



Evaluation of the NAQFC driven by the NOAA Global Forecast System (version 16): comparison with the WRF-CMAQ during the summer 2019 FIREX-AQ campaign

Youhua Tang^{1,2}, Patrick C. Campbell^{1,2}, Pius Lee¹, Rick Saylor¹, Fanglin Yang³, Barry Baker¹, Daniel Tong^{1,2}, Ariel Stein¹, Jianping Huang^{3,4}, Ho-Chun Huang^{3,4}, Li Pan^{3,4}, Jeff McQueen³, Ivanka Stajner³, Jose Tirado-Delgado^{5,6}, Youngsun Jung⁵, Melissa Yang⁷, Ilann Bourgeois^{8,9}, Jeff Peischl^{8,9}, Tom Ryerson⁹, Donald Blake¹⁰, Joshua Schwarz⁹, Jose-Luis Jimenez⁸, James Crawford¹¹, Glenn Diskin⁷, Richard Moore⁷, Johnathan Hair⁷, Greg Huey¹¹, Andrew Rollins⁹, Jack Dibb¹², and Xiaoyang Zhang¹³

¹NOAA Air Resources Laboratory, College Park, MD, USA

²Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA

³NOAA National Centers for Environmental Prediction, College Park, MD, USA

⁴I.M. Systems Group Inc., Rockville, MD, USA

⁵Office of Science and Technology Integration, NOAA National Weather Service, Silver Spring, MD, USA

⁶Science & Technology Corporation, Hampton, VA, USA

⁷NASA Langley Research Center, Hampton, VA, USA

⁸Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA

⁹NOAA Chemical Sciences Laboratory, Boulder, CO, USA

¹⁰Department of Chemistry, University of California at Irvine, Irvine, CA, USA

¹¹School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

¹²Earth Systems Research Center, University of New Hampshire, Durham, NH, USA

¹³Department of Geography and Geospatial Sciences, South Dakota State University, Brookings, SD, USA

Correspondence: Youhua Tang (youhua.tang@noaa.gov)

Received: 18 May 2022 – Discussion started: 9 June 2022

Revised: 28 September 2022 – Accepted: 10 October 2022 – Published: 7 November 2022

Abstract. The latest operational National Air Quality Forecast Capability (NAQFC) has been advanced to use the Community Multiscale Air Quality (CMAQ) model (version 5.3.1) with the CB6r3 (Carbon Bond 6 revision 3) AERO7 (version 7 of the aerosol module) chemical mechanism and is driven by the Finite-Volume Cubed-Sphere (FV3) Global Forecast System, version 16 (GFSv16). This update has been accomplished via the development of the meteorological preprocessor, NOAA-EPA Atmosphere–Chemistry Coupler (NACC), adapted from the existing Meteorology–Chemistry Interface Processor (MCIP). Differing from the typically used Weather Research and Forecasting (WRF) CMAQ system in the air quality research community, the interpolation-based NACC can use various meteorological outputs to drive the CMAQ model (e.g., FV3-GFSv16), even though they

are on different grids. In this study, we compare and evaluate GFSv16-CMAQ and WRFv4.0.3-CMAQ using observations over the contiguous United States (CONUS) in summer 2019 that have been verified with surface meteorological and AIRNow observations. During this period, the Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) field campaign was performed, and we compare the two models with airborne measurements from the NASA DC-8 aircraft. The GFS-CMAQ and WRF-CMAQ systems show similar performance overall with some differences for certain events, species and regions. The GFSv16 meteorology tends to have a stronger diurnal variability in the planetary boundary layer height (higher during daytime and lower at night) than WRF over the US Pacific coast, and it also predicted lower nighttime 10 m winds. In summer 2019,

the GFS-CMAQ system showed better surface ozone (O_3) than WRF-CMAQ at night over the CONUS domain; however, the models' fine particulate matter ($PM_{2.5}$) predictions showed mixed verification results: GFS-CMAQ yielded better mean biases but poorer correlations over the Pacific coast. These results indicate that using global GFSv16 meteorology with NACC to directly drive CMAQ via interpolation is feasible and yields reasonable results compared to the commonly used WRF approach.

1 Introduction

Traditionally, mesoscale meteorological models such as the Weather Research and Forecasting (WRF) model (Powers et al., 2017) are used as the meteorological drivers for air quality models (AQMs) on the same (“native”) model grid, such as the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006). The National Air Quality Forecast Capability (NAQFC) of the NOAA National Weather Service (NWS) has historically used a different approach in which the hourly meteorological outputs from prior operational models, such as the North American Mesoscale (NAM) model, are interpolated to the CMAQ grid to drive its air quality prediction. Prior to this work, a “PREMAQ” coupler (Otte et al., 2004) combined both meteorological processing and Sparse Matrix Operator Kernel Emissions (SMOKE) (Houyoux et al., 2000) processes, such as point-source plume rise effects. However, since the release of CMAQ version 5, the meteorology-dependent plume rise, sea salt and dust emission processes are included as in-line modules in CMAQ; thus, the corresponding emission processes are no longer needed in PREMAQ. Furthermore, PREMAQ has no built-in interpolator and, thus, relies on external interpolators to remap the non-native-grid meteorological inputs, such as NAM, to the targeted CMAQ domain, although it does perform vertical layer collapsing/interpolation to reduce vertical layers for CMAQ. The interpolation approach allows more flexibility for using different meteorological data (i.e., besides just WRF) to drive CMAQ; however, this may cause mass-consistency issues between models. It should be noted that mass-consistency issues may also exist using native-grid couplers (Byun, 1999a, b) and can stem from the original mass-inconsistent meteorological model outputs or arise due to the temporal interpolation of the meteorological data. The well-developed offline AQMs, such as CMAQ, have already considered such mass-consistency treatments using different meteorological inputs (Byun and Ching, 1999).

To upgrade the NAQFC system with the latest CMAQ model and NOAA operational meteorology, we developed an updated interpolation-based meteorological coupler, the NOAA-EPA Atmosphere–Chemistry Coupler (NACC) (Campbell et al., 2022), adapted from version 5 of the US EPA's Meteorology–Chemistry Interface Processor (MCIP)

(Otte and Pleim, 2010; <https://github.com/USEPA/CMAQ>, last access: 24 October 2022). The NACC system effectively couples the Global Forecast System version 16 (GFSv16) (Yang et al., 2020; Harris et al., 2021) with the Finite-Volume Cubed-Sphere (FV3) dynamical core to CMAQv5.3.1 (hereafter referred to as GFS-CMAQ). Campbell et al. (2022) described the development and application of the GFS-CMAQ system using NACC (referred to as “NACC-CMAQ” in their work) and provided a comprehensive comparison between the current (GFS-CMAQ since 20 July 2021) and previous (NAM-CMAQv5.0.2) operational NAQFC model performance.

In this study, we analyze the impacts of the meteorological model drivers, and we compare GFS-CMAQ using NACC interpolation to the commonly used native-grid WRF-CMAQ application and its impact on air quality predictions in summer 2019. Yu et al. (2012a, 2012b) had previously compared the CMAQ performance driven by Weather Research and Forecasting (WRF) with two dynamic cores – the Non-hydrostatic Mesoscale Model (NMM) (Janjic, 2003) and the Advanced Research WRF (ARW) (Skamarock et al., 2005) – during the 2006 Texas Air Quality Study/Gulf of Mexico Atmospheric Composition and Climate Study (TexAQS/GoMACCS) field campaign, and they found that the NMM-CMAQ and ARW-CMAQ showed overall similar performance with some differences for certain events, chemical species and regions. Similarly, this study focuses on the comparison of GFS-CMAQ with WRF-CMAQ (see Sect. 2) and verifies the model performance against the aircraft observations from the Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) field experiment during summer 2019 (Sect. 4). Surface verification is also performed using AIRNow data for August 2019 (Sect. 3), serving as a benchmark for the new NAQFC versus the traditional WRF-CMAQ used in the air quality modeling community. This study focuses on the period of summer 2019, and Campbell et al. (2022) evaluated the GFS-CMAQ for longer periods.

2 Methodology

Here, we compare the two CMAQ (version 5.3.1) runs driven by the interpolated GFSv16 meteorology (GFS-CMAQ) and WRF meteorology (WRF-CMAQ). All other settings, such as emissions and lateral boundary conditions are the same. The meteorology-related physics is discussed in the following sections to address the models' performance discrepancies. Both the GFS-CMAQ and WRF-CMAQ simulations are run for a period covering 12 July to 31 August 2019, each using the last 10 d in July as the model spin-up periods that are not included in the analyses.

2.1 GFS meteorological inputs

The GFSv16 is the current operational global forecast system in NOAA/NCEP using the FV3 dynamical core. Its detailed configuration can be found in Campbell et al. (2022) and Yang et al. (2020). Compared with the previous version (v15), GFSv16 updated many physical schemes (Table 1) and added the parameterization for sub-grid-scale nonstationary gravity-wave drag. To use the GFS's meteorology to drive CMAQ, a meteorological coupler, NACC, is developed (Campbell et al., 2022). Differing from the original MCIP, which was developed to process WRF/ARW meteorology for CMAQ, the NACC coupler interpolates non-native-grid meteorology to a user-defined grid and has parallel processing capability, which drastically reduces its run time for operational forecasts (Campbell et al., 2022). Currently, NACC employs two horizontal interpolation methods: bilinear and nearest neighbor. In this study, we use the nearest-neighbor method for categorical (discontinuous) variables that include land use types, vegetation fraction, terrain elevation, Monin–Obukhov length, friction velocity and soil temperatures, whereas the bilinear interpolation is used for mainly smoothly varying (continuous) meteorological variables that include wind fields, temperature, pressure and specific humidity. The CMAQ model is defined in the Arakawa C-grid (Arakawa and Lamb, 1977); thus, the GFSv16 horizontal wind components (U , V) need to be interpolated to the perpendicular cell faces instead of the cell center (Otte and Pleim, 2010) after rotation to the defined map projection. The scalar variables are defined in the cell centers of the targeted grid; thus, their interpolations are more straightforward: from GFSv16 A-grid to CMAQ A-grid. The NACC coupler can either use the native layers or interpolate inputs to a set number of user-defined CMAQ vertical layers. The GFSv16 has 127 vertical layers with global coverage at roughly a 13 km horizontal resolution, where the targeted CMAQ domain is in 12 km \times 12 km grids over the contiguous United States (CONUS) with 35 vertical layers (Campbell et al., 2022). Here, we use 24 h GFSv16 forecasts starting at 12:00 UTC each day.

Most variables needed by CMAQ are directly interpolated from the GFSv16 hourly outputs. The NACC processor has options to calculate diagnostic variables, such as the planetary boundary layer (PBL) height, if they are needed. In this study, we use the interpolated GFSv16 PBL height instead of the diagnostic one. It also has an option to import the externally provided land surface variables. Here, we import the updated the 2018–2020 climatological averaged leaf area index (LAI) and NOAA near-real-time (NRT) greenness vegetation fraction (GVF) from satellite-based Visible Infrared Imaging Radiometer Suite (VIIRS) retrievals (Campbell et al., 2022). The updated satellite-based LAI and GVF impact CMAQ's biogenic emissions and dry-deposition processes, which were described in detail in Campbell et al. (2022).

2.2 WRF meteorology

For comparison with GFSv16 meteorology processed by NACC, a corresponding WRF version 4.0.3 (Skamarock et al., 2021) simulation is run covering the NAQFC's native grid, which is a 12 km horizontal resolution with Lambert conformal map projection over the CONUS. Table 1 shows the WRF configuration, which is commonly employed in CONUS meteorological and air quality studies in the community, versus the current NOAA/NWS operational global model GFSv16. GFSv16 uses the NOAA/NCEP's Global Data Assimilation System (GDAS) (https://www.emc.ncep.noaa.gov/data_assimilation/data.html, last access: 24 October 2022) for its initial conditions and runs on its own global dynamics and physics without any other constraints. The regional WRF simulation uses GFSv16 for its initial conditions. In this study, GFSv16 was re-initialized with GDAS every 24 h, and WRF conducted the continuous run after its spin-up period. Furthermore, WRF also takes its lateral boundary conditions from GFSv16 every 6 h. For the WRF run, we have enabled 4D data assimilation (FDDA) for the horizontal winds, temperature and humidity (Table 1) every 6 h, thereby nudging towards GFSv16. This nudging method used in WRF runs can help reduce the difference between two meteorological models, although its effect may vary depending on events because WRF and GFS use different physics.

WRF and GFSv16 have similar settings for the land surface model, surface layer and radiation schemes; however, their microphysics and PBL schemes are different (Table 1). Compared with the 35-layer WRF model with a 100 hPa domain top, GFSv16 has a much higher domain top (0.2 hPa) and 127 vertical layers, which are interpolated by NACC to 35σ layers up to 14 km for CMAQ. We use NACC (inherited from MCIP version 5.0) to process WRF hourly meteorology while also maintaining the 35-vertical-layer structure. Thus, in contrast to GFS-CMAQ, the WRF-CMAQ system uses the native grid without interpolation.

2.3 CMAQ configuration

Here, CMAQ version 5.3.1 (Appel et al., 2021) is used with the Carbon Bond 6 revision 3 (CB6r3; Yarwood et al., 2010, 2014; Luecken et al., 2019) chemical mechanism and AERO7 (version 7 of the aerosol module) treatment of secondary organic aerosols (CB6r3_AE7_AQ). CMAQv5.3.1 includes a series of scientific updates from the previous version (Appel et al., 2021), such as the updated air–surface exchange and deposition modules, which showed a significant impact on ozone (O_3) prediction compared with the previous NAQFC (Campbell et al., 2022). We also include the bidirectional NH_3 (BIDI- NH_3) exchange model for NH_3 surface fluxes. An updated Biogenic Emissions Landuse Dataset v5 (BELD5) is used in this study to drive the in-line Biogenic Emissions Inventory System (BEIS) version 3.61. The an-

Table 1. The two meteorological datasets used in this study.

Model settings	FV3-GFSv16/NACC	WRF-ARW/MCIP
Domain	Global C768L127 (~ 13 km horizontal resolution in six cubic spherical tiles with 127 vertical layers up to 80 km), interpolated to the 12 km CONUS domain with 35 layers up to about 14 km (60 hPa)	The 12 km CONUS domain with 35 vertical layers up to 100 hPa
Dynamic core	Finite-Volume Cubed-Sphere (FV3), non-hydrostatic (Putman and Lin, 2007)	WRF-ARW dynamic in hybrid vertical coordinate (Skamarock et al., 2021)
Initial condition	FV3-GFSv16 analysis (GDAS) using the local ensemble Kalman filter (Ott et al., 2004) with 4D incremental analysis update	FV3-GFSv16 analysis (GDAS)
Lateral boundary condition	n/a	FV3-GFSv16 analysis (GDAS)
Cloud microphysics	GFDL six-category cloud microphysics scheme (Lin et al., 1983; Lord et al., 1984; Krueger et al., 1995; Chen and Lin, 2011, 2013)	Morrison two-moment scheme (Morrison et al., 2009)
PBL physics scheme	Scale-aware (sa) turbulent kinetic energy (TKE) based moist eddy diffusivity mass flux (EDMF) scheme (sa-TKE-EDMF) (Han and Bretherton, 2019)	Yonsei University (YSU) scheme (Hong et al., 2006)
Shallow/deep cumulus parameterization	Simplified Arakawa–Schubert scheme (Han and Pan, 2011; Han et al., 2017)	Kain–Fritsch multiscale (Kain, 2004)
Shortwave and longwave radiation	Rapid Radiative Transfer Model for General Circulation Models (RRTMG) (Mlawer et al., 1997; Clough et al., 2005; Iacono et al., 2008)	RRTMG (Iacono et al., 2008).
Land surface model	Noah land surface model (Chen and Dudhia 2001; Ek et al., 2003; Tewari et al., 2004)	Noah (Tewari et al., 2004)
Surface layer	Monin–Obukhov (Monin and Obukhov, 1954; Grell et al., 1994; Jimenez et al., 2012)	Jimenez et al. (2012)
Other treatment		FDDA nudging is enabled for temperature and specific humidity over the whole domain as well as for wind components (U , V) outside the PBL.

n/a denotes not applicable.

thropogenic emissions are provided by the National Emissions Inventory Collaborative (NEIC) with base year 2016 version 1 (NEIC, 2019). We replace the US EPA default CMAQ dust emissions model with an in-line windblown dust model known as “FENGSHA” (Fu et al., 2014; Huang et al., 2015; Dong et al., 2016). The FENGSHA dust scheme uses the sediment supply map and magnitude of the friction velocity (USTAR) compared with a threshold friction velocity (UTHR) to calculate the potential of dust emission flux. The UTHR depends on the land cover, soil type (clay fraction) and soil moisture (Campbell et al., 2022).

We have updated the wildfire emissions system in CMAQv5.3.1 based on the Blended Global Biomass Burn-

ing Emissions Product (GBBEPx) (Zhang and Kondragunta, 2006; Zhang et al., 2011). The GBBEPx uses satellite-detected fire radiative power (FRP) to estimate wildfire smoke emissions for a number of species: CO (carbon monoxide), NO_x (nitrogen oxides), SO₂ (sulfur dioxide), elemental carbon, primarily emitted organic aerosols and fine particulate matter (PM_{2.5}). We derive the wildfire volatile organic compound (VOC) emissions from GBBEPx CO emissions based on the emission ratios of Fire INventory from NCAR (FINN) data (Wiedinmyer et al., 2011), and the splitting factors from the SMOKE model (Baker et al., 2016) are used to further split the total fire VOC emissions to specified hydrocarbon emissions. The satellite FRP is estimated

from the satellite brightness temperature anomaly, and the GBBEPx processor assumes that the wildfire emissions are proportional to the FRP over certain land use types in certain regions. The GBBEPx emissions are based on polar orbiting satellites: MODIS (Aqua and Terra satellites) and VIIRS (Suomi NPP and NOAA-20 satellites) instruments, which are updated once per day. A wildfire emission preprocessor converts the GBBEPx emissions to CMAQ-ready input files using emission speciation and diurnal profiles (high during daytime and low at night), which are adopted from US EPA-based profiles (Baker et al., 2016), and a daily scaling factor. Here, we classify the wildfire into either a long-lasting fire (longer than 24 h) or a short-term fire (shorter than 24 h) based on land use types and regions. As historic statistics show that most fires (> 95 %) in the region east of 110° W last less than 24 h, only fires west of 110° W that have a model grid cell total forest fraction > 0.4 are assumed to be long-lasting fires. All other short-term GBBEPx fires are assumed to have smoke emissions for 24 h (i.e., day 1 only). Burning area could be highly uncertain, as GBBEPx data do not include this information. One grid cell could have multiple fires, and some big fires could appear in several grids. Here, we carry the previous NAQFC's method and apply a constant ratio, 10 % of the grid cell, as the burning area (Pan et al., 2020), according to Rolph et al. (2009). CMAQ treats wildfire emissions as point sources that undergo in-line plume rise to distribute the smoke vertically. The default CMAQ plume rise used here is based on Briggs (1965) and is driven by fire heat flux (converted from FRP with a ratio of 1) and fixed burning area (assumed to be 10 % of the $0.1^\circ \times 0.1^\circ$ grid cell).

3 Model evaluations over the US for August 2019

In this section, in order to first gain a general picture and compare the overall GFS-CMAQ and WRF-CMAQ model performance, we evaluate the near-surface meteorological and air quality predictions during the FIREX-AQ August 2019 period against NOAA's METeorological Aerodrome Report (METAR; https://madis.ncep.noaa.gov/madis_metar.shtml, last access: 24 October 2022) and the US EPA's AIRNow (<https://www.airnow.gov>, last access: 24 October 2022) observation networks. All of the comparisons of meteorological variables are for those actually used in CMAQ. For GFS-CMAQ, it refers to the interpolated GFS data. Campbell et al. (2022) included the detailed comparisons before and after interpolation, showing that the interpolated meteorology was very consistent with the original meteorology. In this study, the model results are spatiotemporally interpolated to the corresponding observation locations for comparison.

3.1 Domain-wide meteorology against the METAR network

Figure 1 shows the mean bias (MB) of interpolated GFS- and WRF-predicted surface meteorological variables compared to METAR data during August 2019. Both meteorological models have a cool bias over the western and northeastern US and a warm bias over the western Rocky Mountains region and southeastern US (Fig. 1a, b). Similar temperature predictions are expected because WRF uses the FDDA method, nudging toward GFS data. However, GFS tends to be cooler than WRF over the Rocky Mountains and in the central and northeastern USA due to their different dynamics and physics. The GFSv16 cold bias in the lower troposphere is impacted by excessive evaporative cooling from rainfall (personal communication with NOAA/NCEP, 2021). Campbell et al. (2022) provides detailed discussions about GFSv16 biases.

Both the GFSv16 and WRF models have similar and rather significant dry biases for specific humidity (SH) predictions across the CONUS domain (Fig. 1c, d). Qian et al. (2020) investigated this common dry bias in many models and found that neglecting an irrigation contribution could cause this dry bias. Besides this issue, WRF's dry bias could also be affected by its nudging toward GFS, as GFS has widespread dry biases (Campbell et al., 2022). Their biases also have some noticeable differences over certain regions. For instance, WRF has less dry bias over southern Texas than GFS.

Both models underestimate the mean 10 m wind speeds compared with METAR stations over the western US: WRF has stronger underpredictions over the Rocky Mountains and overpredictions over northeastern US, whereas GFS has stronger underpredictions over the Appalachian Mountains and overpredictions over Texas and Oklahoma. GFSv16's operational verification (https://www.emc.ncep.noaa.gov/gmb/emc.glopara/vsdb/v16rt2/g2o/g2o_00Z/index.html, last access: 24 October 2022) also shows that it tends to underpredict the 10 m wind speed over the western US during both daytime and nighttime, but it shows overpredictions over the eastern US. Besides the difference in physical schemes (Table 1), for example, other possible reasons causing this surface wind difference could be effect of gravity-wave drag (GFSv16 includes it, but the WRF run here does not) and vertical resolutions (GFS's 127 layers versus WRF's 35 layers, although they have similar vertical layers below 1 km) (Campbell et al., 2022). Some studies (Skamarock et al., 2019) have revealed the necessity for a fine vertical resolution for atmospheric simulations, especially within the PBL, near the tropospheric top and during convective events. Insufficient vertical resolution could also cause plume dilution in chemical transport modeling (Zhuang et al., 2018). The gravity-wave drag is also known to influence the synoptic-scale dynamics of the atmospheric flow over irregularities at the Earth's surface, such as mountains and valleys, and the uneven distribution of diabatic heat sources associated with

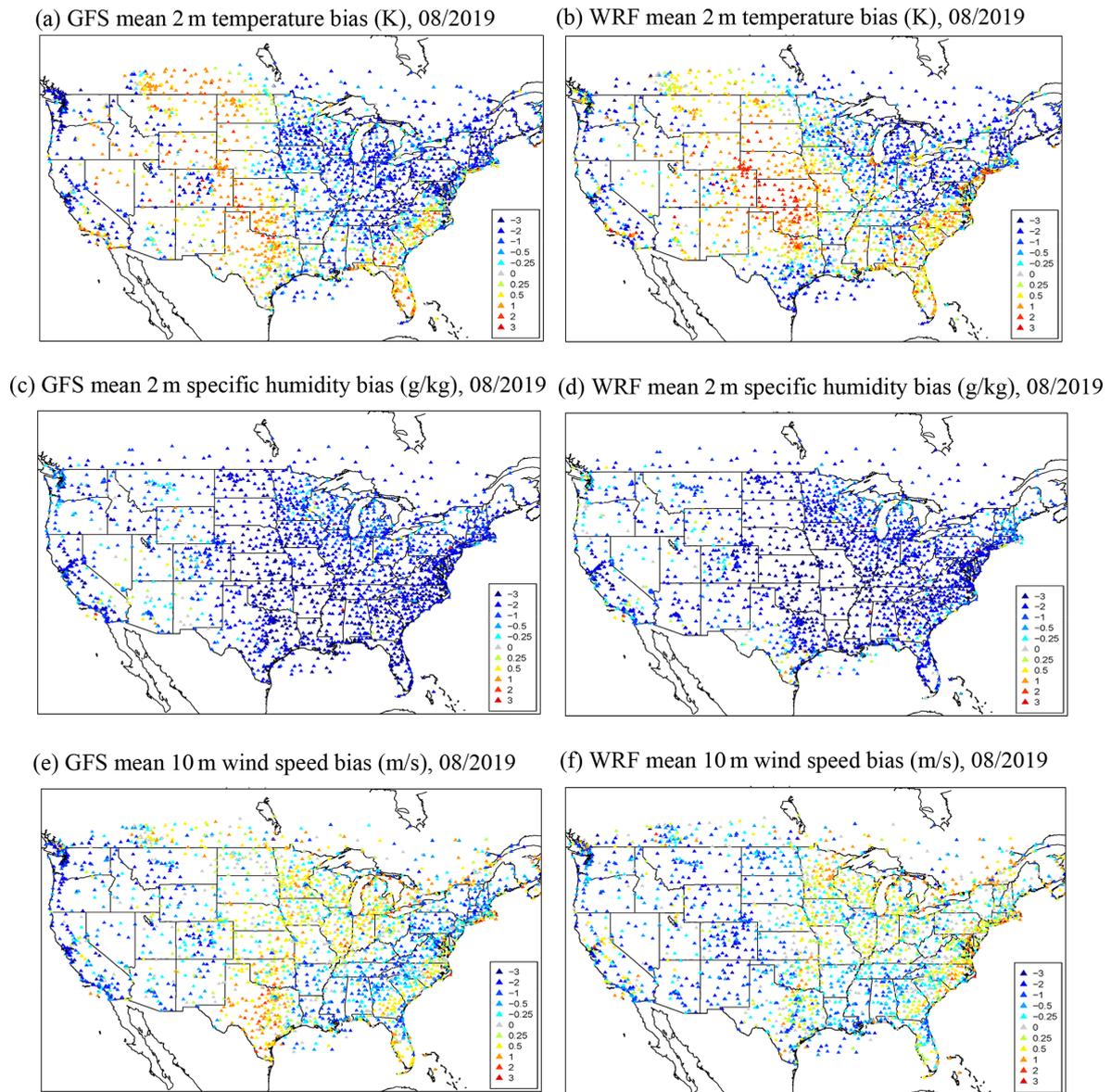


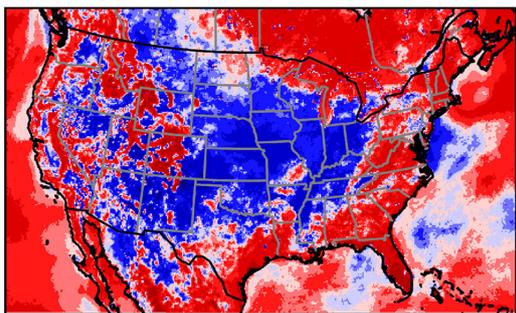
Figure 1. GFS and WRF surface meteorological biases for METAR (METeorological Aerodrome Report) stations averaged over August 2019.

convective systems (Kim et al., 2003). Its parameterization is needed for large-scale models.

There are strong regional variabilities in the monthly mean PBL height differences between GFS and WRF during normal daytime (represented by 18:00 UTC) and nighttime (represented by 06:00 UTC) (Fig. 2). During daytime, GFS has a higher PBL height compared with WRF over the US Pacific coast, northern Rocky Mountains, and northeastern and southeastern US, but it becomes lower over the central US (e.g., Texas, Oklahoma and Kansas). At night, however, most of these regional differences between GFS and WRF are reversed. This diurnal difference is mainly driven by the different PBL schemes employed in GFS (Han and Bretherton,

2019) and WRF (i.e., YSU) and the associated other physical suites, including the land surface data. Hence, this PBL difference has strong regional variations depending on geographic differences. The GFS's PBL height has a strong diurnal variation over these regions, including the western and northeastern US, and it shows a sharp rise and collapse after sunrise and sunset, respectively (Campbell et al., 2022). These two selected times (18:00 and 06:00 UTC) are not in the transition periods for fast diurnal changes in the PBL.

(a) Monthly Mean PBL Difference (GFS-WRF) at 18UTC



(b) Monthly Mean PBL Difference (GFS-WRF) at 06UTC

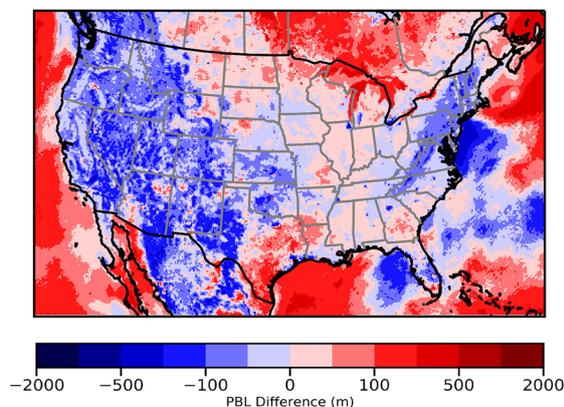


Figure 2. Monthly mean PBL height difference (GFS – WRF) for daytime (a) and nighttime (b) in August 2019.

3.2 Evaluation of regional meteorology and air quality against the AIRNow Network

The US EPA AIRNow network provides hourly observations of near-surface O_3 , $PM_{2.5}$ and meteorology. Campbell et al. (2022) showed detailed verification of GFS-CMAQ with the surface AIRNow data. Here, we focus on the difference between the interpolation-based GFSv16 and WRF downscaling as well as the impacts on meteorological and chemical model performance. Figure 3 shows a comparison of these two models over two specific regions, the western US (CA, OR and WA) and the northeastern states (CT, DE, MA, MD, ME, NH, NJ, NY, PA, RI, VT and the District of Columbia) (Fig. S1 in the Supplement), where the two models have relatively large differences for some meteorological variables. GFS and WRF predict very similar 2 m temperatures over the Pacific coast states: WA, OR and CA, and both of them had a similar cool bias (around 1 K), R value and RMSE (Fig. 3a). However, these two models show significant differences with respect to the 10 m wind speed prediction over the Pacific coast (Fig. 3c), where WRF overpredicts the wind speed, especially at night and in later August. Most AIRNow stations are located near urban or suburban areas, which generally have a weaker 10 m wind speed

than those at the METAR aviation weather stations near airports. For this reason, although Fig. 1e and f show that GFS and WRF underpredict the monthly mean wind speed over the METAR stations in the west, they still tend to overpredict wind (Fig. 3c) over AIRNow stations, especially for the WRF 10 m wind speed at night. Considering that the model grid cells represent $12 \times 12 \text{ km}^2$ averages, the true model–observation comparisons likely fall somewhere between the urban/suburban AIRNow stations and METAR stations, depending on the land use fractions of each grid. Obviously the representation characteristics of observations could affect the verification results. Compared with AIRNow observations, GFSv16 has overall better scores for surface wind speed predictions over the western US, where the WRF’s higher surface wind speed overprediction is associated with its PBL height predictions (Fig. 3e, f). During the nighttime, GFS has a lower PBL height (10%–50% lower than WRF) and weaker vertical mixing, which tends to bring less momentum flux from the upper layers to the surface, leading to lower nighttime wind and better agreements with the AIRNow wind speed observation.

Over the northeastern US, the mean bias (MB) of GFS temperature is about -1 K , whereas the WRF model has a slightly warm MB of about 1.53 K (Fig. 3b). However, the GFS’s temperature prediction has a better correlation coefficient (R) and RMSE, implying that it better captures some events, such as the storm on 28–29 August. Both models overpredict 10 m wind speeds in the northeast, but the GFS model yields better results due to a slightly lower PBL height at night (Fig. 3f) compared with WRF, which had significant overpredictions, especially during 25–29 August (Fig. 3d) when tropical storm Erin approached this region. Especially on 28 August, when the storm was centered near the east coast of North Carolina, WRF significantly underpredicted the 2 m temperature (Fig. 3b) and overpredicted the 10 m wind speed (Fig. 3d). In the west around the same period, tropical storm Ivo appeared southwest of the Baja California Peninsula, bringing heavy rainfall to Mexico. Associated with this storm, a low-pressure system expanded over most of the western US. Differing from GFSv16, which is designed for the operational meteorological forecast, the WRF configuration used in this study is normal for driving CMAQ, but it is not tuned for storm weather prediction.

Figure 4a and b show the O_3 predictions of the two models over these two regions, and GFS-CMAQ yields predominantly lower O_3 than WRF-CMAQ, especially at night. Over the west, the lower O_3 in GFS-CMAQ is associated with their PBL height difference. First, with a certain dry-deposition velocity between the models, it is easier to deplete O_3 given the smaller volume of a shallower PBL. Second, the shallower PBL results in higher surface NO_x concentrations (not shown) and O_3 titration rates near NO_x source regions, consequently resulting in lower O_3 in these areas at night. Last, the lower PBL could decouple from the residual layer and result in weaker or no vertical O_3 exchange with the resid-

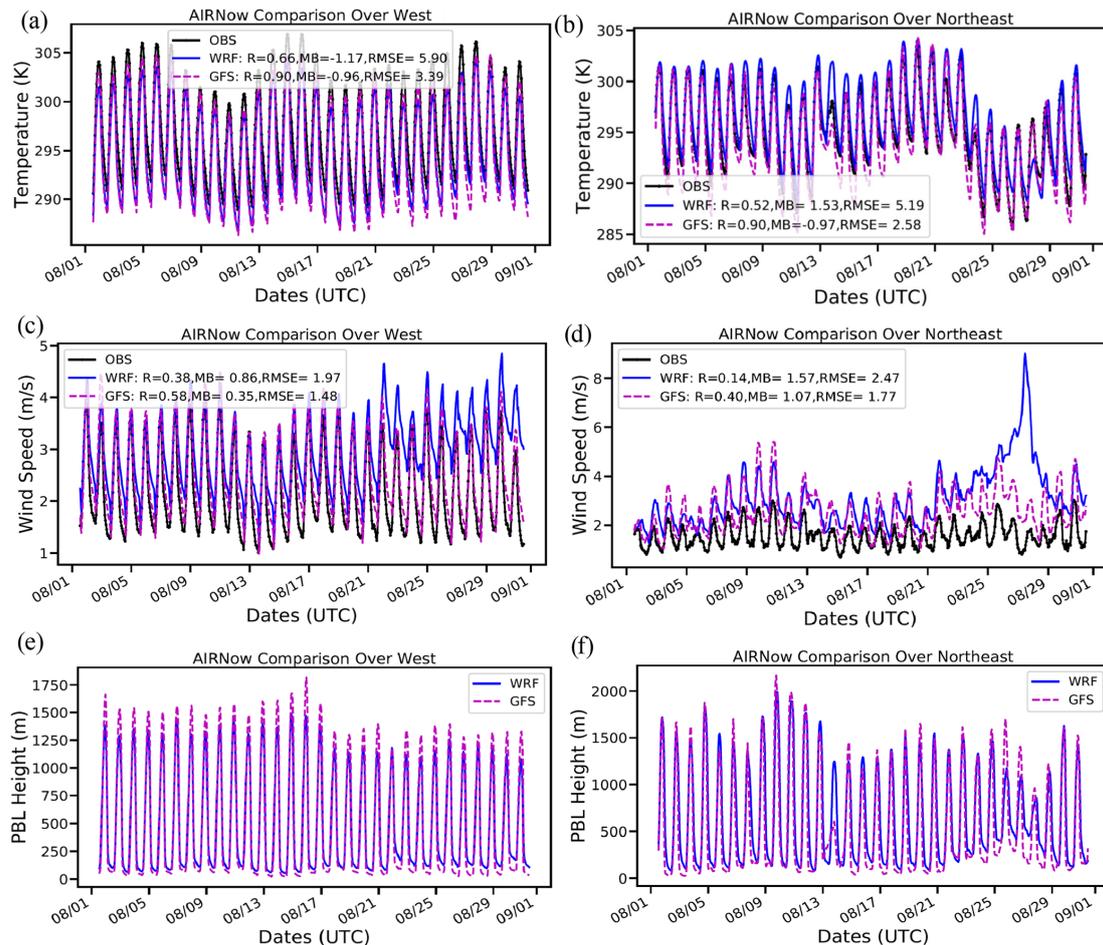


Figure 3. The WRF and GFS time series comparison for AIRNow stations over the western and northeastern US for 2 m temperature (a, b), 10 m wind speed (c, d) and PBL height (e, f).

ual layer at night (Caputi et al., 2019). All of these factors contribute to the lower nighttime O_3 of GFS-CMAQ compared with WRF-CMAQ. As GFS-CMAQ already underpredicts O_3 due to combined meteorological factors, such as the temperature underprediction (Fig. 4a), the GFS-CMAQ's further O_3 reduction (possibly due to its lower PBL height at night) exacerbates its low bias. However, over the northeastern US, the similar impacts help the GFS-CMAQ yield a much better MB due to its better agreement with the observed nighttime low O_3 over this region. Over the entire CONUS domain, the situation is similar, and the GFS-CMAQ has a lower ozone MB (1.1 ppb) compared with WRF-CMAQ (4.7 ppb). Figure 5 shows that both models have similar daytime O_3 prediction over the CONUS. However, GFS-CMAQ better captures low nighttime O_3 over the eastern US than WRF-CMAQ (Fig. 5c, d).

GFS-CMAQ has substantially higher $PM_{2.5}$ mean concentrations over the western US but lower mean concentrations over the northeastern US compared with WRF-CMAQ (Fig. 4c, d). These model differences are also related to their interpolated GFSv16 versus downscaled WRF meteorolog-

ical drivers. Because both models use the same emissions under relatively clean background conditions in the west (i.e., prevailing westerly flow from the Pacific Ocean), the PBL and wind speed differences have significant impacts on their near-surface pollutant concentrations, especially at night. Both models show strong $PM_{2.5}$ diurnal variations (high at night and low during daytime), driven by the meteorological diurnal variation (e.g., PBL), which overcomes the emission diurnal variation (usually high during daytime and low at night). Compared with WRF-CMAQ, GFS-CMAQ has a lower nighttime PBL height and a weaker wind speed at night, which lead to weaker vertical mixing and venting, increasing the pollutant concentrations near the surface and yielding higher surface $PM_{2.5}$ over the western US (Fig. 4c). Its higher surface $PM_{2.5}$ could also result in stronger local dry deposition. In contrast to the local vertical mixing and venting effects on $PM_{2.5}$ discussed above, there are strong (and potentially counterbalancing) impacts of model PBL and horizontal wind speed differences on downstream $PM_{2.5}$ concentrations at night. WRF-CMAQ's deeper PBL and stronger wind speeds at night (Fig. 3c, d, e, f) tend

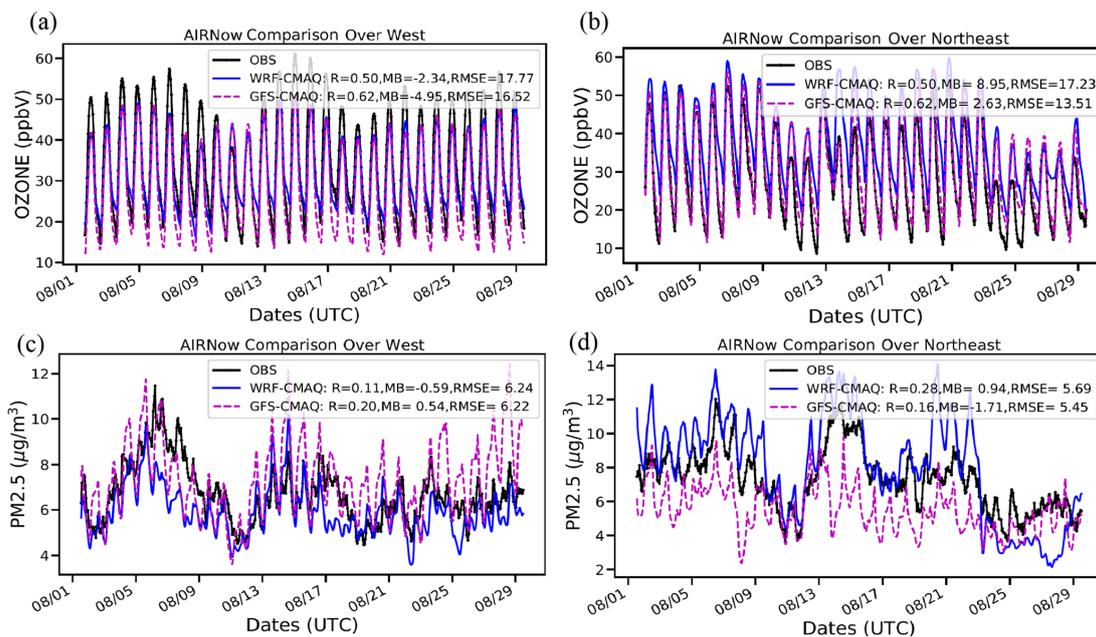


Figure 4. Same as Fig. 3 but for O₃ (a, b) and PM_{2.5} (c, d).

to transport aerosols and their precursors more efficiently downstream via the dominant advection pathway. Figure 6 shows that these monthly mean background PM_{2.5} differences appear east of the Rocky Mountains (WRF-CMAQ is about 2 µg m⁻³ higher) during both daytime and nighttime. This effect is very prominent in the northeastern region. Although both models predicted a similar PM_{2.5} magnitude over the northeastern US, GFS-CMAQ yields an overall PM_{2.5} underprediction, and its monthly mean PM_{2.5} is 2.6 µg m⁻³ lower than the WRF-CMAQ prediction (Fig. 4d). Especially during the 1–9 August period, WRF-CMAQ had about a 4 µg m⁻³ higher surface PM_{2.5} background than that of GFS-CMAQ. In this case, the WRF-CMAQ model shows better agreement with observations (Fig. 4d). It is possible that the GFS-CMAQ's nighttime PBL heights (wind speeds) are too shallow (weak) in this case, which does not allow enough transport of pollutants downstream (to the eastern USA). Overall, GFS-CMAQ and WRF-CMAQ show mixed performance with respect to PM_{2.5} predictions during the August 2019 period: GFS-CMAQ has better PM_{2.5} prediction over the western US, and WRF-CMAQ yields better results over the region east of the Rocky Mountains (Fig. 6).

4 Model comparisons against the FIREX-AQ aircraft data

From late July to early September 2019, the joint NOAA–NASA FIREX-AQ field campaign (<https://csl.noaa.gov/projects/firex-aq/>, last access: 24 October 2022) employed a suite of satellites, aircraft, vehicles and ground site platforms aimed at observing, analyzing and characterizing air

pollutants emitted from wildfire sources over the CONUS (Ye et al., 2021). The FIREX-AQ airborne measurements provide a 3D dataset from various meteorological, gas and aerosol instruments that can be used to verify the GFS-CMAQ and WRF-CMAQ model performance while also elucidating reasons for any model differences. Here, the focus of the FIREX-AQ model comparison and verification is against observations taken primarily from the NASA DC-8 aircraft, which include meteorological variables, gaseous and aerosol concentrations, and aerosol optical properties merged at a 1 min temporal resolution. The model data are spatiotemporally interpolated to the flight paths for comparison. The majority of the FIREX-AQ flights were over the western US, and they sampled within environments that both were and were not (see Sect. 4.1) influenced by wildfire emissions (https://daac.ornl.gov/MASTER/guides/MASTER_FIREX_AQ_JulySept_2019.html, last access: 24 October 2022). During a cluster of major wildfire events (see Sect. 4.2), the DC-8 sampled both near-source and aged smoke plumes between 2 and 8 August 2019 (i.e., the Williams Flats, Snow Creek and Horsefly fires) across the states of ID, WA and MT.

4.1 Comparison of the 22 July non-wildfire event over the central California Valley

On 22 July, the DC-8 aircraft flew from California to Boise, ID, while maintaining a relatively low altitude (< 1 km a.s.l., above sea level) over the California Central Valley (Fig. 7). This flight was not impacted by any major wildfire event and was mainly controlled by anthropogenic emissions and local meteorological conditions. Figure 7 shows that the GFSv16

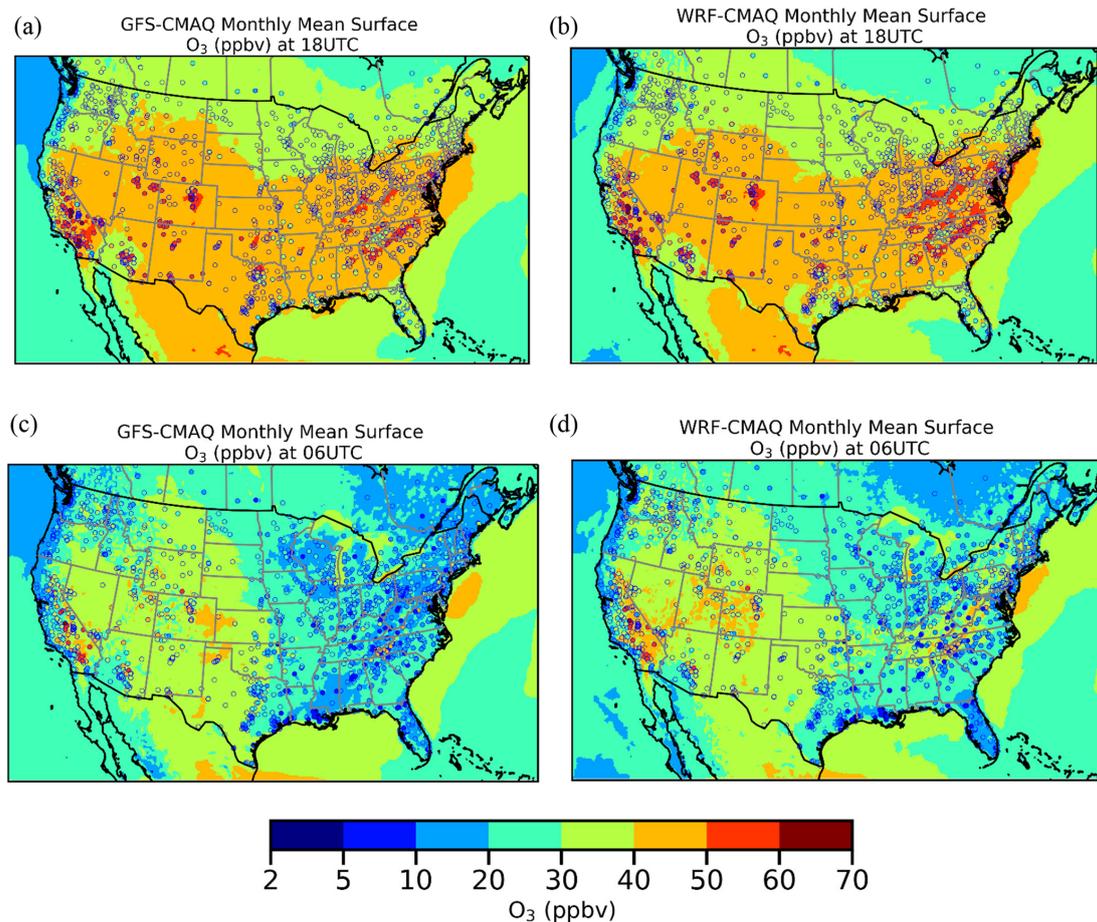


Figure 5. Monthly mean surface O_3 predictions by GFS-CMAQ (a, c) and WRF-CMAQ (b, d) for daytime (a, b) and nighttime (c, d) compared with the corresponding AIRNow observations for August 2019.

and WRF models had similar meteorological temperature and humidity predictions and that both models have dry and warm biases over the Central Valley at lower altitudes (Fig. 7d, e) (Qian et al., 2022). GFS's horizontal wind speeds tend to have a stronger variability than WRF (Fig. 7b), especially in high altitudes. With respect to wind direction, WRF shows a better prediction than GFS around 20:00 and 24:00 UTC (Fig. 7c).

Both GFS-CMAQ and WRF-CMAQ underestimate the vertical wind (W) variability by at least 1 order of magnitude, and WRF-CMAQ has weaker W variability than that of GFS-CMAQ, especially at high altitudes (Fig. 7f). The model vertical velocities are not directly from the GFS nor the WRF model; rather, they are re-diagnosed in CMAQ to conserve mass (Otte and Pleim, 2010) and, thus, represent the whole layer's vertical movement across the $12\text{ km} \times 12\text{ km}$ grid cell. With its flight speed of around 80 to 240 m s^{-1} , the DC-8 aircraft's 1 min average sampling frequency results in an approximate 4.8 to 14 km horizontal scale, respectively, which is comparable with the 12 km CMAQ model resolution. The aircraft observations, however, include turbulence

effects during its 1 min averages, which may not be temporally resolved by the models at this resolution. Thus, both the GFS-CMAQ and WRF-CMAQ vertical velocities are much lower and have almost no correlation with the aircraft observations.

Although both GFS-CMAQ and WRF-CMAQ have reasonable comparisons for most meteorological variables, including the horizontal winds, it continues to be a challenge to compare them with the observed vertical velocities. Thus to further elucidate the model–observation differences in vertical motion, Fig. 8 shows a curtain plot of vertical velocities along the flight path from the two models. As WRF-CMAQ remains on a native grid, its wind fields tend to be more balanced and have lower variability compared with the GFS-CMAQ wind fields. The stronger variability in W for GFS-CMAQ could be caused by GFS's non-hydrostatic dynamics or CMAQ's effort to counteract mass-inconsistency effects from the interpolated horizontal wind fields (Byun, 1999b). Our comparison shows that the first factor should be the major one (Fig. S2) for this event, as the GFS-CMAQ-diagnosed W is very similar to that from the original GFSv16 around

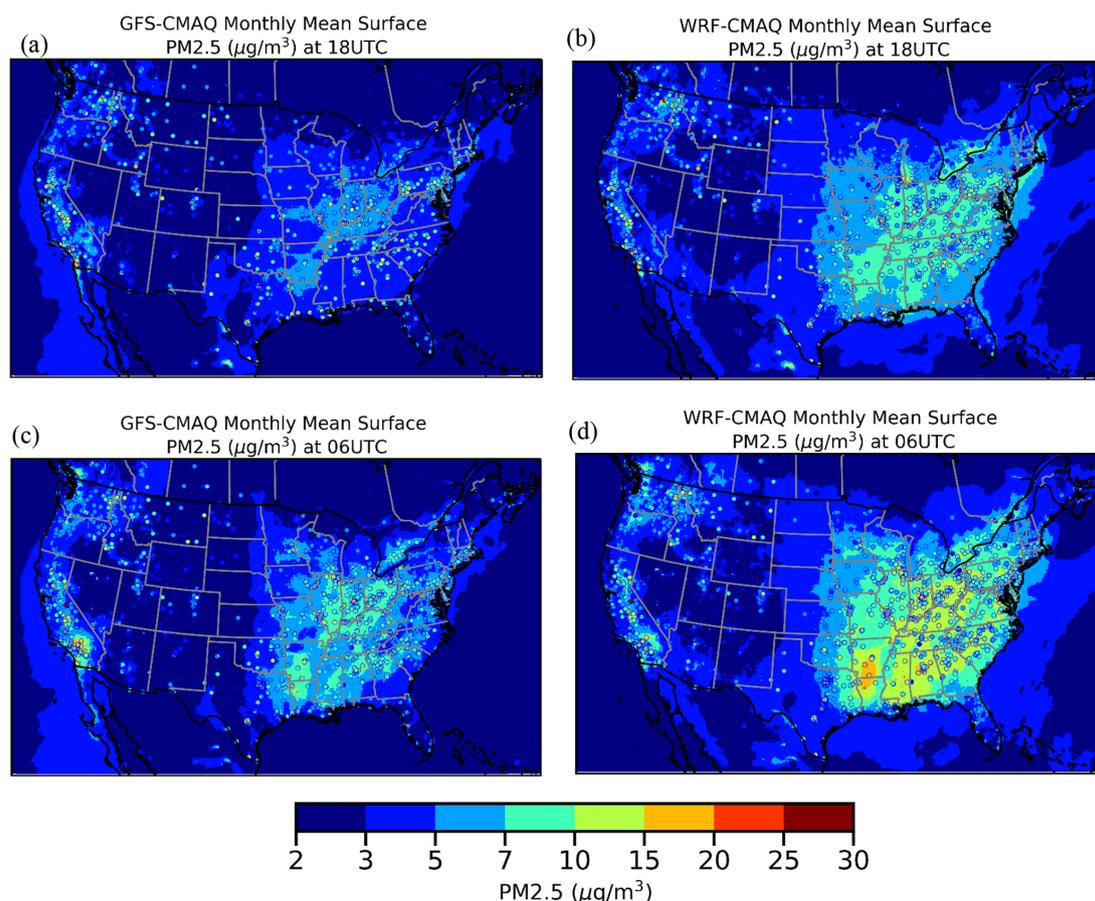


Figure 6. Same as Fig. 5 but for surface PM_{2.5}.

1 km a.g.l. (above ground level). As the original GFSv16 also has similar stronger vertical velocities compared with the original WRF, the W difference between GFS-CMAQ and WRF-CMAQ is unlikely to be due to interpolation error in the horizontal winds.

GFS-CMAQ and WRF-CMAQ yield similar results overall for specific chemical species during this DC-8 flight (Fig. 9). Both models underestimate CO, O₃ and ethane (C₂H₆) concentrations over the lower altitudes in the California Central Valley. Over the same flight segment, they have better NO_x (NO + NO₂) and ethene (C₂H₄) predictions, implying that the emissions of these two species have better accuracy than those of CO and C₂H₆. Figure 9f shows that the two models also underestimate NO_z (NO_y–NO_x), or the oxidized nitrogen species besides NO_x, indicating that photochemical O₃ production may also be underestimated. NO_z is a good indicator of the O₃ photochemical formation (Sillman et al., 1997), where the O₃/NO_z ratio represents the O₃ photochemical efficiency per NO_x oxidation product. Thus, NO_z and O₃ are typically highly correlated over regions with active photochemical production. Our later analysis shows that the models tend to underestimate certain hydrocarbons, such as C₂H₆, which is likely linked to O₃ and NO_z underes-

timations, as the hydrocarbons are photochemical precursors of O₃ and NO_z.

The two models show slight differences in peak values of CO, C₂H₄ and NO_x around 23:30 UTC: the GFS-CMAQ-predicted concentrations are slightly higher and closer to observations (Fig. 9). These differences are due to their PBL predictions (both from the corresponding meteorological model outputs): GFS-CMAQ has a lower PBL height and weaker emission vertical dilution compared with WRF-CMAQ (Fig. 8). However, GFS-CMAQ tends to underpredict O₃ more (Fig. 9b) due to its higher NO_x titration. This implies that the effects of the transport and nonlocal transformation of O₃ could be stronger than that of local precursor emissions. WRF-CMAQ has higher NO_z (Fig. 9f) but lower NO_x compared with GFS-CMAQ due to the time lag of O₃ and NO_z photochemical formation. Consequently, the peak O₃ values may not be well correlated with the emitted precursors, such as NO_x and VOCs. Furthermore, the modeled peak C₂H₆ and C₂H₄ concentrations do not occur at the same time around 23:30 UTC, whereas observations indicate that these two species should be highly correlated in this region. This model mismatch implies that the VOC speciation fac-

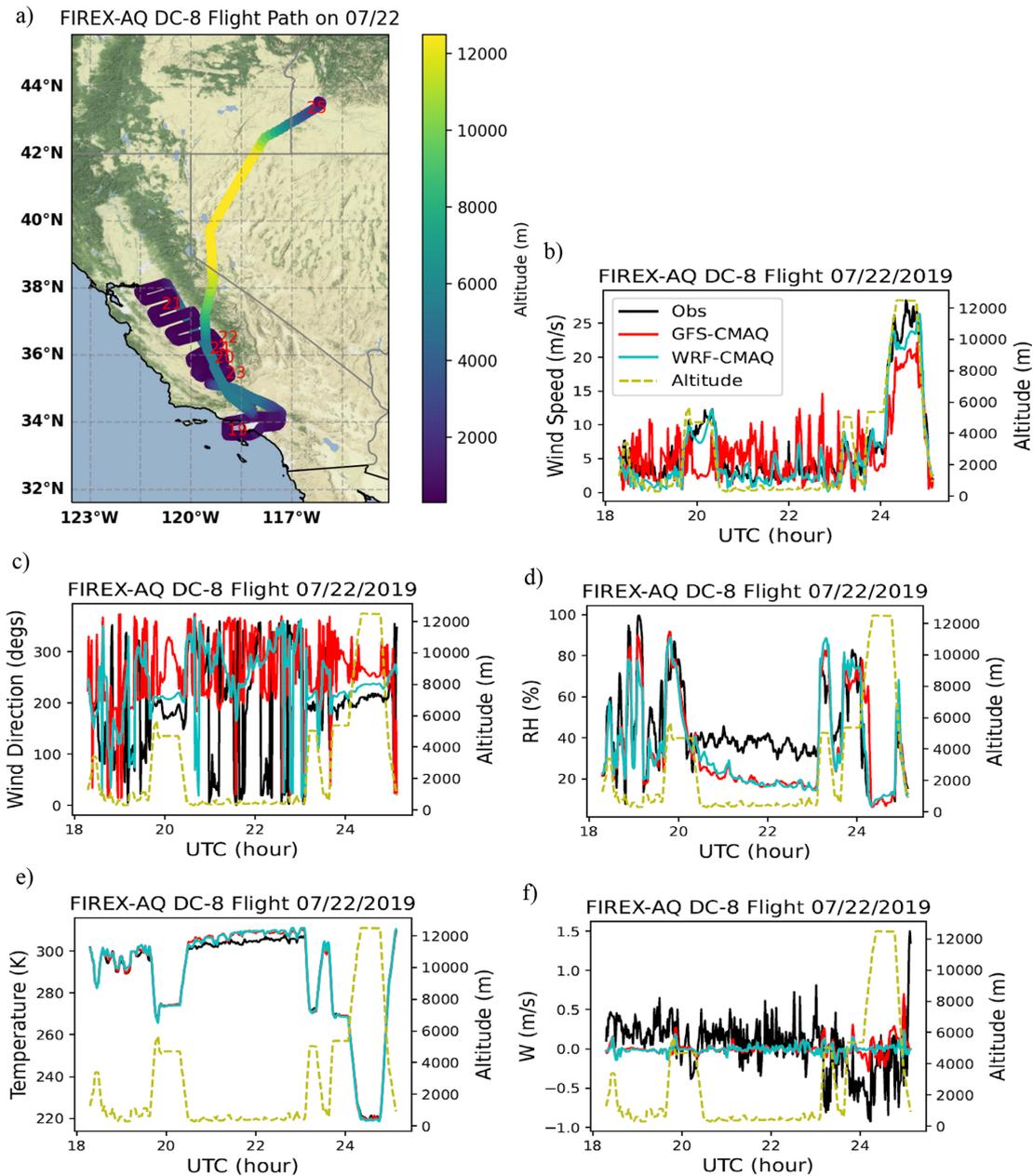


Figure 7. Modeled meteorological variables compared with observations for the DC-8 flight on 22 July 2019 (b–f). Panel (a) shows the flight path color-coded by altitude above sea level with the UTC time given in red text. Base map credits: © OpenStreetMap contributors 2022. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

tors for a certain area or emission sector need to be improved over Southern California.

4.2 Comparison of the 6 August wildfire events over the northwestern US

On 6 August, the DC-8 observed a cluster of three wildfires: the Williams Flats fire (47.98° N, 118.624° W; 80 km to the northwest of Spokane, WA), the Snow Creek fire (47.703° N, 113.4° W; 32 km northeast of Condon, MT) and the Horse-

fly fire (46.963° N, 112.441° W; 24 km east of Lincoln, MT). Figure 10a shows the flight path on that date: the DC-8 aircraft departed from Boise, ID; flew over the Williams Flats fire region; flew to Montana to sample the Snow Creek and Horsefly fires (i.e., Montana fires); and finally returned to the Boise base. The aircraft flew below 8 km for most flight segments near the fire plumes. Figure S3 shows the corresponding GOES-16 satellite true-color image, where these 6 August fires and the associated smoke plumes are visible and can be distinguished from the cloud bands to the south

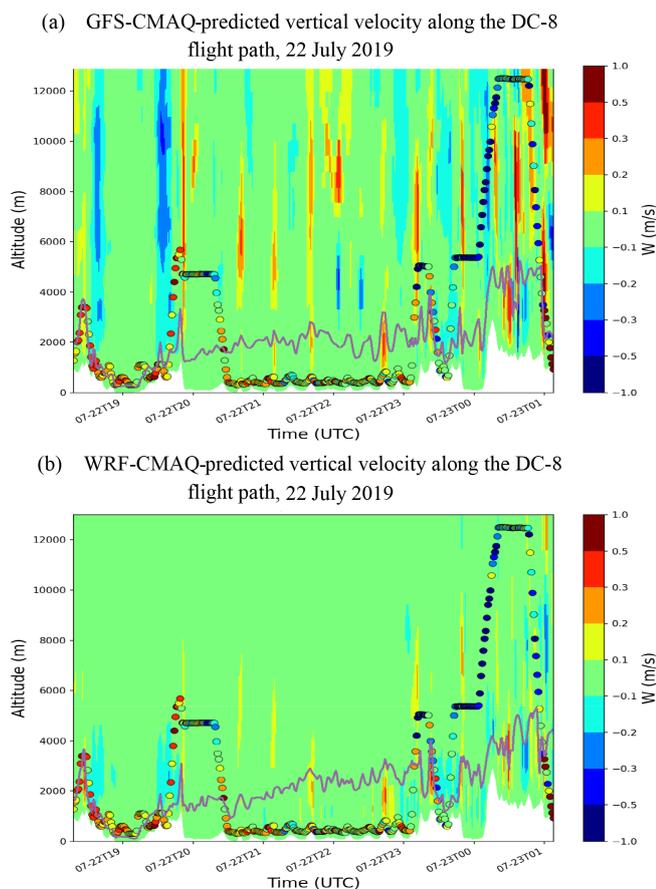


Figure 8. Curtain plots of the vertical velocity (W) predicted by GFS-CMAQ (a) and WRF-CMAQ (b) along the DC-8 flight on 22 July 2019. The colored dots show the DC-8-measured vertical velocities, and the solid lines show the predicted PBL heights of these two models.

that move northward later that day (Fig. S3). The Williams Flats fire was ignited by lightning and was the largest fire event sampled during the FIREX-AQ campaign, burning from about 2–8 August 2019.

Both models significantly underpredicted CO (Fig. 10c), submicron (aerosol diameter $< 1 \mu\text{m}$) organic aerosol (OA) (Fig. 10e) and the aerosol optical extinction coefficient (AOE) (Fig. 10f), suggesting an issue with the GBBEPx gas and aerosol emissions. The models performed well for NO_2 during the Williams Flats and Montana fires below 6 km a.s.l., but there were prominent underestimations for the high-altitude flight segments (Fig. 10d). However, as the NO_2 instrument (the NOAA NO_yO_3 four-channel chemiluminescence instrument) had an NO_2 detection limit of around 0.01 ppb (https://airbornescience.nasa.gov/sites/default/files/documents/NOAANOyO3_SEAC4RS.pdf, last access: 24 October 2022), the models might not truly underestimate NO_2 for these flight segments with extremely low NO_2 . WRF-CMAQ predicted higher O_3 values than the GFS-CMAQ, which generally agreed better with observations for

the Williams Flats fire (Fig. 10b). However, for the Montana fires ($\sim 23:00\text{--}24:00$ UTC), WRF-CMAQ has higher O_3 biases, and GFS-CMAQ yields better results. The difference in O_3 is largely driven by the regional background concentration difference between the two models: WRF-CMAQ tends to have higher domain-wide O_3 concentrations than GFS-CMAQ due to the meteorological effects discussed in Sect. 3, even though they used the same lateral boundary conditions.

Figure S4 shows the spatial overlay comparison of vertically averaged GFS-CMAQ predictions at 21:00 UTC and the DC-8 flight observations for the altitude of 1–3 km a.g.l. on 6 August 2019. The peak NO_2 observation around 48°N , 118.5°W indicates the general location of the Williams Flats fire. The GBBEPx emissions and GFS-CMAQ prediction showed shifted peak-value locations driven by the westerly modeled winds. For this flight, the GBBEPx had stronger NO_x fire emissions over two Montana locations than that over Williams Flats. The model overpredicts the column-averaged NO_2 concentrations, especially over the Montana fires, which can not be reflected by the point-by-point NO_2 comparison result in Fig. 10d. For this flight, the mean GFS-CMAQ NO_2 along the flight path for 1–3 km a.g.l. is about 0.125 ppbv compared with the observed mean NO_2 of 0.169 ppbv, and the model indeed showed an NO_2 underprediction along the flight path. However, in this case, the flight path did not encounter the locations with modeled peak NO_2 values, as the model misplaced the plumes, especially over the Montana fires, leading to this inconsistency. With respect to the O_3 comparison (Fig. S4b), this inconsistency could also exist, although it may not be as significant as the inconsistency for the high-gradient NO_2 concentrations. In the GFS-CMAQ prediction, the high O_3 concentrations are almost co-located with high NO_2 concentrations (Fig. S4b), but the observations did not show this feature. Instead, some high- O_3 flight segments had relatively low NO_2 concentrations, such as those circled in the black rectangle in Fig. S4b. The observed NO_x titration was not able to be produced by the 12 km models. Wang et al. (2021) used a 100 m horizontal resolution large-eddy simulation and demonstrated the capability of using such techniques to capture some high-resolution features of fire plumes and the associated chemical behavior. While such high-resolution techniques are not currently feasible for the operational NAQFC, they demonstrate the limitation of using regional-scale ($12 \text{ km} \times 12 \text{ km}$) models to capture such fine-scale features of the fire plume.

GFS-CMAQ has higher wildfire-related CO, NO_2 , OA and AOE values that are closer to observations than WRF-CMAQ for the Montana fires between 23:00 and 24:00 UTC at flight altitudes of $\sim 4\text{--}5$ km (Fig. 10c, d, e, f). As these two models use the same GBBEPx emissions and wildfire plume rise algorithm (Briggs, 1965), the differences should be due to other reasons. To help explain these model differences, Fig. 11a and b show the aerosol backscatter coefficients (ABCs) retrieved by the differential absorption high-spectral-resolution lidar (DIAL-HSRL) aboard the DC-8 air-

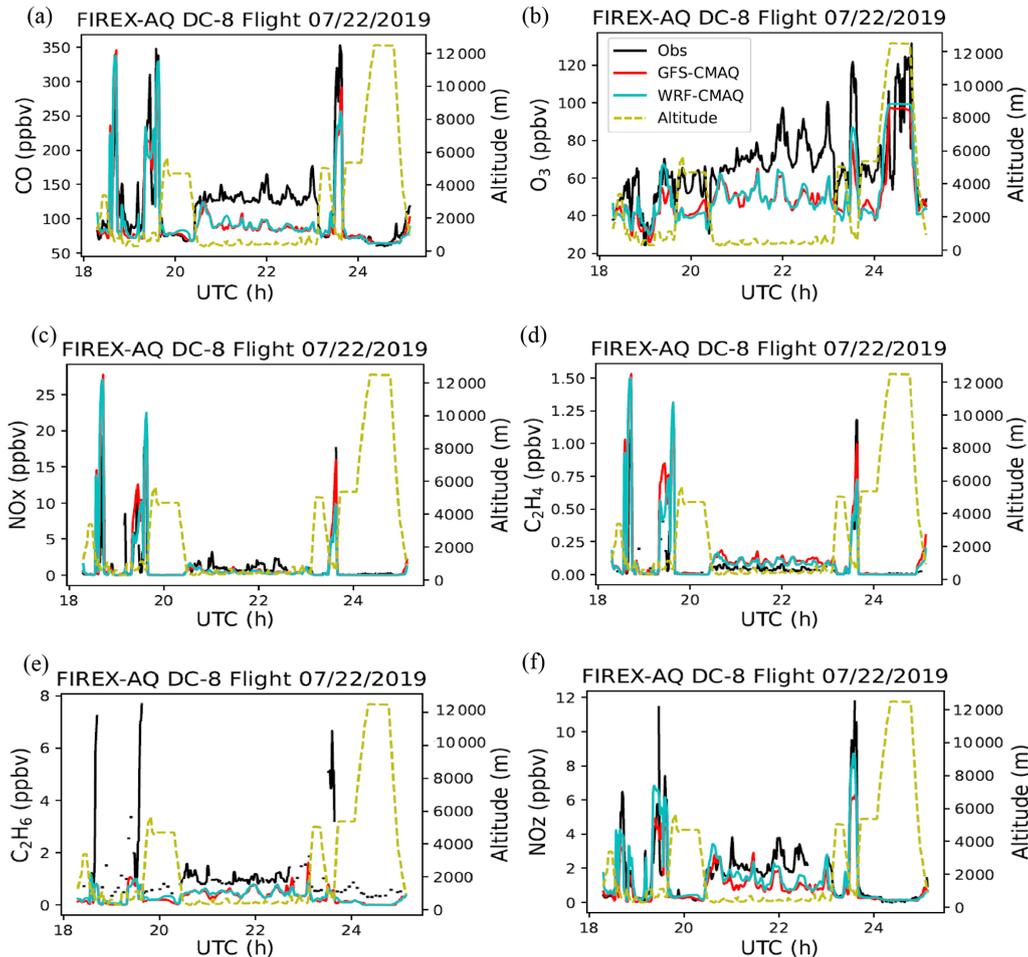


Figure 9. Model-predicted chemical concentrations compared with observations along with the DC-8 flight on 22 July 2019.

craft without and with cloud screening, respectively. It shows that the major fire plumes of the William Flats fire were below 4 km ($\sim 19:00\text{--}22:00$ UTC), but the Montana fires ($\sim 23:00\text{--}24:00$ UTC) extended from the surface up to 6 km, with some detached plumes reaching 10 km. The model-predicted AOE's have an overall similar pattern, with major plumes below 4 km for the Williams Flats fire (Fig. 11c, d). Over the Montana fires, the GFS-CMAQ predicts a slightly higher PBL height, thereby allowing the fire plume to reach a higher height near the DC-8 cruising altitude. In contrast, the WRF-CMAQ wildfire plumes are slightly lower than the aircraft flight path around $23:00\text{--}24:00$ UTC, which leads to underpredictions of the fire-emitted species (Fig. 11d).

An interesting feature in the DIAL observations is the detached plume from 8 to 10 km altitude (Fig. 11a): some cirrus clouds existed in this region, and the DIAL retrieval could not distinguish whether they were pure clouds or clouds mixed with elevated aerosols above 8 km. The cloud-screened image (Fig. 11b) mainly showed the enhanced aerosols below 7 km and some scattered signals near the high cloud edges. Cloud mixing with aerosols was usual for fire-induced

clouds, or pyrocumulonimbus (Peterson et al., 2021). Although, in this event, the middle-sized fires did not show evidence of inducing high-altitude clouds, the indicators of mixed clouds and aerosols at high altitudes still existed: both OA measured in situ (Fig. 10e) and the AOE (Figs. 10f; 11c, d) showed the enhanced aerosols around 01:00 UTC of the next day above 8 km. This elevated plume was generally captured by the GFS-CMAQ simulation, although its strength was underestimated (Fig. 11c); however, this feature was completely missed in WRF-CMAQ (Fig. 11d). Considering the altitude range of the detached plume, the major model disparities are likely due to model convection differences in the free troposphere. To further investigate this impact, Fig. 11e and f show curtain plots of relative humidity (RH) predicted by the two models. GFS-CMAQ yields higher RH at such altitudes (10 km) compared with WRF-CMAQ around $23:00\text{--}24:00$ UTC, indicating that GFS-CMAQ has a stronger convection. The CMAQ model uses input meteorology to diagnose convection activity and drive its Asymmetric Convective Model, version 2 (ACM2) convection scheme. This convective activity is apparent in GOES-16 satellite im-

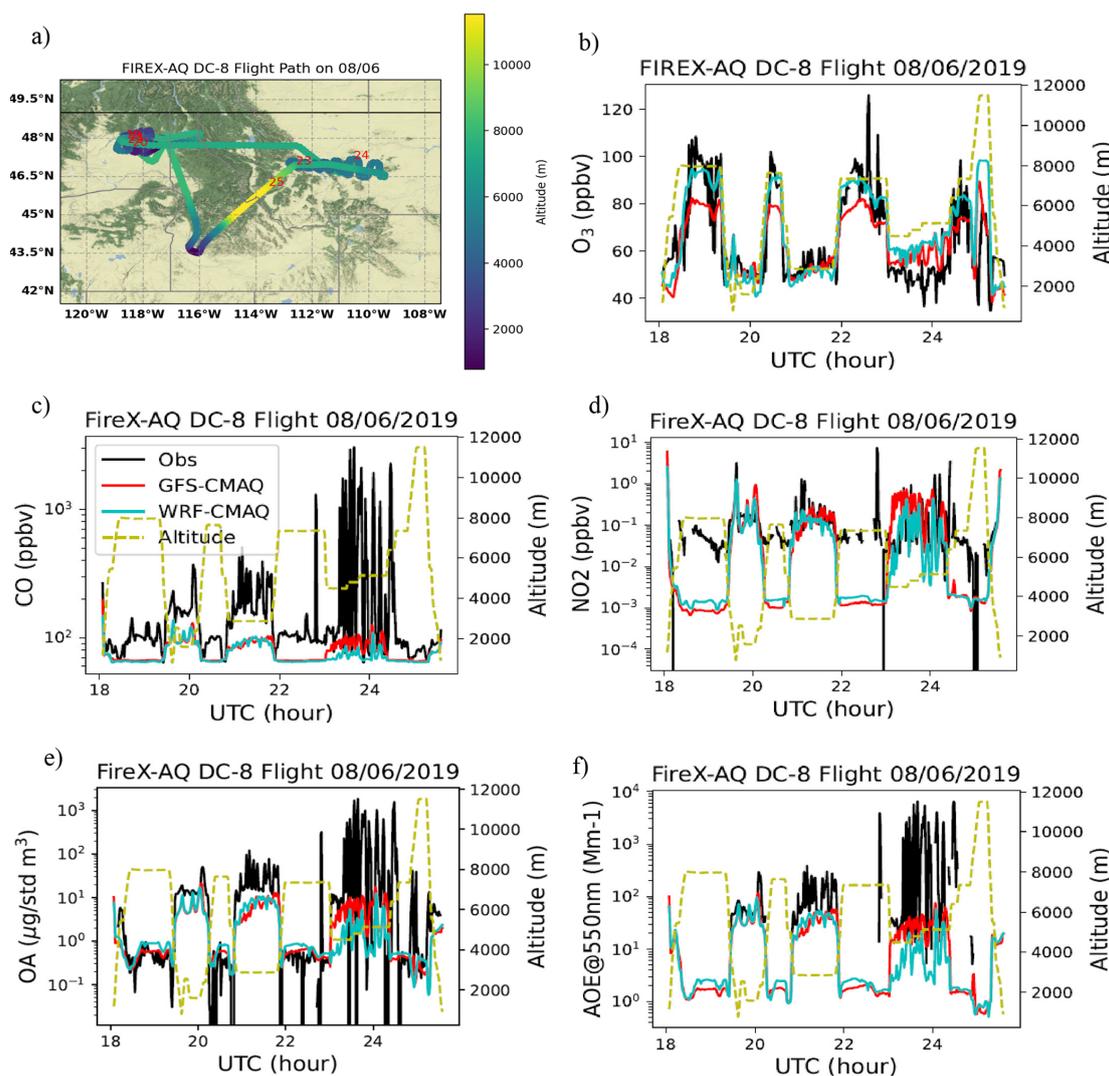


Figure 10. The DC-8 flight path (a) as well as model–observation comparisons for O₃ (b), CO (c), NO₂ (d), submicron organic aerosol (OA) (e) and the aerosol optical extinction coefficient (AOE) at a wavelength of 550 nm (f) on 6 August 2019. Base map credits: © OpenStreetMap contributors 2022. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

ages (Fig. S3), as more fractional clouds appeared ahead of the northward-moving frontal band. Both the GFSv16 and WRF models used here *do not* consider the fire heat feedback effect; thus, their predicted convection and clouds are only driven by the synoptic weather conditions. If such synoptic-to-mesoscale weather models consider wildfire heat feedback effects, their predictions may result in stronger convection and help correct their underpredictions of PBL heights.

4.3 Statistical results of model performance for FIREX-AQ

4.3.1 Meteorological statistics

During the FIREX-AQ field campaign, the DC-8 aircraft performed more than 20 flights over the CONUS with detailed

observations of various chemical compounds. Tables 2 and 3 show the statistical results of the mean bias (MB), normalized mean bias (NMB), root-mean-square error (RMSE), correlation coefficient (R) and linear regression/slopes for the two models' performance over the western US (west of 110° W) only at low altitudes (< 3 km a.s.l.) for both non-fire and fire flight segments. The FIREX-AQ aircraft data included the smoke flag to mark the sampling times associated with fire plumes, identified by CO and aerosol enhancement over background levels in downwind areas of specific fires. This smoke flag is used to distinguish the flight segments with and without fire influences. Most of these flights departed from Boise, ID, except for the 22 July flight that flew from California to Idaho. As a result, they mainly flew over Idaho and its surrounding regions. The GFS tends to have a slightly higher wind speed with a positive MB, whereas

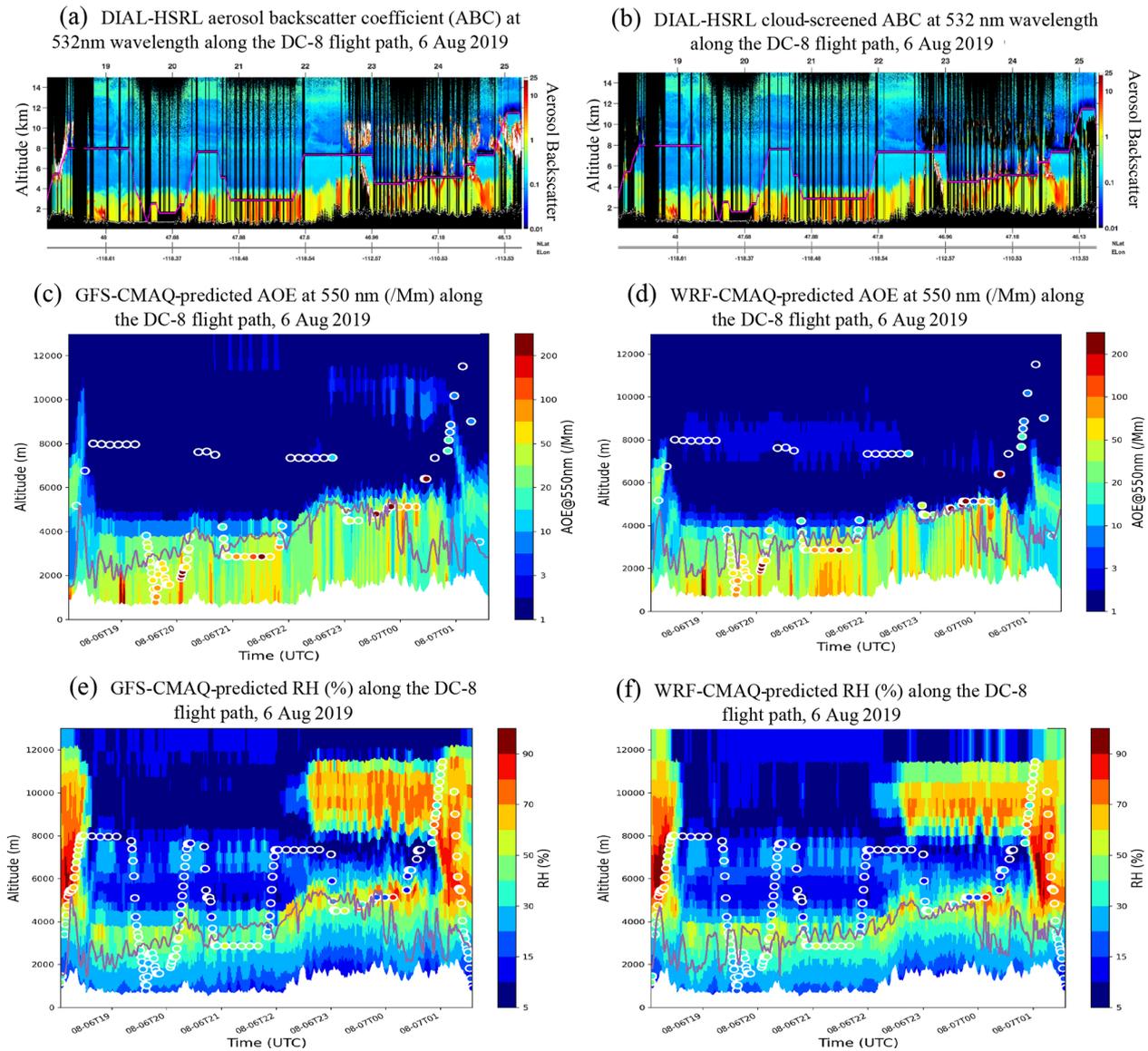


Figure 11. The differential absorption high-spectral-resolution lidar (DIAL-HSRL) retrieved aerosol backscatter coefficients (ABCs) at a 532 nm wavelength in steradians per kilometer (a), the cloud-screened ABCs (b), curtain plots of the AOE (b, c), and relative humidity (RH) predicted by (d) GFS-CMAQ and (e) WRF-CMAQ along the DC-8 flight on 6 August 2019. The colored dots show the corresponding measured values, and the solid lines show the predicted PBL heights of these two models.

WRF has a small negative wind speed bias. Most of the DC-8 flights are during the daytime, and the GFS has a higher daytime wind speed than WRF at low altitudes. The GFS and WRF have very similar temperature predictions. For the RH, the GFS predictions are slightly drier than those of WRF, especially for non-fire events. The meteorological models do not consider wildfire heat effects and, thus, may have (in part) led to slightly warm MBs for the non-fire events (Table 2) and slightly cool MBs for the fire events (Table 3). Because both the GFSv16 and WRF models have similar MB shifts from an average temperature overprediction (Table 2; non-fire events) to an underprediction (Table 3; wildfire events),

we can estimate that the fire effects cause roughly a 1–2 K temperature enhancement to the background along the DC-8 flight paths below 3 km. This estimate assumes that the model temperature biases are generally representative of the western US (west of 110° W) and are independent of the averaged flight segments that have different locations and periods in Tables 2 and 3. Correspondingly, the air masses are drier in the sampled wildfire plumes, as shown by the large reduction in the RH underpredictions (i.e., negative MBs) from Tables 2 to 3.

Table 2. Statistics of the two models compared to the observations for DC-8 flight segments without fire influences below 3 km a.s.l. over the region west of -100° W. All aerosols have a diameter of less than 1 μm . The normalized mean bias (NMB) is given as a percentage.

Variables	Obs mean	GFS-CMAQ					WRF-CMAQ				
		MB	NMB	RMSE	R	Slope	MB	NMB	RMSE	R	Slope
Temperature (K)	295	0.979	0.332	2.04	0.988	1.13	1.16	0.393	2.28	0.989	1.17
RH (%)	35.6	-7.3	-20.5	11.8	0.781	0.717	-6.05	-17	12.6	0.677	0.598
Wind speed (m s^{-1})	4.81	0.758	15.8	3.25	0.432	0.473	-1.11	-23.1	2.4	0.666	0.524
O ₃ (ppbv)	57.9	-10.7	-18.5	15	0.651	0.34	-10.4	-17.9	14.1	0.717	0.413
CO (ppbv)	134	-37.6	-28	53.2	0.654	0.573	-37.1	-27.7	52.9	0.652	0.572
NO _x (ppbv)	1.11	0.507	45.6	2.9	0.704	1.15	0.345	31.1	2.86	0.695	1.12
NO _y (ppbv)	2.56	-0.0418	-1.63	3.07	0.743	0.892	0.055	2.15	3.14	0.724	0.86
NO _z (ppbv)	1.63	-0.465	-28.6	1.17	0.782	0.553	-0.125	-7.66	1.08	0.788	0.721
HONO (ppbv)	0.00432	0.012	279	0.0438	0.379	0.444	0.0134	311	0.0487	0.358	0.48
HNO ₃ (ppbv)	0.291	0.154	53.1	0.421	0.683	1.34	0.337	116	0.65	0.708	1.89
PAN (ppbv)	0.399	-0.251	-63	0.416	0.675	0.221	-0.222	-55.6	0.386	0.681	0.284
NH ₃ (ppbv)	3.55	-0.801	-22.6	5.26	0.0481	0.038	-1.58	-44.5	4.37	0.304	0.155
C ₂ H ₄ (ppbv)	0.121	0.0582	48.1	0.189	0.702	0.869	0.0385	31.9	0.187	0.682	0.836
C ₂ H ₂ (ppbv)	0.146	-0.0734	-50.3	0.137	0.784	0.496	-0.0696	-47.7	0.137	0.771	0.494
SO ₂ (ppbv)	0.342	-0.235	-68.8	0.567	0.0238	0.00835	-0.221	-64.5	0.568	-1.26×10^{-3}	-0.00047
Acetone (ppbv)	2.74	-2.28	-83.1	2.45	0.686	0.192	-2.2	-80.4	2.38	0.668	0.199
HCHO (ppbv)	2.1	-0.972	-46.4	1.26	0.559	0.447	-0.909	-43.4	1.25	0.513	0.442
CH ₃ CHO (ppbv)	0.736	-0.326	-44.2	0.538	0.647	0.386	-0.349	-47.4	0.554	0.643	0.38
Benzene (ppbv)	0.0449	-0.0193	-43	0.057	0.398	0.385	-0.0191	-42.6	0.0564	0.397	0.375
Toluene (ppbv)	0.039	0.0409	105	0.153	0.759	1.74	0.0352	90.1	0.14	0.762	1.63
Isoprene (ppbv)	0.073	0.0361	49.4	0.174	0.6	0.838	0.00661	9.06	0.145	0.648	0.797
EC ($\mu\text{g m}^{-3}$, STP)	0.108	0.191	177	0.572	0.518	2.09	0.228	211	0.609	0.455	1.88
OA ($\mu\text{g m}^{-3}$, STP)	10.9	-7.15	-65.7	9.72	0.565	0.263	-6.48	-59.5	9.45	0.495	0.243
Sulfate ($\mu\text{g m}^{-3}$, STP)	1.31	-0.781	-59.7	1.11	0.0856	0.0188	-0.773	-59	1.11	0.0322	0.00677
NH ₄ ⁺ ($\mu\text{g m}^{-3}$, STP)	0.745	-0.615	-82.5	0.805	0.416	0.103	-0.596	-79.9	0.778	0.509	0.145
Nitrate ($\mu\text{g m}^{-3}$, STP)	1.22	-1.08	-88.1	1.49	0.562	0.229	-1.04	-85.3	1.45	0.57	0.279
AOE (Mm^{-1})	54.5	-29.3	-53.8	47	0.593	0.227	-27.4	-50.2	45.9	0.588	0.227

STP denotes standard temperature and pressure.

4.3.2 Chemical statistics for flight segments without fire influences

For most chemical species, the two models also have similar performance, indicating that the emissions and chemistry are major driving forces. For flight segments not encountering fire plumes, both models overpredict NO_x, HNO₃, toluene, elemental carbon (EC) and ammonium (NH₄⁺), but they underestimate peroxyacetyl nitrate (PAN), benzene, C₂H₂, SO₂, and submicron sulfate and OAs (Table 2). The SO₂ and submicron sulfate underprediction may be impacted by underestimated NEIC2016v1 SO₂ emissions over the western US. As point sources, including power plant emissions, are the SO₂ sources, this comparison implies that the point sources for 2019 events have large uncertainties.

Although the models agree well with NO_y observations, they disproportionately underestimate NO_z (non-NO_x reactive nitrogen species, or NO_y-NO_x), as shown by the regression slopes and MBs. The NO_x and NO_y observations have different missing data, and NO_z is calculated when both NO_x and NO_y observations are available at certain sampling times. Due to the different sample number issues, their observed averages may not exactly match well (averaged NO_x + NO_z \neq NO_y in observations), although their corre-

sponding modeled relationship are well balanced. Gaseous NO_z species can be split into inorganic NO_z (e.g., HNO₃, HONO, HNO₄, NO₃, ClONO₂ and N₂O₅) and organic NO_z (e.g., PAN, methyl peroxy acetyl nitrate – MPAN and the other organic nitrate – RNO₃). The precursors of organic NO_z include hydrocarbons. One of the important organic NO_z species is PAN, and both models underestimate PAN for the flight segments without fire influences (Table 2). The carbonyl precursors of PAN include acetaldehyde (CH₃CHO) (44 % of the global source), methylglyoxal (30 %), acetone (7 %), and a suite of other isoprene and terpene oxidation products (19 %) (Fischer et al., 2014). CH₃CHO and acetone are also underestimated (Table 2), which helps to explain the underestimation of PAN. For the oxidized hydrocarbons, like aldehydes (e.g., HCHO and CH₃CHO), their main atmospheric sources come from the oxidation of highly reactive VOCs, including alkanes, alkenes and aromatics, instead of direct emissions (Parrish et al., 2012). Therefore, the underestimations of HCHO and CH₃CHO are associated with the underestimations of their precursor hydrocarbons, including anthropogenic and biogenic VOCs. Our internal comparison with some limited surface VOC observations indicated that BEIS tends to underpredict biogenic emissions over the

Table 3. Table 3 is the same as Table 2 except that it displays the wildfire-affected flight segments.

Variables	Obs mean	GFS-CMAQ					WRF-CMAQ				
		MB	NMB	RMSE	R	Slope	MB	NMB	RMSE	R	Slope
Temperature (K)	287	-0.389	-0.135	0.702	0.995	1.01	-0.688	-0.24	0.863	0.997	1.04
RH (%)	27.8	-0.761	-2.74	7.84	0.712	0.553	4.3	15.5	11.1	0.556	0.534
Wind speed (m s ⁻¹)	5.42	0.766	14.1	2.16	0.612	0.616	-0.811	-15	2.12	0.604	0.556
O ₃ (ppbv)	55.7	-6.61	-11.9	11.8	0.587	0.262	-7.01	-12.6	11.5	0.653	0.346
CO (ppbv)	486	-377	-77.6	873	0.596	0.0347	-383	-78.8	883	0.442	0.0242
NO _x (ppbv)	2.63	0.06	2.28	6.41	0.465	0.231	-0.619	-23.5	7.02	0.31	0.153
NO _y (ppbv)	7.32	-4.19	-57.3	13.3	0.507	0.123	-4.66	-63.7	14.2	0.31	0.073
NO _z (ppbv)	5.7	-4.8	-84.3	10.2	-0.189	-0.0106	-4.68	-82	10.2	-0.204	-0.0121
HONO (ppbv)	0.283	-0.274	-96.8	1.18	0.355	0.0043	-0.274	-96.8	1.18	0.291	0.00457
HNO ₃ (ppbv)	0.148	0.148	99.7	0.256	0.532	1.07	0.179	121	0.28	0.402	0.768
PAN (ppbv)	0.971	-0.793	-81.7	1.63	0.27	0.0195	-0.765	-78.8	1.61	0.279	0.026
NH ₃ (ppbv)	17.7	-12.3	-69.3	28.3	0.379	0.0654	-13.7	-77.4	29.6	0.232	0.0386
C ₂ H ₄ (ppbv)	4.5	-4.34	-96.3	10.2	0.421	0.00498	-4.36	-96.8	10.2	0.14	0.0018
C ₂ H ₂ (ppbv)	1.04	-1.01	-96.9	2.08	0.534	0.00866	-1.01	-97	2.09	0.363	0.00623
SO ₂ (ppbv)	0.699	-0.322	-46.1	1.38	0.589	0.198	-0.392	-56.1	1.5	0.429	0.132
Acetone (ppbv)	3.54	-3.2	-90.3	4.56	0.13	0.00862	-3.18	-89.7	4.55	0.135	0.0112
HCHO (ppbv)	8.17	-7.13	-87.3	17.8	0.232	0.0062	-7.19	-88	17.8	0.119	0.00303
CH ₃ CHO (ppbv)	3.65	-3.18	-87.4	9.13	0.186	0.00547	-3.21	-88	9.2	-0.027	-0.00097
Benzene (ppbv)	0.683	-0.67	-98.1	1.84	0.54	0.00432	-0.672	-98.3	1.84	0.367	0.00275
Toluene (ppbv)	0.451	-0.436	-96.6	1.36	0.402	0.00491	-0.438	-97	1.36	0.195	0.00245
Isoprene (ppbv)	0.095	-7.9 × 10 ⁻³	-8.29	0.234	0.123	0.0579	-0.033	-34.7	0.242	-0.014	-0.00541
EC (µg m ⁻³ , STP)	1.89	-0.53	-28	3.28	0.612	0.295	-0.787	-41.6	3.7	0.448	0.195
OA (µg m ⁻³ , STP)	156	-146	-93.4	420	0.612	0.0174	-147	-94.2	423	0.472	0.0122
Sulfate (µg m ⁻³ , STP)	0.791	-0.116	-14.7	0.676	0.415	0.184	-0.214	-27.1	0.728	0.322	0.13
NH ₄ ⁺ (µg m ⁻³ , STP)	1	-0.591	-59.1	0.931	0.767	0.351	-0.615	-61.5	0.956	0.729	0.359
Nitrate (µg m ⁻³ , STP)	1.7	-0.56	-32.9	1.47	0.805	0.613	-0.634	-37.2	1.59	0.774	0.599
AOE (Mm ⁻¹)	391	-350	-89.3	994	0.688	0.027	-357	-91.1	1010.	0.532	0.0152

STP denotes standard temperature and pressure.

western US (e.g., isoprene in Table 2). In this comparison, most anthropogenic hydrocarbons are disproportionately underestimated, except toluene, implying a VOC speciation issue in the NEIC2016v1 anthropogenic emissions (Table 2). A previous study discovered that a model overprediction in toluene was also related to the toluene speciation in the NEIC emission inventory (Lu et al., 2020). In this comparison, both models tend to underpredict organic NO_z, which is likely caused by the underestimation of certain VOCs.

Submicron ammonium (NH₄⁺) and nitrate ion (NO₃⁻) are also underestimated by both models during non-fire events (Table 2), suggesting there are NH₃ underestimates due to either insufficient NH₃ emissions or exaggerated NH₃ removal processes. There are, however, overpredictions in the intermediate species nitric acid (HNO₃), indicating a shift in the gas–aerosol equilibrium partitioning of the nitrate ion. This implies that HNO₃ accumulates in the atmosphere because the modeled NO_x and inorganic NO_z (such as NO₃) pathways toward the nitrate ion and organic NO_z are reduced due to underestimations of their other precursors (NH₃ and VOCs).

There are underestimations in the VOC and CO concentrations that contribute to the O₃ underestimation during non-fire flight segments (Table 2). These non-fire com-

parisons also highlight that both models have similar biases due to similar meteorology (Sect. 4.3.1) as well as the use of the same anthropogenic emissions (NEIC2016v1), BEIS biogenic emissions and chemical models/mechanisms (i.e., CMAQv5.3.1). The differences in the two models' bias, error and correlation/slope are much smaller than their individual magnitudes. As discussed above, VOC speciation in the emission inventory could be one issue, as the model tends to overpredict C₂H₄ but underestimate species such as C₂H₂ and C₂H₆. Some common biases over certain regions could be related to certain common issues. For instance, some power plants that were supposed to shut down in the original NEIC2016 inventory might still have been emitting pollutants during the flight observations, leading to the disagreement with respect to SO₂.

4.3.3 Chemical statistics for flight segments with fire influences

The WRF-CMAQ and GFS-CMAQ models significantly underestimate CO, VOC, HONO and OA for fire-influenced flight segments at low altitudes (< 3 km) over the western US (Table 3). In conjunction with underestimated GBBEPx emissions during these wildfire events, other possible causes for the average statistical underprediction are the CMAQ

model's 12 km horizontal resolution and the flight sampling coverage. Most of the fires that are averaged in the statistics, such as the Horsefly (5.5 km² burning area) and Snow Creek fires (7.3 km² burning area), are at a much finer scale than the model grid. Only the largest Williams Flats fire, with a total burning area of 180 km² (Ye et al., 2021), had a comparable horizontal scale to the model resolution.

The DC-8 aircraft had many flight segments near wildfire sources during the fire events in Table 3; thus, dilution of the emissions due to the relatively coarse model resolution may lead to underestimations in the predicted slope for most wildfire-emitted pollutants, such as CO and OA (Table 3). The O₃ concentrations are also underestimated; however, the O₃ underpredictions are reduced from the non-fire (Table 2) to fire events (Table 3). Abundant amounts of wildfire-emitted NO_x can titrate O₃ near the fire source region, and the models likely underestimate these titration effects due to the 12 km model resolution (Fig. S4). Thus, the models cannot capture the strong spatial O₃ variability that is observed due to both reduction near source regions and enhancement in downstream areas. Again, for this fire event comparison, both models showed similar behavior, and their differences were relatively small compared with the overall model biases.

5 Summary and discussion

The operational NOAA/NWS National Air Quality Forecast Capability (NAQFC) recently underwent a major upgrade on 20 July 2021. The advanced NAQFC includes the recent Community Multiscale Air Quality (CMAQ) model version 5.3.1 with the CB6 (carbon bond version 6) AERO7 (version 7 of the aerosol module) chemical mechanism, and it is driven by the latest operational Finite-Volume Cubed-Sphere (FV3) Global Forecast System, version 16 (GFSv16) (Campbell et al., 2022). Here, we analyze the impacts of the driving meteorological models on CMAQ model performance with the new GFSv16 interpolation-based meteorology versus the commonly used native-grid Weather Research and Forecasting (WRF) model version 4.0.3 meteorology. The meteorological and chemical analysis includes both 2D ground-based and 3D aircraft measurements during summer 2019, which encompasses the joint NOAA–NASA Fire Influence on Regional to Global Environments and Air Quality (FIREX-AQ) campaign. As CMAQ has existing mass conservation via adjustments of the contravariant vertical velocity (Otte and Pleim, 2010), the NACC interpolated GFSv16 wind field can be well handled in CMAQ (i.e., GFS-CMAQ).

The different NOAA/NWS operational GFS and commonly chosen WRF physics schemes employed in this study (Table 1) clearly have impacts on temperature, horizontal/vertical wind fields, PBL heights and the corresponding CMAQ model predictions. During this study period over the western US, both models showed a moisture dry bias and a

temperature warm bias at low altitudes, which could be due to the issue mentioned by Qian et al. (2020): the irrigation contribution being neglected (Sect. 3.1) as well as impacts from soil moisture deficits on surface fluxes in both models. Due to their different physics, GFS has a stronger diurnal variation in the PBL height (lower at night and higher during daytime) over the western and northeastern US. The differences in the GFS and WRF physics result in a larger impact than the difference between interpolated and native grids on the models' meteorological and air quality predictions, despite using FDDA to nudge WRF simulation toward the GFSv16 data. Nudging toward observations or including data assimilation may yield different results for the WRF run, although this is not used here. In this study, FDDA nudging was used in WRF to avoid growing errors across a continuous 1-month simulation. We note that if this method would have been turned off, the differences between GFSv16 and WRF predictions would have been even greater. This would further substantiate the dominance of using different model physics and their impacts on CMAQ model predictions. Campbell et al. (2022) present detailed comparisons for interpolated and original fields, and they are very consistent. In this study, we further compare the CMAQ vertical velocity diagnosed from the interpolated GFS horizontal wind, which is very consistent with the original GFS vertical velocity. Overall, the results of this study further corroborate the use of the GFSv16 data and NACC interpolation-based methods (Campbell et al., 2022) for regional CMAQ model applications in the scientific community.

Over the CONUS, GFS-CMAQ demonstrated lower mean surface O₃ (by about 3 ppb) and PM_{2.5} (by about 1 µg m⁻³) than WRF-CMAQ in August 2019 (Sect. 3). In the western US, the GFS has a stronger diurnal variability in the PBL height and a better performance with respect to nighttime 10 m wind speeds compared with the WRF model. The nighttime difference between these two models tends to be more significant than the corresponding daytime difference. Their difference is also impacted by both vertical/convective (mainly daytime) and upstream advective transport differences in GFS-CMAQ and WRF-CMAQ, which somewhat confounds the impact of different meteorological physics on chemical predictions from region to region. This transport effect is more significant on PM_{2.5} than that on O₃, as O₃ has a shorter lifetime and is more sensitive to local emissions in summer. In this study, neither GFS-CMAQ nor WRF-CMAQ show an overwhelming performance advantage over the other, similar to the NMM-CMAQ and ARW-CMAQ comparison in Yu et al. (2012a, b).

GFS-CMAQ and WRF-CMAQ demonstrated rather similar performance for major chemical variables during both FIREX-AQ non-fire (Table 2) and fire (Table 3) events. Both models showed similar biases, indicating that other factors, such as emissions, model resolution and chemistry, could be more important for the model predictions compared with the meteorological differences. The aircraft data compari-

son reveals many common issues in both model systems. One critical issue is whether the flight sampling coverage is comparable to the 12 km model resolution, especially for high-gradient fire emission, such as the case of the 6 August flight (Fig. S4). The observation representation issue also exists in other places, such as the near-surface meteorological comparison between AIRNow stations and METAR stations. Emissions are the driving force for atmospheric composition concentrations. The comprehensive aircraft measurements help verify that the anthropogenic NEIC2016v1 inventory is reasonable overall, except for SO₂, NH₃ and certain hydrocarbons. The wildfire emissions have larger uncertainties, including the emission intensities, pollutant specification and plume rise, as shown by the both models' results.

The NACC interpolation method is advantageous, as it enables one to use the original meteorological driver directly via interpolation, and it avoids running another model such as WRF to drive regional CMAQ applications. It is also faster and more consistent with the original meteorological model (GFS) than using WRF (even with nudging), as WRF's own physics could have a stronger impact. In the current NOAA/NCEP operational GFS-CMAQ system, NACC only takes less than 5 min to process 72 h of data, which saves enough time for CMAQ to forecast an extra 24 h. These aspects can simultaneously benefit real-time forecasting and retrospective air quality applications in the scientific community. NACC can also adapt to quickly use any regional domain globally and may also use other global meteorological data including reanalysis products. This helps mitigate the confounding factors of using different model configurations across the myriad of WRF physics options while also alleviating the difficulty in understanding their impacts on air quality predictions. The operational GFSv16 and associated reanalysis products are well vetted and evaluated across different global agencies and laboratories; thus, they are well suited for regional CMAQ applications using NACC. In fact, there is an ongoing project at NOAA to migrate both the GFSv16 data and NACC software to the Amazon Web Services (AWS) Cloud platform to provide a streamlined product for the user to generate the model-ready meteorological data for any regional CMAQ application globally.

Finally, we note that the current operational GFSv16 has all of the required meteorological variables to drive CMAQ, and users have the option to supply other data (e.g., fractional land use and LAI). GFSv16's C768 grid has a horizontal resolution from 10.21 to 14.44 km, which is close to the NAQFC's 12 km horizontal resolution. One barrier to using this NACC approach is that the original-resolution GFS data files with all of the required variables are very big, even with compression (about 8 GB per time step), and may not be accessible to community users. There is an ongoing effort toward using cloud storage to solve this issue and making this method available to the community. Traditional WRF-CMAQ usually starts from commonly available global meteorological data, such as NCEP or ECMWF reanalysis data,

which have a relatively coarse resolution, and uses WRF to generate all of the meteorological variables needed by CMAQ on the native grid. In some cases, WRF may become the only available method to drive a finer-scale CMAQ model application. WRF's various physics can also be customized for CMAQ simulation over certain regions or under certain meteorological conditions. Both methods have their pros and cons. As shown in this study, GFS and WRF showed mixed performance for driving CMAQ, although they were similar overall.

Code and data availability. The FIREX-AQ field campaign data used in this study are available from <https://www-air.larc.nasa.gov/cgi-bin/ArcView/firexaq> (last access: 16 May 2022) and https://doi.org/10.5067/ASDC/FIREXAQ_Analysis_Data_1 (NASA/LARC/SD/ASDC, 2021). The NACC code used in this study is publicly available from <https://doi.org/10.5281/zenodo.5507489> (Campbell, 2021a) and via GitHub from <https://github.com/noaa-oar-arl/NACC.git> (last access: 5 April 2022). The modified CMAQv5.3.1 for GFS-CMAQ is available from <https://doi.org/10.5281/zenodo.5507511> (Campbell, 2021b) and via GitHub from <https://github.com/noaa-oar-arl/NAQFC> (last access: 5 April 2022).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/gmd-15-7977-2022-supplement>.

Author contributions. YT contributed to the project conceptualization, model runs, software development, data analysis, visualization, investigation and writing the original draft of the paper. PCC was responsible for software development, model runs, data analysis, investigation and draft revision. DT and XZ contributed to the wildfire emissions data. BB was responsible for software development and funding acquisition. FY, JH and HH provided the GFS model data. LP provided the global aerosol model for the lateral boundary condition. PL, RS, AS, JF, IS, JTD and YJ undertook project supervision, project administration and funding acquisition. MY, IB, JF, TR, DB, JS, JLJ, JC, GD, RM, JH, GH, AR and JD contributed to the FIREX-AQ aircraft data.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Financial support. This research has been funded by NOAA's National Air Quality Forecasting Capability (NAQFC) at the National Weather Service Office of Science and Technology Integration (NWS/OSTI) and by NOAA Cooperative Institutes (award

no. NA19NES4320002) at the Cooperative Institute for Satellite Earth System Studies.

Review statement. This paper was edited by Jason Williams and reviewed by three anonymous referees.

References

- Appel, K. W., Bash, J. O., Fahey, K. M., Foley, K. M., Gilliam, R. C., Hogrefe, C., Hutzell, W. T., Kang, D., Mathur, R., Murphy, B. N., Napelenok, S. L., Nolte, C. G., Pleim, J. E., Pouliot, G. A., Pye, H. O. T., Ran, L., Roselle, S. J., Sarwar, G., Schwede, D. B., Sidi, F. I., Spero, T. L., and Wong, D. C.: The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation, *Geosci. Model Dev.*, 14, 2867–2897, <https://doi.org/10.5194/gmd-14-2867-2021>, 2021.
- Arakawa, A. and Lamb, V.: Computational design of the basic dynamical processes of the UCLA general circulation model, *Meth. Comput. Phys.*, 17, 173–265, 1977.
- Baker, K. R., Woody, M. C., Tonnesen, G. S., Hutzell, W., Pye, H. O. T., Beaver, M. R., Pouliot, G., and Pierce, T.: Contribution of regional-scale fire events to ozone and PM_{2.5} air quality estimated by photochemical modeling approaches, *Atmos. Environ.*, 140, 539–554, <https://doi.org/10.1016/j.atmosenv.2016.06.032>, 2016.
- Briggs, G. A.: A Plume Rise Model Compared with Observations, *Journal of the Air Pollution Control Association*, 15, 433–438, <https://doi.org/10.1080/00022470.1965.10468404>, 1965.
- Byun, D. and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system, *Appl. Mech. Rev.*, 59, 51–77, <https://doi.org/10.1115/1.2128636>, 2006.
- Byun, D. W.: Dynamically Consistent Formulations in Meteorological and Air Quality Models for Multiscale Atmospheric Studies. Part I: Governing Equations in a Generalized Coordinate System, *J. Atmos. Sci.*, 56, 3789–3807, [https://doi.org/10.1175/1520-0469\(1999\)056<3789:DCFIMA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<3789:DCFIMA>2.0.CO;2), 1999a.
- Byun, D. W.: Dynamically consistent formulations in meteorological and air quality models for multi-scale atmospheric applications: Part II. Mass conservation issues, *J. Atmos. Sci.* 56, 3808–3820, [https://doi.org/10.1175/1520-0469\(1999\)056<3808:DCFIMA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<3808:DCFIMA>2.0.CO;2), 1999b.
- Byun, D. W. and Ching, J. K. S.: Science algorithms of the EPA models-3 Community Multiscale Air Quality (CMAQ) modeling system, EPA/600/R-99/030, US EPA, 1999.
- Campbell, P. C.: The NOAA-EPA Atmosphere-Chemistry Coupler (NACC) (v1.3.2), Zenodo [data set], <https://doi.org/10.5281/zenodo.5507489>, 2021a.
- Campbell, P. C.: The Advanced National Air Quality Forecast Capability (NAQFC) (v1.1.0), Zenodo [code], <https://doi.org/10.5281/zenodo.5507511>, 2021b.
- Campbell, P. C., Tang, Y., Lee, P., Baker, B., Tong, D., Saylor, R., Stein, A., Huang, J., Huang, H.-C., Strobach, E., McQueen, J., Pan, L., Stajner, I., Sims, J., Tirado-Delgado, J., Jung, Y., Yang, F., Spero, T. L., and Gilliam, R. C.: Development and evaluation of an advanced National Air Quality Forecasting Capability using the NOAA Global Forecast System version 16, *Geosci. Model Dev.*, 15, 3281–3313, <https://doi.org/10.5194/gmd-15-3281-2022>, 2022.
- Caputi, D. J., Faloona, I., Trousdell, J., Smoot, J., Falk, N., and Conley, S.: Residual layer ozone, mixing, and the nocturnal jet in California's San Joaquin Valley, *Atmos. Chem. Phys.*, 19, 4721–4740, <https://doi.org/10.5194/acp-19-4721-2019>, 2019.
- Chen, F. and Dudhia, J.: Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, 129, 569–585, [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2), 2001.
- Chen, J.-H. and Lin, S.-J.: The remarkable predictability of inter-annual variability of atlantic hurricanes during the past decade, *Geophys. Res. Lett.*, 38, L11804, <https://doi.org/10.1029/2011GL047629>, 2011.
- Chen, J.-H. and Lin, S.-J.: Seasonal predictions of tropical cyclones using a 25-km-resolution general circulation model, *J. Climate*, 26, 380–398, <https://doi.org/10.1175/JCLI-D-12-00061.1>, 2013.
- Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., and Brown, P. D.: Atmospheric radiative transfer modeling: A summary of the AER codes, *J. Quant. Spectrosc. Ra.*, 91, 233–244, <https://doi.org/10.1016/j.jqsrt.2004.05.058>, 2005.
- Dong, X., Fu, J. S., Huang, K., Tong, D., and Zhuang, G.: Model development of dust emission and heterogeneous chemistry within the Community Multiscale Air Quality modeling system and its application over East Asia, *Atmos. Chem. Phys.*, 16, 8157–8180, <https://doi.org/10.5194/acp-16-8157-2016>, 2016.
- Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, 108, 8851, <https://doi.org/10.1029/2002JD003296>, 2003.
- Fischer, E. V., Jacob, D. J., Yantosca, R. M., Sulprizio, M. P., Millet, D. B., Mao, J., Paulot, F., Singh, H. B., Roiger, A., Ries, L., Talbot, R. W., Dzepina, K., and Pandey Deolal, S.: Atmospheric peroxyacetyl nitrate (PAN): a global budget and source attribution, *Atmos. Chem. Phys.*, 14, 2679–2698, <https://doi.org/10.5194/acp-14-2679-2014>, 2014.
- Fu, X., Wang, S. X., Cheng, Z., Xing, J., Zhao, B., Wang, J. D., and Hao, J. M.: Source, transport and impacts of a heavy dust event in the Yangtze River Delta, China, in 2011, *Atmos. Chem. Phys.*, 14, 1239–1254, <https://doi.org/10.5194/acp-14-1239-2014>, 2014.
- Grell, G. A., Dudhia, J., and Stauffer, D. R.: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5), NCAR technical Note NCAR TN-398-1-STR, 117 pp., 1994.
- Han, J. and Bretherton, C. S.: TKE-Based Moist Eddy-Diffusivity Mass-Flux (EDMF) Parameterization for Vertical Turbulent Mixing, *Weather Forecast.*, 34, 869–886, <https://doi.org/10.1175/WAF-D-18-0146.1>, 2019.
- Han, J. and Pan, H.-L.: Revision of Convection and Vertical Diffusion Schemes in the NCEP Global Forecast System, *Weather Forecast.*, 26, 520–533, <https://doi.org/10.1175/WAF-D-10-05038.1>, 2011.
- Han, J., Wang, W., Kwon, Y. C., Hong, S.-Y., Tallapragada, V., and Yang, F.: Updates in the NCEP GFS Cumulus Convection Schemes with Scale and Aerosol Awareness, *Weather*

- Forecast., 32, 2005–2017, <https://doi.org/10.1175/WAF-D-17-0046.1>, 2017.
- Harris, L., Chen, X., Putman, W., Zhou, L., and Chen, J. H.: A Scientific Description of the GFDL Finite-Volume Cubed-Sphere Dynamical Core, NOAA technical memorandum OAR GFDL, 2021-001, <https://doi.org/10.25923/6nhs-5897>, 2021.
- Hong, S. Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, *Mon. Weather Rev.*, 134, 2318–2341, <https://doi.org/10.1175/MWR3199.1>, 2006.
- Houyoux, M. R., Vukovich, J. M., Coats, C. J., Wheeler, N. J. M., and Kasibhatla, P. S.: Emission inventory development and processing for the seasonal model for regional air quality (SMRAQ) project, *J. Geophys. Res.*, 105, 9079–9090, <https://doi.org/10.1029/1999JD900975>, 2000.
- Huang, M., Tong, D., Lee, P., Pan, L., Tang, Y., Stajner, I., Pierce, R. B., McQueen, J., and Wang, J.: Toward enhanced capability for detecting and predicting dust events in the western United States: the Arizona case study, *Atmos. Chem. Phys.*, 15, 12595–12610, <https://doi.org/10.5194/acp-15-12595-2015>, 2015.
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *J. Geophys. Res.*, 113, D13103, <https://doi.org/10.1029/2008JD009944>, 2008.
- Janjic, Z. I.: A nonhydrostatic model based on a new approach, *Meteorol. Atmos. Phys.*, 82, 271–285, 2003.
- Jimenez, P. A., Dudhia, J., Gonzalez-Rouco, J. F., Navarro, J., Montavez, J. P., and Garcia-Bustamante, E.: A revised scheme for the WRF surface layer formulation, *Mon. Weather Rev.*, 140, 898–918, <https://doi.org/10.1175/MWR-D-11-00056.1>, 2012.
- Kain, J. S.: The Kain–Fritsch convective parameterization: An update, *J. Appl. Meteor.*, 43, 170–181, [https://doi.org/10.1175/1520-0450\(2004\)043<0170:TKCPAU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2), 2004.
- Kim, Y. J., Eckermann, S. D., and Chun, H. Y.: An overview of the past, present and future of gravity-wave drag parameterization for numerical climate and weather prediction models, *Atmos.-Ocean*, 41, 65–98, <https://doi.org/10.3137/ao.410105>, 2003.
- Krueger, S. K., Fu, Q., Liou, K. N., and Chin, H. N. S.: Improvement of an ice-phase microphysics parameterization for use in numerical simulations of tropical convection, *J. Appl. Meteorol.*, 34, 281–287, <https://doi.org/10.1175/1520-0450-34.1.281>, 1995.
- Lin, Y.-L., Farley, R. D., and Orville, H. D.: Bulk parameterization of the snow field in a cloud model, *J. Clim. Appl. Meteorol.*, 22, 1065–1092, [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2), 1983.
- Lord, S. J., Willoughby, H. E., and Piotrowicz, J. M.: Role of a parameterized ice-phase microphysics in an axisymmetric, nonhydrostatic tropical cyclone model, *J. Atmos. Sci.*, 41, 2836–2848, [https://doi.org/10.1175/1520-0469\(1984\)041<2836:ROAPIP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1984)041<2836:ROAPIP>2.0.CO;2), 1984.
- Lu, Q., Murphy, B. N., Qin, M., Adams, P. J., Zhao, Y., Pye, H. O. T., Efsthathiou, C., Allen, C., and Robinson, A. L.: Simulation of organic aerosol formation during the CalNex study: updated mobile emissions and secondary organic aerosol parameterization for intermediate-volatility organic compounds, *Atmos. Chem. Phys.*, 20, 4313–4332, <https://doi.org/10.5194/acp-20-4313-2020>, 2020.
- Luecken, D. J., Yarwood, G., and Hutzell, W. T.: Multipollutant modeling of ozone, reactive nitrogen and HAPs across the continental US with CMAQ-CB6, *Atmos. Environ.*, 201, 62–72, <https://doi.org/10.1016/j.atmosenv.2018.11.060>, 2019.
- Mlawer, E. J., Taubman S. J., Brown P. D., Iacono M. J., and Clough S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res.*, 102, 16663–16682, <https://doi.org/10.1029/97JD00237>, 1997.
- Monin, A. S. and Obukhov, A. M.: Basic laws of turbulent mixing in the surface layer of the atmosphere, *Contribution Geophysics Institute, Academy of Sciences USSR*, 151, 163–187, 1954 (in Russian).
- Morrison, H., Thompson, G., and Tatarskii, V.: Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes, *Mon. Weather Rev.*, 137, 991–1007, <https://doi.org/10.1175/2008MWR2556.1>, 2009.
- National Emissions Inventory Collaborative (NEIC): 2016v1 Emissions Modeling Platform, <http://views.cira.colostate.edu/wiki/wiki/10202> (last access: 24 October 2022), 2019.
- NASA/LARC/SD/ASDC: FIREX-AQ Analysis and Supplementary Data, NASA Langley Atmospheric Science Data Center DAAC, [data set], https://doi.org/10.5067/ASDC/FIREXQA_Analysis_Data_1, 2021.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A.: A local ensemble Kalman filter for atmospheric data assimilation, *Tellus A*, 56, 415–428, <https://doi.org/10.3402/tellusa.v56i5.14462>, 2004.
- Otte, T. L. and Pleim, J. E.: The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1, *Geosci. Model Dev.*, 3, 243–256, <https://doi.org/10.5194/gmd-3-243-2010>, 2010.
- Otte, T. L., Pleim, J. E., and Pouliot, G.: PREMAQ: A new pre-processor to cmaq for air-quality forecasting, presented at 2004 Models-3 Conference, Chapel Hill, NC, 18–20 October 2004.
- Pan, L., Kim, H., Lee, P., Saylor, R., Tang, Y., Tong, D., Baker, B., Kondragunta, S., Xu, C., Ruminski, M. G., Chen, W., McQueen, J., and Stajner, I.: Evaluating a fire smoke simulation algorithm in the National Air Quality Forecast Capability (NAQFC) by using multiple observation data sets during the Southeast Nexus (SENEX) field campaign, *Geosci. Model Dev.*, 13, 2169–2184, <https://doi.org/10.5194/gmd-13-2169-2020>, 2020.
- Parrish, D. D., Ryerson, T. B., Mellqvist, J., Johansson, J., Fried, A., Richter, D., Walega, J. G., Washenfelder, R. A., de Gouw, J. A., Peischl, J., Aikin, K. C., McKeen, S. A., Frost, G. J., Fehsenfeld, F. C., and Herndon, S. C.: Primary and secondary sources of formaldehyde in urban atmospheres: Houston Texas region, *Atmos. Chem. Phys.*, 12, 3273–3288, <https://doi.org/10.5194/acp-12-3273-2012>, 2012.
- Peterson, D. A., Fromm, M. D., McRae, R. H., Campbell, J. R., Hyer, E. J., Taha, G., Camacho, C. P., Kablick, G. P., Schmidt, C. C., and DeLand, M. T.: Australia’s Black Summer pyrocumulonimbus super outbreak reveals potential for increasingly extreme stratospheric smoke events, *npj Clim. Atmos. Sci.*, 4, 1–16, <https://doi.org/10.1038/s41612-021-00192-9>, 2021.

- Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., and Grell, G. A.: The weather research and forecasting model: Overview, system efforts, and future directions, *B. Am. Meteorol. Soc.*, 98, 1717–1737, <https://doi.org/10.1175/BAMS-D-15-00308.1>, 2017.
- Putman, W. M. and Lin, S.-J.: Finite-volume transport on various cubed-sphere grids, *J. Comput. Phys.*, 227, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>, 2007.
- Qian, Y., Yang, Z., Feng, Z., Liu, Y., Gustafson, W. I., Berg, L. K., Huang, M., Yang, B., and Ma, H. Y.: Neglecting irrigation contributes to the simulated summertime warm-and-dry bias in the central United States, *npj Clim. Atmos. Sci.*, 3, 31, <https://doi.org/10.1038/s41612-020-00135-w>, 2020.
- Rolph, G. D., Draxler, R. R., Stein, A. F., Taylor, A., Ruminiski, M. G., Kondragunta, S., Zeng, J., Huang, H. C., Manikin, G., McQueen, J. T., and Davidson, P. M.: Description and Verification of the NOAA Smoke Forecasting System: The 2007 Fire Season, *Weather Forecast.*, 24, 361–378, <https://doi.org/10.1175/2008waf2222165.1>, 2009.
- Sillman, S., He, D., Cardelino, C., and Imhoff, R. E.: The use of photochemical indicators to evaluate ozone-NO_x-hydrocarbon sensitivity: Case studies from Atlanta, New York, and Los Angeles, *J. Air Waste Manage.*, 47, 1030–1040, <https://doi.org/10.1080/10962247.1997.11877500>, 1997.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 2. NCAR Technical Note NCAR/TND468+STR, <http://www.mmm.ucar.edu/wrf/users/docs/arwv2.pdf> (last access: 24 October 2022), 2005.
- Skamarock, W. C., Snyder, C., Klemp, J. B., and Park, S. H.: vertical resolution requirements in atmospheric simulation, *Mon. Weather Rev.*, 147, 2641–2656, <https://doi.org/10.1175/MWRD-19-0043.1>, 2019.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., and Huang, X.-Y.: A Description of the Advanced Research WRF Model Version 4. NCAR Tech Note, NCAR/TN-556+STR, <https://doi.org/10.5065/1dfh-6p97>, 2021.
- Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R. H.: Implementation and verification of the unified NOAA land surface model in the WRF model, in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 1115, 2165–2170, 2004.
- Wang, S., Coggon, M. M., Gkatzelis, G. I., Warneke, C., Bourgeois, I., Ryerson, T., Peischl, J., Veres, P. R., Neuman, J. A., Hair, J., and Shingler, T.: Chemical Tomography in a Fresh Wildland Fire Plume: a Large Eddy Simulation (LES) Study, *J. Geophys. Res.-Atmos.*, 126, e2021JD035203, <https://doi.org/10.1029/2021JD035203>, 2021.
- Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, *Geosci. Model Dev.*, 4, 625–641, <https://doi.org/10.5194/gmd-4-625-2011>, 2011.
- Yang, F., Tallapragada, V., Kain, J. S., Wei, H., Yang, R., Yudin, V. A., Moorthi, S., Han, J., Hou, Y. T., Wang, J., Treadon, R., and Kleist, D. T.: Model Upgrade Plan and Initial Results from a Prototype NCEP Global Forecast System Version 16, 2020 AMS Conference, Boston, MA, <https://ams.confex.com/ams/2020Annual/webprogram/Paper362797.html> (last access: 24 October 2022), 2020.
- Yarwood, G., Jung, J., Whitten, G. Z., Heo, G., Mellberg, J., and Estes, M.: Updates to the Carbon Bond mechanism for version 6 (CB6), in: 9th Annual CMAS Conference, Chapel Hill, NC, 11–13, 2010.
- Yarwood, Y., Sakulyanontvittaya, T., Nopmongcol, O., and Koo, K.: Ozone depletion by bromine and iodine over the Gulf of Mexico, final report for the Texas Commission on Environmental Quality, <https://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/pm/5821110365FY1412-20141109-enviro-bromine.pdf> (last access: 3 May 2021), November 2014.
- Ye, X., Arab, P., Ahmadov, R., James, E., Grell, G. A., Pierce, B., Kumar, A., Makar, P., Chen, J., Davignon, D., Carmichael, G. R., Ferrada, G., McQueen, J., Huang, J., Kumar, R., Emmons, L., Herron-Thorpe, F. L., Parrington, M., Engelen, R., Peuch, V.-H., da Silva, A., Soja, A., Gargulinski, E., Wiggins, E., Hair, J. W., Fenn, M., Shingler, T., Kondragunta, S., Lyapustin, A., Wang, Y., Holben, B., Giles, D. M., and Saide, P. E.: Evaluation and intercomparison of wildfire smoke forecasts from multiple modeling systems for the 2019 Williams Flats fire, *Atmos. Chem. Phys.*, 21, 14427–14469, <https://doi.org/10.5194/acp-21-14427-2021>, 2021.
- Yu, S., Mathur, R., Pleim, J., Pouliot, G., Wong, D., Eder, B., Schere, K., Gilliam, R., and Rao, S. T.: Comparative evaluation of the impact of WRF/NMM and WRF/ARW meteorology on CMAQ simulations for PM_{2.5} and its related precursors during the 2006 TexAQSGoMACCS study, *Atmos. Chem. Phys.*, 12, 4091–4106, <https://doi.org/10.5194/acp-12-4091-2012>, 2012a.
- Yu, S., Mathur, R., Pleim, J., Pouliot, G., Wong, D., Eder, B., Schere, K., Gilliam, R., and Rao, S. T.: Comparative evaluation of the impact of WRF-NMM and WRF-ARW meteorology on CMAQ simulations for O₃ and related species during the 2006 TexAQSGoMACCS campaign, *Atmos. Pollut. Res.*, 3, 149–162, <https://doi.org/10.5094/APR.2012.015>, 2012b.
- Zhang, X. and Kondragunta, S.: Estimating forest biomass in the USA using generalized allometric model and MODIS land data, *Geographical Research Letter*, 33, L09402, <https://doi.org/10.1029/2006GL025879>, 2006.
- Zhang, X., Kondragunta, S., and Quayle, B.: Estimation of biomass burned areas using multiple-satellite-observed active fires, *IEEE T. Geosci. Remote.*, 49, 4469–4482, <https://doi.org/10.1109/TGRS.2011.2149535>, 2011.
- Zhuang, J., Jacob, D. J., and Eastham, S. D.: The importance of vertical resolution in the free troposphere for modeling intercontinental plumes, *Atmos. Chem. Phys.*, 18, 6039–6055, <https://doi.org/10.5194/acp-18-6039-2018>, 2018.