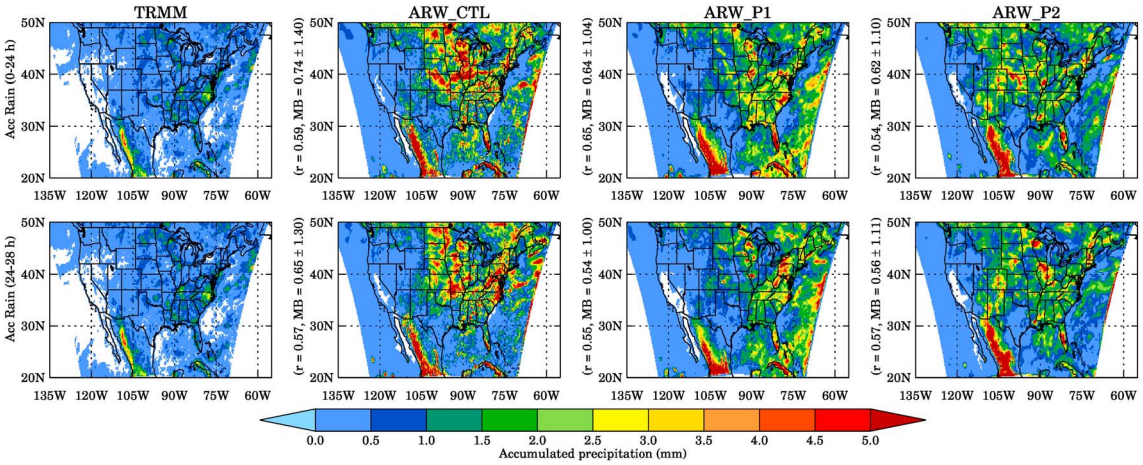

Figure 7. Same as Figure 6 but for July 2016.

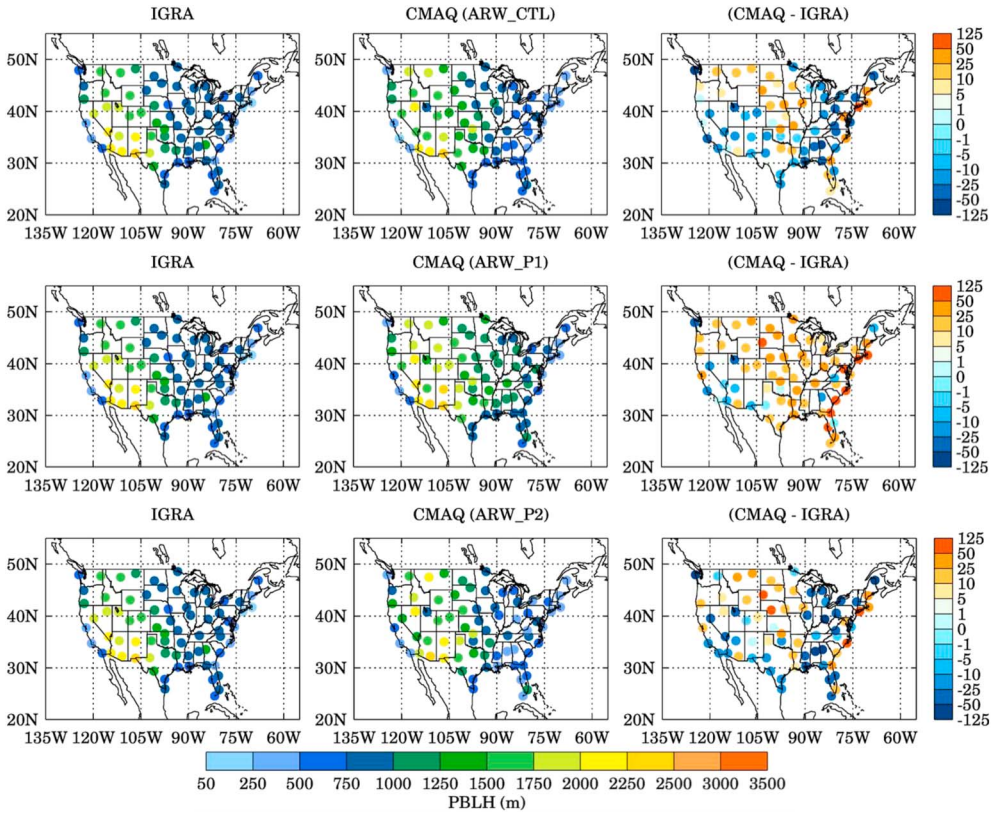

Figure 8. Comparison of Integrated Global Radiosonde Archive (IGRA)-derived planetary boundary layer height (PBLH) with arw_ctl, arw_p1, and arw_p2 simulated PBLH values. Percentage difference between Weather Research and Forecasting (WRF) and IGRA PBLH is shown in the rightmost panel.

the first to second day as indicated by a small decrease in the correlation coefficient. Arw_ctl shows the highest mean bias and variability in both the first- and second-day precipitation forecasts.

Figure 8 compares the PBLH simulated by arw_ctl, arw_p1, and arw_p2 with the IGRA-derived estimates. Most of the IGRA-derived PBLH estimates are available for 0000 UTC, and thus, this comparison is just a snapshot of the model performance at 0000 UTC. We have combined the data for all the 4 months because the IGRA estimates are relatively sparse in time and comparison for individual months do not provide for a statistically meaningful comparison. We notice that arw_p1 mostly overestimates the IGRA PBLH, while arw_ctl and arw_p2 show a mixed behavior. Averaged PBL heights in both IGRA and WRF are higher over the western United States and lower over the eastern United States because 0000 UTC corresponds to late afternoon hours over the western United States and evening hours over the eastern United States. The difference between WRF and IGRA PBLH varies between -80% and 107%, but the differences are within ±50% at about 89%, 83%, and 82% of the sites for arw_ctl, arw_p1, and arw_p2, respectively.

To understand the meteorological variability among the three WRF simulations in nonobservation areas, the monthly averaged spatial distributions of arw_ctl-simulated 2-m temperature and water vapor mixing ratios, 10-m wind speed, and PBLH are analyzed (Figures S13-S16). Average differences between arw_ctl and arw_p1 and arw_ctl and arw_p2 are also analyzed (Figures S13-S16). Arw_p1 and arw_p2 show mixed deviation from the arw_ctl as they simulate lower temperatures in some parts of the domain (e.g., western United States in winter for arw_p2) and higher temperatures in other parts (e.g., midwestern and northeastern United States and Canada in arw_p1). The largest deviations in surface temperature are seen during winter with values ranging from -5.8 to 3.9 K. Domain-wide average ± standard deviation values of arw_ctl-simulated 2-m temperature during January, April, July, and October 2016 are estimated to be 277±12 K, 286±8 K, 296±5 K, and 289±7 K, respectively. Domain-wide-averaged values of arw_ctl-arw_p1 are estimated to be -0.3±1.3 K, -0.1±0.4 K, -0.4±0.4 K, and -0.01±0.4 K, respectively, while the corresponding values of arw_ctl - arw_p2 are 0.2±0.8 K, -0.2±0.4 K, -0.5±0.6 K, and -0.1±0.4 K, respectively.

The arw_ctl-simulated 2-m water vapor mixing ratios generally decrease over land in the arw_p1 and arw_p2 simulations, except slight increases in some areas during winter. The largest deviations over land are seen in arw_p1 during summer, with differences ranging between -8.0 and 1.5 g/kg. Domain-wide average ± standard deviation values of arw_ctl-simulated 2-m water vapor mixing ratios during January, April, July, and October 2016 are estimated to be 5.4±3.7 g/kg, 7.0±3.7 g/kg, 12.5±4.5 g/kg, and 9.0±4.3 g/kg, respectively. Domain-wide mean of arw_ctl - arw_p1 are estimated to be -0.2±0.4 g/kg, -0.05±0.4 g/kg, -0.6±0.7 g/kg, and 0.02±0.5 g/kg, respectively, while the corresponding values of arw_ctl - arw_p2 are -0.3±0.6 g/kg, -0.6±0.5 g/kg, -0.8±0.5 g/kg, and -0.7±0.7 g/kg, respectively.

Not surprisingly, the arw_ctl-simulated 10-m wind speed shows higher values over the ocean and lower values over the land. We note faster winds over land in arw_p1 and arw_p2 compared to arw_ctl except over the mountainous regions in the western United States. The deviations in arw_p2 are slightly larger than arw_p1. Domain-wide average values of arw_ctl-simulated 10-m wind speed during January, April, July, and October 2016 are estimated to be 5.0±2.6 m/s, 4.7±2.2 m/s, 3.8±1.8 m/s, and 4.6±2.4 m/s, respectively. Domain-wide mean of arw_ctl-arw_p1 are estimated to be 0.3±1.2 m/s, -0.6±1.2 m/s, -0.4±1.0 m/s, and -0.3±1.2 m/s, respectively, while the corresponding values of arw_ctl-arw_p2 are -0.7±1.3 m/s, -0.9±1.4 m/s, -0.8±1.2 m/s, and -0.7±1.3 m/s, respectively.

The seasonal variations in PBLH are similar to temperature, with the highest values in summer and the lowest values in winter. Both arw_p1 and arw_p2 simulate a shallower PBL in some parts of the domain (e.g., western United States in April for arw_p2) and a deeper PBL in other parts (e.g., midwestern and northeastern United States and Canada in winter under arw_p1). Domain-wide average values of arw_ctl-simulated PBLH during January, April, July, and October of 2016 are estimated to be 454±292 m, 664±263 m, 723±325 m, and 576±269 m, respectively. Domain-wide mean of arw_ctl-arw_p1 are estimated to be -90±106 m, -62±142 m, -35±131 m, and -48±100 m, respectively, while the corresponding values of arw_ctl-arw_p2 are -41±246 m, -45±203 m, 51±198 m, and 42±180 m, respectively.

In summary, none of the members is consistently better than the others in comparison to the observations, but all the simulations show moderate to very good performance for different meteorological variables providing confidence in using these WRF simulations for driving CMAQ simulations. Large bias in the wind speed especially in arw_p1 and arw_p2 can be reduced via assimilation of wind speed observations, but we have not assimilated any observations here to be consistent with the operational SREF output.

4.2. Effect of Meteorological Variability on Processes Controlling $PM_{2.5}$

This section discusses how different processes controlling $PM_{2.5}$ respond to the variability in meteorology and identifies which process contributes the most to changes in $PM_{2.5}$ due to changes in meteorology. The monthly averaged AMASSJ tendency due to EMIS, VDIF, AERO, CLDS, ADV (HADV + ZADV), and DDEP during January, April, July, and October 2016 in arw_ctl are depicted in the left panels of Figures 9–12, respectively. We focus on AMASSJ because accumulation mode aerosols contribute more than 80% to $PM_{2.5}$ mass concentrations. The contribution of HDIF is an order of magnitude smaller than other processes and thus is not shown. Average changes in AMASSJ tendencies in arw_p1 (middle panel, Figures 9–12)- and arw_p2 (right panel, Figures 9–12)-driven simulations compared to the arw_ctl simulation are also analyzed. Positive values in the left panels of Figures 9–12 indicate that the process is a source for AMASSJ and negative values mean that it is a sink.

As expected, emissions represent the most important source of $PM_{2.5}$ everywhere in the domain during all the months. Meteorological variability leads to larger changes in the emissions over the oceanic regions likely because of dependency of sea-salt emissions on the winds. To examine how sea-salt emissions change between different configurations, we performed a 48-hr CMAQ simulation on 6 January 2016. This 48-hr simulation is used to understand the differences in wind speed and sea-salt emissions among the three simulations. We did not save sea-salt emissions for all the simulations due to storage limitations. Table S3 shows the sea-salt emissions for different species in the three meteorological configurations. The correlation coefficient between the 10-m wind speed and sea-salt emissions are in the range of 0.82-0.85 for all the species presented in Table S3. Average 10-m wind speed in arw_ctl, arw_p1, and arw_p2 for this 48-hr simulation period are estimated to be 6.95 ± 3.37 m/s, 6.77 ± 3.10 m/s, and 6.64 ± 3.24 m/s, respectively. The emissions of all the species are the highest in arw_ctl because it has the highest mean wind speed

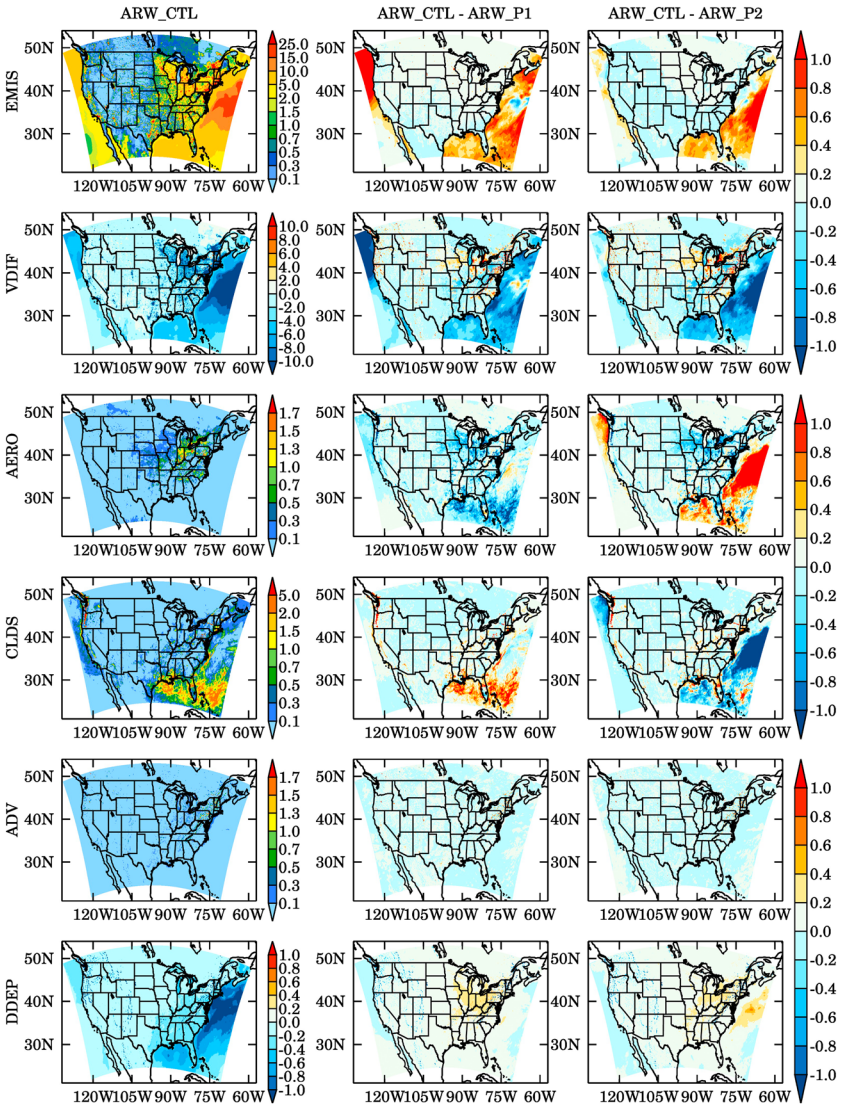

Figure 9. (left panel) Monthly averaged contribution of emissions (EMIS), vertical diffusion (VDIF), aerosol processes (AERO), cloud processes (CLDS), advection (ADV), and dry deposition (DDEP) to hourly AMASSJ for January 2016 in arw_ctl-driven Community Multiscale Air Quality (CMAQ) simulation. The changes in the contribution of these processes in arw_p1 and arw_p2-driven CMAQ simulations relative to arw_ctl-driven CMAQ simulation are shown in the middle and right panels.

and variability (standard deviation). Significant changes in the emissions are also discerned in several inland areas, especially in the urban areas during January and April, and over the southeastern and central United States in July and October in both the arw_p1 and arw_p2 simulations compared to arw_ctl simulation. We notice the largest changes in emissions over the land areas during October.

The VDIF acts as a sink for aerosols over most parts of the domain during January, April, and October because it transports the emissions out of the surface layer. However, VDIF acts a source for accumulation mode aerosol particles over the southeastern United States and parts of the northeastern United States, the Four Corners region, and Canada in July, indicating vertical transport of aerosols from the model layers aloft. Relative to arw_ctl, VDIF in arw_p1 and arw_p2 leads to mixed changes in AMASSJ tendencies with the majority of grid boxes experiencing a stronger diffusion in arw_p1 and arw_p2 as represented by blue colors in the middle and right panels of Figures 9–12. However, some parts of the United States especially the northeastern United States in January, the southeastern United States in July, and the central,

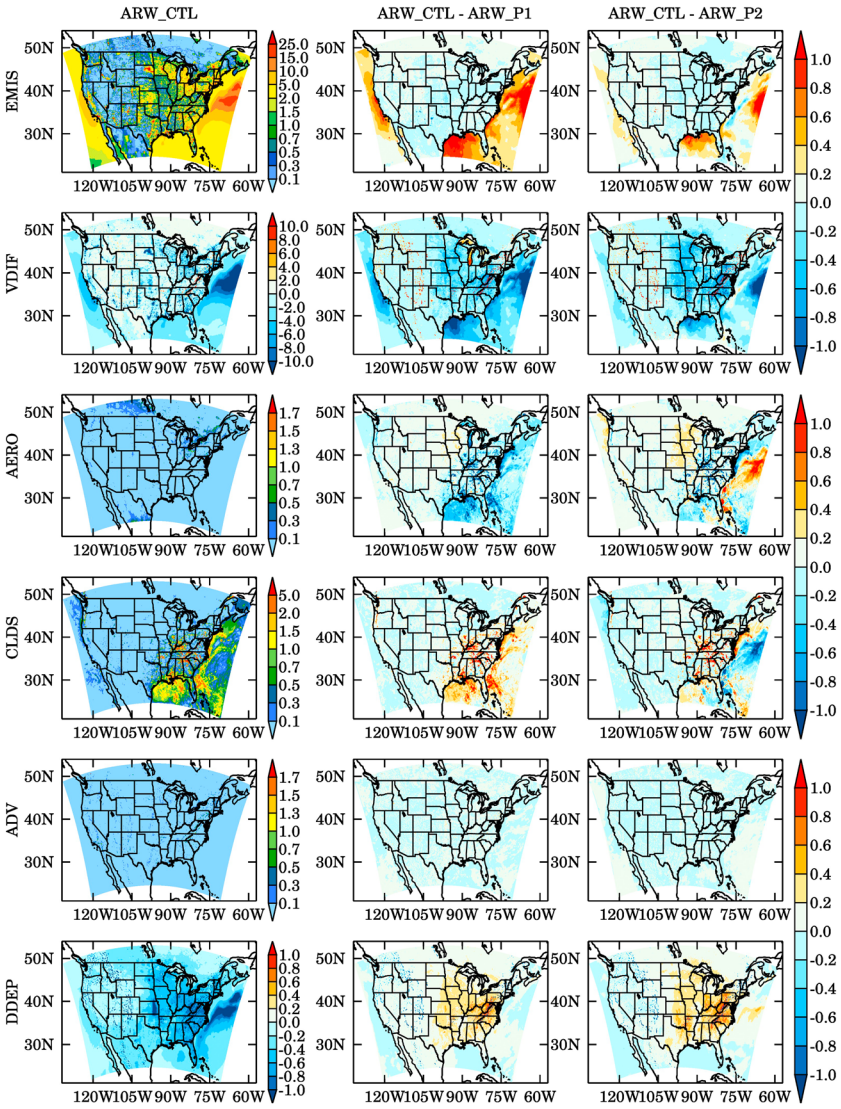

Figure 10. Same as Figure 9 but for April 2016.

southeastern, and parts of the western United States in October experience much stronger VDIF in the arw_ctl simulations.

AERO have a larger effect on AMASSJ in January (over the eastern United States) and July (over the southeastern United States) compared to April and October. Meteorological changes in arw_p1 and arw_p2 compared to arw_ctl affect the AERO more over the eastern United States and in the oceanic regions. CLDS also have a larger impact on AMASSJ over the oceanic regions, eastern United States, and parts of the Pacific Northwest. The strongest changes in AMASSJ due to AERO over the oceanic regions are seen in January when the strength of CLDS decreases in the arw_p1 simulation relative to arw_ctl but increases in the arw_p2 simulation. Similar behavior of CLDS is observed in April and to some extent in October.

The advection process leads to smaller absolute changes (within $\pm 0.2\ \mu g/m^3$) in AMASSJ compared to other processes. The DDEP always acts as a sink for aerosols and thus is always negative. The removal of aerosols via DDEP unsurprisingly increases from January through April to July and then decreases again in October. Similarly, the changes in DDEP in the arw_p1 and arw_p2 simulations compared to the arw_ctl simulations are also the highest in July. Interestingly, the DDEP tends to become weaker in the northeastern United States and stronger in the southeastern United States in both arw_p1 and arw_p2 simulations during all the months, except in July when DDEP tends to be weaker even in the southeastern United States.

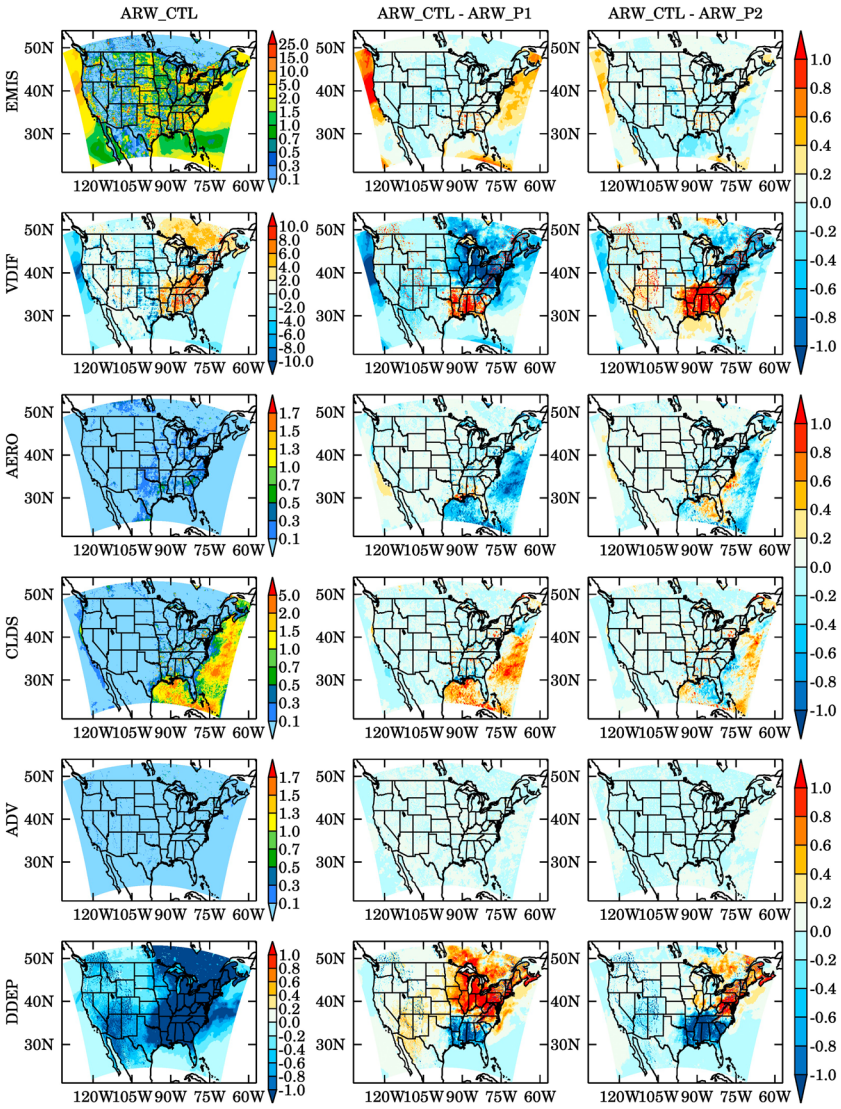

Figure 11. Same as Figure 9 but for July 2016.

Table 3 shows the monthly domain average and standard deviation (representing the spatial variability) of AMASSJ tendency due to AERO, CLDS, DDEP, EMIS, VDIF, HADV, and ZADV for January, April, July, and October 2016 in the three simulations. On average monthly domain-wide scale, emissions, and CLDS act as sources of AMASSJ, while all other processes act as sinks of AMASSJ in all the simulations except ZADV in January and VDIF in arw_p1 in July. AERO, CLDS, EMIS, and VDIF diffusion have a larger effect on AMASSJ than DDEP, HADV, HDIF, and ZADV. The highest values for EMIS and VDIF are higher than those for other processes by a factor of 10.

We have also estimated correlation coefficient between arw_ctl and arw_p1 (r1) and arw_ctl and arw_p2 (r2) to understand if meteorological variability changes the spatial distribution of AMASSJ tendency due to different processes. As expected, we see strong to perfect correlation (0.97–1) between the three simulations for EMIS as meteorological variations do not affect the location of emission sources. Lower correlation coefficients for HADV and ZADV (0.81-0.89) than VDIF (0.95-1) indicate that meteorological variability perturbs spatial structure of advection processes more than diffusion. The meteorological variability substantially changes the spatial structure of AERO and CLDS as indicated by moderate correlation coefficient values in the range of 0.52-0.76. This is consistent with the large heterogeneity in surface temperature distribution as shown in Figure S13.

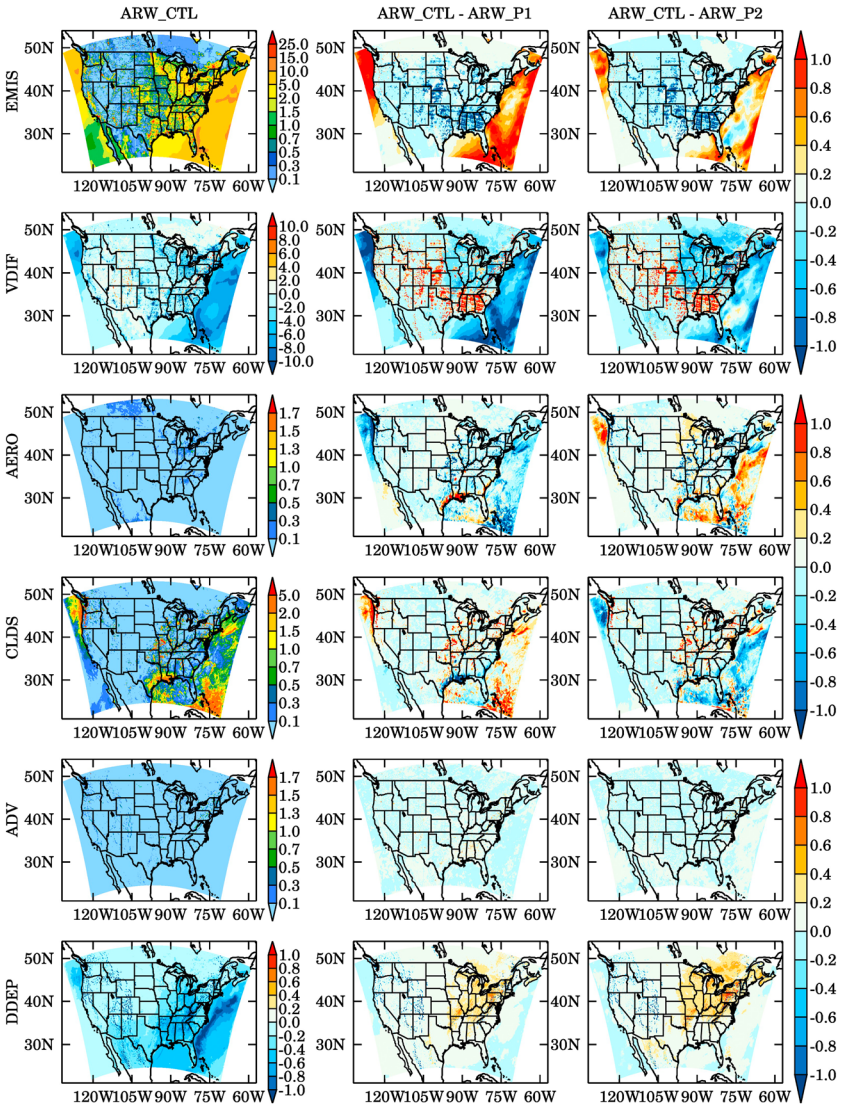

Figure 12. Same as Figure 9 but for October 2016.

We also identify which process contributes the most to the changes in AMASSJ tendency when the meteorological configuration is changed from arw_ctl to arw_p1 and arw_p2 (Figure 13). The corresponding percentage of grid points dominated by each process is shown in Table 4. The changes in AMASSJ tendency due to changes in meteorological configuration over the oceanic regions is dominated by EMIS (orange color), VDIF (dark green color), AERO (red color), and CLDS (light green color). However, the spatial heterogeneity is larger over the oceanic regions when arw_ctl is changed to arw_p2. HADV (brown color) and ZADV (pink color) advection contributes the most to the changes in AMASSJ tendency over the mountainous regions in the western United States in all the months. Both the advection processes dominate 6-12% of the total number of grid points in different months. HDIF does not dominate any grid point in any month. During January, changes in AMASSJ tendency in most parts of the midwest, northeast, and southeast United States are dominated by AERO followed by DDEP and VDIF. Across the domain, VDIF (27-29%) and AERO (20-26%) dominate the maximum number of grid points in January. VDIF dominates most of the areas in eastern parts of domain in April, and thus, the percentage of VDIF-dominated grid points increases to 41%. AERO-dominated grid points decrease to 10-12% in April, but EMIS-dominated grid points increase to 12-19%. DDEP becomes the second most important process after VDIF in July as DDEP-dominated grid points increase to 20-23%. In October, EMIS ranks as the second most important process with a percentage of 20-24%, while all other processes remain at less than 10%.

Table 3
Monthly Average and Standard Deviation in Tendency of AMASSJ Due to AERO, CLDS, DDEP, EMIS, VDIF, HADV, and ZADV for January, April, July, and October 2016

	Mem.	AERO	CLDS	DDEP	EMIS	VDIF	HADV	ZADV
Jan 2016	arw_ctl	-0.09±0.33	0.08±0.30	-0.006±0.09	0.09±1.21	-0.07±1.14	-0.001±0.11	0.0006±0.11
	arw_p1	-0.09±0.27	0.08±0.24	-0.008±0.08	0.08±1.17	-0.07±1.10	-0.003±0.13	0.0007±0.13
	arw_p2	-0.09±0.30	0.08±0.24	-0.009±0.08	0.09±1.17	-0.07±1.10	-0.003±0.14	0.0007±0.13
	Min, Max	-27.5, 9.6	-2.0, 25.9	-24.3, 0	0, 281.2	-276.3, 19.5	-23.9, 9.8	-11.0,10.0
	r1, r2	0.76, 0.73	0.74, 0.67	0.68, 0.68	1, 1	0.99, 0.99	0.82, 0.81	0.84, 0.83
Apr 2016	arw_ctl	-0.08±0.28	0.07±0.26	-0.009±0.12	0.06±0.84	-0.04±0.77	-0.002±0.09	-0.0009±0.09
	arw_p1	-0.08±0.21	0.06±0.19	-0.012±0.10	0.05±0.80	-0.03±0.72	-0.002±0.11	-0.0019±0.09
	arw_p2	-0.06±0.18	0.05±0.14	-0.012±0.11	0.06±0.82	-0.03±0.74	-0.002±0.11	-0.0015±0.09
	Min, Max	-21.5, 5.0	-0.7, 21.7	-31.1, 0	0, 186.8	-182.2, 30.6	-20.0, 7.0	-8.5, 5.9
	r1, r2	0.73, 0.72	0.70, 0.67	0.66, 0.65	0.99, 1	0.99, 1	0.89, 0.89	0.88, 0.88
Jul 2016	arw_ctl	-0.07±0.33	0.06±0.31	-0.034±0.43	0.05±0.69	-0.007±0.71	-0.003±0.13	-0.0016±0.13
	arw_p1	-0.07±0.22	0.06±0.21	-0.039±0.42	0.04±0.67	0.002±0.67	-0.003±0.13	-0.0027±0.12
	arw_p2	-0.04±0.19	0.03±0.16	-0.030±0.34	0.05±0.69	-0.011±0.67	-0.004±0.12	-0.0021±0.11
	Min, Max	-18.2, 10.0	-3.4, 18.2	-60.0, 0	0.0, 175.7	-171.6, 57.9	-23.3, 8.7	-11.8, 10.2
	r1, r2	0.68, 0.59	0.64, 0.52	0.86, 0.85	0.99, 1	0.95, 1	0.86, 0.85	0.86, 0.85
Oct 2016	arw_ctl	-0.10±0.40	0.09±0.38	-0.009±0.15	0.08±0.88	-0.053±0.80	-0.003±0.12	-0.0002±0.11
	arw_p1	-0.11±0.29	0.09±0.28	-0.011±0.08	0.07±0.86	-0.047±0.78	-0.003±0.14	-0.0011±0.13
	arw_p2	-0.08±0.25	0.07±0.21	-0.012±0.10	0.08±0.90	-0.049±0.80	-0.004±0.14	-0.0009±0.14
	Min, Max	-45.6, 4.3	-2.6, 45.0	-41.0, 0	0, 177.4	-173.7, 39.2	-12.8, 5.9	-5.8, 8.2
	r1, r2	0.66, 0.68	0.61, 0.63	0.53, 0.52	0.97, 0.98	0.96, 0.96	0.87, 0.88	0.88, 0.88

Note. All values are in $\mu g \cdot m^{-}3 \cdot hr^{-}1$. Average horizontal diffusion values are in the range of 10^{-7} to 10^{-8} $\mu g \cdot m^{-}3 \cdot hr^{-}1$ and thus are not shown here. Standard deviation represents the spatial variability across the domain. Minimum and maximum values across the three simulations are also shown along with spatial correlation between arw_ctl and arw_p1 (r1) and arw_ctl and arw_p2 (r2). AERO, aerosol processes; CLDS, cloud processes; DDEP, dry deposition; EMIS, emissions; HADV, horizontal advection; VDIF, vertical diffusion; ZADV, vertical advection.

To examine the effect of meteorological variability on aerosol chemical composition, the percentage contribution of monthly averaged sulfate (SO_4), nitrate (NO_3), ammonium (NH_4), elemental carbon, organic carbon (OC), soil, and all other aerosol chemical components to monthly averaged $PM_{2.5}$ mass concentrations over the entire domain in arw_ctl, arw_p1, and arw_p2 configurations are examined (Figure 14). The contribution of each aerosol chemical component to $PM_{2.5}$ mass concentration is estimated using the sharp-cut $PM_{2.5}$ inlet method (Jiang et al., 2006). During all the months, SO_4 dominates the $PM_{2.5}$ chemical composition in all the three configurations with the highest contribution in arw_p2 simulation. The percentage contribution of NO_3 in July is less than 5%, while it is around or more than 10% in other months in all the configurations. Similar decrease is seen in NH_4 concentrations in July. In contrast, the percentage contribution of OC and others increase in July compared to the other months. The contribution of elemental carbon remains constant at around 3% in all the months and simulations, and soil shows larger variability among the three simulations in October. The meteorological induced variability among the simulations is generally within ±5% on a monthly average scale.

In summary, the meteorological variability considerably affects the strength of all processes that control the spatial and temporal distribution of aerosol mass concentrations. The meteorology-induced changes in $PM_{2.5}$ over 55-73% of the domain are dominated by emissions, VDIF, and AERO except in July when DDEP dominates over 20-23% of the domain. Meteorological variability leads to the largest spatial variability in aerosol and CLDS.

4.3. Variability in $PM_{2.5}$ Forecasts

The maps of monthly averaged $PM_{2.5}$ mass concentrations over the model domain in January, April, July, and October 2016 in arw_ctl, arw_p1, and arw_p2 configurations are shown in Figure 15. We see higher $PM_{2.5}$ mass concentrations over the eastern United States. Arw_p1 shows the highest concentrations in most parts of the domain in January except for some areas in the northeastern part of the domain where arw_p2 shows the highest concentrations. $PM_{2.5}$ mass concentrations decrease in April and July relative to January especially over the eastern United States and increase again in October. In contrast, $PM_{2.5}$ mass concentrations increase over the western United States in July likely because of the wildfires.

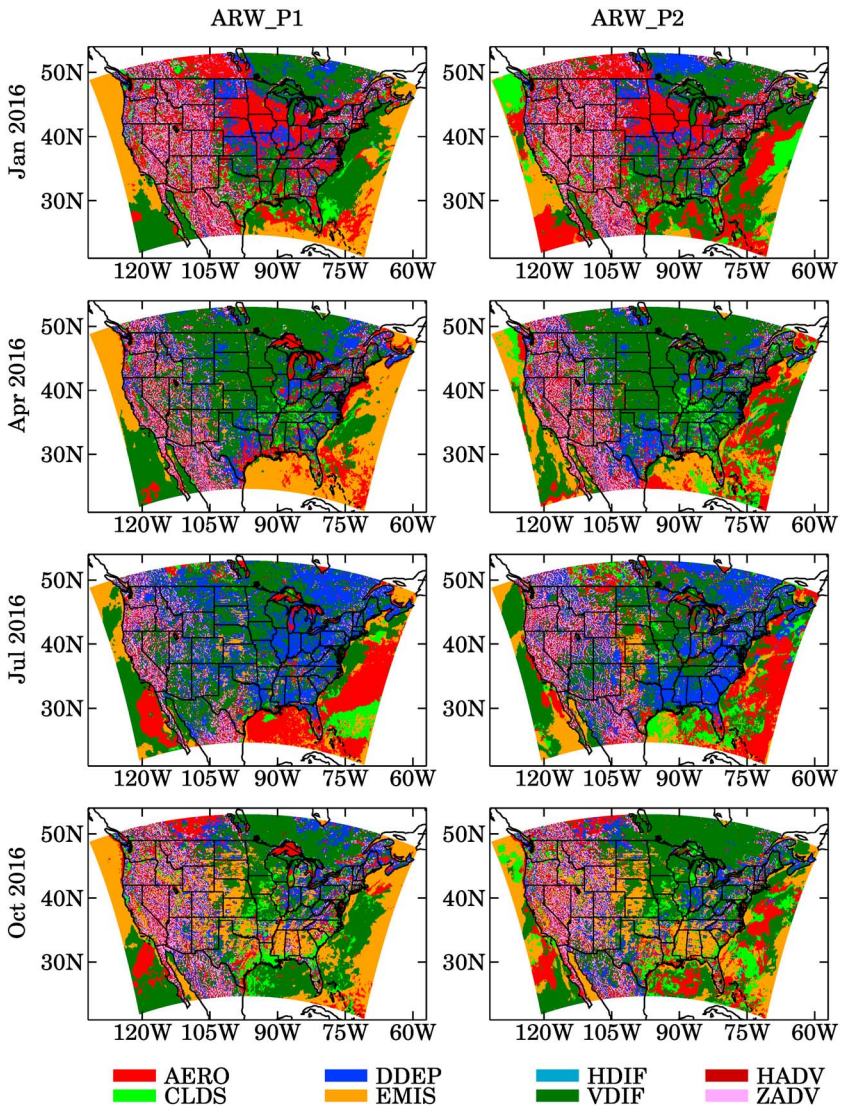

Figure 13. Maps showing which process contribute the most to changes in AMASSJ tendency when meteorological configuration is changed from arw_ctl to arw_p1 (left panel) and arw_p2 (right panel) in January, April, July, and October 2016.

Table 4
Percentage of Grid Points Dominated by Different Processes Contributing to the Changes in AMASSJ Tendency Due to Changes in Meteorological Configuration From arw_ctl to arw_p1 and arw_p2

	January 2016		April 2016		July 2016		October 2016	
Process	arw_p1	arw_p2	arw_p1	arw_p2	arw_p1	arw_p2	arw_p1	arw_p2
AERO	20	26	10	12	18	16	9	12
CLDS	5	9	4	8	4	7	6	9
DDEP	9	9	11	10	23	20	9	9
EMIS	14	7	19	12	8	9	24	20
VDIF	29	27	41	41	34	31	33	34
HDIF	0	0	0	0	0	0	0	0
HADV	12	12	8	9	6	9	9	8
ZADV	10	11	7	8	7	9	9	8

Note. AERO, aerosol processes; CLDS, cloud processes; DDEP, dry deposition; EMIS, emissions; HADV, horizontal advection; HDIF, horizontal diffusion; VDIF, vertical diffusion; ZADV, vertical advection.

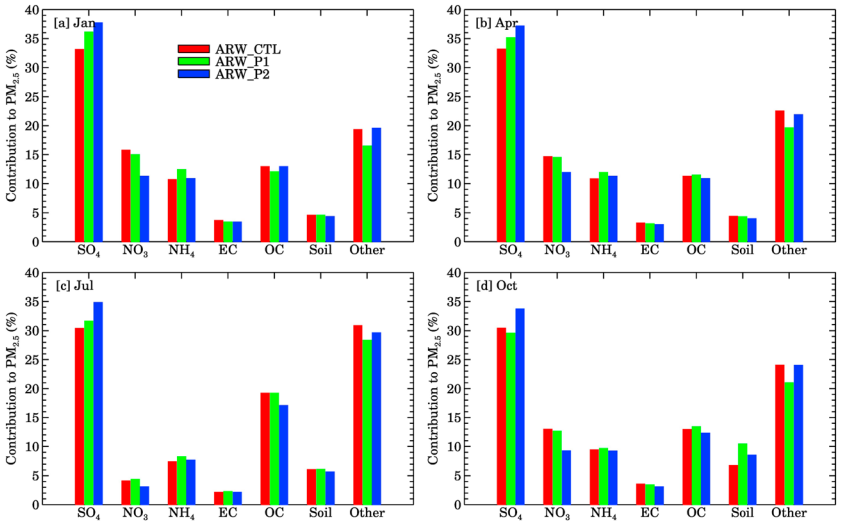

Figure 14. Percentage contribution of monthly averaged sulfate (SO_4), nitrate (NO_3), ammonium (NH_4), elemental carbon (EC), organic carbon (OC), soil, and all other aerosol chemical components to monthly averaged particulate matter 2.5 ($PM_{2.5}$) mass concentrations over the entire domain in arw_ctl, arw_p1, and arw_p2 configurations during (a) January, (b) April, (c) July, and (d) October 2016.

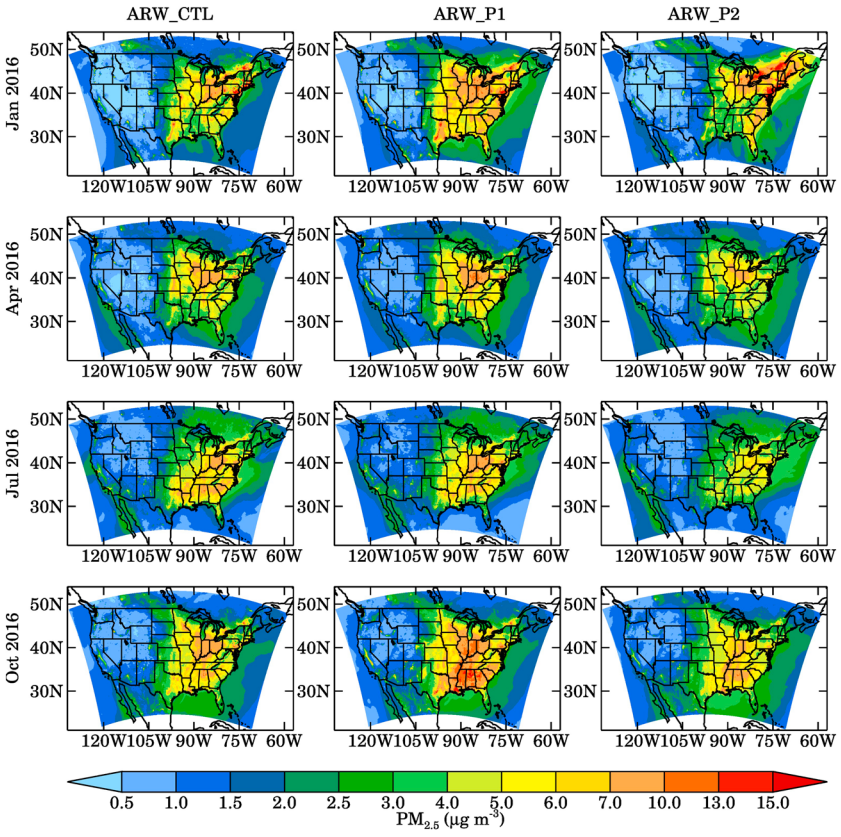

Figure 15. Spatial distribution of monthly averaged particulate matter 2.5 ($PM_{2.5}$) mass concentrations over the model domain in January, April, July, and October 2016 in arw_ctl, arw_p1, and arw_p2 configurations.

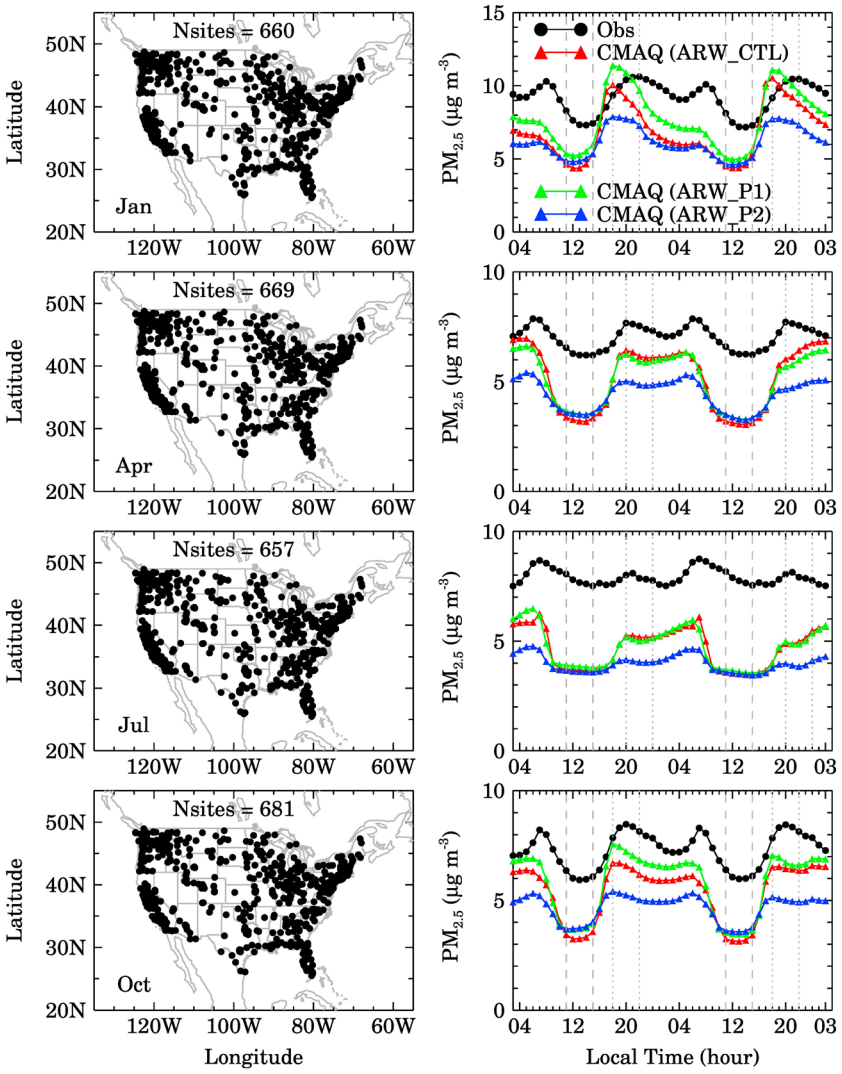

Figure 16. Spatial distribution of Environmental Protection Agency (EPA) AirNOW sites used for evaluation of Community Multiscale Air Quality (CMAQ) meteorology-based ensemble members during January, April, July, and October 2016 (left panel). The comparison of observed and modeled 48-hr particulate matter 2.5 ($PM_{2.5}$) forecasts averaged over all the sites in the four months studied here. The standard deviation in average observed values range from 4.5 to 12.4 μg/m^3 and those in the average CMAQ value range from 2.5 to 17.0 μg/m^3. Standard deviation in the average values is not shown for brevity. Note the different ordinate axis range for $PM_{2.5}$ plots. The dashed and dotted lines in the right panel enclose the time periods used for calculation of day and night meteorological ensemble variability.

The observed and modeled 48-hr $PM_{2.5}$ forecasts averaged over all the AirNOW sites during the months of January, April, July, and October 2016 are compared in Figure 16 for all three simulations. CMAQ consistently underestimates the observed $PM_{2.5}$ levels during all the months in all the three simulations except for a few hours in arw_p1 and arw_ctl during January. Both the observations and model show higher nighttime $PM_{2.5}$ levels and lower daytime levels. Among the three CMAQ simulations, arw_p2 shows the lowest $PM_{2.5}$ mass concentrations during nighttime in all the months. Nighttime $PM_{2.5}$ mass concentrations are the highest in arw_p1 during January and October and in arw_ctl during April and July.

The three simulations show larger variability during nighttime and agree well during daytime. The diurnal variability, that is, the difference between maximum and minimum $PM_{2.5}$ levels, also decreases in both the model and observations from January through April to July but increases again in October (Table 5). The diurnal variability in arw_p2 is smaller than the observations in January and October but higher in April and similar in July. Both arw_ctl and arw_p1 show a larger diurnal variability than the observations in all the months.

Table 5
Diurnal Variability of $PM_{2.5}$ Levels in the Observations and Three CMAQ Members

Month	Obs	arw_ctl	arw_p1	arw_p2
January	3.4	5.0	5.5	2.7
April	1.6	3.8	3.4	2.1
July	1.2	2.8	3.1	1.3
October	2.5	3.1	3.8	1.7

Note. CMAQ, Community Multiscale Air Quality; $PM_{2.5}$, particulate matter 2.5.

To understand why nighttime $PM_{2.5}$ mass concentrations exhibit higher variability than the daytime variability, we analyze how the contribution of different processes to AMASSJ tendency changes from the day (the area enclosed by dashed lines on the right panels in Figure 16) to night periods (the area enclosed by dotted lines on the right panels in Figure 16). The changes in the contribution of a process to changes in AMASSJ from daytime to nighttime are estimated using the following equation:

$$\mathrm{AMASSJ}_p = \overline{\mathrm{AMASSJ}}_{p,\mathrm{night}} - \overline{\mathrm{AMASSJ}}_{p,\mathrm{day}} \quad (5)$$

where p represents the processes AERO, CLDS, DDEP, EMIS, HADV, HDIF, VDIF, and ZADV and $\overline{\mathrm{AMASSJ}}_{p,\mathrm{night}}$ and $\overline{\mathrm{AMASSJ}}_{p,\mathrm{day}}$ represent the average AMASSJ tendencies due to process p during night and day periods, respectively. The overbars denote the averages over the day and night periods.

Figure 17 depicts AMASSJ averaged over all the sites for all the processes during January, April, July, and October 2016. Emissions and VDIF have the strongest influence on AMASSJ except during July when the role of DDEP is larger compared to VDIF. During January, emissions increase from daytime to nighttime but decrease in other months. The amount of changes in emissions from daytime to nighttime is nearly the same in all the three simulations, except during October when the nighttime decrease in emissions in arw_ctl simulation is smaller. The variability in average AMASSJ tendency due to emissions in arw_p1 and arw_p2 relative to arw_ctl is less than 3.1% in all the months. However, the variability in VDIF among different simulations is higher (up to 19%) especially in April and July. While aerosol, clouds, and advection processes have a smaller absolute contribution to average AMASSJ tendency, there is larger variability (up to 300%) in these processes among different simulations. The day and night variability in DDEP among the three simulations during day and night periods ranges from -6 to 114% in July 2016.

At each site, we have also identified the process that leads to the highest changes in AMASSJ tendency during day and night periods when arw_ctl is changed to arw_p1 and arw_p2 to understand how the dominance of different processes changes between the day and night periods. Percentage of sites dominated by each process during the day and night periods are shown in Table 6. The number of sites dominated by AERO, CLDS, HADV, and ZADV increase from day to night periods except during January when AERO dominate the same number of sites in day and night periods. In contrast, the number of sites dominated by emissions, VDIF, and DDEP decrease from day to night periods.

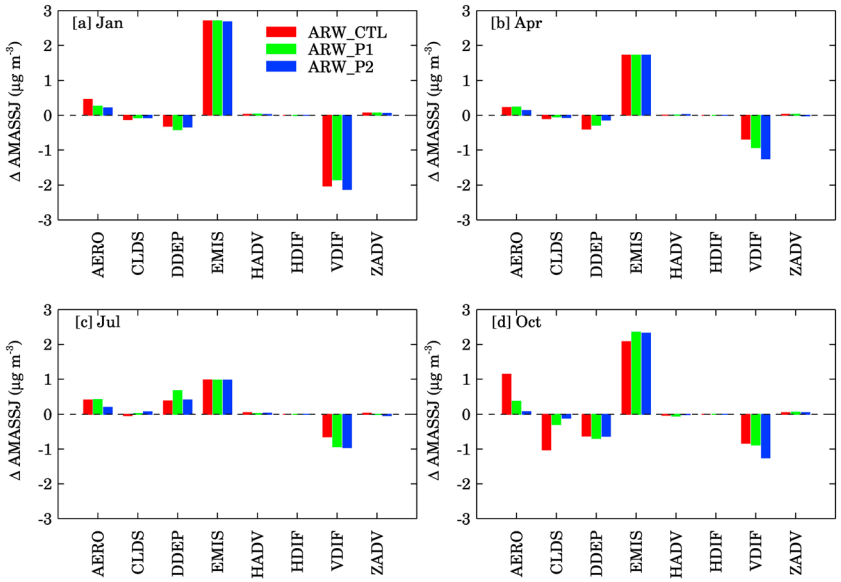

Figure 17. Monthly average values of AMASSJ for (a) January, (b) April, (c) July, and (d) October 2016.

Table 6
Percentage of Grid Points Dominated by Different Processes Contributing to the Changes in AMASSJ Tendency Due to Changes in Meteorological Configuration From arw_ctl to arw_p1 and arw_p2 at All the Observation Sites in January, April, July, and October 2016

	January 2016		April 2016		July 2016		October 2016	
Process	Day	Night	Day	Night	Day	Night	Day	Night
AERO	32	32	9	22	3	11	5	19
CLDS	6	8	4	12	1	9	4	11
DDEP	13	3	19	4	44	11	16	4
EMIS	3	2	4	1	3	1	18	9
HADV	12	26	10	23	11	29	10	21
VDIF	29	16	48	21	29	16	40	22
ZADV	6	13	5	17	8	23	9	15

Note. Horizontal diffusion does not dominate at any site either during the day or night period and thus is not shown. AERO, aerosol processes; CLDS, cloud processes; DDEP, dry deposition; EMIS, emissions; HADV, horizontal advection; VDIF, vertical diffusion; ZADV, vertical advection.

The performance of three CMAQ simulations is assessed compared to the observations in terms of correlation coefficient, mean bias, and RMSE in Figure 18. Data from all the sites at each lead time have been used to calculate these statistical parameters. The correlation coefficient is lower than 0.4 in all the months for all the simulations. The arw_p1 and arw_p2 shows slightly higher correlation coefficients during January and October than the arw_ctl simulation. All three simulations show nearly similar performance in terms of the correlation coefficient during April and July. Similar to the mean $PM_{2.5}$ mass concentrations, the mean bias and RMSE also show larger diurnal variability in January. The mean bias ranges from -5 to 2 $\mu g/m^3$ in January, -3.5 to 0 $\mu g/m^3$ in April and October, and -5 to -1 $\mu g/m^3$ in July. The RMSE ranges from 5 to 13 $\mu g/m^3$, except during January when the nighttime values exceed 15-20 $\mu g/m^3$ in the arw_ctl and arw_p1 simulations.

Finally, we estimate the $PM_{2.5}$ variability across the three CMAQ simulations at each observation site using equation (6) to identify the regions that are affected the most by the meteorological variability.

$$PM_{2.5}\ \text{variability} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\frac{1}{3}\sum_{m=1}^{3}\left(\overline{PM_{2.5}}^{t} - PM_{2.5}^{m,t}\right)^2\right)} \quad (6)$$

where $t = 1, 2, \ldots, N$ represents the CMAQ forecast lead times across all the days, $\overline{PM_{2.5}}^{t}$ represent the $PM_{2.5}$ mass concentrations averaged over the three CMAQ simulations ($m = 1, 2, 3$) at time t, and $PM_{2.5}^{m,t}$ represents the $PM_{2.5}$ mass concentration for the CMAQ simulation corresponding to the member m. $PM_{2.5}$ variability is estimated separately for day and night periods to examine whether or not the larger nighttime variability seen in the average picture presented above prevails across all the sites.

The $PM_{2.5}$ variability for both the daytime and nighttime periods in all the 4 months is shown in Figure 19. $PM_{2.5}$ variability varies from as low as 0.08 $\mu g/m^3$ to as high as 24 $\mu g/m^3$ across all the months. Consistent with the average picture, most of the sites show larger nighttime variability. $PM_{2.5}$ variability is higher in the eastern United States both during the daytime and nighttime, but many sites in the western United States also show variability similar to the eastern United States during nighttime. The northeastern states

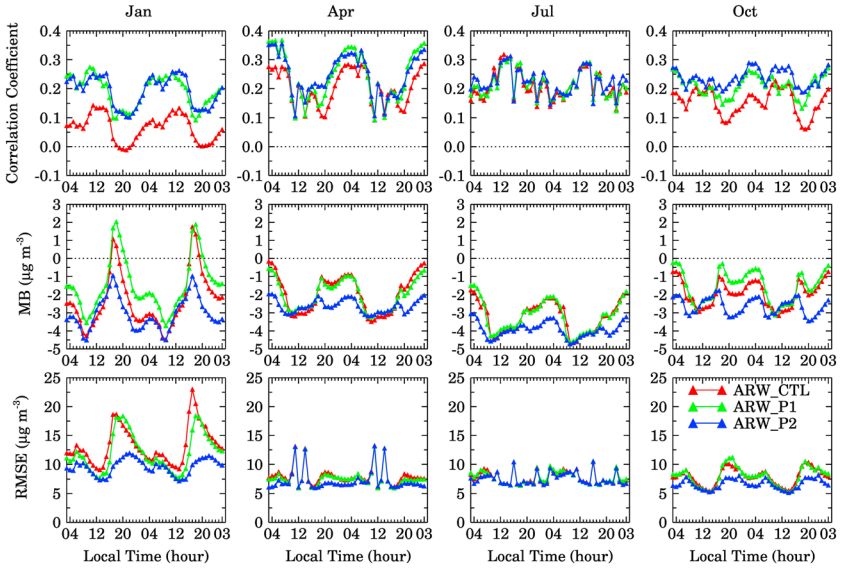

Figure 18. Time variation in correlation coefficient, mean bias (MB), and the root-mean-square error (RMSE) estimated using collocated model-observation pairs from all the Environmental Protection Agency (EPA) AirNOW sites for each ensemble member during January, April, July, and October 2016.

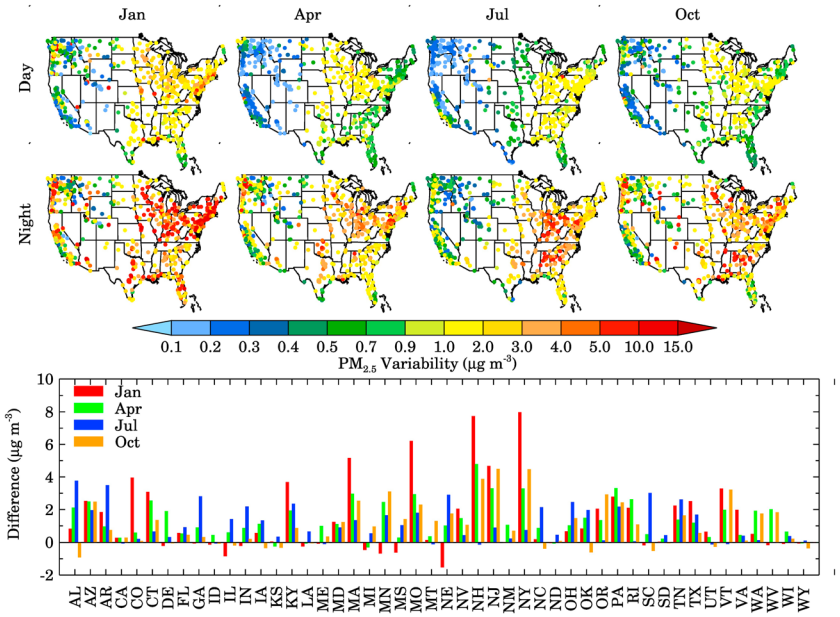

Figure 19. Particulate matter 2.5 ($PM_{2.5}$) variability for the day and night periods in January, April, July, and October 2016 are shown for each AirNOW site in the top and middle panels, respectively. The bottom panel shows the difference between day and night variability for each state.

show the highest variability in all the months except during July when southeastern states also show variability similar to northeastern states. To understand when the daytime and nighttime $PM_{2.5}$ variability is the most different at state level, we estimated the difference in $PM_{2.5}$ variability by subtracting the daytime values from the corresponding nighttime values at each site and averaging the difference variability over all the sites belonging to individual states. This difference in $PM_{2.5}$ variability is shown in bottom panel of Figure 19. We find that Massachusetts, Missouri, New Hampshire, New Jersey, and New York show the largest average difference (>4 $\mu g/m^3$) in $PM_{2.5}$ variability during January.

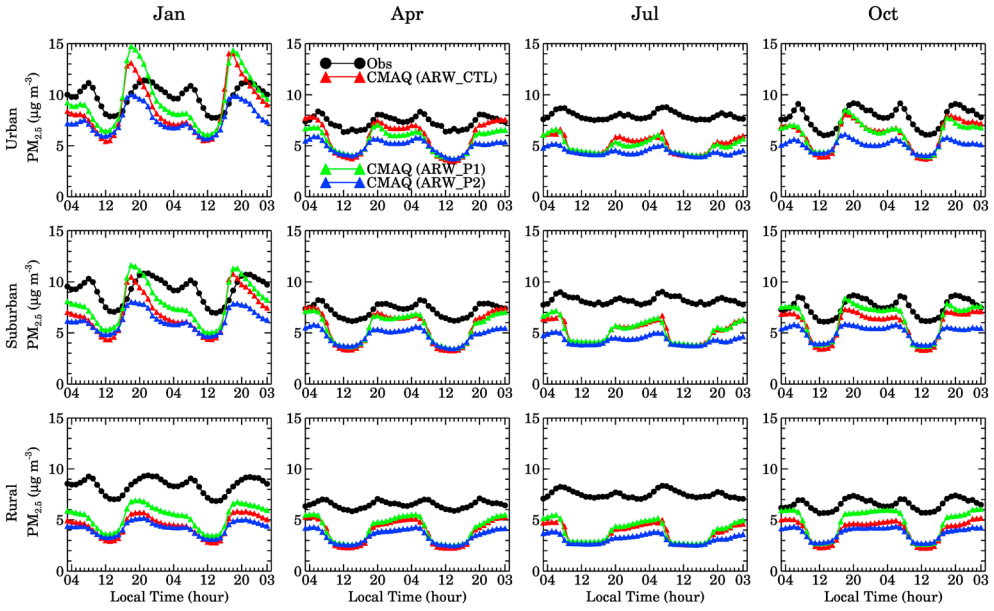

Figure 20. Community Multiscale Air Quality (CMAQ) 48-hr forecasts of monthly averaged particulate matter 2.5 ($PM_{2.5}$) mass concentrations for urban, suburban, and rural sites for January, April, July, and October 2016.

To understand how model performance vary with site type (urban, suburban, and rural), we compared the 48-hr CMAQ forecasts of $PM_{2.5}$ mass concentrations for urban, suburban, and rural sites for January, April, July, and October 2016 (Figure 20). Not surprisingly, we find that $PM_{2.5}$ mass concentrations are the highest at the urban sites followed by the suburban and rural sites. $PM_{2.5}$ simulated using the arw_p2 meteorological configuration is the lowest at all types of the sites in all the months likely because of the largest overestimation of wind speed in arw_p2 configuration. Both arw_ctl and arw_p1 yield similar concentrations in April and July, but arw_p1 simulates higher concentration at suburban and rural sites in October, and at all types of sites in January. Similar to the average picture (Figure 16), we see the largest variability among the three simulations in January. The overestimation seen in nighttime hours of January in arw_p1 and arw_ctl (Figure 16) is mostly from the overestimation at the urban sites with a very small contribution from the suburban sites. All CMAQ simulations underestimate the observed $PM_{2.5}$ at rural sites in January. At all types of sites, the nighttime variability is higher than the daytime variability. At the rural sites, the model shows the largest underestimation of $PM_{2.5}$, which can partially be attributed to the use of static LBCs in our configuration.

5. Summary

This study examines the impact of meteorological variability on the simulations of $PM_{2.5}$ mass concentrations over the CONUS. The NOAA SREF system is used to represent the meteorological variability in the CMAQ simulations of $PM_{2.5}$. To reduce the computational burden, an HCA technique is applied to down-select a subset of the SREF members that objectively represents the overall meteorological forecast variability of SREF. Three SREF members, namely, arw_ctl, arw_p1, and arw_p2, are selected to drive CMAQ simulations. These three members are found to be an adequate subset of the full SREF ensemble in terms of having comparable statistical consistency. Replicating the configuration of the three selected members, we run our own "SREF-like" 48-hr WRF-ARW forecasts that are then processed through the MCIP for use in CMAQ simulations. The processed meteorological fields are also fed to SMOKE to capture meteorology-induced changes in emissions input to CMAQ. CMAQ 48-hr forecasts are initialized daily at 0300 UTC throughout the months of January, April, July, and October 2016 to understand the impact of meteorological variability on $PM_{2.5}$ simulations in each season.

The WRF-ARW forecasts of 2-m temperature, 10-m wind speed, precipitation, and PBLH are evaluated against in situ (EPA and IGRA) and satellite (TRMM) observations. All the WRF configurations simulate the spatial and temporal variability of 2-m temperature very well but overestimate 10-m wind speed and precipitation. All the WRF simulations overestimate PBLH at many sites as well. However, the PBLH evaluation was limited only to later afternoon/evening hours. More comprehensive PBLH evaluation should be performed for daytime hours as well to better understand the causes of daytime $PM_{2.5}$ underprediction. WRF-ARW and CMAQ forecasts are analyzed to understand the effect of meteorological variability on $PM_{2.5}$ mass concentrations as well as different processes controlling $PM_{2.5}$. Across the three simulations, meteorological parameters (temperature, water vapor, wind speed, and PBLH) show mixed changes with increases over some parts of the domain and decreases over others. The meteorological variability affects all the processes that contribute to variability in $PM_{2.5}$ mass concentrations, but meteorology-induced changes in accumulation mode aerosol mass concentration, that is, AMASSJ over 55-73% of the domain, are dominated by emissions, VDIF, and AERO except in July when DDEP dominates over 20-23% of the domain. We also find that meteorological variability leads to the largest spatial variability in aerosol and CLDS followed by advection. In all seasons and configurations, SO_4 dominates the $PM_{2.5}$ chemical composition, but we also notice an increase in the contribution of OC in July.

Averaged over CONUS, CMAQ simulations driven by all three meteorological configurations underestimate the observed $PM_{2.5}$ mass concentrations with a good agreement between the simulated $PM_{2.5}$ during daytime and large variability during nighttime. All the CMAQ simulations show larger diurnal variability in January and October, with arw_p1 and arw_ctl showing diurnal variability even larger than observations. Process analysis for the selected day and night periods showed that changes in the strength of VDIF from day to night plays an important role in determining the day to night increase in $PM_{2.5}$ mass concentrations, except during October when changes in both the emissions and VDIF are important. Our results show that meteorological variability leads to variations of 0.08–24 $PM_{2.5}$ $\mu g/m^3$ over the CONUS, with the eastern

United States showing larger variability compared to the western United States. The variability is similar across the urban, suburban, and rural sites.

This study highlights the necessity of capturing meteorological variability in air quality ($PM_{2.5}$) forecasting over the CONUS. Uncertainties in other processes such as emissions, and inadequate representation of some of the processes, for example, secondary aerosol formation, dry and wet deposition, and initial and boundary conditions, can also lead to large variability in simulated $PM_{2.5}$ mass concentrations. Future work will combine the simulations presented here with additional CMAQ simulations designed to capture other sources of major uncertainties, particularly emissions and SOA formation, to generate a novel dynamical ensemble for probabilistic predictions of $PM_{2.5}$ mass concentrations over the CONUS.

Acknowledgments
We gratefully acknowledge the funding from NOAA OAR Office of Weather and Air Quality (grant NA16OAR4590116) for this study. We would like to acknowledge high-performance computing support from Cheyenne (Computational and Information Systems Laboratory, 2017). We also gratefully acknowledge the assistance of Jun Du in helping us obtain the operational NCEP SREF data archived on NOAA HPC. Data supporting the conclusions of this paper can be obtained here (https://zenodo.org/record/2603347#.XJVZqRNKiro). The National Center for Atmospheric Research is sponsored by the National Science Foundation. We thank the three anonymous reviewers for their constructive comments on the manuscript.

References

Alhamed, A., Lakshivarahan, S., & Stensrud, D. J. (2002). Cluster analysis of multimodel ensemble data from SAMEX. *Monthly Weather Review*, *130*, 226–256. https://doi.org/10.1175/1520-0493(2002)130<0226:CAOMED>2.0.CO;2

Aligo, E., Ferrier, B., Carley, J., Rogers, E., Pyle, M., Weiss, S. J., & Jirak, I. L. (2014). Modified microphysics for use in high-resolution NAM forecasts. *27th Conf. on Severe Local Storms*, Madison, WI. *Amer. Meteor. Soc.*, https://ams.confex.com/ams/27SLS/webprogram/Paper255732.html

Appel, K. W., Pouliot, G. A., Simon, H., Sarwar, G., Pye, H. O. T., Napelenok, S. L., et al. (2013). Evaluation of dust and trace metal estimates from the Community Multiscale Air Quality (CMAQ) model version 5.0. *Geoscientific Model Development*, *6*(4), 883–899. https://doi.org/10.5194/gmd-6-883-2013

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., et al. (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Monthly Weather Review*, *144*, 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Binkowski, F. S., & Roselle, S. J. (2003). Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description. *Journal of Geophysical Research*, *108*(D6), 4183. https://doi.org/10.1029/2001JD001409

Brown, B., Bullock, R., Fowler, T., Halley Gotway, J., Newman, K., & Jensen, T. (2016). Model Evaluation Tools version 5.2 (METv5.2) user's guide. Developmental Testbed Center, Boulder, CO, 329 pp., www.dtcenter.org/met/users/docs/users_guide/MET_Users_Guide_v5.2.pdf

Burnett, R. T., & Coauthors (2014). An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental Health Perspectives*, *122*, 397–403. https://doi.org/10.1289/ehp.1307049

Byun, D. W. (1999). Dynamically consistent formulations in meteorological and air quality models for Multiscale atmospheric studies. Part I: Governing equations in a generalized coordinate system. *Journal of the Atmospheric Sciences*, *56*(21), 3789–3807. https://doi.org/10.1175/1520-0469(1999)056<3789:DCFIMA>2.0.CO;2

Byun, D. W., & Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) model. *Applied Mechanics Reviews*, *59*, 51–77. https://doi.org/10.1115/1.2128636

Carlton, A. G., Bhave, P. V., Napelenok, S. L., Edney, E. O., Sarwar, G., Pinder, R. W., et al. (2010). Model representation of secondary organic aerosol in CMAQv4.7. *Env. Environmental Science & Technology*, *44*(22), 8553–8560. https://doi.org/10.1021/es100636q

Computational and Information Systems Laboratory (2017). Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). Boulder, CO, National Center for Atmospheric Research, https://doi.org/10.5065/D6RX99HX

Dawson, J. P., Adams, P. J., & Pandis, S. N. (2007). Sensitivity of $PM_{2.5}$ to climate in the Eastern US: A modeling case study. *Atmospheric Chemistry and Physics*, *7*, 4295–4309. https://doi.org/10.5194/acp-7-4295-2007

Di, Q., Wang, Y., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., et al. (2017). Air pollution and mortality in the Medicare population. *New England Journal of Medicine*, *376*, 2513–2522. https://doi.org/10.1056/NEJMoa1702747

Du, J., G. DiMego, B. Zhou, D. Jovic, B. Ferrier, & B. Yang (2015). Regional ensemble forecast systems at NCEP. *27th Conf. on Weather Analysis and Forecasting/23rd Conf. on Numerical Weather Prediction*, Chicago, IL, Amer. Meteor. Soc., https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273421.html

Environmental Modeling Center (2003). The GFS atmospheric model. NCEP Office Note 442, 14 pp. Retrieved from http://www.emc.ncep.noaa.gov/officenotes/newernotes/on442.pdf

Fann, N., Lamson, A. D., Anenberg, S., Wesson, K., Risley, D., & Hubbell, B. J. (2012). Estimating the national public health burden associated with exposure to ambient $PM_{2.5}$ and ozone. *Risk Analysis*, *32*, 81–95. https://doi.org/10.1111/j.1539-6924.2011.01630.x

Fountoukis, C., & Nenes, A. (2007). ISORROPIA II: A computational efficient thermodynamic equilibrium model for K+-Ca2+-Mg2+-NH4+-Na+-SO42—NO3—Cl—H2O aerosols. *Atmospheric Chemistry and Physics*, *7*(17), 4639–4659. https://doi.org/10.5194/acp-7-4639-2007

Fraley, C., Raftery, A. E., & Gneiting, T. (2010). Calibrating multimodel forecast ensemble with exchangeable and missing members using Bayesian model averaging. *Monthly Weather Review*, *138*, 190–202. https://doi.org/10.1175/2009MWR3046.1

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*, 2318–2341. https://doi.org/10.1175/MWR3199.1

Im, U., & Coauthors (2018). Assessment and economic valuation of air pollution impacts on human health over Europe and the United States as calculated by a multi-model ensemble in the framework of AQMEII3. *Atmospheric Chemistry and Physics*, *18*, 5967–5989. https://doi.org/10.5194/acp-18-5967-2018

Jaramillo, P., & Muller, N. Z. (2016). Air pollution emissions and damages from energy production in the U.S.: 2002–2011. *Energy Policy*, *90*, 202–211. https://doi.org/10.1016/j.enpol.2015.12.035

Jhun, I., Coull, B. A., Schwartz, J., Hubbell, B., & Koutrakis, P. (2015). The impact of weather changes on air quality and health in the United States in 1994–2012. *Environmental Research Letters*, *10*, 84009. https://doi.org/10.1088/1748-9326/10/8/084009

Jiang, W., Smyth, S., Giroux, E., Roth, H., & Yin, D. (2006). Differences between CMAQ fine mode particle and $PM_{2.5}$ concentrations and their impact on model performance evaluation in the Lower Fraser Valley. *Atmospheric Environment*, *40*(26), 4973–4985. https://doi.org/10.1016/j.atmosenv.2005.10.069

Jiménez, P. A., & Dudhia, J. (2012). Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. *Journal of Applied Meteorology and Climatology*, *51*, 300–316. https://doi.org/10.1175/JAMC-D-11-084.1

Johnson, A., Wang, X., Xue, M., & Kong, F. (2011). Hierarchical cluster analysis of a convection-allowing ensemble during the Hazardous Weather Testbed 2009 Spring Experiment. Part II: Ensemble clustering over the whole experiment period. *Monthly Weather Review*, *139*, 3694–3710. https://doi.org/10.1175/MWR-D-11-00016.1

Kelly, J. T., Bhave, P. V., Nolte, C. G., Shankar, U., & Foley, K. M. (2010). Simulating emission and chemical evolution of coarse sea-salt particles in the Community Multiscale Air Quality (CMAQ) model. *Geosci. Model Dev.*, *3*(1), 257–273. https://doi.org/10.5194/gmd-3-257-2010

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Wu, W.-S., & Lord, S. J. (2009). Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather and Forecasting*, *24*, 1691–1705. https://doi.org/10.1175/2009WAF2222201.1

Kumar, R., Delle Monache, L., Bresch, J., Saide, P. E., Tang, Y., Liu, Z., et al. (2019). Toward improving short-termpredictions offine particulate matterover the United States via assimilationof satellite aerosol optical depth retrie-vals. *Journal of Geophysical Research: Atmospheres*, *124*, 2753–2773. https://doi.org/10.1029/2018JD029009

Lee, J. A. (2012). Techniques for down-selecting numerical weather prediction ensembles. Ph.D. dissertation, The Pennsylvania State University, 131 pp., https://doi.org/10.13140/RG.2.2.29268.81280

Lee, J. A., Haupt, S. E., & Young, G. S. (2016). Down-selecting numerical weather prediction multi-physics ensembles with hierarchical cluster analysis. *Journal Climatol Weather Forecasting*, *4*, 156. https://doi.org/10.4172/2332-2594.1000156

Lee, P., & Coauthors (2017). NAQFC developmental forecast guidance for fine particulate matter ($PM_{2.5}$). *Weather and Forecasting*, *32*, 343–360. https://doi.org/10.1175/WAF-D-15-0163.1

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*, 243–256. https://doi.org/10.5194/gmd-3-243-2010

Pleim, J. E. (2007). A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *Journal of Applied Meteorology and Climatology*, *46*, 1383–1395. https://doi.org/10.1175/JAM2539.1

Reff, A., Bhave, P. V., Simon, H., Pace, T. G., Pouliot, G. A., Mobley, J. D., & Houyoux, M. (2009). Emissions inventory of $PM_{2.5}$ trace elements across the United States. *Environmental Science & Technology*, *43*(15), 5790–5796. https://doi.org/10.1021/es802930x

Rogers, E., Black, T., Ferrier, B., Y. Lin, Parrish, D., & DiMego, G. (2001). Changes to the NCEP Meso Eta Analysis and Forecast System: Increase in resolution, new cloud microphysics, modified precipitation assimilation, modified 3DVAR analysis. *NWS Tech. Procedures Bull*, http://www.emc.ncep.noaa.gov/mmb/mmbpll/eta12tpb/

Rogers, E., & Coauthors (2009). The NCEP North American Mesoscale modeling system: Recent changes and future plans. 23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction, Omaha, NE *Amer. Meteor. Soc.*, http://ams.confex.com/ams/pdfpapers/154114.pdf

Simon, H., & Bhave, P. V. (2012). Simulating the degree of oxidation in atmospheric organic particles. *Environmental Science & Technology*, *46*(1), 331–339. https://doi.org/10.1021/es202361w

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2008). A description of the Advanced Research WRF Version 3. *NCAR Tech. Note NCAR/TN-475+STR*, 113 pp., https://doi.org/10.5065/D68S4MVH

Solazzo, E., Riccio, A., Kioutsioukis, I., & Galmarini, S. (2013). Pauci ex tanto numero: Reducing redundancy in multi-model ensembles. *Atmospheric Chemistry and Physics*, *13*, 8315–8333. https://doi.org/10.5194/acp-13-8315-2013

Tai, A. P. K., Mickley, L. J., & Jacob, D. J. (2010). Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to climate change. *Atmospheric Environment*, *44*, 3976–3984. https://doi.org/10.1016/j.atmosenv.2010.06.060

Tang, Y., Lee, P., Tsidulko, M., Huang, H. C., McQueen, J. T., DiMego, G. J., et al. (2009). The impact of lateral boundary conditions on CMAQ predictions over the continental United States. *Environmental Fluid Mechanics*, *9*(1), 43–58. https://doi.org/10.1007/s10652-008-9092-5

Tang, Y., Pagowski, M., Chai, T., Pan, L., Lee, P., Baker, B., et al. (2017). A case study of aerosol data assimilation with the Community Multi-scale Air Quality Model over the contiguous United States using 3D-Var and optimal interpolation methods. *Geoscientific Model Development*, *10*, 4743–4758. https://doi.org/10.5194/gmd-10-4743-2017

Wei, M., Toth, Z., Wobus, R., & Zhu, Y. (2006). Initial perturbations based on the ensemble transform (ET) technique in the NCEP Global Ensemble Forecast System. *NCEP Office Note 453*, 33 pp., http://www.lib.ncep.noaa.gov/ncepofficenotes/files/on453.pdf

Whitten, G. Z., Heo, G., Kimura, Y., McDonald-Buller, E., Allen, D. T., Carter, W. L., & Yarwood, G. (2010). A new condensed toluene mechanism for Carbon Bond: CB05-TU. *Atmospheric Environment*, *44*(40), 5346–5355. https://doi.org/10.1016/j.atmosenv.2009.12.029

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences* (2nd ed., Vol. 626). Academic Press.

Yussouf, N., Stensrud, D. J., & Lakshmivarahan, S. (2004). Cluster analysis of multimodel ensemble data over New England. *Monthly Weather Review*, *132*, 2452–2462. https://doi.org/10.1175/1520-0493(2004)132<2452:CAOMED>2.0.CO;2

References From the Supporting Information

Beljaars, A. C. M. (1994). The parameterization of surface fluxes in large-scale models under free convection. *Quarterly Journal of the Royal Meteorological Society*, *121*, 255–270. https://doi.org/10.1002/qj.49712152203

Chou, M.-D. & Suarez, M. J. (1999). A solar radiation parameterization for atmospheric studies. *NASA Tech. Report NASA/TM-1999-104606*, 51 pp., http://www2.mmm.ucar.edu/wrf/users/phys_refs/SW_LE/New_Goddard.pdf

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, *108*(D22), 8851. https://doi.org/10.1029/2002JD003296

Fels, S. B., & Schwarzkopf, M. D. (1981). An efficient, accurate algorithm for calculating CO_2 15 μm band cooling rates. *Journal of Geophysical Research*, *86*, 1205–1232. https://doi.org/10.1029/JC086iC02p01205

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, *29*(14), 1693. https://doi.org/10.1029/2002GL015311

Han, J., & Pan, H. (2011). Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Weather and Forecasting*, *26*, 520–533. https://doi.org/10.1175/WAF-D-10-05038.1

Hong, S. Y., & Lim, J.-O. J. (2006). The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). *Journal of the Korean Meteorological Society*, *42*, 129–151. http://www2.mmm.ucar.edu/wrf/users/phys_refs/MICRO_PHYS/WSM6.pdf

Hong, S.-Y., & Pan, H.-L. (1996). Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Monthly Weather Review*, *124*, 2322–2339. https://doi.org/10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD009944

Iacono, M. J., Mlawer, E. J., Clough, S. A., & Morcrette, J.-J. (2000). Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. *Journal of Geophysical Research*, *105*, 14,873–14,890. https://doi.org/10.1029/2000JD900091

Janjić, Z., & Gall, R. L. (2012). Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on the B grid (NMMB). Part 1: Dynamics. *NCAR Tech. Note NCAR/TN-489+STR*, 80 pp., https://doi.org/10.5065/D6WH2MZX

Janjić, Z. I. (1994). The Step-Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, *122*, 927–945. https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2

Janjić, Z. I. (1996a). The Mellor-Yamada level 2.5 scheme in the NCEP Eta Model. Preprints, 11th Conf. on Numerical Weather Prediction, Norfolk, VA, Amer. Meteor. Soc., 333–334.

Janjić, Z. I. (1996b). The surface layer parameterization in the NCEP Eta Model. Preprints, 11th Conf. on Numerical Weather Prediction, Norfolk, VA, Amer. Meteor. Soc., 354–355.

Janjić, Z. I. (2002). Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model. *NCEP Office Note No. 437*, 61 pp., http://www2.mmm.ucar.edu/wrf/users/phys_refs/SURFACE_LAYER/eta_part4.pdf

Kain, J. S. (2004). The Kain-Fritsch convective parameterization: An update. *Journal of Applied Meteorology*, *43*, 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*, 16,663–16,682. https://doi.org/10.1029/97JD00237

Nakanishi, M., & Niino, H. (2006). An improved Mellor-Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Boundary-Layer Meteorology*, *119*, 397–407. https://doi.org/10.1007/s10546-005-9030-8

Pan, H.-L., & Mahrt, L. (1987). Interaction between soil hydrology and boundary layer developments. *Boundary-Layer Meteorology*, *38*, 185–202. https://doi.org/10.1007/BF00121563

Pan, H.-L., & Wu, W.-S. (1995). Implementing a mass flux convection parameterization package for the NMC Medium-Range Forecast Model. *NMC Office Note No. 409*, 40 pp., www2.mmm.ucar.edu/wrf/users/phys_refs/CU_PHYS/Old_SAS.pdf

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., et al. (2017). The Weather Research and Forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, *98*, 1717–1737. https://doi.org/10.1175/BAMS-D-15-00308.1

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., et al. (2004). Implementation and verification of the unified Noah land surface model in the WRF model. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., http://www2.mmm.ucar.edu/wrf/users/phys_refs/LAND_SURFACE/noah.pdf

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Monthly Weather Review*, *136*, 5095–5115. https://doi.org/10.1175/2008MWR2387.1