Evaluation of the offline-coupled GFSv15-FV3-CMAQv5.0.2 in support of the next-generation National Air Quality Forecast Capability over the contiguous United States

Xiaoyang Chen¹, Yang Zhang¹, Kai Wang¹, Daniel Tong², Daniel Tong², Daniel Tong², Pius Lee³, Youhua Tang³, Jianping Huang⁴, Patrick C. Campbell⁴, Jeff Mcqueen⁵, Havaa O.T. Pye⁷, Benjamin N. Murphy⁷, and Daowen Kang⁷

¹Department of Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA
²Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, VA 22030, USA
³Center for Spatial Information Science and System, George Mason University, Fairfax, VA 22030, USA
⁴Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD 20740, USA
⁵National Oceanic and Atmospheric Administration/National Centers for Environmental Prediction/Environmental Modeling Center, College Park, MD 20740, USA
⁶IM Systems Group, Rockville, MD 20852, USA
⁷Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Correspondence: Yang Zhang (ya.zhang@northeastern.edu)
Abstract

The next-generation National Air Quality Forecast Capability (NAQFC) will use the meteorology from Global Forecast System with the new Finite Volume Cube-Sphere dynamic core (GFS-FV3) to drive the chemical evolution of gases and particles described by the Community Multiscale Air Quality modelling system version 5.3 (CMAQ v5.3). CMAQ v5.0.2, a historical version of CMAQ, has been coupled with the North American Mesoscale Forecast System (NAM) model in the current operational NAQFC. An experimental version of the NAQFC based on the offline-coupled GFS-FV3 version 15 with CMAQv5.0.2 modeling system (GFSv15-CMAQv5.0.2), has been developed by the National Oceanic and Atmospheric Administration (NOAA) to provide real-time air quality forecasts over the contiguous United States (CONUS) since 2018. In this work, comprehensive region-specific, time-specific, and categorical evaluations are conducted for meteorological and chemical forecasts from the offline-coupled GFSv15-CMAQv5.0.2 for the year 2019. The forecast system shows good overall performance in forecasting meteorological variables with the annual mean biases of -0.2 °C for temperature at 2-m, 0.4% for relative humidity at 2-m, and 0.4 m s⁻¹ for wind speed at 10-m against the METeorological Aerodrome Reports (METAR) dataset. Larger biases occur in seasonal and monthly mean forecasts, particularly in spring. Although the monthly accumulated precipitation forecasts show generally consistent spatial distributions with those from the remote sensing and ensemble datasets,
moderate-to-large biases exist in hourly precipitation forecasts against the Clean Air Status and Trends Network (CASTNET) and METAR. While the forecast system performs well in forecasting ozone (O\textsubscript{3}) throughout the year and fine particles with a diameter of 2.5 μm or less (PM\textsubscript{2.5}) for warm months (May-September), it significantly overpredicts annual-mean concentrations of PM\textsubscript{2.5}. This is due mainly to the high predicted concentrations of fine fugitive, coarse-mode, and nitrate particle components. Underpredictions in the southeastern U.S. and California during summer are attributed to missing sources and mechanisms of secondary organic aerosol formation from biogenic volatile organic compounds (VOCs) and semi- or intermediate-VOCs. This work identifies possible underlying causes for systematic region- and time-specific model biases, which will provide a scientific basis for further development of the next-generation NAQFC, in particular, derivation of the science-based bias correction methods to improve forecasting skill for O\textsubscript{3} and PM\textsubscript{2.5}.

1. Introduction

Three-dimensional air quality models (3-D AQMs) have been widely applied in real time air quality forecasting (RT-AQF) since the 1990s in the U.S. (Stein et al., 2000; McHenry et al., 2004; Zhang et al., 2012a). The developments and applications of the national air quality forecasting systems based on 3-D AQMs were conducted in the 2000s (Kang et al., 2005; Otte et al., 2005; McKeen et al., 2005, 2007, 2009). Since then,
improvements and significant progress have been achieved in RT-AQF through the further development of AQMs and the use of advanced techniques. For example, more air pollutants in the products, more detailed gas-phase chemical mechanisms and aerosol chemistry, and the implementation of chemical data assimilation were available (Zhang et al., 2012b; Lee et al., 2017). Various AQMs, coupled with meteorological models in either an online or offline manner, were developed and applied in RT-AQF (e.g., Chuang et al., 2011; Lee et al., 2011; Žabkar et al., 2015; Ryan, 2016). The early version of the National Air Quality Forecast Capability (NAQFC) was jointly developed by the U.S. National Oceanic and Atmospheric Administration (NOAA) and the U.S. Environmental Protection Agency (EPA) to provide forecasts of ozone (O₃) over the northeastern U.S. (Eder et al., 2006). Since the first operational version over the contiguous United States (CONUS) (Eder et al., 2009), the NAQFC has been continuously updated and developed to provide more forecasting products (including O₃, smoke, dust, and particulate matter with a diameter of 2.5 μm or less (PM₂.₅)) with increasing accuracy (Mathur et al., 2008; Stajner et al., 2011; Lee et al., 2017).

The forecast skill of a historical NAQFC, which was based on the North American Mesoscale Forecast System (NAM) model (Black, 1994) and the Community Multiscale Air Quality Modeling System version 4.6 (CMAQ v4.6), over CONUS during year 2008 was evaluated by Kang et al. (2010a) for operational O₃ and experimental PM₂.₅ products. Overall, maximum 8-h O₃ was slightly overpredicted over the CONUS.
during the summer, with the mean bias (MB), normalized mean bias (NMB), and correlation coefficient (Corr) of 3.2 ppb, 6.8 %, and 0.65, respectively. The performance of predicted daily mean PM$_{2.5}$ varied: with an underprediction during the warm season and an overprediction in the cool season. The MBs and NMBs during warm/cool seasons were -2.3/4.5 µg m$^{-3}$ and -19.6%/45.1%, respectively. The current version of the U.S. NOAA’s operational NAQFC has provided the air quality forecast to the public for O$_3$ and PM$_{2.5}$ at a horizontal grid resolution of 12 km over CONUS since 2015. It is currently based on the CMAQ v5.0.2 (released May 2014) (U.S. EPA, 2014) coupled offline with the NAM model. Daily mean PM$_{2.5}$ was underpredicted during warm months (May and July 2014) and overpredicted during a cool month (January 2015) over CONUS still persist (Lee et al., 2017).

Efforts have been made to reduce the seasonal and region-specific biases in the historical and current NAQFC. Development and implementation of an analog ensemble bias correction approach was applied to the operational NAQFC to improve forecast performance in PM$_{2.5}$ predictions (Huang et al., 2017). Kang et al. (2008, 2010) investigated the Kalman Filter (KF) bias-adjustment technique for operational use in the NAQFC system. The KF bias-adjusted forecasts showed significant improvement in both O$_3$ and PM$_{2.5}$ for discrete and categorical evaluations. However, limitations in the underlying models and the bias correction/adjustment approaches need further improvement. Characterizing the current NAQFC forecasting skill and identifying the
underlying causes for region- and time-specific biases can result in further development of the NAQFC system and improved pollutant predictions.

In the next generation NAQFC, the NAM will be replaced by the Finite Volume Cube-Sphere Dynamical Core (FV3), the dynamical core in Global Forecast System (GFS). To support this new development, a prototype of the offline-coupled GFS version 15 (v15) with CMAQv5.0.2 (GFSv15-CMAQv5.0.2) has been developed and applied by the NOAA for RT-AQF over CONUS since 2018 (Huang et al., 2018, 2019, 2020). In this work, the meteorological and air quality forecasts from the offline-coupled GFSv15-CMAQv5.0.2 are comprehensively evaluated for the year of 2019. The main objectives of this work are to: (1) evaluate the forecast skills of the experimental prototype of the GFSv15-CMAQv5.0.2 system; (2) identify the major model biases, in particular, systematic biases and persistent region- and time-specific biases in major species; (3) investigate underlying causes for the biases to provide a scientific basis for improving the model representations of chemical processes and developing science-based bias correction methods for $O_3$ and PM$_{2.5}$ forecasts. This work will support NAQFC’s further development and improvement through enhancing its forecasting abilities and generating a benchmark for the operational version of next-generation NAQFC that is being developed by NOAA based on the offline-coupled GFS-FV3 v16 with CMAQ v5.3 (Campbell et al., 2020). Eventually, the latest version of CMAQ (version 5.3), which has updates in gas-phase chemistry (Yarwood et al., 2010; Emery et al., 2015; Luecken et al.,...
lightning nitric oxide (LNO) production schemes (Kang et al., 2019a, 2019b), and secondary aerosol formation (in particular, secondary organic aerosol) (e.g., Pye et al., 2013, 2017; Murphy et al., 2017) among others, will be coupled with GFS-FV3 v16 and be implemented into NAQFC.

2. Model system and evaluation protocols

2.1 Description and configuration of offline-coupled GFSv15-CMAQv5.0.2

FV3 is a dynamical core for atmospheric numerical models developed by the Geophysical Fluid Dynamics Laboratory (GFDL) (Putman and Lin, 2007). It is a modern and extended version of the original FV core with a cubed-sphere grid design and more computationally efficient solvers. It was selected for implementation into the GFS as the next generation dynamical core in 2016 (Zhang et al., 2019a). The GFS-FV3 v15 (GFS v15) has been operational since June 2019. The GFS v15 uses the Rapid Radiative Transfer Method for GCMs (RRTMG) scheme for shortwave/longwave radiation (Mlawer et al., 1997; Iacono et al., 2000; Clough et al., 2005), the Hybrid eddy-diffusivity mass-flux (EDMF) scheme for Planetary Boundary Layer (PBL) (National Centers for Environmental Prediction, 2019a), the Noah Land Surface Model (LSM) scheme for land surface option (Chen et al., 1997), the Simplified Arakawa-Schubert (SAS) deep convection for cumulus parameterization (Arakawa et al., 1979).
1974; Grell, 1993), and a more advanced GFDL microphysics scheme for microphysics (National Centers for Environmental Prediction, 2019b). An interface preprocessor has been developed by NOAA to interpolate data, transfer coordinates, and convert the GFS v15 outputs into the data format required by CMAQv5.0.2 (Huang et al., 2018, 2019). The original outputs from GFS v15, which have a horizontal grid with 13-km resolution and a Lagrangian vertical coordinate with 64 layers in NEMSIO format, are processed to Lambert-Conformal Conic projection by PREMAQ, a preprocessor, to recast the meteorological fields for CMAQ into an Arakawa C-staggering grid (Arakawa and Lamb, 1977) with a 12-km horizontal resolution and 35 vertical layers (Table 1). The first 72 hours in 12:00 UTC forecast cycles from GFS v15 are used to drive the air quality forecast by the offline-coupled GFSv15-CMAQv5.0.2 system.

CMAQ has been continuously developed by the U.S. EPA since the 1990s (Byun and Schere, 2006) and has been significantly updated in many atmospheric processes since then. Chemical boundary conditions for the GFSv15-CMAQv5.0.2 system are mainly from the global 3-D model of atmospheric chemistry driven by meteorological input from the Goddard Earth Observing System (GEOS-Chem). The lateral boundary condition for dust is from the outputs of NEMS GFS Aerosol Component (NGAC) (Lu et al., 2016). The area sources from National Emissions Inventory of year 2014 version 2 (NEI 2014v2), point sources from NEI 2005 with projected sulfur dioxide (SO2) and nitrogen oxide (NOX) to year 2019, and U.S. EPA’s MOVES 2014 mobile sources along
with the biomass burning emission inventory from the Blended Global Biomass Burning Emissions Product system (GBBEPx) (Zhang et al., 2019b) are processed by Sparse Matrix Operator Kerner Emissions (SMOKE) model and the PREMAQ for CMAQ. Biogenic emissions are calculated inline by Biogenic Emission Inventory System (BEIS) version 3.14 (Schwede et al., 2005). Sea-salt emission is parameterized within CMAQ v5.0.2. While the deposition velocities are calculated inline, the fertilizer ammonia bi-directional flux for in-line emissions and deposition velocities is turned off. Detailed configurations of photolysis, gas-phase chemistry, aqueous chemistry, and aerosol chemistry for CMAQ v5.0.2 are listed in Table 1.

2.2 Datasets and evaluation protocols

Comprehensive evaluation of the GFSv15-CMAQv5.0.2 forecasting system is conducted for both meteorological and chemical variables for year 2019, including discrete, categorical, and region-specific evaluations. The products in the first 24-hour of each 72-hour forecast cycle are extracted and combined as a continuous, annual forecast. The evaluation of meteorological variables is carried out for those results from PREMAQ in GFSv15-CMAQv5.0.2 system. Detailed information for datasets used in this study is listed in Table S1. Observed hourly temperature at 2-meters (T2), relative humidity at 2-meters (RH2), precipitation (Precip), wind direction at 10-meters (WD10), and wind speed at 10-meters (WS10) are obtained from the Clean Air Status and Trends Network (CASTNET) and the METeorological Aerodrome Reports (METAR) datasets. The
majority of CASTNET sites are suburban and rural sites. Approximately 1900 METAR sites over CONUS are used in this study (Fig. S1). For evaluation of precipitation, a threshold of ≥0.1 mm is used for valid records because the CASTNET and METAR have different definitions of 0.0 mm values. In CASTNET, the records without any precipitation are filled as 0.0 mm, the same as those records with negligible precipitation. However, in METAR, the records without any precipitation are left as blank, the same as an invalid record. The negligible precipitation is recorded as 0.0 mm.

The air quality forecasting products are evaluated include hourly $O_3$, hourly PM$_{2.5}$, maximum daily 8-hour average $O_3$ (MDA8 $O_3$), and daily average PM$_{2.5}$ (24-h avg PM$_{2.5}$) for chemical forecast. The AIRNow dataset is used for observed hourly $O_3$ and PM$_{2.5}$. It is a near real time (NRT) dataset which has preliminary quality control (QC). Many abnormal records are not quality controlled completely. To filter the abnormal records, the thresholds of 120 ppb and 100 µg m$^{-3}$ for $O_3$ and PM$_{2.5}$ are used, respectively. Remote sensing data from the Global Precipitation Climatology Project (GPCP) and the Climatology-Calibrated Precipitation Analysis (CCPA) (Hou et al., 2014; Zhu and Luo, 2015) datasets are also used for evaluation of precipitation. GPCP is a global precipitation dataset with a spatial resolution of 0.25 degree and a monthly temporal resolution. The CCPA uses linear regression and downscaling techniques to generate analysis product of precipitation from two datasets: the National Centers for Environmental Prediction (NCEP) CPC Unified Global Daily Gauge Analysis and the
NCEP EMC Stage IV multi-sensor quantitative precipitation estimations (QPEs). The CCPA product with a spatial resolution in 0.125 degree and temporal resolution of an hour is used in this study. Satellite-based Aerosol Optical Depth (AOD) at 550 nm from Moderate Resolution Imaging Spectroradiometer (MODIS) Terra platform (Levy et al., 2015) is used for the evaluation of monthly AOD. The statistic measures such as mean bias, the root mean square error (RMSE), the normalized mean bias, the normalized mean error (NME), and the correlation coefficient are used, more details about evaluation protocols are referring to Zhang et al. (2009, 2016). The Taylor diagram (Taylor, 2001), which includes the correlations, NMBs, and the normalized standard deviations (NSD), is used to present the overall performance (Wang et al., 2015). The NMBs ≤ 15% and NMEs ≤ 30% by Zhang et al. (2006) and NMBs (≤ 15% and ≤ 30%), NMEs (≤ 25% and ≤ 50%), and Corr (>0.5 and >0.4) for MDA8 O$_3$ and 24-h PM$_{2.5}$, respectively, by Emery et al. (2017) are considered as performance criteria. Monthly, seasonal, and annual statistics and analysis are included. Seasonal analysis for O$_3$ is separated into ozone season (May-September) and non-ozone season (January-April and October-December). Analysis for ten CONUS regions, defined by U.S. EPA (www.epa.gov/aboutepa), are included and listed in Fig. S1c..

3. Evaluation of model forecast skills

3.1 Evaluation of meteorological forecasts
Discrete performance evaluation is conducted for post-processed meteorological fields from the GFSv15-CMAQv5.0.2 system (Table 2). The GFS v15 can predict well the boundary layer meteorological variables. It has overall cold biases and wet biases for annual T2 and RH2 in 2019, respectively. It also overpredicts WS10, and underpredicts hourly precipitation. Despite CASTNET siting being slightly different from that of METAR, the annual and most of the seasonal performance for the model show similar pattern in terms of bias for both the CASTNET and METAR networks. Mean biases of T2 are mostly within ±0.5 degree Celsius except those in February and March against CASTNET (Table S2). Underprediction is generally larger against CASTNET than METAR. For spatial distribution of MB for seasonal T2 against METAR (Fig. 1), cold biases are mainly found in the Midwest and West U.S. where most of the CASTNET sites are located. GFS v15 usually underpredicts T2 in West Coast, the Mountain States, and the Midwest. Overpredictions of T2 in the states of Kansas, Oklahoma, the areas near the East Coast, and the Gulf Coast offset some underpredictions, resulting in smaller mean biases but similar RMSE for the model against METAR compared to that against CASTNET. The difference between observed T2 from the two datasets is larger in cooler months than warmer months. The largest underpredictions occur in the spring (MAM) season. In general, GFS v15 underpredicts T2 for both CASTNET and METAR, consistent with cold biases found in other studies using GFS v15 (e.g., Yang, 2019). Such underpredictions will affect chemical forecasts, especially the forecast of O3. Consistent
with the overall underpredictions of T2, GFS v15 overpredicts RH2 in general. The largest overprediction is found in spring (MBs of 3.4% and 2.7% with CASTNET and METAR, respectively), corresponding to the largest underprediction of T2 in spring (MBs of -0.5 °C and -0.4 °C with CASTNET and METAR, respectively). GFS v15 shows moderately good performance predicting wind. The annual MB and NMB of WS10 against METAR are 0.4 m s⁻¹ and 10.7 %, respectively. A larger overprediction of WS10 is found with CASTNET than other datasets (Zhang et al., 2016). GFSv15-CMAQv5.0.2 also gives higher overpredictions for CASTNET compared to METAR. The largest biases in wind speed are found in summer. GFSv15-CMAQv5.0.2 gives the largest cold biases, wet biases in spring, indicating the necessity of improving model performance in such seasons in future GFS-FV3 development.

By adopting the threshold of ≥0.1 mm, performance against the CASTNET and METAR show similar results: a large underprediction in hourly precipitation. Predicted monthly accumulated precipitation (Fig. S2) shows consistency in spatial distribution with observations from CCPA (Fig. S3) and GPCP (Fig. S4). The high precipitation in the Southeast are captured well in spring while the high precipitation in the Midwest and South are captured well in other seasons. It indicates that GFSv15-CMAQv5.0.2 has good performance in capturing the spatial distributions of accumulated precipitation but has poor performance in predicting hourly precipitation. In the current version of the experimental GFSv15-CMAQv5.0.2 system, the precipitation from original GFS v15
output is artificially spread out over time during the preprocessing by the interface preprocessor due to the interpolation using a temporal allocation algorithm. Short rains are interpolated into adjacent time steps (Fig. S5). Such an algorithm leads the model and measurements being more consistent for monthly accumulated precipitation than for discrete hourly precipitation from GFS v15 (which will be resolved by NOAA in the next version of NAQFC based on the GFSv16-CMAQ forecasting system).

An overall comparison of performance with CASTNET and METAR datasets is performed using a Taylor diagram (Fig. 2). The normalized standardized deviations (NSDs), Corrs, and NMBs are considered. The NSDs are ratios of variance of predicted values to variance of observed values, following the equations by Wang et al. (2015). The NSDs represent the amplitude of variability. With the NSDs closer to 1, the predicted values have closer variance as the observed values. Consistent with other analysis in this section, larger biases and lower correlation in model wind speed and wind direction are found for CASTNET compared to METAR. The amplitude of variability of WS10 against CASTNET is overpredicted (with the NSD larger than 1), while it is underpredicted against METAR. Because of the post-processing smearing of hourly precipitation, the variance of predicted precipitation is smaller than the observed one, leading to very small NSDs for precipitation. The location of the T2 and RH2 points near the REF marker in the Taylor diagram indicates that the GFSv15-CMAQv5.0.2 is capturing the magnitude and variability of these variables well.
3.2 Evaluation of chemical forecast over the CONUS

Performance of chemical forecasts (i.e. O₃ and PM₂.₅) are evaluated on monthly, seasonal, and annual timescales for the studied period of 2019. Performance of the MDA8 O₃ and the 24-h average PM₂.₅ (24-h avg PM₂.₅) are considered as the primary objectives. Categorical performance evaluations for MDA8 O₃ and 24-h avg PM₂.₅ are also conducted. Table 3 shows the discrete statistics of predicted MDA8 O₃ and 24-h avg PM₂.₅ against AIRNow.

The GFSv15-CMAQv5.0.2 has good performance for MDA8 O₃ on a seasonal and annual basis with MBs ≤ ±1 ppb, NMB ≤ 2.5 %, and NME ≤ 20%. The monthly NMBs/NMEs are within ±15 %/25 %, respectively. Moderate overpredictions and underpredictions are found in both seasons with MB of 0.9 and -0.9 ppb, respectively. The largest underprediction is found in spring months, especially in March. Underprediction of MDA8 O₃ in spring months is consistent with the largest underprediction of T2 in spring. The ozone temperature relationship was found and studied by previous researches (S. Sillman and Samson, 1995; Sillman, 1999). O₃ is expected to increase with increasing temperature within specific range of temperature (Bloomer et al., 2009; Shen et al., 2016). It indicates biases in predicted T2 could be one of the reasons for the corresponding biases in O₃ prediction. Predicted MDA8 O₃ is lower than observed values in major parts of the Midwest and West regions during the O₃
season (Fig. 3 and S7), which is consistent with underprediction of T2 in summer. But GFSv15-CMAQv5.0.2 gives very high O3 in the southeastern U.S., especially in areas near the Gulf Coast. Such overpredictions compensate for moderate underpredictions in Midwest and West, causing an overall overprediction in overall CONUS. Prediction and simulation of O3 in coastal or marine areas are impacted by halogens chemistry and emissions (Adams and Cox, 2002; Sarwar et al., 2012; Liu et al., 2018), including bromine and iodine chemistry (Foster et al., 2001; Sarwar et al., 2015; Yang et al., 2020) and oceanic halogen emissions (Watanabe, 2005; Tegtmeier et al., 2015; He et al., 2016). CMAQ v5.0.2 has only simple chlorine chemistry for CB05 mechanisms, and the reduction of O3 by reaction with bromine and iodine is not included in CMAQ v5.0.2. Iodide-mediated O3 deposition over seawater and detailed marine halogen chemistry has been found to reduce O3 by 1-4 ppb near the coast (Gantt et al., 2017), suggesting the missing halogen chemistry and O3 deposition processes contribute to overpredicted O3 in coastal and marine areas seen here. Coastal and marine areas are also impacted by air-sea interaction processes, which are simply represented in the current meteorological models without coupling oceanic models (He et al., 2018; Zhang et al., 2019c,d). For example, coastal O3 mixing ratios are impacted by predicted sea surface temperatures and land-sea breezes through their influence on chemical reaction conditions and diffusion processes. As discussed in Section 3.1, T2 is significantly overestimated near the Gulf Coast during summer, which could contribute to biases in O3 predictions directly or indicate missing
land-sea breezes and thus missing transport effects in the GFSv15-CMAQv5.0.2 air quality forecasting system. In the non-O_3 season, GFSv15-CMAQv5.0.2 can forecast well the spatial variations of MDA8 O_3 with overall underpredictions in the Northeast.

Unlike the good performance for O_3, GFSv15-CMAQv5.0.2 gives significant overpredictions for 24-h avg PM$_{2.5}$ with annual MB, NMB, and NME of 2.2 µg m$^{-3}$, 29.0%, and 65.3%, respectively (Table 3). The MBs and NMBs range from -0.2 µg m$^{-3}$ to 5.0 µg m$^{-3}$, and -2.6 % to 59.7 % across the four seasons. With the exception of California and the Southeast, predicted 24-h avg PM$_{2.5}$ shows overprediction during most of the year in spring, autumn, and winter (Fig. 4). Using the historical emission inventories from NEI 2011 and NEI 2014 instead of the latest version of NEI 2017 is one of the reasons for the overpredictions of PM$_{2.5}$ concentrations in 2019. Moderate underpredictions of PM$_{2.5}$ are found in California in spring, autumn, and summer.

Murphy et al. (2017) found that secondary organic aerosols (SOA) generated from anthropogenic combustion emissions were important missing PM sources in California prior to CMAQ v5.2. Higher predicted PM$_{2.5}$, typically SOA, in California is expected in the future using GFS-FV3-CMAQv5.3. The largest underpredictions of PM$_{2.5}$ occur in the Southeast in summer. Biogenic volatile organic compounds (BVOCs) and biogenic SOA (BSOA) are most active in Southeast region in summer. Many missing sources and mechanisms for SOA formation from BVOCs have been identified in recent years (Pye et al., 2013, 2015, 2017; Xu et al., 2018) and have resulted in significant improvements in
predicting SOA in the Southeast using CMAQ v5.1 through v5.3. Anthropogenic emissions and aerosol inorganic compounds were found to have impacts on BSOA (Carlton et al., 2018; Pye et al., 2018, 2019). Such interactions and mechanisms are not represented sufficiently in CMAQ v5.0.2, further enhancing the biases in predicted PM$_{2.5}$ in the Southeast. In general, updating NAQFC with CMAQ v5.3 is expected to reduce the biases in California and the Southeast.

Categorical evaluation is conducted to quantify the accuracy of the GFSv15-CMAQv5.0.2 system in predicting events in which the air pollutants exceed moderate or unhealthy categories for the U.S. air quality index (AQI) (www.airnow.gov). The metrics of False Alarm Ratio (FAR) and the Hit Rate (H) are used (Kang et al., 2005; Barnes et al., 2009). The scatter plots for predicted and observed MDA8 O$_3$ and 24-h avg PM$_{2.5}$ are shown in Fig. 5a and Fig. 5b, respectively. The plots are divided into four areas based on whether the predicted and/or observed data fall above or below the AQI thresholds: (a) observed values ≤ thresholds and predicted values > thresholds; (b) observed and predicted values > thresholds; (c) observed and predicted values ≤ thresholds; (d) observed values > thresholds and predicted values ≤ thresholds. The FAR and H are defined in Eq. (1) and Eq. (2):

$$FAR = \frac{a}{a+b} \times 100\%$$

(1)

$$H = \frac{b}{b+d} \times 100\%$$

(2)
Numbers of the scatters in the four areas (a) to (d) are indicated in the equations. The higher the FAR is, the more GFSv15-CMAQv5.0.2 overpredicts the AQI leading to false air quality warnings. The higher the H is, exceedances are more successfully captured by the GFSv15-CMAQv5.0.2 system. In this study, the thresholds for two categories of “Moderate” and “Unhealthy for Sensitive Groups” are considered. Since 2018, they are defined as 55 ppb and 70 ppb for MDA8 O₃ and 12 µg m⁻³ and 35.5 µg m⁻³ for 24-h avg PM₂.₅. For comparison with previous studies, the historical thresholds are also included into the evaluation: 60 ppb and 75 ppb for MDA8 O₃ and 15 µg m⁻³ and 35 µg m⁻³ for 24-h avg PM₂.₅. The metrics in four categories, corresponding to four thresholds, are shown in Fig. 5c. Categorical performance under stricter AQI standards is better than under historical standards. For example, the FAR decreases from 47.8 % to 41.1 %, and the H increases from 40.3 % to 43.9 % with the “Moderate” thresholds change from 60 ppb to 55 ppb. It could be due to the better performance of the forecast system for values closer to the annual average level (~40 ppb). The scatters are more discrete for extreme values (Fig. 5a). When the thresholds of MDA8 O₃ are closer to the average level, the categorical performance increases. The categorical performance of GFSv15-CMAQv5.0.2 in predicting MDA8 O₃ is close to the performance of the previous NAQFC (Kang et al., 2010). Similar improvement in the FAR and H for predicting categorical 24-h avg PM₂.₅ can be found when the threshold changes from 15 µg m⁻³ to 12 µg m⁻³: the FAR decreases from 79.7 % to 70.1 %, and the H increases from
51.9 % to 57.0 %. However, the FAR is high (over 90%) and the H is much lower under the threshold of 35.5 µg m$^{-3}$. It is because most of the false alarms occur when observed 24-h avg PM$_{2.5}$ are lower than 20 µg m$^{-3}$ and the predicted values are higher than 20 µg m$^{-3}$ (Fig. 5b). It shows the poorer performance in correctly capturing the category of “Unhealthy for Sensitive Groups” due to the significant overprediction of PM$_{2.5}$ in cooler months.

Evaluation of predicted AOD against observations from MODIS is shown in Fig. 6. High predicted AOD in the Midwest during cooler months show consistency with MODIS and correspond to high surface PM$_{2.5}$ predictions. High predicted AOD are missing in California, corresponding to underprediction of surface PM$_{2.5}$ in California. In summer months, AOD is largely underpredicted in California and the Southeast region, which may be caused by the previously mentioned missing sources of SOA.

To further analyze the major source for spatial and temporal biases in predicted PM$_{2.5}$, key chemical components of PM$_{2.5}$ in January, July, and August are depicted in Fig. 7. Extremely high particulate sulfate and organic carbon, generated by large wildfires, are carried in from the north boundary in July. The forecast spatial pattern agrees well with the observed AOD in July. High concentrations of PM$_{2.5}$ associated with soil components, unspecified coarse mode components, and high particulate NO$_3^{-}$ concentrations are major contributors to the high PM$_{2.5}$ in the Midwest. The soil components are estimated using the Interagency Monitoring of Protected Visual
Environments (IMPROVE) equation and specific constituents (Appel et al., 2013). These high concentrations are caused by large emissions of anthropogenic primary PM$_{2.5}$, primary coarse PM, ammonia (NH$_3$), and NO$_x$ in the Midwest (Fig. S6). The large emissions of anthropogenic primary course PM, as well as the wind-blown dust are the major sources for soil components and unspecified coarse mode components. Appel et al. (2013) also indicated CMAQ overpredicts soil components, sources of which include fugitive and wind-blown dust, in the eastern United States.

### 3.3 Region-specific evaluation

As discussed in section 3.2, biases in predicted O$_3$ and PM$_{2.5}$ vary from region to region. To further analyze the region-specific performance of the GFSv15-CMAQv5.0.2 system, evaluation for 10 regions within CONUS is conducted. By identifying the detailed characteristics of region-specific biases and indicating the underlying causes for such biases, this section aims to help the NAQFC to improve its forecast ability for specific regions. A science-based bias correction method will be developed for the operational GFS-FV3-CMAQ system in the future. This section can also contribute to hypotheses that may serve as a scientific basis for future bias correction methods.

Figure 8 shows the annual model performance for MDA8 O$_3$ and 24-h avg PM$_{2.5}$ in the 10 CONUS regions. In section 3.2, a slight underprediction of MDA8 O$_3$ on annual
basis was found over the CONUS. MDA8 O₃ is overpredicted in most of the regions except regions 2, 4, and 6 (Fig. 8a). The overpredictions in regions 4 and 6 are mostly from the large biases near the coast area during O₃ season. Correlations between predictions and observations in most of the regions are higher than 0.6, except for 0.55 in region 4 and 0.50 in region 7. Poor performance in regions 4 and 7 is illustrated by the Taylor Diagram (Fig. 8b). Small Corr and NSD, result in the markers of regions 4 and 7 laying farthest from the reference point. The amplitude of variability of the predicted MDA8 O₃ are smaller than observed values in all the regions, especially in regions 4 and 7. The performance in region 2 is the best, with smallest MB/NMB, highest Corr, and similar variability in predictions and observations. The time series of the MDA8 O₃ for the 10 regions during 2019 is shown in Fig. S7. Regions 1, 2, 4, and 6 show different results for the O₃ season and non-O₃ season: GFSv15-CMAQv5.0.2 tends to overpredict MDA8 O₃ during the O₃ season and underpredicts during the non-O₃ season. The underprediction during spring months, which is indicated in section 3.2, can be also found in most of the regions with obvious gaps between observed and predicted curves in March and April. The lowest O₃ predictions occur at 5 am local standard time (LST) in most of the regions (Fig. S8). For regions 4 and 6, significant overprediction occurs not only during the O₃ season for MDA8 O₃ (which mainly occurs during the daytime) but also during the nighttime. During the non-O₃ season, the biases in predicting MDA8 O₃ for regions 4 and 6 are small and consistent with good daytime predictions. However, O₃
is still overpredicted during the nighttime in these regions, associated with the collapse of
the boundary layer and difficulty in simulating its time and magnitude (Hu et al., 2013;
Cuchiara et al., 2014; Pleim et al., 2016).

Consistent with the analysis in section 3.2, PM$_{2.5}$ is significantly overpredicted in
most of the regions except in regions 4, 6, and 9 (Fig. 8c). The underprediction during
warmer months, likely due to missing sources and mechanisms for BSOA, compensate
for the annual biases in regions 4 and 6, leading to smaller MBs/NMBs but low
correlations in these regions. The variability in predictions is much larger than in
observations, with the NSDs >1 for all regions (Fig. 8d). The forecast system has best
performance in region 9 with an NSD of 1.2, an NMB of -12.0%, and a Corr of 0.40. As
discussed in section 3.2, the performance of predicted PM$_{2.5}$ in region 9 is expected to be
further improved with the updates in CMAQ v5.3, specifically the representation of
anthropogenic SOA.

Figure S9 shows the time series of 24-h avg PM$_{2.5}$ in the 10 CONUS regions. The
gaps between observed and predicted curves are large in cooler months, but the
GFSv15-CMAQv5.0.2 system has relatively good performance in warmer months for
most of the regions. Less overprediction is found in regions 6, 8, and 9 during cooler
months, and those regions generally show the best performance (see Taylor Diagram).
The different biases across the regions further indicate that multiple factors likely
contribute to them. To further analyze the underlying causes for varied patterns and
performance on season- and region-specific basis, diurnal evaluations for PM$_{2.5}$ and chemical components of PM$_{2.5}$ during O$_3$ season and non-O$_3$ season are shown in Fig. 9. The GFSv15-CMAQv5.0.2 has a large seasonal variation in diurnal PM$_{2.5}$, inconsistent with the observation. While PM$_{2.5}$ is underpredicted during daytime in regions 4, 6, 8, and 9 during O$_3$ season, PM$_{2.5}$ is always overpredicted across the day during non-O$_3$ season except for region 9. Increased OC, particulate nitrates, soil and unspecified coarse mode components contribute to most of the increase in predicted total PM$_{2.5}$. The general cold biases over CONUS, especially in region 5, could make the GFSv15-CMAQv5.0.2 system predict higher nitrate particulates, leading to larger increase in PM$_{2.5}$ from O$_3$ season to non-O$_3$ season. Emissions vary from month to month in the year (Fig. S10). Larger emissions for NH$_3$, NO$_x$, VOC, primary coarse PM, and primary PM$_{2.5}$ are in O$_3$ season compared to non-O$_3$ season. Primary organic carbons (POC) emissions are higher in O$_3$ season. Changes in emissions are not fully consistent with the changes in PM$_{2.5}$ components, indicating other biases or uncertainty could also contribute to the significant overprediction during non-O$_3$ season. For example, the implementation of bidirectional flux of NH$_3$ and the boundary layer mixing processes under more stable condition (during non-O$_3$ season) in GFSv15-CMAQv5.0.2 system need to be further studied. Pleim et al., (2013, 2019) found that the NH$_3$ fluxes and concentrations could be better simulated and the monthly variations of NH$_3$ concentrations were larger comparing to the raw model by implementing the bidirectional flux of NH$_3$. The absolute biases for diurnal PM$_{2.5}$ are
It is generally larger during nighttime in most of the regions, except for region 9. It is consistent with the analysis by Appel et al. (2013), which suggested that the efforts of improving nighttime mixing in CMAQ v5.0 be further needed, further indicating the need for improvements of CMAQ in predicting dispersion and mixing of air pollutants under stable boundary layer conditions.

4. Conclusion

In this work, the air quality forecast for the year 2019 predicted by the offline-coupled GFSv15-CMAQv5.0.2 system is comprehensively evaluated. The GFSv15-CMAQv5.0.2 system is found to perform well in predicting surface meteorological variables (temperature, relative humidity, and wind) and O\textsubscript{3} but has mixed performance for PM\textsubscript{2.5}. Moderate cold biases and wet biases are found in spring season, especially in March. While the GFSv15-CMAQv5.0.2 system can generally capture the monthly accumulated precipitation compared to remote sensing and ensemble datasets, temporal distributions of hourly precipitation show less consistency with in-situ monitoring data, which is attributed to the interpolation and post-processing in the offline-coupling interface preprocessor.

MDA8 O\textsubscript{3} is slightly overpredicted and underpredicted in ozone and non-ozone seasons, respectively. The cold biases of T2 contribute to the underprediction of MDA8.
O₃ in spring. The significant overprediction near the Gulf Coast, which is caused by missing halogen chemistry and the missing of model representation of the air-sea interaction processes, compensates for underprediction in the West and Midwest in O₃ season for nation-wide metrics. GFSv15-CMAQv5.0.2 has poorer performance in predicting PM₂.₅, comparing to the performance for O₃. Significant overpredictions are found in spring, autumn, and winter, with the largest overprediction in the Midwest, WA, and Oregon, due mainly to high concentrations of predicted soil, unspecified coarse mode and nitrate components. The overall cold biases in the Region 5/Midwest could contribute to higher predicted nitrate particulate matter but overprediction of PM₂.₅ in the region is likely driven by sources containing trace metals such as anthropogenic fugitive dust and wind-blown dust. The forecasting system may be improved through updating the emissions inventory used (i.e., NEI 2014) to NEI 2016v2 or NEI 2017 which are more presentative to the year of 2019 in the next development of next-generation NAQFC.

Categorical evaluation indicates that the GFSv15-CMAQv5.0.2 can capture well the air quality classification of “Moderate” described by the AQI. However, the categorical performance is poorer for PM₂.₅ at the “unhealthy for sensitive groups” threshold due mainly to the significant overprediction during the cooler months. Region-specific evaluation further discusses the biases and underlying causes in the 10 USEPA defined regions in CONUS. An update from CMAQ v5.0.2 to v5.3.1 is expected to alleviate potential errors in missing sources and mechanisms for SOA formation. The
variations of performance in between O₃ and non-O₃ seasons, as well as during the
daytime and nighttime, indicate further studies need to be conducted to improve boundary
layer mixing processes within GFSv15-CMAQv5.0.2. The varied region-specific
performance indicates that improvements, such as bias corrections, should be considered
individually from region to region in the following development of the next generation
NAQFC.

We have used bias analyses in this work to identify several areas of weakness for
further improvement and development of next-generation NAQFC. Further studies are
still needed for improving the emissions, the aerosol chemistry, and the boundary layer
mixing for the future GFS-FV3-CMAQ system. Our work and the further studies can
provide information and scientific basis for the development and implement of a
science-based bias correction method in next-generation NAQFC.

Supplement

The supplement related to this article is available in
gmd-2020-272_supplement.pdf

Code and data availability
The documentation and source code of CMAQ v5.0.2 are available at doi:10.5281/zenodo.1079898. The GFS forecasts in grib2 format are available at https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system -gfs. The GFS forecast inputs in binary (nemsio) format and the coupler used in this study for the GFSv15-CMAQv5.0.2 system are available upon request. The AIRNow data is available for download through the AirNow-Tech website (http://www.airnowtech.org).

The CASTNET data is available for download from https://java.epa.gov/castnet/clearsession.do. The METAR data is available for download from https://madis.ncep.noaa.gov. The GPCP data is available through NOAA website (https://www.ncei.noaa.gov/data/global-precipitation-climatology-project-gpcp-monthly).

The CCPA precipitation is available from https://www.nco.ncep.noaa.gov/pmb/products/gens. The MODIS_MOD04 dataset is available at dx.doi.org/10.5067/MODIS/MOD04_L2.006. The data processing and analysis scripts are available upon request.

Author contribution
YZ and DT defined the scope and focus of the manuscript and designed the model simulations. XC and YZ developed the paper outline and structure. PL, JH, YT, and JM performed the forecast simulations. YT generated the emissions and PC generated the lateral boundary conditions for the model simulations. XC performed the model
evaluation and drafted the manuscript. XC and KW developed postprocessing and
statistical scripts. HP, BM, and DK assisted in analysis of region-specific biases. YZ, HP,
DK, BM, JH, PC, PL, DT, and KW reviewed the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This project is sponsored by NOAA Office of Weather and Air Quality through
grant #NA19OAR4590084 at North Carolina State University, #NA20OAR4590259 at
Northeastern University and #NA19OAR4590085 at George Mason University. Thanks
to Fanglin Yang for providing information regarding GFS v15. High performance
computing at Northeastern University was support by the Stampede XSEDE high
performance computing support under the NSF ACI 1053575.

Disclaimer

The scientific results and conclusions, as well as any views or opinions expressed
herein, are those of the author(s) and do not necessarily reflect the views of NOAA or the
Department of Commerce. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the U.S. EPA. EPA does not endorse any products or commercial services mentioned in this publication.

References


Grell, G. A.: Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations,


Huang, J., McQueen, J., Wilczak, J., Djalalova, I., Stajner, I., Shafran, P., Allured, D., Lee, P., Pan, L., Tong, D., Huang, H.-C., DiMego, G., Upadhayay, S. and Delle Monache, L.: Improving NOAA NAQFC PM 2.5 Predictions with a Bias Correction


Putman, W. M. and Lin, S. J.: Finite-volume transport on various cubed-sphere grids, J.


Stajner, I., Davidson, P., Byun, D., McQueen, J., Draxler, R., Dickerson, P. and Meagher, J.: US National Air Quality Forecast Capability: Expanding Coverage to Include


to the Texas Commission on Environmental Quality, Work Order No. 878-7-84005-FY10-26, 2010.


Tables and Figures

Table 1. Configuration of GFSv15-CMAQv5.0.2 system

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Model Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast period</td>
<td>Jan.-Dec., 2019</td>
</tr>
<tr>
<td>Domain</td>
<td>Continental U.S.</td>
</tr>
<tr>
<td>Resolution</td>
<td>Horizontal: 12 km (442×265); Vertical: 35 layers</td>
</tr>
</tbody>
</table>

**Physical Options**
- Shortwave/longwave radiation: The Rapid Radiative Transfer Method for GCMs
- Planetary boundary layer (PBL): Hybrid eddy-diffusivity mass-flux (EDMF) PBL
- Land surface: Noah Land Surface Model (LSM)
- Microphysics: A more advanced GFDL microphysics scheme
- Cumulus: The Simplified Arakawa-Schubert (SAS) deep convection

**Chemical Options**
- Photolysis: In-line method (Binkowski et al., 2007)
- Gas-phase chemistry: The Carbon Bond mechanism version 5 with active chlorine chemistry and updated toluene mechanism (CB05tucl) (Yarwood et al., 2005; Sarwar et al., 2012)
- Aqueous-phase chemistry: AQCHEM (Sarwar et al., 2011)
- Aerosol module: AERO6 with nonvolatile POA (Carlton et al., 2010; Simon et al., 2012; Appel et al., 2013)

Table 2. Performance statistics of meteorological forecasts

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Variable</th>
<th>Period</th>
<th>Mean Obs.</th>
<th>Mean Sim.</th>
<th>MB</th>
<th>RMSE</th>
<th>NMB, NME, %</th>
<th>Corr</th>
<th>Mean Obs.</th>
<th>Mean Sim.</th>
<th>MB</th>
<th>RMSE</th>
<th>NMB, NME, %</th>
<th>Corr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASTNET</td>
<td>T2, °C</td>
<td>DJF</td>
<td>-0.1</td>
<td>-0.5</td>
<td>-0.4</td>
<td>2.6</td>
<td>-588 -2850</td>
<td>0.96</td>
<td>2.7</td>
<td>2.6</td>
<td>-0.1</td>
<td>2.5</td>
<td>-3.1</td>
<td>69.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAM</td>
<td>9.9</td>
<td>9.4</td>
<td>-0.5</td>
<td>2.4</td>
<td>-5.2 18.2</td>
<td>0.97</td>
<td>12.3</td>
<td>11.9</td>
<td>-0.4</td>
<td>2.3</td>
<td>-3.0</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JJA</td>
<td>21.5</td>
<td>21.4</td>
<td>-0.2</td>
<td>2.4</td>
<td>-0.8  8.6</td>
<td>0.93</td>
<td>23.4</td>
<td>23.1</td>
<td>-0.3</td>
<td>2.3</td>
<td>-1.2</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SON</td>
<td>11.5</td>
<td>11.3</td>
<td>-0.2</td>
<td>2.6</td>
<td>-2.0 16.1</td>
<td>0.97</td>
<td>13.8</td>
<td>13.8</td>
<td>-0.2</td>
<td>2.3</td>
<td>-1.3</td>
<td>12.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annual</td>
<td>10.9</td>
<td>10.6</td>
<td>-0.3</td>
<td>2.5</td>
<td>-3.0 17.0</td>
<td>0.98</td>
<td>13.2</td>
<td>13.0</td>
<td>-0.2</td>
<td>2.3</td>
<td>-1.3</td>
<td>13.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DJF</td>
<td>69.1</td>
<td>71.9</td>
<td>2.8</td>
<td>14.3</td>
<td>4.0 15.1</td>
<td>0.74</td>
<td>74.1</td>
<td>74.4</td>
<td>0.4</td>
<td>13.3</td>
<td>0.5</td>
<td>13.4</td>
</tr>
<tr>
<td>METAR</td>
<td>RH2, %</td>
<td>MAM</td>
<td>62.7</td>
<td>66.1</td>
<td>3.4</td>
<td>14.2</td>
<td>5.4 16.6</td>
<td>0.82</td>
<td>67.4</td>
<td>70.1</td>
<td>2.7</td>
<td>13.8</td>
<td>4.0</td>
<td>15.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JJA</td>
<td>55.0</td>
<td>53.3</td>
<td>-1.7</td>
<td>12.2</td>
<td>-3.2 16.4</td>
<td>0.89</td>
<td>67.0</td>
<td>67.3</td>
<td>0.3</td>
<td>13.1</td>
<td>0.5</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SON</td>
<td>59.0</td>
<td>57.6</td>
<td>-1.4</td>
<td>13.0</td>
<td>-2.4 16.1</td>
<td>0.87</td>
<td>68.7</td>
<td>67.0</td>
<td>-1.7</td>
<td>13.2</td>
<td>-2.5</td>
<td>14.5</td>
</tr>
</tbody>
</table>
Table 3. Performance statistics of chemical variables against AIRNow dataset

<table>
<thead>
<tr>
<th>Period</th>
<th>Mean Obs.</th>
<th>Mean Sim.</th>
<th>MDA8 $O_3$, ppb</th>
<th>24-h avg PM$_{2.5}$, µg m$^{-3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>MB RMSE</td>
<td>NMB, %</td>
</tr>
<tr>
<td>Jan</td>
<td>32.4</td>
<td>32.0</td>
<td>-0.3</td>
<td>7.9</td>
</tr>
<tr>
<td>Feb</td>
<td>36.7</td>
<td>35.7</td>
<td>-1.1</td>
<td>8.4</td>
</tr>
<tr>
<td>Mar</td>
<td>45.1</td>
<td>40.4</td>
<td>-4.7</td>
<td>8.9</td>
</tr>
<tr>
<td>Apr</td>
<td>46.6</td>
<td>43.1</td>
<td>-3.5</td>
<td>8.0</td>
</tr>
<tr>
<td>May</td>
<td>44.3</td>
<td>42.7</td>
<td>-1.6</td>
<td>7.9</td>
</tr>
<tr>
<td>Jun</td>
<td>45.9</td>
<td>43.9</td>
<td>-2.0</td>
<td>11.2</td>
</tr>
<tr>
<td>Jul</td>
<td>44.5</td>
<td>46.6</td>
<td>2.1</td>
<td>9.7</td>
</tr>
<tr>
<td>Aug</td>
<td>43.9</td>
<td>46.9</td>
<td>3.0</td>
<td>9.5</td>
</tr>
</tbody>
</table>

T2: temperature at 2-m; RH2: relative humidity at 2-m; WS10: wind speed at 10-m; WD10: wind direction at 10-m; Precip: precipitation; DJF: winter; MAM: spring; JJA: summer; SON: autumn; MB: mean bias; RMSE: root mean square error; NMB: normalized mean bias; NME: normalized mean error; Corr: correlation coefficient; Obs.: Observation; Sim.: Prediction.
<table>
<thead>
<tr>
<th>Season</th>
<th>MDA8</th>
<th>O3-</th>
<th>PM2.5</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept</td>
<td>42.7</td>
<td>2.9</td>
<td>8.1</td>
<td>0.78</td>
</tr>
<tr>
<td>Oct</td>
<td>37.2</td>
<td>3.1</td>
<td>8.0</td>
<td>0.77</td>
</tr>
<tr>
<td>Nov</td>
<td>34.3</td>
<td>0.5</td>
<td>8.4</td>
<td>0.64</td>
</tr>
<tr>
<td>Dec</td>
<td>30.7</td>
<td>0.5</td>
<td>9.0</td>
<td>0.49</td>
</tr>
<tr>
<td>DJF</td>
<td>44.3</td>
<td>0.9</td>
<td>9.4</td>
<td>0.67</td>
</tr>
<tr>
<td>MAM</td>
<td>45.1</td>
<td>1.6</td>
<td>2.0</td>
<td>16.0</td>
</tr>
<tr>
<td>JJA</td>
<td>38.2</td>
<td>9.4</td>
<td>-2.3</td>
<td>16.4</td>
</tr>
<tr>
<td>SON</td>
<td>37.4</td>
<td>0.68</td>
<td>7.5</td>
<td>10.0</td>
</tr>
<tr>
<td>Annual</td>
<td>41.1</td>
<td>-0.1</td>
<td>8.9</td>
<td>16.2</td>
</tr>
</tbody>
</table>

MDA8: maximum daily average 8-hour ozone; 24-h avg PM2.5: 24-hour average PM2.5.

**Figures**

Figure 1. Spatial distribution of MBs for forecasted seasonal T2 by GFSv15-CMAQv5.0.2 against observations from METAR.

Figure 2. Taylor diagram with variance, Corr, and NMB for meteorological variables (T2, RH2, WS10, WD10, and Precip) against CASTNET and METAR dataset.

Figure 3. Spatial distribution of forecasted MDA8, MB, and NMB during O3 and winter season. Observation from AIRNow is shown as filled circles in the overlay plots of concentrations.

Figure 4. Forecasted seasonal daily PM2.5 by GFSv15-CMAQv5.0.2 overlaid observations from AIRNow and MB against observations from AIRNow.

Figure 5. Categorical evaluation of MDA8 and 24-h avg PM2.5.
Figure 6. Monthly AOD from MODIS (left), predicted AOD from GFSv15-CMAQv5.0.2 (middle), and predicted surface 24-h avg PM$_{2.5}$ (right)

Figure 7. Monthly average concentrations of PM$_{2.5}$ components

Figure 8. Annual performance of MDA8 in 10 CONUS regions (a); Taylor Diagram for annual performance of MDA8 (b); Annual performance of 24-h avg PM$_{2.5}$ in 10 CONUS regions (c); Taylor Diagram for annual performance of 24-h avg PM$_{2.5}$. Outliers represent regions with NSDs >3.5 (d)

Figure 9. Diurnal PM$_{2.5}$ in: (a) O$_3$ season for regions 1 to 5; (b) Non-O$_3$ season for regions 1 to 5; (c) O$_3$ season for regions 6 to 10; (d) Non-O$_3$ season for region 6 to 10. Solid curves are observed values and dash curves are predicted values. Average of predicted PM$_{2.5}$ and components of PM$_{2.5}$ within CONUS in: (e) O$_3$ season, and (f) Non-O$_3$ season
Figure 1. Spatial distribution of MBs for forecasted seasonal mean T2 by GFSv15-CMAQv5.0.2 against observations from METAR.
Figure 2. Taylor diagram (Taylor, 2001) with Normalized Standardized Deviations (NSD), Corr, and NMB for meteorological variables (T2, RH2, WS10, WD10, and Precip) against CASTNET and METAR dataset. The REF marker at x-axis represents a referred perfect performance. The closer each variable is to the REF marker, the better performance the forecast system has for that variable.
Figure 3. Spatial distribution of forecasted MDA8, MB, and NMB during O₃ and non-O₃ season. Observation from AIRNow is shown as filled circles in the overlay plots of concentrations.
Figure 4. Forecasted seasonal daily PM$_{2.5}$ by GFSv15-CMAQv5.0.2 overlaid observations from AIRNow and MB against observations from AIRNow.
Figure 5. Categorical evaluation of MDA8 and 24-h avg PM$_{2.5}$: (a) scatter plot of predicted and observed MDA8. The scatters are categorized into 4 areas using the threshold of 55 ppb for both observation and prediction; (b) scatter plot of predicted and observed 24-h avg PM$_{2.5}$. The scatters are categorized into 4 areas using the threshold of 12 µg m$^{-3}$ for both observation and prediction; (c) False Alarm Ratio (FAR) and Hit Rate (H) in 4 categories for forecasts of MDA8 and 24-h avg PM$_{2.5}$. 

https://doi.org/10.5194/gmd-2020-272
Preprint. Discussion started: 2 September 2020
© Author(s) 2020. CC BY 4.0 License.
Figure 6. Monthly AOD from MODIS (left), predicted AOD from GFSv15-CMAQv5.0.2 (middle), and predicted surface 24-h avg PM$_{2.5}$ (right)
Figure 7. Monthly average concentrations of PM$_{2.5}$ components in Jan, Jul, and Aug.

PM$_{2.5}$ _SOIL: soil components and unspecified coarse mode components; PM$_{2.5}$ _NO$_3^-$: nitrate components; PM$_{2.5}$ _SO$_4^{2-}$: sulfate components; PM$_{2.5}$ _OC: organic carbon components.
Figure 8. Annual performance of MDA8 in 10 CONUS regions (a); Taylor Diagram for annual performance of MDA8 (b); Annual performance of 24-h avg PM$_{2.5}$ in 10 CONUS regions (c); Taylor Diagram for annual performance of 24-h avg PM$_{2.5}$. Outliers represent regions with NSDs $>$3.5 (d)
Figure 9. Diurnal PM$_{2.5}$ in: (a) O$_3$ season for regions 1 to 5; (b) Non-O$_3$ season for regions 1 to 5; (c) O$_3$ season for regions 6 to 10; (d) Non-O$_3$ season for region 6 to 10.

Solid curves are observed values and dash curves are predicted values. Average of
predicted PM$_{2.5}$ and components of PM$_{2.5}$ within CONUS in: (e) O$_3$ season, and (f) Non-O$_3$ season.