JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD033588

Key Points:

- Pleim-Xiu land-surface model, Asymmetric Convective Model version 2 (ACM2) planetary boundary layer model, and Pleim surface layer scheme from the Weather Research and Forecasting model (WRF) were added as options to the Model for Prediction Across Scales (MPAS) model
- Errors over the United States were substantially reduced in MPAS by using the Pleim-Xiu Land Surface Model (P-X LSM), ACM2 and Pleim surface layer configuration
- The magnitude of errors in MPAS are comparable to WRF, so MPAS is suitable for driving retrospective air quality models

Supporting Information:

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

Correspondence to:

R. C. Gilliam,
gilliam.robert@epa.gov

Citation:

Gilliam, R. C., Herwehe, J. A., Bullock, Jr, O. R., Pleim, J. E., Ran, L., Campbell, P. C., & Foroutan, H. (2021). Establishing the suitability of the model for prediction across scales for global retrospective air quality modeling. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033588. https://doi.org/10.1029/2020JD033588

Received 7 AUG 2020
Accepted 24 DEC 2020

Author Contributions:

Conceptualization: Jonathan E. Pleim, Hosein Foroutan
Formal analysis: Jerold A. Herwehe, Limei Ran, Patrick C. Campbell
Methodology: Jerold A. Herwehe, O. Russell Bullock, Jonathan E. Pleim, Limei Ran, Patrick C. Campbell, Hosein Foroutan
Resources: Limei Ran, Patrick C. Campbell

Establishing the Suitability of the Model for Prediction Across Scales for Global Retrospective Air Quality Modeling

Robert C. Gilliam[1], Jerold A. Herwehe[1], O. Russell Bullock Jr[1], Jonathan E. Pleim[1], Limei Ran[1,2], Patrick C. Campbell[3,4], and Hosein Foroutan[5]

[1]Center for Environmental Measurements and Modeling, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA, [2]Natural Resources Conservation Service, United States Department of Agriculture, Greensboro, NC, USA, [3]Center for Spatial Information Science and Systems/Cooperative Institute for Satellite Earth System Studies, George Mason University, Fairfax, VA, USA, [4]Department of Atmospheric and Oceanic Science/Cooperative Institute for Climate and Satellites-Maryland, University of Maryland, ARL/NOAA Affiliate, College Park, MD, USA, [5]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA

Abstract The U.S. EPA (United States Environmental Protection Agency) is leveraging recent advances in meteorological modeling to construct an air quality modeling system to allow consistency from global to local scales. The Model for Prediction Across Scales-Atmosphere (MPAS-A or MPAS) has been developed by the National Center for Atmospheric Research (NCAR) as a global complement to the Weather Research and Forecasting model (WRF). Patterned after a regional coupled system with WRF, the Community Multiscale Air Quality (CMAQ) modeling system has been coupled within MPAS to explore global-to-local chemical transport modeling. Several options were implemented into MPAS for retrospective applications. Nudging-based data assimilation was added to support continuous simulations of past weather to minimize error growth that exists with a weather forecast configuration. The Pleim-Xiu land-surface model, the Asymmetric Convective Model 2 boundary layer scheme, and the Pleim surface layer scheme were added as the preferred options for retrospective air quality applications with WRF. Annual simulations were conducted using this EPA-enhanced MPAS configuration on two different mesh structures and compared against WRF. MPAS generally compares well with WRF over the conterminous United States. Errors in MPAS surface meteorology are comparable to WRF throughout the year. Precipitation statistics indicate MPAS performs slightly better than WRF. Solar radiation in MPAS is higher than WRF and measurements, suggesting fewer clouds in MPAS than WRF. Upper-air meteorology is well-simulated by MPAS, but errors are slightly higher than WRF. These comparisons lend confidence to use MPAS for retrospective air quality modeling and suggest ways it can be further improved in the future.

Plain Language Summary The US EPA analyses and performs research on the past, present and future air quality of the United States using the Community Multiscale Air Quality model (CMAQ). Historically, the modeling was focused on the U.S. as regulations and impact are first order local issues. Global modeling is becoming more attainable and common as computer potential has increased, modeling systems advanced and air quality viewed as a global issue. This research demonstrates that we now have a meteorological modeling system that is capable of modeling air quality from global to local scales. The more comprehensive air quality modeling will directly address research issues on the link between air quality and human/ecological health. Specific results presented with this work indicate that the meteorology from the new global model is on par with the meteorology we have been using for the last decade. This suggests that the global air quality simulations will not be significantly restricted by poor driving meteorology and we can move forward with some confidence.

1. Introduction

Over more than two decades, the United States Environmental Protection Agency (U.S. EPA) has led the development of the Community Multiscale Air Quality (CMAQ; Byun & Schere, 2006) modeling system to inform regulatory decision-making, air quality research, forecasting, and other aspects of air quality science

Software: O. Russell Bullock, Jonathan E. Pleim, Limei Ran, Hosein Foroutan
Validation: O. Russell Bullock
Writing – review & editing: Jerold A. Herwehe, O. Russell Bullock, Jonathan E. Pleim, Limei Ran, Patrick C. Campbell, Hosein Foroutan

and application. CMAQ was initially designed to be driven by meteorology from the Mesoscale Model version 5 (MM5; Grell et al., 1994). It was then adapted for the Weather Research and Forecasting model (WRF; Skamarock et al., 2008) around 2007. Part of the adaptation of WRF for CMAQ applications was the implementation of four-dimensional data assimilation (FDDA) for retrospective modeling studies (Deng et al., 2007). Another key development in WRF was including physics options from MM5 that were tailored for air quality (AQ) modeling: The Pleim-Xiu Land Surface Model (P-X LSM), the Asymmetric Convective Model version 2 (ACM2) planetary boundary layer (PBL) scheme, and the Pleim surface layer model. This specific WRF model configuration–detailed in Gilliam and Pleim (2010) and Pleim and Gilliam (2009), facilitated the official transition from MM5 to WRF as the primary source of meteorology for CMAQ. This WRF configuration has been used extensively for CMAQ developments and applications over the last decade (Appel et al., 2014, 2017; Foley et al., 2015; Gan et al., 2015; Kang et al., 2020; Matichuk et al., 2017; McNider et al., 2018; Pleim et al., 2019; Pye et al., 2015; Xing et al., 2014).

Driven by MM5 or WRF, CMAQ modeling is typically conducted using limited-area, rectilinear grids on Lambert conformal or polar stereographic projections, requiring both meteorological and chemical boundary conditions from other models or analysis products. Chemical boundary conditions in CMAQ have advanced beyond using static background pollution levels to values that vary in time and space, originating from global air quality models such as the Goddard Earth Observing System model coupled to chemistry (GEOS-Chem; Bey et al., 2001). These global models use different chemical mechanisms and may solve for a different set of chemical species than CMAQ, which creates inconsistencies along the domain boundaries that introduces more uncertainty. Responding to a need to quantify the impact of global boundary conditions on regional scale air quality models (e.g., Schere et al., 2012; Solazzo, Bianconi, et al., 2017, Solazzo, Hogrefe, et al., 2017), Hogrefe et al. (2018) found a substantial difference in CMAQ performance for a limited-area conterminous United States (CONUS) domain when it was driven by chemical boundaries from four different global/hemispheric models. To improve consistency, CONUS simulations with CMAQ have begun to use boundary conditions from CMAQ run on hemispheric-scale domains (Mathur et al., 2017; Xing et al., 2014). However, a global modeling capability with grid refinement would eliminate the need for any lateral boundary conditions.

Global weather modeling has become more standard as technological and scientific advances move the science away from limited area regional modeling. For example, the Model for Prediction Across Scales–Atmosphere (Skamarock et al., 2012; MPAS-A, or MPAS hereafter) is a global complement to WRF. The U.S. EPA has a vision of global-to-local-scale air quality modeling using MPAS as the meteorological driver. The MPAS mesh refinement capability allows for coarser grid cells over large parts of the globe with gradual refinement over a region that is the key focus area like the CONUS in the case of the U.S. EPA. Progress has been made on a coupled MPAS-AQ modeling system (Pleim et al., 2018; Wong et al., 2020) where CMAQ is called inside the meteorological model driver like a weather-based physics parameterization. The CMAQ components are a one-dimensional chemistry column model with all transport performed by the MPAS mass-conservative advection routines.

As MPAS-AQ is developed and refined, preparing MPAS for retrospective meteorological modeling has followed a parallel track. Initially, the WRF analysis-nudging-based FDDA module was implemented in MPAS (Bullock et al., 2018). In this context, FDDA allows for continuous retrospective simulations where errors are mitigated by constraining the model toward meteorological analyses. The only alternative for retrospective modeling applications aside from 4DVAR, which is computationally prohibitive for most air quality applications, is running short-term forecasts with frequent reinitialization. The FDDA method analysis nudging was first implemented in MM5 by Stauffer and Seaman (1990), Stauffer et al. (1991) and Stauffer and Seaman (1994). Otte (2008a, 2008b) evaluated FDDA in MM5 for a CMAQ air quality application. Subsequently, Gilliam and Pleim (2010) was the first documented study comparing the FDDA configuration described by Otte (2008a, 2008b) of the MM5 model with a comparable configuration of WRF. Other studies such as Otte et al. (2012), Bowden et al. (2012), and Bullock et al. (2014) demonstrated this FDDA method as an effective tool to downscale general circulation models to the regional scale.

The next stage in developing MPAS-AQ seeks to further tailor the meteorology simulations for air quality applications. As was done with the MM5 and WRF (Gilliam & Pleim, 2010), the P-X LSM (Pleim and Xiu, 1995, 2003; Xiu & Pleim, 2001), ACM2 PBL scheme (Pleim, 2007a, 2007b), and Pleim surface layer

Table 1
Model Configuration of Main MPAS and WRF Simulations

	Model run identification				
Settings	WRF12	MPAS12	MPAS25	MPASOG	MPASNOAH
CONUS grid spacing (km)	12	12	25	12	12
Global grid spacing (km)	None	46	92	46	46
Microphysics	Morrison	WSM6	WSM6	WSM6	WSM6
Subgrid Convection	Kain-Fritsch	Kain-Fritsch	Kain-Fritsch	Kain-Fritsch	Kain-Fritsch
Radiation (SW/LW)	RRTMG	RRTMG	RRTMG	RRTMG	RRTMG
Boundary Layer	ACM2	ACM2	ACM2	ACM2	YSU
Land-Surface	Pleim-Xiu	Pleim-Xiu	Pleim-Xiu	Pleim-Xiu	NOAH/MOSAIC
FDDA	OG-NAM12	GFS28	GFS28	GFS28	GFS28
Soil Nudging	OG-NAM12	GFS28/NAM12	GFS28/NAM12	GFS28/OG-NAM12	none
Landuse data set	NLCD40	NLCD40	NLCD40	NLCD40	MODIS20

CONUS, conterminous United States; FDDA, four-dimensional data assimilation; MPAS, Model for Prediction Across Scales; RAOB, Radiosonde observations; WRF, Weather Research and Forecasting model.

scheme (Pleim, 2006) were implemented in MPAS. Concurrently, the Noah tiled land cover characterization parameterization was extended to MPAS (Campbell et al., 2020), where land-surface interactions are computed for each land use category in each grid cell and scaled up to a total grid cell value. To test these implementations, MPAS simulations for 2016 were evaluated against several observation platforms (surface, upper-air, precipitation and radiation) and compared with a similarly configured, limited-area, WRF domain over the CONUS. The goal is to establish confidence in MPAS as the meteorological driver for the global retrospective air quality modeling system.

2. Methods

2.1. MPAS/WRF Domain and Physics Configurations

In MPAS, the grid refinement allows for smooth transition from coarse mesh for most of the globe to finer mesh over a focal region (such as a country or an urban area). The National Center for Atmospheric Research (NCAR), which maintains MPAS, is developing software to construct custom Voronoi meshes. While this has not been officially released, we are successfully testing an unreleased version including the limited-area mesh option in a newer version of the model. For this phase of the project, however, the two pre-defined MPAS meshes used by Campbell et al. (2020) were employed: ~92-km global mesh that is refined to ~25-km over the CONUS, and ~46-km global to ~12-km CONUS mesh.

For the comparative WRF simulation a 12-km CONUS domain used for retrospective modeling at the U.S. EPA for the last decade (Appel et al., 2017; Gilliam & Pleim, 2010; U.S. EPA, 2019) was employed. The physics options and other aspects of the simulations were synchronized as much as possible between MPAS and WRF as documented in Table 1. Base versions of the model codes used for the simulations are MPASv5.2 and WRFv4.0. The 50 MPAS and 35 WRF vertical levels are defined in Figure 1. Some early internal testing of MPAS indicated a general lack of sensitivity to the vertical grid structure on near-surface meteorology and precipitation even prior to the FDDA implementation. FDDA by design tends to suppress differences that vertical grid structure could cause, at least in the state variables that are nudged, so it is not expected that the different vertical level structure between the models would impact results.

Both models use the RRTMG shortwave and longwave radiation schemes (Iacono et al., 2008). The Kain-Fritsch 2 (KF2, Kain, 2004) convective parameterization scheme was used in both, and MPAS was modified to include the same subgrid cloud feedback to the radiation (Alapaty et al., 2012; Herwehe et al., 2014) used in WRF. Scale-aware convective schemes are gaining traction in multiscale modeling systems such as MPAS, but we elected to use the KF2 scheme because it has been a routine setting. Furthermore, initial

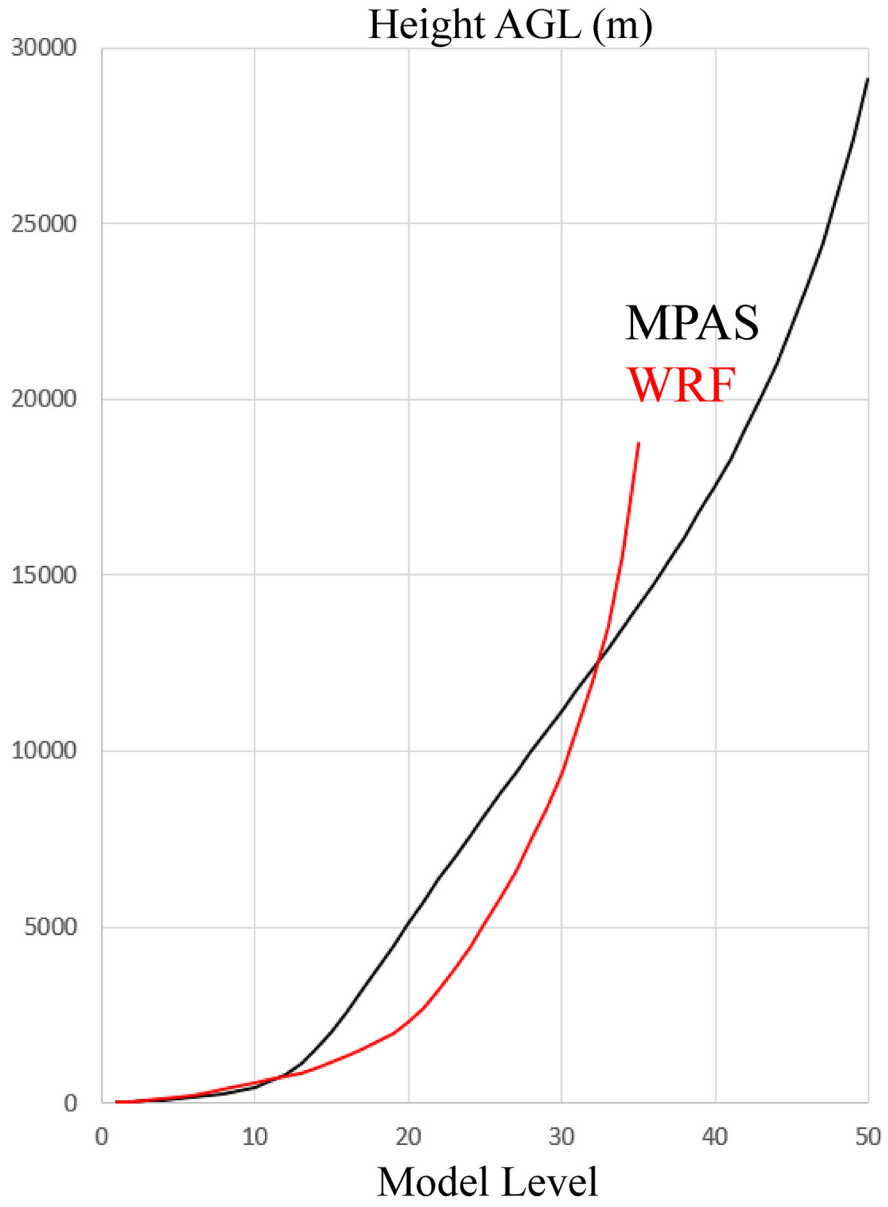

Figure 1. Approximated model vertical level heights (m) at sea level.

testing of convective schemes and microphysics in MPAS (Herwehe et al., 2018) identified the KF2 with Weather Research Forecasting single-moment 6-class microphysics schemes (WSM6) as the combination with the least error and bias in terms of surface meteorology, precipitation and surface shortwave radiation.

The most notable difference in the physics choices between MPAS and WRF was the choice of microphysics. The Morrison double-moment microphysics (Morrison et al., 2009) has been used for many years in WRF, but it was not available in MPAS, so the single-moment WSM6 (Hong and Lim, 2006) was chosen, again based on Herwehe et al. (2018). We do present a supplemental sensitivity using WRF that details the impact that the Morrison and WSM6 options likely have on model performance.

The P-X LSM, ACM2 PBL, and Pleim surface layer schemes were identical in both modeling systems after the implementation into MPAS. Four main MPAS sensitivities (MPAS12, MPAS25, MPASOG, and MPASNOAH) and one baseline WRF simulation (WRF12) will be the focus of the evaluation (Table 1). MPAS12 and MPAS25 are the same except the mesh is coarser in the MPAS25. MPAS12 and MPASOG are the same except the soil nudging inputs in MPASOG are sourced directly from WRF12 as will be described

in Section 2.3. The MPASNOAH simulation is from Campbell et al. (2020) with physics options: Noah LSM (Chen & Dudhia, 2001) with tiled land use, Yonsei University PBL (YSU; Hong et al., 2006) and standard Monin-Obukhov surface layer similarity schemes. This was included as the most likely alternative LSM/PBL choice available in MPAS.

In this study, MPAS was modified to leverage a land use data set developed by the U.S. EPA for WRF modeling. In this data set (NLCD40), the 20-class, high resolution 2011 National Land Cover Data set (NLCD) was applied over the CONUS, while the global 20-class MODIS land use was used elsewhere. NLCD40 was used for all simulations including supplementary sensitivities.

2.2. FDDA Configuration

For the 2016 annual MPAS simulation, FDDA based on analysis nudging (Bullock et al., 2018) leveraged analyses from the National Centers for Environmental Prediction (NCEP) Global Forecast System's (GFS) Data Assimilation System (GDAS). GDAS starts with a 6-h forecast and blends observations using three-dimensional variational data assimilation (3DVar) with Gridpoint Statistical Interpolation (GSI) to produce an analysis where most observations fit within specified error bounds. GSI essentially provides a statistically optimized description of the atmosphere on a specified grid (Wu et al., 2002). As Table 1 outlines, the nudging/FDDA of all MPAS simulations (MPAS12, MPAS25, MPASOG, and MPASNOAH) used the ~28-km GSI/GFS-base analyses simply labeled GFS28. Horizontal wind, temperature, and water vapor mixing ratio were nudged above the PBL with nudging coefficients of 1.0×10^{-14} s^{-1}, 1.0×10^{-14} s^{-1}, and 1.0×10^{-15} s^{-1}, respectively, for all MPAS and WRF simulations (Spero et al., 2018). These nudging coefficients have been standard in almost all U.S. EPA modeling of the last decade. MPAS simulations used 6-h GFS28 analyses.

Table 1 indicates that WRF12 was nudging toward 3-h NAM-based 12-km analyses (NAM12) using the label OG-NAM12. A WRF pre-processor tool *Obsgrid* (OG) was leveraged as typically done for U.S. EPA simulations. *Obsgrid* ingests an analysis product like NAM12 that has been interpolated to the WRF grid and reintroduces surface and upper-air observations resulting in the reanalysis OG-NAM12. This *Obsgrid* utility is not available in MPAS, so the lack thereof will have some impact on the evaluation of upper-air meteorology in Section 3.4. In order to elucidate this impact, a supplementary sensitivity was done within WRF where the GFS28 and OG-NAM12 FDDA inputs and effect on model performance is detailed (Figure S4).

2.3. Indirect Soil Nudging and Snow Cover

The most powerful aspect of the P-X LSM for retrospective air quality modeling is the indirect soil moisture and temperature nudging (Pleim & Gilliam, 2009; Pleim & Xiu, 2003). Gridded analyses of 2-m temperature and moisture are compared to the model-estimated temperature and moisture at each time step for the indirect soil nudging calculations. This nudging alters the soil moisture and temperature, which repartitions surface fluxes and adjusts the near-surface meteorology toward the guiding analyses.

Table 1 outlines the soil nudging settings of all simulations. Since the GFS analysis was resolved at ~28-km and the smallest MPAS polygon used was ~12-km, the soil nudging target files for MPAS used a process that blends two analyses: a coarse analysis was used on the global MPAS grid (GFS28), while a higher-resolution regional analysis (NAM12) was applied in the focal area that resulted in merged GFS28/NAM12 global analysis. Both MPAS12 and MPAS25 used the same GFS28/NAM12 soil nudging inputs to the P-X LSM. For the WRF12 simulation, Table 1 shows an analysis OG-NAM12 was used, which is the surface nudging file from *Obsgrid* that blended surface observations with the NAM12 analysis on the WRF12 grid.

Gilliam and Pleim (2010) stressed that more refined soil nudging inputs via *Obsgrid*, or any other analysis product for that matter, will improve the simulated 2-m temperature. As a result, the difference in soil nudging inputs between WRF and MPAS will have some impact on results. We were able to test the impact of *Obsgrid* on simulated near-surface meteorology in MPAS that also demonstrates the effectiveness of the PX LSM soil nudging scheme. This MPASOG simulation (Table 1) was achieved by producing soil nudging inputs using nearest-neighbor interpolation of the WRF12 inputs from *Obsgrid* (OG-NAM12) to the MPAS12 grid points that fell within the WRF12 domain.

P-X LSM does not explicitly simulate snow cover. Rather, snow cover was specified by using an analysis of snow depth that typically accompanies the indirect soil nudging fields. Other LSMs like Noah track snow processes with physical formulations that consider accumulation, melting, sublimation, etc. To accommodate the P-X LSM soil nudging and snow cover, MPAS input processing for time-varying sea surface temperature was replicated to produce a time-varying input file with 2-m temperature, 2-m moisture, and snow depth. For the MPAS12, MPAS25, MPASOG simulations, the GFS28 analysis was the source of snow cover. For the WRF12 simulation and all supplementary sensitivities the snow cover was extracted from the NAM12 analysis. A visual comparison of the snow cover fields did not uncover differences we expect would impact the model performance statistics.

2.4. Model Evaluation

The Atmospheric Model Evaluation Tool (AMET; Appel et al., 2011; Gilliam et al., 2005) has been used to evaluate meteorology models used in air quality simulations at the U.S. EPA. AMET was originally developed to provide routine evaluations of the MM5, then adapted for WRF and, more recently, for the MPAS model. The tool uses observations from the Meteorological Assimilation and Data Ingest System (MADIS) which archives these meteorological data for the U.S. and more recently across the globe. The Parameter-elevation Relationships on Independent Slopes Model (Daly et al., 2008; PRISM Climate Group, 2015) data set was used by AMET to evaluate the model precipitation. The bilinearly interpolation option in AMET was used to pair MADIS surface data with the models. Upper-air soundings (i.e., Radiosonde observations (RAOB)) and precipitation were paired using nearest neighbor. For this investigation, we used AMET to examine hourly surface meteorology, hourly global downward shortwave radiation at the surface, seasonal precipitation, and upper-air meteorology using the standard twice-daily RAOB soundings.

3. Results

3.1. Surface Meteorology

Four key meteorological variables are examined in this analysis (Table 2): 2-m temperature (T2), 2-m water vapor mixing ratio (Q2), 10-m wind speed (WS10) and direction (WD10). Since WRF12 uses a limited-area CONUS 12-km domain, the MPAS12, MPAS25, MPASOG and MPASNOAH errors are computed using only observation sites within the WRF12 domain. Furthermore, the statistics shown here are limited to root-mean squared error (RMSE) or mean absolute error (MAE) because bias can be misleading when high bias in some areas, or times of the day, can offset low bias as an example.

In our experience, T2 is a primary indicator of model performance. Table 2 shows that the monthly T2 RMSE of the benchmark WRF12 aligns with accepted error levels in retrospective modeling (U.S. EPA, 2019). WRF12 domain averaged errors reach a maximum of around 2.35 K during colder months and fall below 2.00 K during warmer months. T2 RMSE by region and season in WRF12 are generally comparable to the U.S. EPA report although the data there was subset by climate region. The two reasons U.S. EPA (2019) was not explicitly used here as a benchmark: It was (1) an older version of WRF (v3.8) and (2) used lightning assimilation that has not been implemented in MPAS, which was shown to significantly improve surface meteorology and precipitation (Heath et al., 2016).

MPAS12 has higher T2 error every month: An increase over WRF12 of ~0.20 K in cooler months to 0.10 K or less during the warmer months. A main configuration difference between MPAS12 and WRF12 is the absence of *Obsgrid* in MPAS as detailed in Section 2.3. The MPASOG sensitivity incrementally improves upon MPAS12 and compares favorably to WRF12. Again, when the same *Obsgrid* inputs from the WRF12 simulation are used in the MPASOG over the CONUS, T2 errors are slightly lower than WRF12 during the six warmer months of the year and only slightly higher the other six months. MPASOG has a consistently lower RMSE of T2 than MPAS12 by 0.10–0.15 K confirming the efficacy of *Obsgrid* and response of the soil nudging scheme to refined inputs.

MPAS25, where the finest resolution over the CONUS is 25-km, has a slightly higher T2 error than MPAS12, an increase of about 0.10 K on average each month. This indicates that the finer resolution of the 12-km mesh offers some limited benefit. The implementation of the P-X LSM and ACM PBL in MPAS12 greatly

Table 2
Monthly Error Statistics for Simulations Identified in 1.

T2 RMSE (K)						Q2 RMSE (g kg^{-1})					
Month	WRF12	MPAS12	MPAS25	MPASOG	MPASNOAH	Month	WRF12	MPAS12	MPAS25	MPASOG	MPASNOAH
1	2.30	2.51	2.59	2.40	3.45	1	0.69	0.74	0.79	0.73	0.75
2	2.28	2.47	2.52	2.33	3.56	2	0.80	0.90	0.90	0.88	0.86
3	2.14	2.28	2.34	2.17	3.00	3	1.02	1.14	1.16	1.11	1.07
4	2.02	2.15	2.22	2.06	2.63	4	1.15	1.28	1.26	1.23	1.28
5	2.05	2.16	2.27	2.08	2.63	5	1.39	1.46	1.43	1.43	1.53
6	2.05	2.11	2.21	2.02	2.78	6	1.73	1.88	1.85	1.84	2.11
7	2.08	2.17	2.24	2.06	2.97	7	1.92	2.09	2.05	2.04	2.43
8	1.96	2.03	2.11	1.93	2.91	8	1.76	1.93	1.88	1.87	2.43
9	1.93	2.00	2.18	1.88	2.68	9	1.53	1.66	1.57	1.65	2.13
10	2.01	2.09	2.17	1.99	2.47	10	1.17	1.21	1.17	1.21	1.46
11	2.11	2.23	3.31	2.10	2.59	11	0.97	0.97	0.97	1.00	1.13
12	2.35	2.55	2.60	2.41	3.15	12	0.79	0.80	0.80	0.79	0.86
WS10 RMSE (m s^{-1})						WD10 MAE (°)					
Month	WRF12	MPAS12	MPAS25	MPASOG	MPASNOAH	Month	WRF12	MPAS12	MPAS25	MPASOG	MPASNOAH
1	1.73	1.75	1.79	1.73	1.87	1	26	26	26	26	26
2	1.80	1.81	1.79	1.80	1.99	2	25	25	27	25	26
3	1.80	1.81	1.82	1.79	1.93	3	27	27	27	27	28
4	1.75	1.76	1.78	1.74	1.86	4	27	27	28	27	28
5	1.69	1.69	1.69	1.68	1.77	5	31	30	31	30	32
6	1.64	1.64	1.66	1.64	1.74	6	32	31	32	31	33
7	1.64	1.65	1.66	1.64	1.75	7	34	33	34	34	35
8	1.56	1.56	1.58	1.56	1.67	8	34	33	34	33	35
9	1.56	1.56	1.63	1.56	1.69	9	30	30	31	30	31
10	1.67	1.66	1.70	1.65	1.78	10	27	26	27	26	27
11	1.65	1.70	1.70	1.65	1.81	11	26	26	27	26	27
12	1.84	1.85	1.87	1.86	2.01	12	25	25	26	25	26

Errors are provided for 2-m temperature (T2), 2-m water vapor mixing ratio (Q2), 10-m wind speed (WS10), and 10-m wind direction (WD10). RMSE is listed for T2, Q2, and WS10, and mean absolute error (MAE) for WD10.
MPAS, Model for Prediction Across Scales; RSME, Root Mean Square Error; WRF, Weather Research and Forecasting model.

improves the verification of T2 compared with MPASNOAH configuration, which have monthly T2 errors that are 0.50–1.00 K higher depending on the time of year. This shows that soil nudging in P-X LSM is an effective option in reducing T2 errors, as was shown in similar comparison of WRF and MM5 in Gilliam et al. (2010). The Flux-adjusting Surface Data Assimilation Scheme (FASDAS; Alapaty et al., 2008) is a relatively new option linked to the NOAH LSM in WRF, but it is unavailable in MPAS. While it is a fundamentally different nudging scheme (direct nudging of the atmosphere) than the P-X LSM indirect soil nudging, FASDAS is sure to improve MPAS performance of retrospective simulations if implemented. While the soil nudging is likely the main attribution of differences in model performance, we do not discount that the PBL and surface layer scheme are different and could have some impact, although not explored.

Figure 2 provides another examination of the T2 RMSE using a daily timeseries for 2016. The MPASNOAH is clearly the least precise model, while all P-X LSM-based simulations with soil nudging are more similar in terms of performance. Again, the primary reason is the lack of surface/soil nudging (e.g., FASDAS) in MPASNOAH as previously indicated. The MPASNOAH is closer to the P-X LSM simulations in the

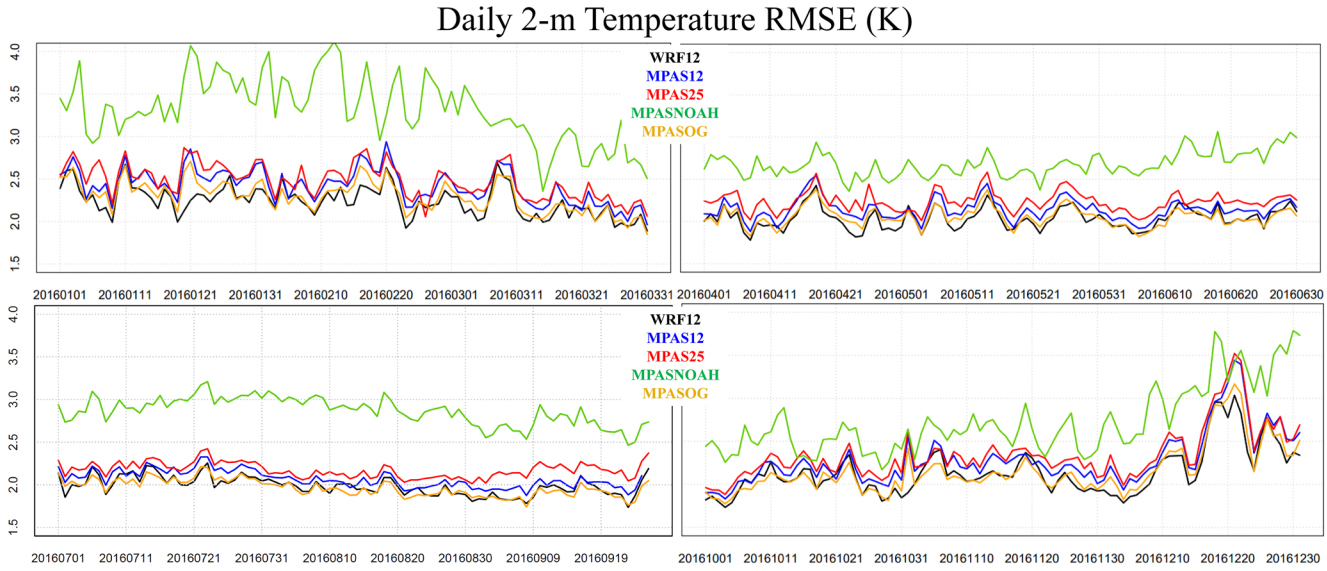

Figure 2. Daily RSME of 2-m temperature (K) for each model simulation using observational data over the CONUS. Each panel represents a three-month period: January–March (top left), April–June (top right), July–September (bottom left), and October–December (bottom right). CONUS, conterminous United States; RSME, Root Mean Square Error.

transition seasons but has notably larger error in the winter and summer. WRF12 and MPAS12 are most similar during warmer months. Using *Obsgrid* inputs (MPASOG) clearly lowers daily RMSE and approaches WRF12 or lower. Starting in May and ending in November, MPASOG is consistently at WRF12 levels or lower in terms of daily error. MPAS12 verifies slightly better than MPAS25 throughout the year except for a few days, again showing a positive response of the modeling to a reduction in grid size.

Q2 is another proxy of overall model performance (Table 2 and Figure 3). As with T2, the WRF12 has lower Q2 RMSE than MPAS12. Differences between all simulations including MPASNOAH are less during some of the cooler months because water vapor mixing ratio is naturally smaller in the winter on average. In the summer, the WRF12 has a lower RMSE by 0.10–0.15 g kg^{-1}, when compared to MPAS12 and those difference decreases slightly when *Obsgrid* inputs are used (MPASOG). The July–September period represents

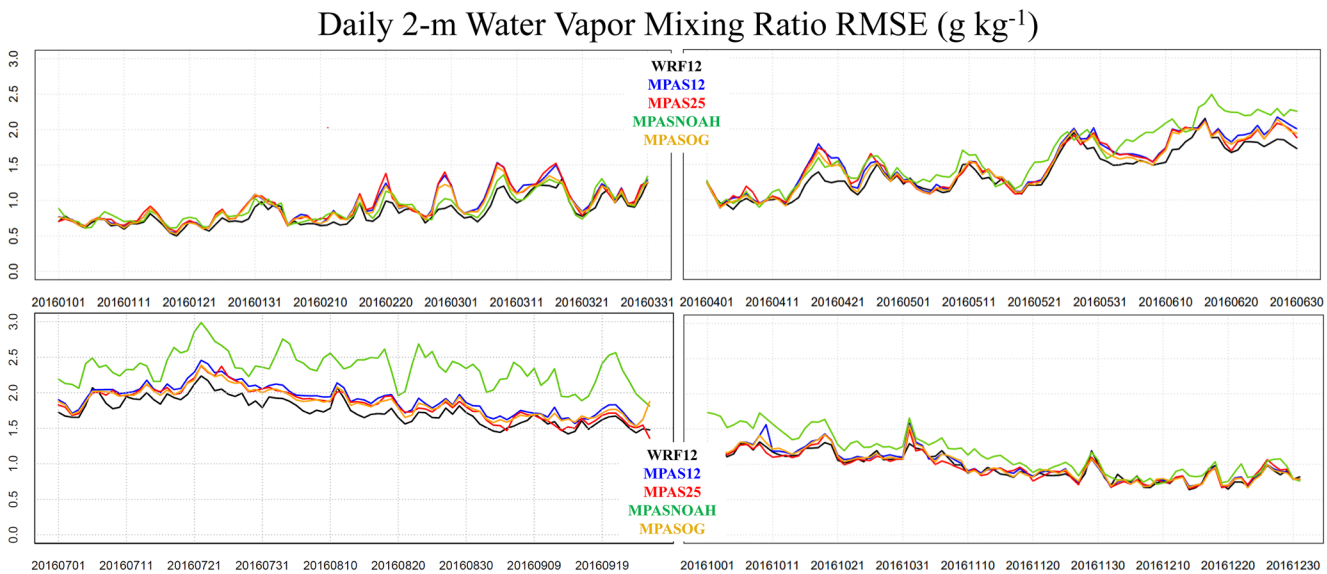

Figure 3. Same as Figure 2, but for 2-m water vapor mixing ratio (g kg^{-1}).

the largest difference in Q2 RMSE between the WRF12 and MPAS12/MPASOG simulations. MPASNOAH is the outlier in the summer months with errors 0.50 g kg^{-1} greater than the other simulations that have the advantage of indirect soil moisture and temperature nudging.

Monthly RMSE of 10-m wind speed (WS10) and MAE of wind direction (WD10) are more comparable between the simulations (Table 2). MAE is used for WD10 because large wind direction deviations in frequent light wind conditions skew the RMSE. Throughout 2016, the difference in RMSE of WS10 between WRF12 and MPAS12 is capped at 0.05 ms^{-1} in November, but essentially the same for the other months. MPASOG essentially has the same level of wind error as WRF12. Reducing MPAS mesh size over the CONUS from 25 to 12 km (MPAS25 vs. MPAS12) negligibly impacted these statistics. MPASNOAH has ~0.10 ms^{-1} higher RMSE than the MPAS P-X configurations. WD10 MAE is nearly identical across all simulations. As one would expect, the WD10 errors increase in the summer as the large-scale flow weakens and surface winds decrease, yielding more instances of light and variable winds.

The spatial distribution of errors in the winter (January–March) and summer months (July–September) of the MPASOG simulation along with the RMSE difference with the WRF12 (center row) and MPAS12 (bottom row) are shown in Figure 4. MPASOG errors are presented in the top row since it most closely matches the soil nudging in WRF12, and as previous analysis showed has the lowest errors of the MPAS configurations. As already established, the MPASOG T2 errors are generally within expected performance levels when all CONUS observations are considered. Error difference plots shows how model performance differs regionally between simulations.

In the winter, MPASOG errors are between 1.5 and 2.5 K outside of areas with more complex topography where errors are around 3.0 K (e.g., Intermountain West). MPASOG performs quite well on the West Coast, southern U.S. and Ohio River Valley with RMSE values between 1.5 and 2.0 K. Difference of RMSE T2 between MPASOG and WRF12 in the winter (center-left panel of Figure 4) reveal that WRF12 performs best in the Intermountain West, northern Plains, Midwest and Canada where WRF12 errors are 0.2–0.5 K lower than MPASOG. MPASOG generally performs best in the southern and eastern parts of the CONUS where errors are 0.1–0.4 K lower than WRF12. Overall MPASOG has lower error at a 42% minority of observation sites.

A supplemental sensitivity of microphysics (Morrison vs. WSM6) indicates that at least in WRF the Morrison choice leads to lower shortwave radiation at the surface in January 2016 as shown in Figure S2. The sun angle is much lower in the north-central U.S. during the winter where WRF12 has the lowest T2 error so this difference in shortwave radiation at the surface in Figure S2 is a larger fraction of the average radiation in northern parts of the U.S. This is also an area of persistent snow cover that interrupts the effectiveness of the soil nudging algorithm in the P-X LSM. Figure S3 is the difference in T2 errors from this supplemental sensitivity (Morrison-WSM6). The T2 error difference pattern in Figure S3 makes it clear that the larger MPASOG T2 error in the north-central U.S. relative to WRF12 is driven by the WSM6 microphysics parametrizations.

It is shown in lower-left panel of Figure 4 that the *Obsgrid* inputs used in MPASOG, rather than the NAM12-based inputs in MPAS12, improve temperature error at 71% of observation sites (bottom left panel) during winter months and 64% of sites during the summer months (bottom right panel). MPASOG has errors that are 0.1–0.4 K lower with some scatter sites indicating errors are as much as 1.0 K lower with the improved soil nudging inputs.

During summer, the MPASOG T2 RMSE is below 2 K outside of the complex terrain of the Intermountain West where errors range from 2.0 to 3.0 K. In this simulation, the number of sites with lower T2 error relative to WRF12 increases from 42% in winter to 65% in the summer. Comparing the spatial difference of MPASOG and WRF12 (center right panel of Figure 4) indicates where the *Obsgrid* inputs had the most impact on T2. Improvements in T2 error are most evident in the central and eastern U.S. It may be coincidence, but these areas generally align with the wettest parts of the eastern U.S. as discussed next.

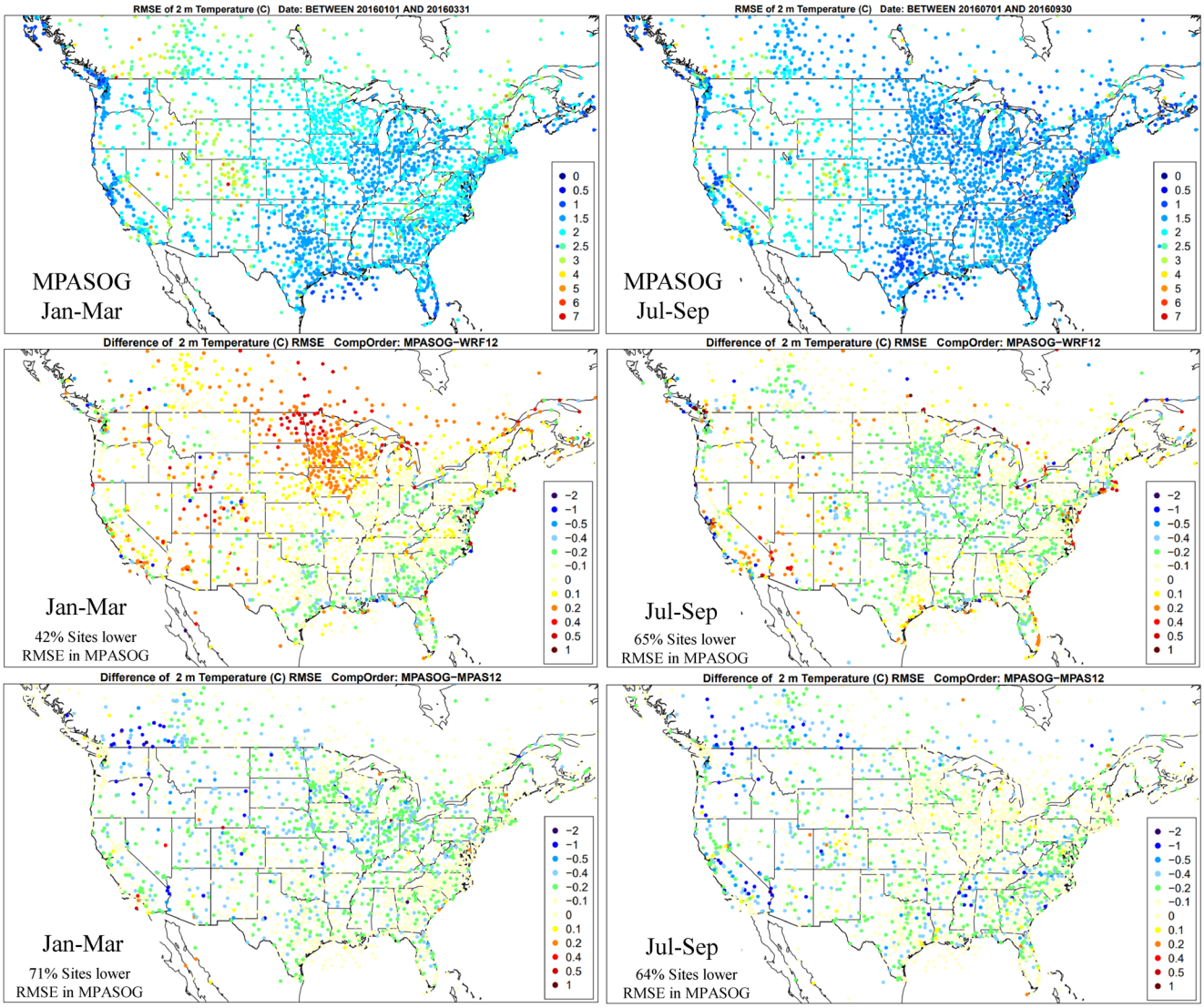

Figure 4. RMSE of 2-m temperature for January–March (top left) and July–September (top right) for the MPASOG simulation. Error difference with comparative simulations are depicted in the two rows below. MPASOG-WRF12 for January–March and July–September (center row). MPASOG-MPAS12 for January–March and July–September (bottom row). MPAS, Model for Prediction Across Scales; RSME, Root Mean Square Error; WRF, Weather Research and Forecasting model.

3.2. Precipitation

Precipitation is another common and important component of an evaluation, especially when migrating to a new model. Poorly simulated precipitation signals that the model has an underlying deficiency. If the precipitation is well characterized, confidence in that model and downstream environmental modeling systems like CMAQ increases because precipitation has historically been a source of large uncertainties.

Figure 5 presents the seasonal observed PRISM precipitation and the model totals of MPAS12 and WRF12. Both models represent the precipitation patterns well for the January–March 2016 quarter of the year. The spatial distribution and accumulations over the wetter parts of the western U.S. where orographic forcing dominates are captured by both models. The highest accumulations in the southeastern U.S. are underestimated by the models but amounts in other areas of the eastern U.S. are comparable to PRISM. Figure 6 and Table 3 provide additional metrics on the model performance with spatial normalized differences (by PRISM), MAE, bias, and correlation based on land-based grid cells in the CONUS. For January–March (Table 3), MPAS12 has a MAE of 40 mm, WRF12 a MAE of 47 mm, and MPAS25 43 mm. MPAS12 has a bias of

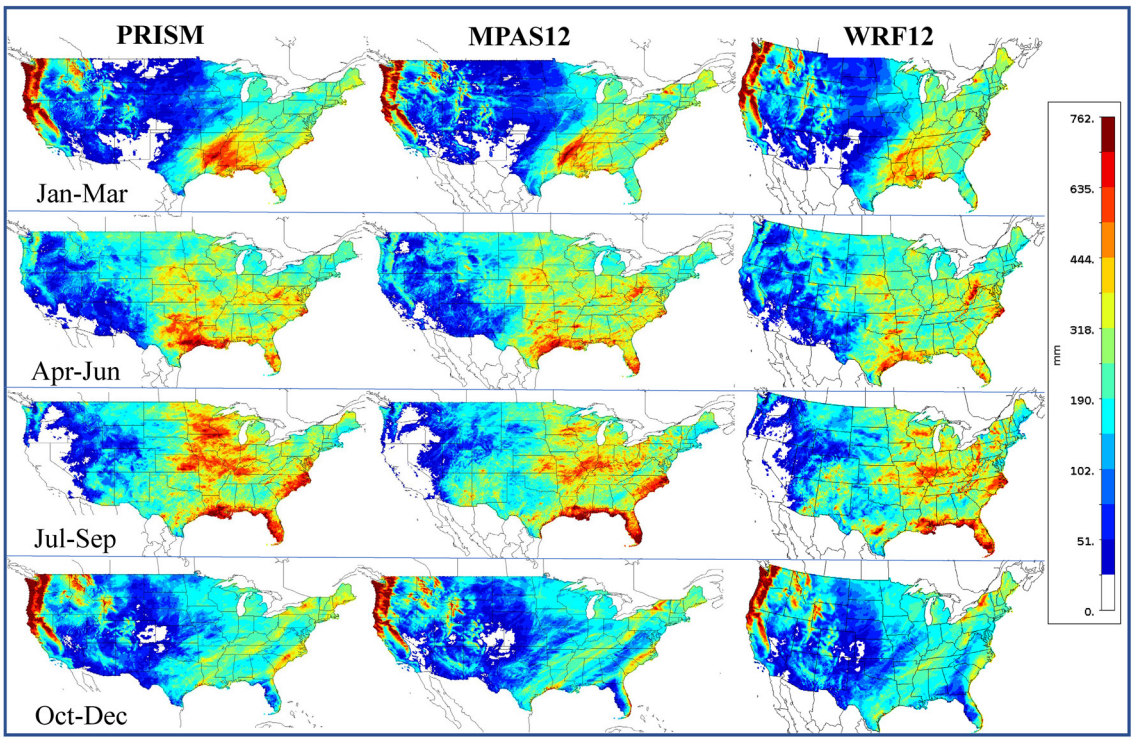

Figure 5. Seasonal precipitation totals (mm) from the observation-based PRISM data (left), MPAS12 (center) and WRF12 (right). MPAS, Model for Prediction Across Scales; PRISM, Parameter-elevation Relationships on Independent Slopes Model; WRF, Weather Research and Forecasting model.

−4 mm, WRF12 + 5 mm and MPAS25 -27 mm. Spatial correlation of this winter period is about the same in all configurations where the correlations greater than 0.90 indicate the models represent precipitation patterns with coherency.

The normalized precipitation bias is presented in Figure 6 as the seasonal difference between the models and PRISM divided by the total PRISM precipitation. This metric needs to be interpreted with caution in dry areas where small differences between the models and observations and/or low observed precipitation can drive high normalized bias. A masked was applied where seasonal PRISM totals are less than 50 mm to eliminate more arid regions based on the upper range suggested by Laity (2009). In areas of higher observed precipitation over the January–March period, the normalized precipitation biases are generally ±20%. Comparing MPAS12 and WRF12 corroborates Table 3, as MPAS12 has a slightly lower bias. One area of the U.S. where MPAS12 clearly has less bias is the intermountain western and north-central U.S.

Over the April–June period, MPAS12 and WRF12 compare well with PRISM. Western U.S. precipitation is well simulated, though rainfall is lower than in winter. High springtime precipitation along the Gulf Coast is represented by both models, as are locally wet areas in West Virginia, coastal North Carolina and Florida. Table 3 shows MPAS12 performs best with the MAE of 51 mm as compared to 58 mm in WRF12. While the correlation was the same between models in the winter, the correlation of MPAS12 in spring is 0.85 while it was 0.80 in WRF12. The grid bias is −20 mm in WRF12, -3 mm in MPAS12 and -27 mm in MPAS25. This bias difference between MPAS12 and WRF12 is evident in the normalized bias in Figure 6. For example, areas that have a negative bias in WRF12 are more biased than in MPAS12 (e.g., Texas).

July–September is the Northern Hemisphere summer when much of the southern and eastern U.S. experiences a more subtropical climate, characterized by high near-surface temperatures, water vapor mixing ratio and frequent diurnally driven convection. This weather regime is more difficult to simulate than precipitation driven by larger-scale mid-latitude weather systems that dominate in winter and transition seasons. The two models generally represent the precipitation well, from minimal precipitation in California to over 600 mm along the coast of the southeastern quadrant of the U.S (Figure 5). Higher observed precipitation

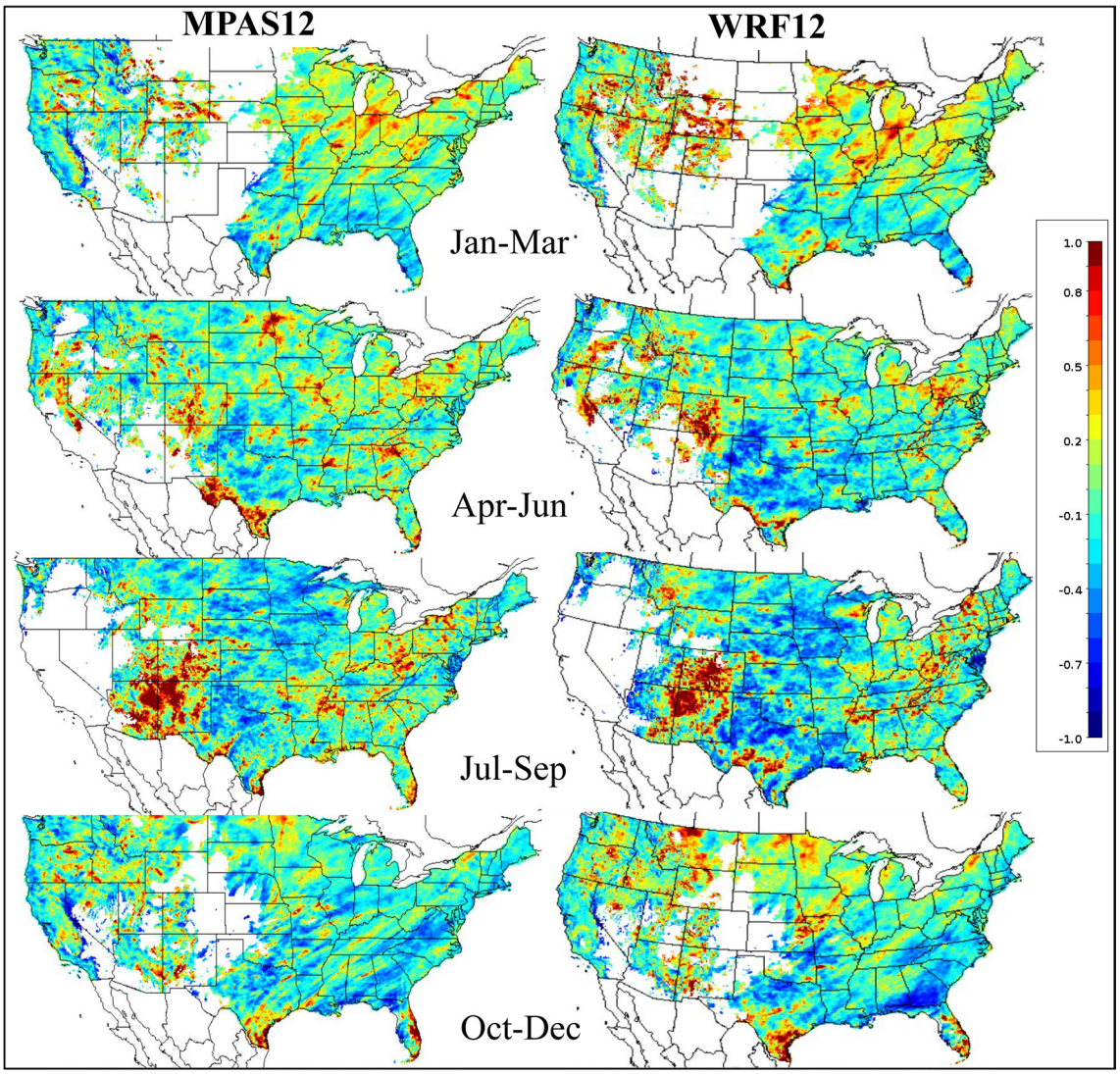

Figure 6. Normalized model bias of MPAS12 (left) and WRF12 (right) seasonal precipitation using observation-based PRISM. Mask applied (white) where PRISM totals are less than 50 mm. MPAS, Model for Prediction Across Scales; PRISM, Parameter-elevation Relationships on Independent Slopes Model; WRF, Weather Research and Forecasting model.

in the U.S. states of Iowa, Minnesota, and Wisconsin is clearly underestimated by the models. This broad area with greater than 500 mm precipitation is underestimated by the models that simulated around 200–350 mm. The normalized precipitation bias (Figure 6) indicates that MPAS12 performs somewhat better with respect to PRISM. Table 3 statistics reveal MPAS12 has a MAE of 69 mm, while WRF12 has a MAE of 77 mm, and MPAS25 has a MAE of 81 mm. The dry bias of MPAS12 is half of that in WRF12 at −15 versus −29 mm, supporting the differences in the normalized bias map (Figure 6). The MPAS12 correlation is higher at 0.82 compared to 0.78 in WRF12 and MPAS25.

Over the last three months of 2016, the models capture the orographically forced precipitation well in the western U.S. as done for the winter period. In October 2016, Hurricane Matthew passed along the southeastern coast of the U.S., depicted by the streak of high precipitation in PRISM (Figure 5). MPAS12 simulated the amount and location of precipitation from Matthew with better detail than WRF12. Both models underestimate precipitation from coastal areas of the Gulf of Mexico to Virginia. The two models simulate around 50–100 mm of precipitation, while more than double that amount was observed (150–200 mm). Normalized precipitation difference (Figure 6) shows that MPAS12 has a generally lower bias across the domain. Table 3

Table 3
Seasonal and Annual Precipitation Statistics for the MPAS12, MPAS25 and WRF12 Simulations Based on the Observation-Based PRISM Data

	Grid mean absolute error (mm)		
Months	WRF12	MPAS12	MPAS25
JFM	47	40	43
AMJ	58	51	56
JAS	77	69	81
OND	46	43	51
Annual	161	137	175
	Grid mean error/Bias (mm)		
Months	WRF12	MPAS12	MPAS25
JFM	5	−4	−27
AMJ	−20	−3	−27
JAS	−29	−15	−24
OND	−14	−21	−39
Annual	−57	−42	−116
	Grid correlation		
Months	WRF12	MPAS12	MPAS25
JFM	0.93	0.93	0.93
AMJ	0.80	0.85	0.84
JAS	0.78	0.82	0.78
OND	0.91	0.92	0.92
Annual	0.89	0.92	0.90

MPAS, Model for Prediction Across Scales; PRISM, Parameter-elevation Relationships on Independent Slopes Model; WRF, Weather Research and Forecasting model.

shows the MAE of the MPAS12 is slightly lower than WRF12 (43 vs. 46 mm, respectively), but over these three months unlike the rest of the year, the grid bias is slightly higher in MPAS12.

Annual statistics (Table 3) provide a higher-level summary of the model performance. Mean absolute error is 137 mm in MPAS12, 161 mm in WRF12, and 175 mm in MPAS25. Bias in MPAS12 is slightly lower, and correlation is slightly higher than WRF12. Reducing the CONUS grid mesh in MPAS from 25 to 12 km improves the annual precipitation as MAE drops from 175 to 137 mm and bias from −116 to −42 mm.

The attribution of this improve precipitation statistics in MPAS is not clear. The microphysics (WSM6 and Morrison) and FDDA drivers (GFS28 vs. OG-NAM12) are two potential differences that may impact precipitation. Table S1 leverages two supplemental sensitivities that test these options within the WRF model. Neither of these differences lead to much contrast in the verification of precipitation. The use of WSM6 and Morrison in WRF has almost no impact on the grid MAE, bias or correlation. The use of GFS28 analyses in FDDA has a slight impact in July, but at least in WRF make the precipitation statistics slightly worse than when the OG-NAM12 drive the FDDA. We have yet to test vertical layering structure, but MPAS does have 50 vertical layers up to 30-km and resolves the upper troposphere and tropopause better as show in Figure 1. Another untestable difference is global nature of MPAS with no lateral boundaries unlike WRF. In any case, these statistics make a clear case that MPAS with these new FDDA and physics options has parity or better with the long-used WRF model in terms of representing historical precipitation totals.

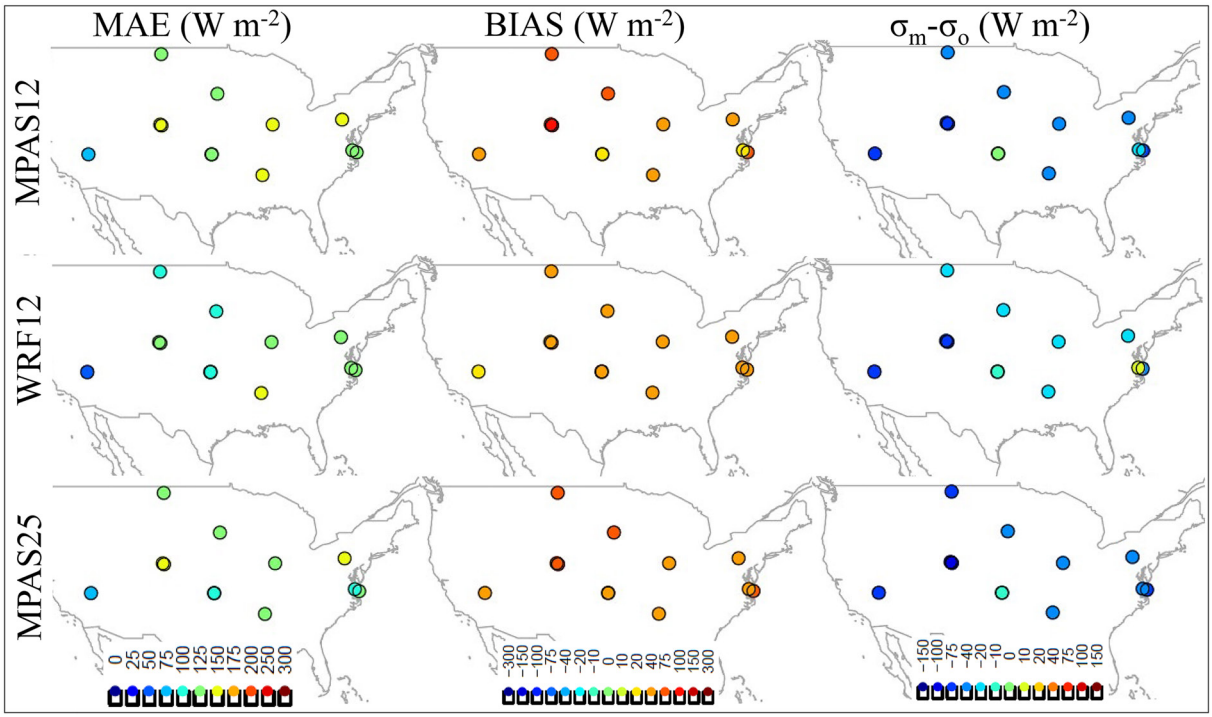

Figure 7. Spatial plots of mean absolute error (MAE), mean error (BIAS), and standard deviation difference (σ_m-σ_o) between the models and the observed downward global shortwave radiation at the surface. Statistics are valid for 2016 and presented for the MPAS12 (top), WRF12 (middle) and MPAS25 (bottom) simulations. MPAS, Model for Prediction Across Scales; WRF, Weather Research and Forecasting model.

3.3. Radiation

Global downward shortwave radiation at the surface is compared against the Baseline Surface Radiation Network (BSRN; Ohmura et al., 1998). The observed BSRN data are acquired and released at a 1 min average but compared to the models using time window averaging (±10 min) centered at the model output time step. Observations from BSRN sites across the CONUS are used to compute error, bias, and variability of the radiation for 2016.

Figures 7 and 8 are provided to show simulated shortwave radiation statistics with spatial and diurnal plots. Statistics presented are error, bias, and variability differences (σ_m-σ_o), where σ_m is the standard deviation of the model timeseries, and σ_o represents the BSRN measurements. WRF12 has the lowest MAE (Figure 7). The two MPAS simulations are similar, but the MPAS12 has a slightly higher error than MPAS25. Diurnal plots in Figure 7 show these differences with more distinction. Individual site statistics are provided along with the average hourly statistics of each model based on the 10 sites in the CONUS. Statistics at each site are adjusted by centering around local solar noon since solar angle has an inherent impact on the metrics. Diurnal MPAS12 errors quickly increase beyond those in MPAS25 and WRF12 until well into the afternoon when the two MPAS simulations are comparable and become slightly lower than WRF12 errors.

Bias of simulated radiation generally follows the performance rankings of the models in terms of error levels. There are several sites where spatial bias indicates MPAS12 is lower than MPAS25, but MPAS12 is generally the poorest performing model. The diurnal bias (Figure 7) shows a positive bias throughout the day in all models, prompting several supplemental sensitivities. The systematic positive model bias in all simulations signifies too few simulated clouds or an optical thickness quality that is not attenuating enough solar radiation. MPAS12 and MPAS25 also have a higher bias early in the day but slightly lower bias compared with WRF12 in the afternoon. This may suggest a difference in how diurnally driven clouds are represented.

Two key differences in MPAS and WRF runs may explain the generally higher bias in MPAS. (1) MPAS12 and WRF12 are driven by FDDA using different analyses (GFS28 vs. OG-NAM12) where the moisture nudging, for example, can have a large impact on clouds (Spero et al., 2014, 2018). 2) The differing microphysics

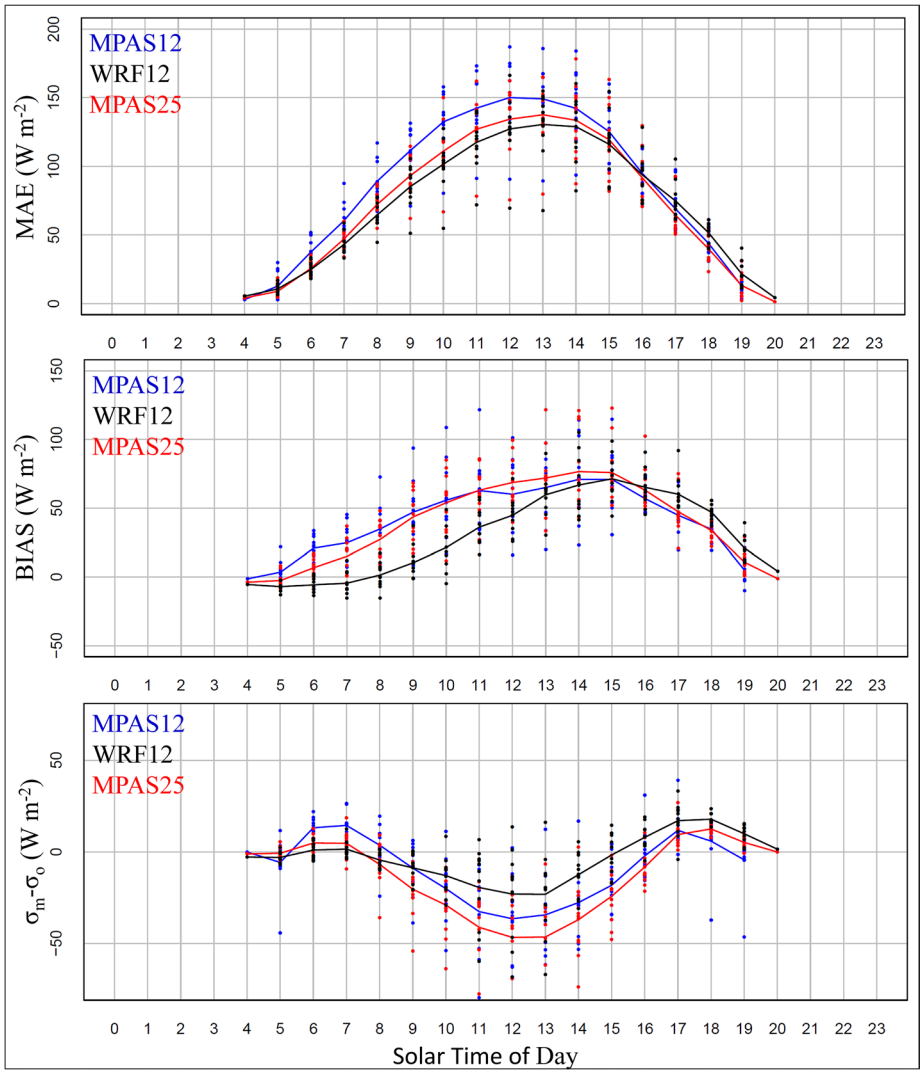

Figure 8. Diurnal shortwave radiation statistics for the MPAS12 (blue), MPAS25 (red), and WRF12 (black) simulations based on 10 observation sites in the CONUS (dots) identified in Figure 7. Average diurnal error, bias and variability difference of each model are plotted (lines). CONUS, conterminous United States; MPAS, Model for Prediction Across Scales; WRF, Weather Research and Forecasting model.

schemes where, again, WRF12 used Morrison and MPAS12 used the WSM6 scheme. The T2 error analysis (Figure S3) identified the choice of microphysics as a main reason for differences in the T2 errors in the winter. Figure S1 shows that averaged downward shortwave radiation at the surface is constantly lower across the CONUS in the WRF12-Morrison simulation than the MPAS12-WSM6. The supplementary sensitivity of the Morrison and WSM6 microphysics in WRF attributes the lower shortwave radiation to a more attenuative Morrison scheme (Figure S2) and that likely extends to MPAS. Figure S4 confirms that the use of WSM6 results in higher shortwave radiation bias at the surface than the Morrison scheme during the first half of the day. This impact on radiation is consistent with Morrison et al. (2009) that found the main difference between the double and single moment schemes was more stratiform precipitation in Morrison. The use of the GFS28 analyses in FDDA has a minor impact, but not as much as the microphysics. Again, Herwehe et al. (2018) tested the microphysics and subgrid convective options available in MPAS. The MPAS12 configuration with WSM6 and KF2 had a high bias but was the least biased of the available microphysics and subgrid convection combinations.

The standard deviation of the radiation timeseries conveys the variability. The difference between the model and observed variability (σ_m–σ_o) indicates a model tendency to over or underestimate variability caused by

clouds. Both models always underrepresent the variability of the observed solar radiation (Figure 7) because 12-km grid models cannot completely capture the shading from small (sub-grid) clouds. In air quality modeling this deficiency will affect the photochemistry, so these results should be considered in ozone predictions, for example. Figure 8 indicates, as did the MAE, the WRF12 simulation has a smaller difference in variability during the middle parts of the day as compared to MPAS12 and MPAS25. The scale reduction in MPAS (MPAS12 vs. MPAS25) appears to increase the variability of radiation except early in the day. As with bias (Figure S4), the variability difference between the WRF and MPAS simulations is largely explained by the different representation in the microphysics schemes. The WSM6 microphysics results in fewer clouds, less attenuation of shortwave radiation and less variability at the surface when compared to observations. We expect a Morrison microphysics option in MPAS would improve the simulation in these respects.

3.4. Upper-Air

The evaluation of the simulated upper-air meteorology uses twice-daily rawinsonde soundings of the tropospheric temperature (TEMP), relative humidity (RH) and wind speed/direction (WS and WD). Spatial analysis of error (Figure 9), error timeseries (Figure 10), and vertical error profiles (Figure 11) are leveraged to understand how these models perform. Figure 9, more specifically, provides the layer average error (1,000–200 hPa) of MPAS12 (CONUS area only) and WRF12. Tropospheric TEMP is well-simulated in most locations, with RMSE of 0.75–1.25 K. MPAS12 has slightly higher error than WRF12, but TEMP errors near 1.0 K are close to the average difference between analyses and observations (e.g., the best that can be expected from a nudged model). Results imply that both MPAS12 and WRF12 approach the level of TEMP error in typical weather analyses. This study expands upon Bullock et al. (2018), only conducted for a winter and a summer month, by showing that MPAS can now be ran continuously with FDDA with no error growth at any level of the atmosphere.

Figure 10 condenses the CONUS-only spatial statistics in Figure 9 to a timeseries comparison of MPAS12 and WRF12 over the year. The statistics are valid for the 1,000–200 hPa layer that roughly represents the troposphere. Focusing on TEMP, the RMSE of the two model runs track each other over the annual cycle. RMSE peaks around 1.25 K during the cold parts of the year when the atmosphere over the CONUS has large temperature variations because of the active synoptic weather patterns. During the warm season, RMSE is minimized at ~0.8 K in WRF12 and just below 1.0 K in MPAS12, which has a consistently larger TEMP error than WRF12.

The lower error in WRF12 may be explained by the difference in nudging inputs. Again, WRF12 used the regional OG-NAM12 analysis every three hours that was further refined on the WRF grid using *Obsgrid* to reintroduce RAOB soundings. MPAS12 used inputs directly from the coarser GFS28 analysis at an interval of six hours, where RAOB soundings were not blended because *Obsgrid* was unavailable in MPAS. Thus, WRF12 had three distinct advantages: 12-km NAM analyses, *Obsgrid*, and 3-hourly FDDA inputs. A supplementary simulation was conducted where WRF FDDA was driven by the same GFS28 as used in MPAS12. Figure S5 shows the results of this sensitivity that indicate OG-NAM12 FDDA inputs lead to lower errors in the upper-air meteorology, consistent with the difference between MPAS12 and WRF12. Nevertheless, MPAS12 error levels provide confidence that tropospheric temperature (TEMP) can be well-simulated within about 1 K of the observations on average.

We concede that the evaluation is not fully independent because the soundings were used in both the evaluation and in the analyses (NAM and GFS) used in FDDA. However, Gilliam et al. (2012) explored the use of three different wind profile datasets in *Obsgrid* and found that when one set was withheld for independent evaluation and the other two included, wind errors aloft relative to the withheld data set were reduced. It is expected that this finding would hold here meaning that blending of RAOB via *Obsgrid* more generally improves the upper-air simulated meteorology, especially considering the even and relatively close spacing of RAOB across the CONUS.

Examining temperature and relative humidity as a function of height also shows the same slightly higher tropospheric temperature error in MPAS12 as the spatial and timeseries analysis and supplementary sensitivity, but illustrates how these errors are distributed throughout the troposphere (Figure 11). Expectedly, because FDDA only operates above the PBL in the free troposphere, both models have higher error near the

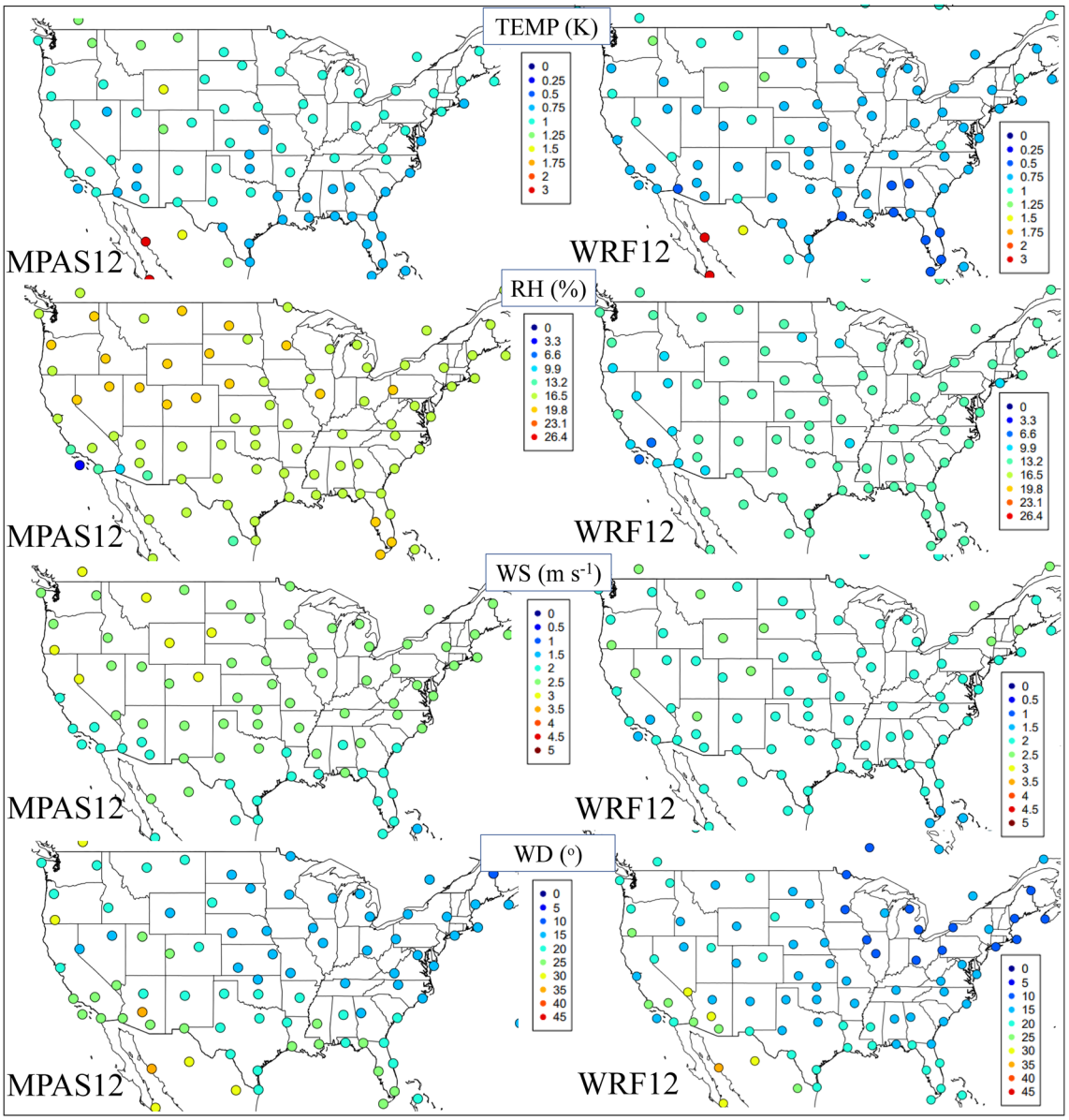

Figure 9. Spatial RMSE of tropospheric (1,000–200 hPa) temperature (TEMP), relative humidity (RH), wind speed (WS), and wind direction (WD) for the MPAS12 (left) and WRF12 (right) simulations, valid for 1 January through 31 December 2016. MPAS, Model for Prediction Across Scales; RSME, Root Mean Square Error; WRF, Weather Research and Forecasting model.

surface and near the tropopause. The models have a well-simulated TEMP RMSE of around 1.25 K near the surface with WRF12 slightly lower. In the middle troposphere, WRF12 has an RMSE that is as low as 0.50 K and MPAS12 around 0.70 K. These errors rise to ~1.50 K around the tropopause and lower stratosphere. The interface between the troposphere and stratosphere is characterized by extreme temperature gradients, so any small error in tropopause height will cause larger errors in TEMP, like near the land-surface where large gradients exist.

The major model performance difference between MPAS12 and WRF12 initially appears to be the RH. MPAS12 has larger RH errors over the CONUS compared to WRF12 as indicated in Figures 9–11. Errors are 10%–15% in WRF12 and 15%–23% in MPAS12 (Figure 9). MPAS12 consistently has a 5% higher RH error throughout the year (Figure 10).

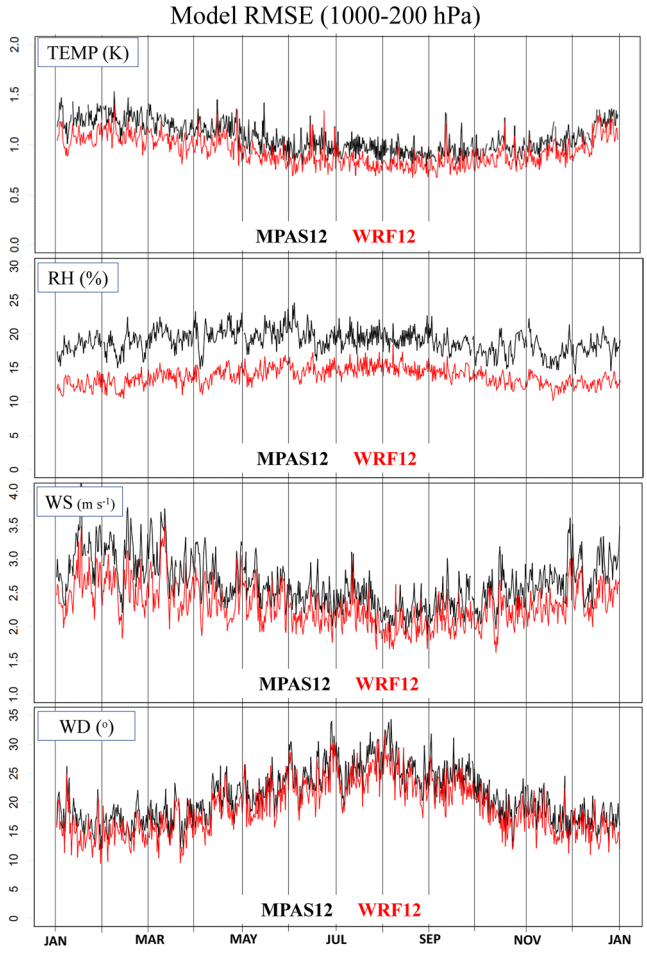

Figure 10. RMSE timeseries of temperature (TEMP), relative humidity (RH), wind speed (WS), and wind direction (WD) based on all RAOB soundings over the CONUS for the 1,000–200 hPa layer of the troposphere. Statistics are provided for the MPAS12 and WRF12 simulations. CONUS, conterminous United States; MPAS, Model for Prediction Across Scales; RAOB, Radiosonde observations; RSME, Root Mean Square Error; WRF, Weather Research and Forecasting model.

Both models verify somewhat similarly in the lower troposphere (Figure 11), but the MPAS12 error peaks at 25% at 300 hPa, while the WRF12 error remains at or below 15%. This significant difference led to a more in-depth investigation into the calculation of relative humidity. In MPAS, the RH is an internal calculation that is available in the model output. WRF does not have this option, so the model evaluation tool used here, as well as other WRF post processing tools, compute saturation vapor pressure from the temperature using Clausius-Clapeyron (C-C calc) and then saturation mixing ratio using the relationship with the vapor pressure. Then, RH as the simple ratio of water vapor mixing ratio and saturation mixing ratio.

The sensitivity of this calculation in the mid-to-upper troposphere to low temperature, pressures and water vapor is evident in a supplementary sensitivity. Figure S6 shows a case where MPAS12 RH was recomputed using the same method as in WRF12 (center profile). The mixing ratio was not available in the original MPAS12 or MPASOG simulations so we could not use that annual simulation. We did have one month (July 2016) from a more recent MPAS simulation configured the same as MPAS12. Using the external C-C calc when computing MPAS errors leads to much different error profiles (MPAS Internal vs. MPAS C-C calc). Furthermore, Figure S6 shows MPAS and WRF are more similar when RH is computed external of the model using the C-C-based calculation. MPAS even has slightly lower errors throughout the profile. It is presumed the internal calculation of MPAS is more representative of grid-scale microphysics and feedback from the subgrid convective scheme, but more investigation is needed. At the very least, any model comparisons should be consistent. We typically look at water vapor mixing ratio since it is independent of temperature, but the MADIS observation data set only has relative humidity, so conversion using the same C-C-based methodology in reverse would still be an issue.

WS and WD errors (Figures 9 and 10) show relatively small differences between MPAS12 and WRF12, but MPAS12 has a slightly higher error. RMSE of wind speed of the MPAS12 simulation is generally 2.0–2.5 ms^{-1} in the summer and 2.5–3.0 ms^{-1} in the winter. This is an acceptable tolerance considering that wind speed averages 30 ms^{-1} at 500 hPa in the winter and 20 ms^{-1} in the summer (Ratner, 1956), and 25–50 ms^{-1} at 300 hPa, depending on season. If normalized by wind speed this is a 5%–10% error. Similar as temperature, the FDDA constrains error in WS to near the levels in analyses. WD errors are ~15° in the winter and ~25° in the weaker wind regimes in summer. MPAS12 tracks with WRF12, but errors are a few degrees higher. As with TEMP, the slightly higher WS and WD errors can be explained by the FDDA inputs (OG-NAM12 vs. GFS28) that was explored in the supplementary sensitivity detailed in Figure S6.

4. Conclusions

Key meteorological variables modeled by MPAS and WRF were compared with observations to determine if MPAS is suitable for retrospective air quality modeling. WRF was used as a benchmark since it is the primary meteorological model used to drive U.S. EPA air quality modeling. The MPAS configuration used here improves upon Bullock et al. (2018) by the implementation of the P-X LSM and its indirect soil nudging scheme, as well as the ACM2 PBL scheme and the Pleim surface layer model. MPAS compares well with WRF over the CONUS for some variables and certain times of the year, but there are some instances where WRF has consistently lower error. Most of the contrast between models is attributed to some options in WRF that are not yet available MPAS. The first is the *Obsgrid* reanalysis tool that enhances the efficacy

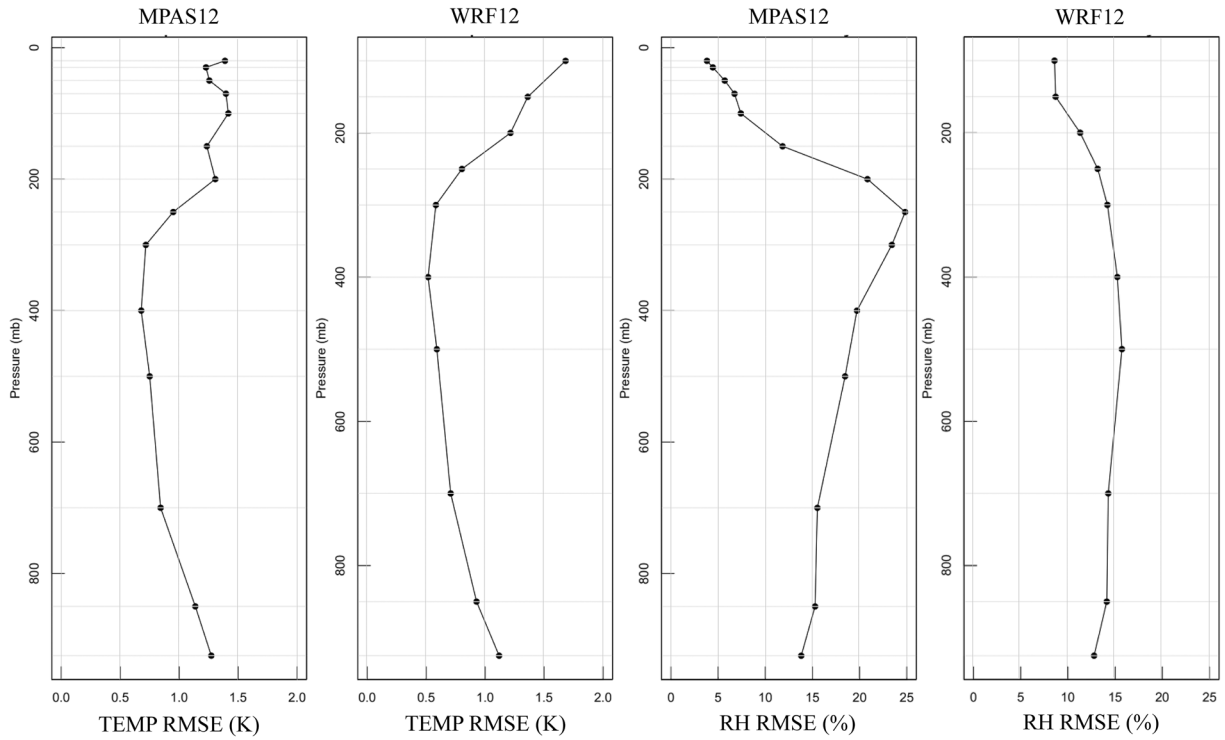

Figure 11. RMSE profiles that represent the same layer average timeseries in Figure 10 for the MPAS12 and WRF12 simulations. Temperature (TEMP; left two panels) and relative humidity (RH; right two panels) RMSE values are based on all RAOB data in the CONUS for January 1, through December 31, 2016. CONUS, conterminous United States; MPAS, Model for Prediction Across Scales; RAOB, Radiosonde observations; RSME, Root Mean Square Error; WRF, Weather Research and Forecasting model.

of soil nudging and FDDA. The second is the Morrison microphysics scheme that improves the simulated clouds, radiation and near-surface temperature. Again, when these are extended to MPAS we expect more incremental improvements in the simulated meteorology.

Even with those differences, MPAS errors are comparable to WRF for several atmospheric fields. Precipitation was well simulated by MPAS with lower error and bias over all seasons when compared to WRF12. Surface meteorology simulated by MPAS had errors close to WRF levels during cooler parts of the year but remained slightly higher. For the warmer half of the year MPAS had temperature errors at or lower than WRF when the same soil nudging inputs were used, but moisture errors remained slightly higher. Wind speed and direction errors were almost identical in both models over the year.

WRF maintained an advantage in surface radiation because of the availability of the Morrison microphysics scheme that verified with lower error, bias and more representative variability. In MPAS, the WSM6 scheme resulted in a larger positive bias relative to observations than WRF has with Morrison. MPAS also had less variability in the radiation with respect to the observations than WRF, but both models underestimated the variability.

Model simulations overall had relatively low error in tropospheric temperature and winds, approaching those of the analyses that drove the FDDA. The slightly larger errors in MPAS originate from that advantage WRF has with the *Obsgrid* reanalysis tool. MPAS seemed to deviate from WRF in the evaluation of relative humidity. MPAS was biased high in the upper troposphere and had errors ~25% throughout the troposphere compared with ~15% in WRF. A more detailed look at the RH calculation indicated the higher error in MPAS was just a difference in the calculations. When both models used the same calculation of RH, the error profiles were similar.

This evaluation and the supplementary sensitivities support using MPAS for retrospective air quality modeling as the level of errors are acceptable for air quality modeling. Microphysics options need to be expanded

in MPAS, such as the Morrison scheme to improve the representation of clouds and impact on radiation. Additionally, the emerging 3DVar methods in MPAS may improve analyses used for FDDA and soil nudging as this modeling matures. MPAS has evolved since the version used here, including porting additional physics schemes from WRF. These may provide opportunities to further improve the retrospective simulations by employing another physics option or improvements of the existing options.

Data Availability Statement

Data used in this analysis are available via the U.S. EPA's Environmental Data set Gateway (https://doi.org/10.23719/1519128).

Acknowledgments
The authors thank Dr. Brian Eder and Chris Misenis for internal technical reviews of this manuscript. Valuable discussions with Bill Skamarock and Michael Duda at NCAR informed some of the code development and analysis. PRISM data were obtained from http://www.prism.oregonstate.edu. Point observation data for AMET were provided by NCEP's MADIS program (https://madis.ncep.noaa.gov). The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

References

Alapaty, K., Herwehe, J. A., Otte, T. L., Nolte, C. G., Bullock, O. R., Mallard, M. S., et al. (2012). Introducing subgrid-scale cloud feedbacks to radiation for regional meteorological and climate modeling. *Geophysical Research Letters*, *39*, L24808. https://doi.org/10.1029/2012gl054031

Alapaty, K., Niyogi, D., Chen, F., Pyle, P., Chandrasekar, A., & Seaman, N. (2008). Development of the flux-adjusting surface data assimilation system for mesoscale models. *Journal of Applied Meteorology and Climatology*, *47*, 2331–2350. https://doi.org/10.1175/2008JAMC1831.1

Appel, K. W., Gilliam, R. C., Davis, N., Zubrow, A., & Howard, S. C. (2011). Overview of the Atmospheric Model Evaluation Tool (AMET) v1.1 for evaluating meteorological and air quality models. *Environmental Modelling and Software*, *26*(4), 434–443. https://doi.org/10.1016/j.envsoft.2010.09.007

Appel, K. W., Napelenok, S. L., Foley, K. M., Pye, H. O. T., Hogrefe, C., Luecken, D. J., et al. (2017). Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1. *Geoscientific Model Development*, *10*, 1703–1732. https://doi.org/10.5194/gmd-10-1703-2017

Appel, W., Gilliam, R., Pleim, J., Pouliot, G., Wong, D., Hogrefe, C., et al. (2014). Improvements to the WRF-CMAQ modeling system for fine-scale air quality simulations. *Equipment Management*, *14*(9), 16–21. Retrieved from http://pubs.awma.org/flip/EM-Sept-2014/appel.pdf

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, *106*, 23073–23096. https://doi.org/10.1029/2001jd000807

Bowden, J. H., Otte, T. L., Nolte, C. G., & Otte, M. J. (2012). Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling. *Journal of Climate*, *25*, 2805–2823. http://dx.doi.org/10.1175/JCLI-D-11-00167.1

Bullock, O. R., Jr, Alapaty, K., Herwehe, J. A., Mallard, M. S., Otte, T. L., Gilliam, R. C., & Nolte, C. G. (2014). An observation-based investigation of nudging in WRF for downscaling surface climate information to 12-km grid spacing. *Journal of Applied Meteorology and Climatology*, *53*, 20–33. http://dx.doi.org/10.1175/JAMC-D-13-030.1

Bullock, O. R., Jr, Foroutan, H., Gilliam, R. C., & Herwehe, J. A. (2018). Adding four-dimensional data assimilation by analysis nudging to the Model for Prediction Across Scales - Atmosphere (version 4.0). *Geoscientific Model Development*, *11*, 2897–2922. https://doi.org/10.5194/gmd-11-2897-2018

Byun, D., & Schere, K. L. (2006). Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Applied Mechanics Reviews*, *59*(2), 51–77. https://doi.org/10.1115/1.2128636

Campbell, P. C., Bash, J. O., Herwehe, J. A., Gilliam, R. C., & Li, D. (2020). Impacts of tiled land cover characterization on global meteorological predictions using the MPAS-A. *Journal of Geophysical Research - D: Atmospheres*, *125*. https://doi.org/10.1029/2019JD032093

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 Modeling System. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*, 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, *28*, 2031–2064. https://doi.org/10.1002/joc.1688

Deng, A., Stauffer, D. R., Dudhia, J., Otte, T. L., & Hunter, G. K. (2007). *Update on analysis nudging FDDA in WRF-ARW*. The 8th Users' Workshop, National Center for Atmospheric Research (NCAR), Boulder, CO.

Foley, K. M., Hogrefe, C., Pouliot, G., Possiel, N., Roselle, S. J., Simon, H., & Timin, B. (2015). Dynamic evaluation of CMAQ part I: Separating the effects of changing emissions and changing meteorology on ozone levels between 2002 and 2005 in the eastern US. *Atmospheric Environment*, *103*, 247–255. https://doi.org/10.1016/j.atmosenv.2014.12.038

Gan, C.-M., Pleim, J., Mathur, R., Hogrefe, C., Long, C. N., Xing, J., et al. (2015). Assessment of long-term WRF-CMAQ simulations for understanding direct aerosol effects on radiation "brightening" in the United States. *Atmospheric Chemistry and Physics*, *15*, 12193–12209. https://doi.org/10.5194/acp-15-12193-2015

Gilliam, R. C., Appel, W., & Philips, S. (2005). *The Atmospheric Evaluation Tool (AMET): Meteorology Module*, presented at: The 4th Annual CMAS Conference, September 26-28, 2005. Retrieved from http://www.cmascenter.org/conference/2005/abstracts/6_1.pdf

Gilliam, R. C., Godowitch, J. M., & Rao, S. T. (2012). Improving the horizontal transport in the lower troposphere with four dimensional data assimilation. *Atmospheric Environment*, *53*, 186–201. https://doi.org/10.1016/j.atmosenv.2011.10.064

Gilliam, R. C., & Pleim, J. E. (2010). Performance assessment of new land surface and planetary boundary layer physics in the WRF-ARW. *Journal of Applied Meteorology and Climatology*, *49*, 760–774. https://doi.org/10.1175/2009jamc2126.1

Grell, G., Dudhia, J. and, Stauffer, D. R., (1994). *A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5)*, (pp138). NCAR Scientific Tech Note. NCAR/TN-398+STR. https://doi.org/10.5065/D60Z716B

Heath, N. K., Pleim, J. C., Gilliam, R., & Kang, D. (2016). A simple lightning assimilation technique for improving retrospective WRF simulations. *Journal of Advances in Modeling Earth Systems*, *8*, 1806–1824. https://doi.org/10.1002/2016MS000735

Herwehe, J. A., Alapaty, K., Spero, T. L., & Nolte, C. G. (2014). Increasing the credibility of regional climate simulations by introducing subgrid-scale cloud-radiation interactions. *Journal of Geophysical Research - D: Atmospheres*, *119*, 5317–5330. https://doi.org/10.1002/2014JD021504

Herwehe, J. A., Gilliam, R. C., Bullock, O. R., & Pleim, J. E. (2018). *Sensitivity Study of Convection Parameterizations in MPAS-A Utilized in Conjunction with USEPA Physics Options*. Poster presented at: WRF Users' Conference; 2018 June 11-14; Boulder, Co. Retrieved from http://www2.mmm.ucar.edu/wrf/users/workshops/WS2018/posters/p7.pdf

Hogrefe, C., Liu, P., Pouliot, G., Mathur, R., Roselle, S., Flemming, J., et al. (2018). Impacts of different characterizations of large-scale background on simulated regional-scale ozone over the continental United States. *Atmospheric Chemistry and Physics*, *18*, 3839–3864. https://doi.org/10.5194/acp-18-3839-2018

Hong, S., & Lim, J. (2006). The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). *Journal of the Korean Meteorological Society*, *42*, 129–151.

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

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

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

Kang, D., Mathur, R., Pouliot, G. A., Gilliam, R. C., & Wong, D. C. (2020). Significant ground-level ozone attributed to lightning-induced nitrogen oxides during summertime over the Mountain West States. *npj Climate and Atmospheric Science*, *3*, 6. https://doi.org/10.1038/s41612-020-0108-2

Laity, J. J. (2009). *Deserts and desert environments* (p. 7). John Wiley & Sons.

Mathur, R., Xing, J., Gilliam, R., Sarwar, G., Hogrefe, C., Pleim, J., et al. (2017). Extending the Community Multiscale Air Quality (CMAQ) modeling system to hemispheric scales: Overview of process considerations and initial applications. *Atmospheric Chemistry and Physics*, *17*, 12449–12474. https://doi.org/10.5194/acp-17-12449-2017

Matichuk, R., Tonnesen, G., Luecken, D., Gilliam, R., Napelenok, S. L., Baker, K. R., & Roselle, S. (2017). Evaluation of the community multiscale air quality model for simulating winter ozone formation in the Uinta Basin. *Journal of Geophysical Research: Atmosphere*, *122*, 13545–13572. https://doi.org/10.1002/2017JD027057

McNider, R. T., Pour-Biazar, A., Doty, K., White, A., Wu, Y., Qin, M., et al. (2018). Examination of the physical atmosphere in the great lakes region and its potential impact on air quality-overwater stability and satellite assimilation. *Journal of Applied Meteorology and Climatology*, *57*, 2789–2816. https://doi.org/10.1175/JAMC-D-17-0355.1

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*, 991–1007. https://doi.org/10.1175/2008mwr2556.1

Ohmura, A., Gilgen, H., Hegner, H., Müller, G., Wild, M., Dutton, E. G., et al. (1998). Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, *79*, 2115–2136. https://doi.org/10.1175/1520-0477(1998)079<2115:BSRNBW>2.0.CO;2

Otte, T. L. (2008a). The impact of nudging in the meteorological model for retrospective air quality simulations. part I: Evaluation against national observation networks. *Journal of Applied Meteorology and Climatology*, *47*, 1853–1867. https://doi.org/10.1175/2007jamc1790.1

Otte, T. L. (2008b). The impact of nudging in the meteorological model for retrospective air quality simulations. part II: Evaluating collocated meteorological and air quality observations. *Journal of Applied Meteorology and Climatology*, *47*, 1868–1887. https://doi.org/10.1175/2007jamc1791.1

Otte, T. L., Nolte, C. G., Otte, M. J., & Bowden, J. H. (2012). Does nudging squelch the extremes in regional climate modeling? *Journal of Climate*, *25*, 7046–7066. http://dx.doi.org/10.1175/JCLI-D-12-00048.1

Pleim, J. E. (2006). A simple, efficient solution of flux-profile relationships in the atmospheric surface layer. *Journal of Applied Meteorology and Climatology*, *45*, 341–347. https://doi.org/10.1175/jam2339.1

Pleim, J. E. (2007a). 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

Pleim, J. E. (2007b). A combined local and nonlocal closure model for the atmospheric boundary layer. part II: Application and evaluation in a mesoscale meteorological model. *Journal of Applied Meteorology and Climatology*, *46*, 1396–1409. https://doi.org/10.1175/jam2534.1

Pleim, J. E., & Gilliam, R. (2009). An indirect data assimilation scheme for deep soil temperature in the Pleim-Xiu land Surface model. *Journal of Applied Meteorology and Climatology*, *48*, 1362–1376. https://doi.org/10.1175/2009jamc2053.1

Pleim, J. E., Ran, L., Appel, W., Shephard, M. W., & Cady-Pereira, K. (2019). New bidirectional ammonia flux model in an air quality model coupled with an agricultural model. *Journal of Advances in Modeling Earth Systems*, *11*(9), 2934–2957. https://doi.org/10.1029/2019ms001728

Pleim, J. E., Wong, D., Gilliam, R., Herwehe, J., Bullock, R., Hogrefe, C., et al. (2018). The new generation of air quality modeling systems. *Air and Waste Management Association EM*, 16–21.

Pleim, J. E., & Xiu, A. (1995). Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models. *Journal of Applied Meteorology and Climatology*, *34*, 16–32. https://doi.org/10.1175/1520-0450-34.1.16

Pleim, J. E., & Xiu, A. (2003). Development of a Land Surface Model. Part II: Data assimilation. *Journal of Applied Meteorology*, *42*, 1811–1822. https://doi.org/10.1175/1520-0450(2003)042<1811:doalsm>2.0.co;2

PRISM Climate Group, Oregon State University [internet]. Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States. Retrieved from http://www.prism.oregonstate.edu/documents/PRISM_datasets.pdf

Pye, H. O. T., Luecken, D. J., Xu, L., Boyd, C. M., Ng, N. L., Baker, K. R., et al. (2015). Modeling the current and future roles of particulate organic nitrates in the southeastern United States. *Environmental Science and Technology*, *49*, 14195–14203. https://doi.org/10.1021/acs.est.5b03738

Ratner, B. (1956). *Upper-air climatology of the United States. Technical paper no. 32*. Office of Climatology, U.S. Weather Bureau. Washington, DC. Retrieved from https://www.nws.noaa.gov/ohd/hdsc/Technical_papers/TP32P3.pdf

Schere, K., Flemming, J., Vautard, R., Chemel, C., Colette, A., Hogrefe, C., et al. (2012). Trace gas/aerosol boundary concentrations and their impacts on continental-scale AQMEII modeling domains. *Atmospheric Environment*, *53*, 38–50. https://doi.org/10.1016/j.atmosenv.2011.09.043

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S.-H., & Ringler, T. D. (2012). A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tesselations and C-grid staggering. *Monthly Weather Review*, *140*, 3090–3105. https://doi.org/10.1175/MWR-D-11-00215.1

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* (p. 125). University Corporation for Atmospheric Research. https://doi.org/10.5065/D68S4MVH

Solazzo, E., Bianconi, R., Hogrefe, C., Curci, G., Tuccella, P., Alyuz, U., et al. (2017a). Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: Multivariable temporal and spatial breakdown. *Atmospheric Chemistry and Physics*, *17*, 3001–3054. https://doi.org/10.5194/acp-17-3001-2017

Solazzo, E., Hogrefe, C., Colette, A., Garcia-Vivanco, M., & Galmarini, S. (2017b). Advanced error diagnostics of the CMAQ and Chimere modeling systems within the AQMEII3 model evaluation framework. *Atmospheric Chemistry and Physics*, *17*, 10435–10465. https://doi.org/10.5194/acp-17-10435-2017

Spero, T. L., Nolte, C. G., Mallard, M. S., & Bowden, J. H. (2018). A maieutic exploration of nudging strategies for regional climate applications using the WRF model. *Journal of Applied Meteorology and Climatology*, *57*, 1883–1906. https://doi.org/10.1175/JAMC-D-17-0360.1

Spero, T. L., Otte, M. J., Bowden, J. H., & Nolte, C. G. (2014). Improving the representation of clouds, radiation, and precipitation using spectral nudging in the Weather Research and Forecasting Model. *Journal of Geophysical Research - D: Atmospheres*, *119*, 11682–11694. https://doi.org/10.1002/2014jd022173

Stauffer, D. R., & Seaman, N. L. (1990). Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Monthly Weather Review*, *118*, 1250–1277. https://doi.org/10.1175/1520-0493(1990)118<1250:UOFDDA>2.0.CO;2

Stauffer, D. R., & Seaman, N. L. (1994). Multiscale four-dimensional data assimilation. *Journal of Applied Meteorology*, *33*, 416–434. https://doi.org/10.1175/1520-0450(1994)033<0416:mfdda>2.0.co;2

Stauffer, D. R., Seaman, N. L., & Binkowski, F. S. (1991). Use of four-dimensional data assimilation in a limited-area mesoscale model part II: Effects of data assimilation within the planetary boundary layer. *Monthly Weather Review*, *119*, 734–754. https://doi.org/10.1175/1520-0493(1991)119<0734:uofdda>2.0.co;2

U.S. EPA. (2019). *Meteorological model performance for annual 2016 simulation WRF v3.8*. Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency. Retrieved from https://www3.epa.gov/ttn/scram/reports/Met_Model_Performance-2016_WRF.pdf

Wong, D., Foroutan, H., Pleim, J., Bullock, R., Gilliam, R., Herwehe, J., et al. (2020). A proof-of-concept for linking the global meteorological model, MPAS-A with the Air Quality Model, CMAQ. In C. Mensink, W. Gong, & A. Hakami (Eds.), *Air quality modeling and its application XXVI*. Springer Proceedings in Complexity. https://doi.org/10.1007/978-3-030-22055-6_6

Wu, W.-S., Purser, R. J., & Parrish, D. F. (2002). Three-dimensional variational analysis with spatially inhomogeneous covariances. *Monthly Weather Review*, *130*, 2905–2916. https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2

Xing, J., Mathur, R., Pleim, J., Hogrefe, C., Gan, C.-M., Wong, D. C., et al. (2014). Observations and modeling of air quality trends over 1990–2010 across the Northern Hemisphere: China, the United States and Europe. *Atmospheric Chemistry and Physics*, *14*, 25453–25501. https://doi.org/10.5194/acp-15-2723-2015

Xiu, A., & Pleim, J. E. (2001). Development of a land-surface model. part I: Application in a Mesoscale Meteorological Model. *Journal of Applied Meteorology*, *42*, 1811–1822. https://doi.org/10.1175/1520-0450(2001)040<0192:DOALSM>2.0.CO;2