Comparison of Chemical Lateral Boundary Conditions for Air Quality Predictions over the Contiguous United States during Pollutant Intrusion Events

Youhua Tang[1,2], Huisheng Bian[3,4], Zhining Tao[3,5], Luke D. Oman[3], Daniel Tong[2,10], Pius Lee[1], Patrick C. Campbell[1,2], Barry Baker[1,2], Cheng-Hsuan Lu[6,7], Li Pan[8,9], Jun Wang[8], Jeffery McQueen[8], Ivanka Stajner[8]

1. NOAA Air Resources Laboratory, College Park, MD, USA
2. Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA
3. NASA Goddard Space Flight Center, Greenbelt, MD, USA
4. University of Maryland at Baltimore County, Baltimore, MD, USA
5. Universitiesf Space Research Association, Columbia, MD, USA
6. University at Albany, State University of New York, Albany, NY, USA
7. Joint Center for Satellite Data Assimilation, Boulder, CO, USA
8. NOAA/NCEP/Environmental Modeling Center, College Park, MD, USA
9. I.M. Systems Group Inc., Rockville, MD, USA
10. Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, Fairfax, VA, USA

Abstract

The National Air Quality Forecast Capability (NAQFC) operated in U.S.'s National Oceanic and Atmospheric Administration (NOAA) provides the operational forecast guidance for ozone and fine particulate matters with aerodynamic diameters less than 2.5μm ($PM_{2.5}$) over the contiguous 48 U.S. states (CONUS) using the Community Multi-scale Air Quality (CMAQ) model. The existing NAQFC uses climatological chemical lateral boundary conditions (CLBCs), which cannot capture pollutant intrusion events originating outside of the model domain. In this study, we developed a model framework to use dynamic CLBCs from the Goddard Earth Observing System Model, version 5 (GEOS) to drive NAQFC. A mapping of the GEOS chemical species to the CMAQ's Carbon Bond 5 (CB05)-Aero6 species was developed. The utilization of the GEOS dynamic CLBCs in NAQFC showed the best overall performance in simulating the surface observations during the Saharan dust intrusion and Canadian wildfire events in summer 2015.The simulated $PM_{2.5}$ was improved from 0.18 to 0.37 and the mean bias was reduced from -6.74 μg/m^3 to -2.96 μg/m^3 over CONUS. Although the effect of CLBCs on the $PM_{2.5}$ correlation was mainly near the inflow boundary, its impact on the background concentrations reached further inside the domain. The CLBCs could affect background ozone concentrations through the inflows of ozone itself and its precursors, such as CO. It was further found that the aerosol optical thickness (AOT) from satellite retrievals correlated well with the column CO and elemental carbon from GEOS. The satellite-derived AOT CLBCs generally improved the model performance for the wildfire intrusion events during a summer 2018 case study, and demonstrated how satellite observations of atmospheric composition could be used as an

alternative method to capture the air quality effects of intrusions when the global model CLBCs, such as GEOS CLBCs, are not available.

1. Introduction

The chemical lateral boundary conditions (CLBCs) are pivotal to the prediction accuracy of regional chemical transport models (CTMs) (Tang et al., 2007; 2009). The CLBCs represent the spatiotemporal distribution of species concentrations along the lateral boundaries of a regional model domain. CLBCs can be either static or dynamic in type, and can significantly affect CTMs predictions. One effect is imposing a constraint with static background concentrations for long-lived pollutants, such as CO and O_3, which is the typical role of climatological CLBCs for non-intrusion events. For example, regional models like the Community Air Quality Multi-scale Model (CMAQ) hemispheric version (Mathur et al, 2017) utilizes static CLBCs that constrain chemical concentrations along the equator. The influences of external pollutant intrusion events, can only be achieved with dynamic (time-varying) CLBCs. Such CLBCs can come from a global model, a regional model that uses a larger domain (Tang et al., 2007), or observed profiles (Tang et al., 2009).. Henderson et al. (2014) compiled a ten year CLBCs database over the Contiguous United States (CONUS) using a global chemical transport model (GEOS-Chem, Bey et al., 2001) and evaluated it against satellite retrieved ozone and CO vertical profiles.

The U.S. National Oceanic and Atmospheric Administration's (NOAA) National Air Quality Forecast Capability (NAQFC), which is currently based on the regional-scale CMAQ model, requires CLBCs for its daily prediction. The current NAQFC uses the dust-only aerosol CLBCs from the NOAA Environmental Modeling System (NEMS) Global Forecast System (GFS) Aerosol Component (NGAC) (Lu et al, 2016; Wang et al, 2018), which is an inline global model coupled with the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol mechanism (Chin et al., 2000, 2002; Colarco et al., 2010). Prior to the implementation of the NGAC CLBCs, NAQFC used the background, static aerosol profiles for the aerosol CLBCs (Lee et al., 2017). For the gaseous species, NAQFC uses modified monthly-averaged CLBCs from a 2006 GEOS-Chem simulation (Pan et al., 2014). To alleviate surface ozone over-predictions, the upper tropospheric ozone CLBCs from GEOS-Chem has been limited $\leq$ 100 parts per billion in volume (ppbv).

Static CLBCs cannot capture the signals of some intrusion events, such as the biomass burning plumes from the outside of the domain, which could affect the prediction of ozone and particulate matter with aerodynamic diameter less than $_{2.5}\mu m$ ($PM_{2.5}$). For non-intrusion events, Tang et al. (2007) investigated the sensitivity of regional CTMs to CLBCs, and found that the background magnitude of the pollutant concentrations was more important than the variation of the CLBCs for the near-surface prediction over polluted areas. Over the contiguous U.S., the northern and western USA are near to the prevailing inflow lateral boundaries where Canadian emissions and long-range transported Asian air-masses can affect the chemical background concentrations. Additionally, the southern and eastern boundaries are subjected to the Saharan

dust intrusions during the summer, which may result in surface $PM_{2.5}$ concentration increases (Lu et al, 2016). CLBCs from global models are needed to fully assess such impacts of intrusion events, and to advance the operational NAQFC. In this study, we extracted the CLBCs from the GEOS global chemical circulation model (Strode et al. 2019; Molod et al., 2012) in both static (monthly average) and dynamic (every three hours) modes. The NAQFC runs using both GEOS and NGAC CLBCs are compared to a NAQFC base case with monthly 2006 GEOS-Chem CLBCs for the summer 2015. During this period, the Canadian wildfires and Sahara dust storms affected the CONUS domain's northern and southern regions, respectively. In addition, we investigate the method of using satellite-derived CLBCs for pollutant intrusion events when global model CLBCs may not be available.

2. Model Configuration and Experiment Design

The operational NAQFC is based on CMAQ version 5.0.2, driven by meteorological forecasts from NOAA/NCEP's North American Mesoscale Model (NAM). The CMAQ configuration includes the CB05 gaseous chemical mechanism (Yarwood et al., 2005) with updated toluene (Whitten et al., 2010) and chlorine chemistry (CB05tucl) (Tanaka et al., 2003; Sarwar et al., 2007), and Aero6 (Carlton et al., 2010; Foley et al., 2010; Sonntag et al., 2014) aerosol module driven by NOAA/NCEP's North American Mesoscale Model (NAM) forecasting. It has a 12×12km horizontal resolution covering CONUS, with 35 vertical layers up to 100 hPa. Anthropogenic area and mobile emissions are based on the on the 2011 U.S. EPA National Emission Inventory (NEI2011v2), and the point source emissions have been updated with the U.S. EPA Continuous Emission Monitoring System (CEMS) for the prediction year (2015). Biogenic emissions are based on the Biogenic Emission Inventory System (BEIS) 3.14 (Pierce et al., 1998). Wildfire emissions originating inside the CONUS domain are estimated using the U.S. Forest Service (USFS) BlueSky fire emissions estimation algorithm, in which the fire location information is provided by the NOAA Hazard Mapping System (HMS). The NOAA HMS is a satellite-based fire detection system that includes manual quality control. The detailed wildfire emission process of this system has been described in Pan et al. (2020).

In this study, we conducted five model runs with different CLBCs (Table 1) over the CONUS domain (Figure 1). The first run is the NAQFC-CMAQ base case (referred to as CMAQ_Base), which uses the modified GEOS-CHEM 2006 monthly gaseous CLBCs and clean aerosol background. The CMAQ_base CLBCs were used in the earlier NAQFC system before NGAC was made available. The second run, NGAC-LBC, is the same as in CMAQ_Base for gaseous CLBCs, but uses NGAC's dynamic aerosol CLBCs. The third run, GEOS-LBC simulation, uses GEOS dynamic CLBCs and has full chemistry and dynamic variation for both gaseous and aerosol species, while the 4^{th} run, GLBC-monthly, tests the GEOS monthly mean CLBCs to gauge the impacts of the CLBCs' temporal variability. The fifth and final run incorporates satellite-based aerosol optical thickness (AOT) for the northern CLBCs (AOT-NLBC). The AOT-NLBC run is the same as the GEOS-monthly run, except that its northern boundary

condition is generated from the relationship of VIIRS (Visible Infrared Imaging Radiometer Suite) AOT and GEOS-LBC for the wildfire intrusion events, which will be described later.

The extraction of the GEOS CLBCs for the NAQFC's domain boundaries is based on the existing global-to-regional interface tool developed by Tang et al (2008, 2007) for MOZART, RAQMS, and NGAC global models with additional enhancements to support GEOS's NetCDF4 format, vertical layers, and chemical species. This tool includes two major functions: spatial mapping and species mapping. Spatially, GEOS's concentrations from its 576×361 grid in a 0.625º×0.5º horizontal resolution with 72 vertical layers are three-dimensionally interpolated into CMAQ's CONUS lateral boundary periphery in a 12 km horizontal resolution. The species mapping is also needed due to the different chemical mechanisms employed in GEOS and CMAQ, as discussed in the following sections.

2.1 Gaseous Species Mapping

The GEOS outputs 122 gaseous chemical species and 15 aerosol species. For species such as O_3, CO, NO, and NO_2, an explicit one-to-one mapping can be achieved. However, some volatile organic compounds (VOCs) need special treatment during the conversions as GEOS uses different lumping approaches compared to what is done in the CMAQ CB05tucl (carbon bond 5 mechanism with toluene and chloride species). Table 2 lists the VOC species mapping used to convert GEOS's gaseous species to CMAQ's CB05tucl species. Two methods were employed for mapping the VOCs species: one was based on the carbon bond structure, e.g., ALK4 → 4 PAR (Table 2), and the other was based on the similarity of the reactions. In GEOS, for example, the products of the isoprene reaction with NO3 are lumped into INO2, an intermediate RO2 radical.

ISOP + NO3 → INO2

The radical INO2 participates in the following reactions (Eastham et al., 2014; Tyndall et al., 2001):

INO2 + NO → 1.10NO2 + 0.80HO2 +0.85HNO3 + 0.05NO2 + 0.10MACR + 0.15CH2O + 0.05MVK

INO2 + HO2 → INPN

INO2 + MO2 → 0.55NO2 + 0.40HO2 + 0.425HNO3 + 0.025NO2 + 0.05MACR + 0.08CH2O + 0.03MVK + 0.25RCHO + 0.75CH2O + 0.25MOH + 0.25ROH + 0.05HO2

INO2 + MCO3 → MO2 + 0.10NO2 + 0.80HO2 + 0.85HNO3 + 0.05NO2 + 0.10MACR + 0.15CH2O + 0.05MVK

INO2+MCO3 → RCHO + ACTA + NO2

The CB05tucl mechanism skips the intermediate INO2, and directly represents it as

ISOP + NO3 → 0.200*ISPD + 0.800*NTR + XO2 + 0.800*HO2 + 0.200*NO2 + 0.800*ALDX + 2.400*PAR

Therefore, the GEOS INO2 species is split into seven CB05tucl species with the corresponding factors, respectively (Table 2). This conversion is just an approximation, and a perfect consistency for mapping these species can not be achieved due to the large differences between these two mechanisms, especially in regards to the complexity of isoprene chemistry. Fortunately, for the CONUS domain, the isoprene chemistry influence on the CONUS CLBCs are less significant when compared to the major intrusion events from wildfire plumes and dust storms. Most biogenic emitted species are short-lived, and their direct impact on CLBCs are relatively weak. A similar situation can also be applied to other short-lived species, such as NOx, which will be discussed later. Biogenic emissions can affect local photochemical processes, however, and subsequently generate relatively long-lived species, such as ozone and NTR. Such species may originate from outside the regional domain, and thus have impacts on CLBCs and downstream chemistry. This issue is mitigated by the fact that most of these secondary long-lived species are explicitly included in both GEOS and CMAQ chemical mechanisms, and can be directly mapped.

Other gaseous species are represented explicitly in the GEOS model, such as methyl vinyl ketone (MVK), which is lumped in the CB05tucl's isoprene product (ISPD). In GEOS, the MVK mainly comes from isoprene, which is consistent with the CMAQ's ISPD source. Some GEOS species can also be mapped to CB05tucl species based on their carbon bonds, e.g. R4N2 (GEOS's C4-5 alkyl nitrates), which can be mapped to NTR + 2.0 PAR in the CB05tucl mechanism. Some of the mapping treatments, such as ALK4 (C4 or higher alkanes) conversion to 4 paraffin carbon bonds (table 2), may have a "truncation error" as it only counted butane isomers. The effect of this truncation error, however, is likely relatively minor for CONUS CLBCs. The GEOS global model also mainly treats ALK4 as butane or Cn where n ~ 4. Although GEOS's ALK4 emission includes some C5 or higher (C5+) alkanes, the relatively shorter lifetime of C5+ alkanes (Helmig et al, 2014) makes them harder to reach CONUS from their major upstream sources, such as East Asia. In this study, wildfire emissions may contribute to the C5+ alkane's impacts on the CONUS CLBCs, but these C5+ emissions are at least one order of magnitude lower than the corresponding wildfire CO, ethane, and propane emissions (Urbanski et al, 2008). Moreover, the impacts of the complex chemistry mapping on the CLBCs for the pollutant intrusion events (mainly wildfire events) are not expected to be significant for the ozone and $PM_{2.5}$ predictions in this study, since the constituents of the major wildfire intrusion from the GEOS global model are CO, NOx, ethane, propane, elemental carbon (EC), and organic carbon (OC).

2.2 Aerosol Species Mapping

The GEOS model uses an updated GOCART aerosol scheme (Bian et al, 2017), compared to NGAC GOCART (Colarco et al 2010, respectively), which includes additional species of ammonium and 3-bin nitrates (NO3an1, NO3an2 and NO3an3). Table 3 lists the aerosol species mapping from GEOS aerosols to CMAQ Aero6 species used in this study. GEOS aerosols have fixed size bins defined by their diameters, while CMAQ aerosols use 3 size modes- Aitken (ATKN), accumulations (ACC) and coarse (COR); or alternatively i, j, and k modes respectively (Appel et al., 2010). Each of these size modes has its own lognormal size distribution (Whitby and McMurry, 1997). To convert the aerosol species from GEOS to CMAQ's Aero6, we need to consider not only the aerosol composition and the GEOS size bins conversion to the CMAQ size modes, but also the size distribution within each CMAQ size mode that is controlled by the CMAQ aerosol number concentrations (the 3rd column of Table 3). GEOS's dust aerosols are mapped to AOTHRJ (other unreactive aerosols in the accumulation mode) and ASOIL (soil particles in the coarse mode) in CMAQ. Although the CMAQ Aero6 has explicit elemental ions such as Ca and Mg, which are possible dust ingredients, we do not consider the reaction effects due to these ions. Tang et al. (2004) studied the dust outflow during the ACE-Asia field experiment and found that only a small portion of cations in dust particles was available for aerosol uptake and reactions, and that this portion is negligible for aged dust air masses.

3. Case Studies for the Summer 2015

During summer 2015, two intrusion events occurred, one in the southeastern USA and one in northern USA, respectively. The southeastern intrusion was due to the long-range transported dust storm from the Saharan desert. The northern intrusion was due to a Canadian wildfire event and its southward transport into the United States. Figure 2 shows the Suomi-NPP satellite's VIIRS AOT retrieval from later June to early July in 2015, and highlights these two intrusion events.

3.1 Dust Storm Events in Summer 2015

As shown in Figure 2, a dust storm originating from the Saharan desert reached the southeastern U.S. via trans-Atlantic transport. The two global models, GEOS and NGAC, captured this dust intrusion, and increased the aerosol CLBCs of the NAQFC. Figure 3 shows the corresponding three CLBCs for ASOIL and AOTHRJ along the model's boundaries on 2 July, 2015. With the exception of CMAQ_Base, all the other three CLBCs showed enhanced ASOIL (the coarse-mode dust) and AOTHRJ (the accumulation-mode dust) near the domain's southeastern corner and the central southern boundary. The GLBC-Monthly represents the monthly average of GEOS-LBC for July 2015, and has the lowest increments for the two types of aerosols. The two dynamic CLBCs, the GEOS-LBC and NGAC-LBC, showed similar aerosol increments along the domain boundaries. However, the NGAC aerosols tended to have a broader spread than those of the GEOS-LBC, especially for ASOIL, which could reach above an altitude of10km with

concentrations greater than 5 $\mu g/m^3$ (Figure 3e). The NGAC-LBC also showed enhanced dust signals over the western boundary, where the GEOS-LBC did not show any dust-related aerosols. Another difference between these two CLBCs was their ratio of AOTHRJ versus ASOIL. The dynamic NGAC-LBC had higher ASOIL, the coarse-mode dust, than that of the GEOS-LBC (Figure 3a, 3e), but its AOTHRJ was lower than the latter (Figure 3b, 3f). This is particularly true over the central southern boundary, where the GEOS-LBC had AOTHRJ up to 30 $\mu g/m^3$. It implies that besides their difference on transport patterns, these two global models also had some differences in their dust size distributions.

Figure 4 shows comparisons of the simulated $PM_{2.5}$ concentrations against the observations from the U.S.EPA AIRNow stations. The CMAQ_Base represented a clear background situation, and has obviously missed this dust intrusion event, and underestimated the $PM_{2.5}$ over the southern and southeastern United States. The two dynamical CLBCs, GEOS-LBC and NGAC-LBC, captured the intrusion signals well and yielded the best model performances. While both GEOS-LBC and NGAC-LBC underpredicted $PM_{2.5}$ over central Florida, their performances were improved compared to the CMAQ_Base. Further downwind over Texas, the GEOS-LBC yielded more widespread and higher $PM_{2.5}$ enhancements compared to the NGAC-LBC, and agreed better with the observations (except for the overpredictions over northern Texas). The GLBC-Monthly run had a moderate $PM_{2.5}$ enhancement but still underestimated the dust intrusion, falling between the CMAQ_Base and two dynamic CLBCs cases in magnitude of $PM_{2.5}$ enhancements. Figure 5 shows a similar story for the scenario of three days later. The GEOS-LBC yielded the best overall model performance, although it still underpredicted the $PM_{2.5}$ concentration over Florida and northern Texas.

A time-series comparison over Florida and Texas showed that, in general, the best model performance in capturing the dust intrusion are, in order, GEOS-LBC, NGAC-LBC, GLBC-Monthly, and CMAQ_Base (Figure 6). An exception, however, is the NGAC-LBC's underprediction for $PM_{2.5}$ concentrations over Florida in June. These comparisons demonstrate the advantage of using dynamic CLBCs for capturing intrusion events. The dynamic CLBCs (GEOS-LBC and NGAC-LBC) still missed some intrusion peaks, such as the one on 30 June over Texas, and also had disagreement with the observed temporal variability, e.g. 1 July over Florida, and 8 July over Texas. It should be noted that the nighttime $PM_{2.5}$ spike on 4 July (5 July in UTC time) was not related to the dust intrusion, but was caused by U.S. Independence Day's fireworks. This firework emission was not included in our anthropogenic emission inventory. Most firework emissions were injected in elevated levels, and the associated pollutants could be transported to extended downstream areas. If the downstream areas were relatively big, its regional averaged effect could appear for a longer time. This is the reason why some $PM_{2.5}$ concentration spikes started from 4 July could last longer than the firework emission durations, e.g. one hour.

3.2 The Wildfire Event in Summer 2015

During the same period of summer 2015, a wildfire event occurred in Canada and the biomass burning plume was transported to the northern U.S., as shown in Figure 2. While the dust storm intrusions mainly affected the aerosol concentrations, the biomass burning plumes also included

gaseous pollutants, such as enhanced levels of CO, NO_x, and VOCs, which could affect the photochemical generation of ozone. For aerosol species, the biomass burning air mass mainly consisted of elemental carbon (EC) and primary organic carbon (POC), which are associated with the AECJ and APOCJ in CMAQ (Table 3). The GEOS-LBC showed the highest aerosol and CO concentrations with AECJ+APOCJ up to 300 $\mu g/m^3$, and CO up to 3000 ppbv along the domains northern boundary (Figure 7). It also showed CO enhancement at elevated altitudes up to 12km (Figure 7b). The monthly averaged CLBCs, GLBC-monthly, had similar patterns to the GEOS-LBC, but with much lower concentrations (Figure 7c, 7d). The NGAC-LBC had similar AECJ+APOCJ profiles to those of GLBC-monthly, but its static CO boundary condition (same as the CMAQ_base) did not reflect the wildfire influence (Figure 7e, 7f).

The enhanced gaseous pollutants in the full-chemistry CLBCs increased the photochemical generation of ozone, and consequently the higher ozone appeared along the northcentral boundary (Figure S1a, S1b), where the GEOS-LBC showed 10 ppbv or higher O_3 concentration compared to the static NGAC-LBC and CMAQ_Base for the altitudes < 4km (Figure S1c). The wildfire induced ozone enhancements appeared not only in the lower troposphere, but also at higher altitudes, e.g. 11km, and were not solely due to downward transport of high stratospheric ozone (Figure S1a). The full-chemistry GEOS-LBC also indicated that the short-lived NOx had less than 1 ppbv increase (Figure S2a) due to the wildfire intrusion. The NOz (sum total of all NOx oxidation products, NOz=NOy-NOx) enhancements, however, could reach 30 ppbv (Figure S2b) along the northern boundary around 10-12km altitude, and co-existed with the CO increments (Figure 7b). NOz is a good indicator for the photochemical formation of ozone (Sillman et al., 1997) while O_3/NOz ratio represents the ozone photochemical efficiency per NOx. The high-altitude CO and NOz increments reflected that the GEOS model had strong fire plume rise and injected wildfire emissions into the upper troposphere. The VOCs also showed increments due to the wildfire plume, such as ethane (Figure S2c) and HCHO (Figure S2d). HCHO is a short-lived species, and an indicator of VOC oxidation (Arlander et al., 1995). With these magnitudes of CO, VOC and NOx increments, the GEOS-LBC mainly provided the VOC- and CO-rich airmass with limited NOx to the regional CMAQ model. When this CO/VOC rich airmass arrived at NOx-rich regions, such as the urban areas, it could contribute to the photochemical generation of ozone.

Figure 8 shows the comparison of $PM_{2.5}$ predictions on 3 July, 2015 at 18:00UTC. The CMAQ_Base missed the intruded biomass burning plumes and the corresponding high $PM_{2.5}$ over North/South Dakota, Montana, and Minnesota (Figure 8a). The GEOS-LBC predicted the highest $PM_{2.5}$ increment (up to 200 $\mu g/m^3$) over these states, and agreed best with the AIRNow observations (Figure 8b). The dynamic NGAC-LBC and static GLBC-Monthly showed similar $PM_{2.5}$ enhancements over the affected states, but were almost one order of magnitude lower than that of GEOS-LBC.

Figure 9 shows similar predictions but for ozone, where the GEOS-LBC yielded the highest ozone increase due to the wildfire plume, but still underestimated the ozone over North Dakota

(Figure 9b). The GLBC-Monthly systematically underestimated the ozone over all of these regions. The CMAQ_Base and NGAC-LBC used the same static gaseous CLBCs, including that for ozone, and gave even larger underestimates. The NGAC-LBC had more wildfire-induced aerosol loadingand consequently a lower photolysis rate compared to the CMAQ_Base. As both of NGAC-LBC and CMAQ_Base had the "cleaner" air mass with low concentrations of ozone precursors over the northern U.S., the photolysis reduction due to aerosols mainly led to the reduced ozone photolytic destruction, such as $O_3 \rightarrow O^1D + O_2$ or $O_3 \rightarrow O^3P + O_2$, instead of its photochemical generation. For the same reason, ozone's lifetime in winter is longer that in summer (Janach, 1989). Over polluted regions, however, the photolysis reduction would cause a lower ozone concentration by limiting its photochemical production. Overall, this effect of photolysis rates on ozone was relatively small.

Figure 10 shows the time-series comparison over the northcentral and northeastern U.S. for surface $PM_{2.5}$ and ozone concentrations, in which the GEOS-LBC showed better $PM_{2.5}$ predictions compared to the other cases, especially from 29 June to 2 July over the northern U.S. The GEOS-LBC still had the systematic $PM_{2.5}$ underestimation on the night of 4 July due to the missed firework emissions, and underestimated $PM_{2.5}$ further downwind in the northwestern U.S.. The GEOS-LBC also better captured the peak ozone concentrations, e.g. 1 to 2 July, though it overpredicted ozone in some instances, especially during nighttime. The small ozone differences (regional averages < 1 ppbv) between the CMAQ_Base and NGAC-LBC reflected the impact of wildfire aerosols on the photolysis rates (Figure 9c, 9d).

3.3 Statistics and Discussion

Table 4 summarizes the $PM_{2.5}$ statistics during the two weeks of the intrusion events over the CONUS domain and sub-regions. The dynamic CLBCs, GEOS-LBC and NGAC-LBC, showed significant improvements for almost all scores over these regions as compared to the CMAQ_Base. The GLBC-Monthly was also better than the base case, though its correlation coefficient (R) and index of agreement (IOA) were lower than those of the dynamics CLBCs, as the time-averaging method removed the temporal variability. Over the further downwind regions of the intrusion events, the CLBCs' impact depended on the regional characteristics of the pollutant concentrations. For instance, since the Rocky Mountain region was relatively clean due to its low local PM emissions, the external influence weighed more, and thus the CLBCs showed more significant impact there. Over more polluted regions where relatively strong local PM emissions existed, such as the Pacific Coast and the northeastern U.S., the CLBCs mainly changed the background concentration for $PM_{2.5}$, and had a very limited impact on R or IOA. Overall, the GEOS-LBC yielded the best scores in term of mean bias (MB), root mean square error (RMSE), R and IOA. The other dynamic CLBCs, NGAC-LBC, had the next best performance, and the CMAQ_Base ranked last in term of the $PM_{2.5}$ prediction.

Table 5 shows the similar statistics for ozone. The CMAQ_Base had a preexisting O_3 overprediction, especially over the south-central U.S., which affected the impacts of the CLBCs

and the corresponding model performance changes. Differing from $PM_{2.5}$, ozone had strong diurnal variation during the summertime, which resulted in relatively less impacts of the CLBCs on R and IOA. The NGAC-LBC did not change any precursor concentrations related to ozone production, and thus only affected the ozone formation by reducing photolysis rates. Therefore, as compared to CMAQ_Base, the NGAC-LBC had very weak influence on O_3 and only reduced the regional O_3 by around 0.2 ppbv, with little to no impact on R or IOA. The GEOS-LBC tended to increase ozone concentrations in most regions, except the south-central USA, where the GEOS-LBC showed general improvement for all statistical metrics. The GEOS-LBC had the weakest impact on ozone over the Pacific Coast and Rocky Mountain regions, or the farther downstream areas. The GLBC-monthly had the largest ozone increase over most regions except the southcentral region, and also had the slightly higher RMSE. This result suggests that averaging the temporal variation of CLBCs may not have a linear effect on ozone predictions. The GEOS-LBC showed the best model performance compared to other runs except the mean bias over most regions, though its improvement for O_3 was not as significant as that for $PM_{2.5}$. As discussed above, the CLBC's impact on ozone inside the domain was realized through changing inflow concentration of the O_3 inflow itself and/or O_3 precursors, such as NOx, VOC or CO. The distance or depth of the CLBC's effective impact from the inflow boundary depended on the lifetime of these species. All these species have a longer lifetime in winter compared to summer. Our other study showed that the CLBC's impact on ozone in winter was stronger than that in summer.

The GEOS-LBC case is further used to illustrates the impact of CLBCs on the prediction statistics and their relations to the distance from the domain boundary during the pollutant intrusion events across the southern (the Saharan dust storm, Figure 11 a, b) and the northern U.S. (wildfire, Figure 11 c, d). The CLBCs have two effects for the regional predictions: 1) they provide a constraint for background concentrations represented by the mean biases, and 2) they introduce a dynamic external influence, represented by the correlation coefficients. The CLBCs impacts on the background and variability both affect the RMSE of predictions. Over the southern U.S., the Saharan dust storm intruded through the states of Texas and Louisiana, or -100°W to -86°W, and moved northward (Figure 4). Figure 11a showed that the GEOS-LBC's improvement on the correlation coefficient (R) for the $PM_{2.5}$ prediction reached the highest near the southernmost near-boundary region, and gradually reduced along the latitude for the inland region. On the other hand, the corresponding MB improvement for $PM_{2.5}$ did not show a significant reduction along the distance from the influenced boundary. The second effect of CLBCs, which constrains $PM_{2.5}$ background concentrations, can exist further inside of the domain. The $PM_{2.5}$ RMSE change reflected the combined changes of MB and R. The improving impact of the GEOS-LBC on RSME also became weaker moving from the boundary because the MB did not vary much and the RMSE changes followed the correlation coefficient's change northward. Contrary to $PM_{2.5}$, the most significant R improvement for O_3 was not near the boundary, but rather for more northward regions (29°N to 32°N) (Figure 11b). Overall, for the

dynamic CLBCs, the improvements in ozone MB and RMSE have similar spatial variability, which is more significant near the inflow boundary and fades in the further inland.

Differences in $PM_{2.5}$ and O_3 statistics arise because O_3 typically has a stronger diurnal variation in summer driven by local photochemical activities in polluted regions, which may impact the correlation more than the external CLBCs. Therefore, the GEOS-LBC's major influence on O_3 prediction for this event was changing O_3 background concentration. The GEOS-LBC MB change for ozone was also variable compared to the CMAQ_Base case northward from the boundary (Figure 11b). The GEOS-LBC had a lower ozone concentration compared to the CMAQ_Base at low altitudes for the southern boundary, but had higher ozone concentrations in the altitudes higher than 14 km (Figure S1). The high ozone concentration could reach the surface after a certain distance of downward transport in the model system with strong vertical mixing (Tang et al., 2009), which results in the higher ozone MB of the GEOS-LBC over the deeper inland region.

There was a similar spatial distribution for the $PM_{2.5}$ statistical differences between GEOS-LBC3 and CMAQ_Base for the wildfire intrusion event over the northern U.S. The most significant R and RMSE improvements for GEOS-LBC appeared near the boundary, and these improvements were reduced farther from the boundary. However, the corresponding MB differences could exist deeper inland. For O_3, the difference between the GEOS-LBC and CMAQ_Base cases became more complex because wildfire plumes also contained the intrusion from O_3 and its precursors. The GEOS-LBC run generally yielded higher O_3, which exacerbated the preexisting model overprediction near the boundary, but helped reduce the ozone underpredictions further inland (Figure 11d). The largest O_3 MB differences were also farther away from the boundary itself, as it took more time for the for ozone precursors to contribute to the photochemical formation of O_3. The spatial variation of O_3 RMSE difference was similar to that of O_3 MB except for the further inland region, such as the south of 43°N, where the GEOS-LBC did not improve the RMSE. A similar issue also appeared for the R difference for the region south of 46°N, implying that the intruded wildfire plume represented by the GEOS-LBC could introduce some spatial or temporal biases for O_3 precursors.

4. AOT Derived Lateral Boundary Conditions

The dynamic CLBCs, such as GEOS-LBC, showed overall better prediction of the pollutant intrusion events by better capturing the spatiotemporal impacts of external gases and aerosols across the regional model domain. However, the full-chemistry CLBCs sometimes are not easy to obtain, especially for a near-real-time forecast. Some event-dependent emissions, including wildfires, may need additional time to retrieve and refine, and thus may lag behind the valid forecast times. In order to represent the intrusion influence when the real-time model CLBCs are not available, we test an alternative CLBCs method based on the historical data adjusted with certain indicators. Here we focus on the wildfire intrusion, since it is more difficult to capture the sudden outbreak of wildfire signals than the long-range transported dust intrusion. Further

alleviating this issue for dust intrusion is the current availability of the operational NGAC dust forecasting for the NAQFC (Wang et al, 2018).

4.1 Development of the CLBCs with VIIRS AOT for Wildfire Plumes

While ground-based AIRNow surface stations are reliable and could be as a historical data indicator to represent intrusion events, their spatial coverage along the wildfire intrusion boundaries are not dense enough for this purpose. VIIRS AOT retrievals, however, well reflected the wildfire intrusion with broad spatial coverage, superior to the sporadic surface stations along the northern boundary of the CONUS domain (Figure 2). Thus, VIIRS AOT may be used as an indicator for wildfire plumes. Figure S3 showed the comparison of extracted VIIRS AOT versus GEOS CO and EC column loading along the northern boundary for June to July 2015, with their correlation coefficients (R) > 0.5. The regression relationship derived out of Figure S3 can be used to resample the historical GEOS-LBC data to derive the new CLBCs for wildfire intrusion events when the corresponding AOT is available. This regression methodology is strengthen by the fact that the domain's northern boundary was relatively clean in most periods of the summer, unless the wildfire events occurred. During the June and July 2015, the VIIRS AOT data were available once or twice per day around the local noontime under cloud-free conditions. To maximize the amount of VIIRS AOT data used along the northern boundary, we relaxed the radius of influence up to 300 km when "nearest-neighbor" pairing the VIIRS AOT geolocation and the northern boundary location. Here we paired the GEOS's northern CLBCs (NLBC) for 18UTC with the daily VIIRS AOT along the same location, and averaged the whole column with AOT interval of 0.2 to build a CLBC database sorted in AOT. We only chose to resample the CLBCs for the primarily emitted species from the wildfire sources, which include POC, EC, CO, NOx, and two NOz species: PAN and HNO_3, but did not include the ozone CLBCs. When VIIRS AOT are available for a NLBC grid in new intrusion events, the whole-column species concentration data from that database are chosen to form the new CLBCs for that grid based on the nearest neighbor VIIRS AOT value.

4.2 A Case Study with VIIRS AOT Derived LBC in August, 2018

In mid to late August 2018, there were dominant high-pressure and dry weather conditions that led to a wildfire outbreak that quickly spread across western Canada (Figure S4).. There was prevailing north to northeast winds, which brought the fire pollutants southward and affected the north-northwestern U.S.. The corresponding VIIRS AOT retrievals for this event showed high AOT values in western Canada as well as the northern and northwestern U.S. (Figure 12a). We used this AOT data to derive new CLBCs along the northern boundary (Figure 12b, c) for CO and wildfire emitted aerosols (AECJ+APOCJ) by resampling the historical GEOS-LBC database from the Jun-Jul, 2015 period. These AOT derived northern CLBCs (AOT-NLBC) were updated once per day due to the VIIRS data availability, while the western, southern, and eastern boundaries came from the climatological monthly-mean GEOS-LBC (averaged from 2011 to 2015). The AECJ+APOCJ increments of the AOT-NLBC mainly existed below 3km, but the CO

enhancement could reach up to the altitude of 10km, due to the elevated CO plume in the original GEOS-LBC, e.g. Figure 7b. The NGAC-LBC (Figure 13d) also showed the enhanced AECJ+APOCJ concentrations along the northern boundary, but it was much lower than that of AOT-NLBC. In addition, unlike the AOT-NLBC's two peaks, the NGAC-LBC mainly just showed one peak near the northwestern boundary.

Figure 13 shows the surface ozone and $PM_{2.5}$ concentrations over this region one day later (08/17/2018). The CMAQ_Base underpredicted both species over this region, and the AOT-NLBC reduced the underprediction by increasing background concentrations from the northern boundary. Since the AOT-NLBC did not include the dynamic ozone boundary conditions, any enhancements in ozone concentration were due to the CO and NOx enhancements transported from the northern boundary, which sometimes caused the overprediction over further downwind areas, such as North Dakota. Overall, the AOT-NLBC showed better $PM_{2.5}$ prediction over southwestern Canada and northwestern USA due to the higher background concentrations. The NGAC-LBC had nearly the same ozone concentration as the CMAQ_Base (Figure 13e), and also had the similar $PM_{2.5}$ background enhancements to that of the AOT-NLBC over the northwestern U.S.. Unlike the AOT-NLBC, the NGAC-LBC did not show the $PM_{2.5}$ increases east of -96°W compared to the CMAQ_base run, as the AOT-NLBC had additional aerosol increment peaks over the north-central boundary. However, that aerosol background enhancement of the AOT-NLBC led to the $PM_{2.5}$ overprediction over Minnesota, implying that the derived CLBCs could incur some errors.

Figure 14 shows the corresponding models vs. AIRNow time-series comparison over EPA region 8 (states of Montana, North and South Dakotas, Wyoming, Colorado, and Utah), region 10 (states of Washington, Idaho, and Oregon), region 5 (states of Minnesota, Wisconsin, Illinois, Indiana, Michigan, and Ohio), and region 9 (states of California, Nevada and Arizona). Both observed and predicted ozone showed strong diurnal variation. The AOT-NLBC showed better skill in capturing daytime ozone maximum for the region 8 and 10 with about 3-10 ppbv higher amounts than the CMAQ_base prediction, though it tended to overpredict ozone at night. Over the EPA region 5 (north-central U.S.), the ozone differences between the AOT-NLBC and CMAQ_base runs became narrower since the major pollutant intrusion from this event occurred in the northwestern U.S.. The AOT-NLBC increased the preexisting high bias for ozone over the region 5. Region 9 (Southwestern USA) was located further downwind from the domain's northern boundary, meaning it should get a much weaker influence from the AOT-NLBC. However, during the period of 08/21-08/25/2018, the impacts of the AOT-NLBC on ozone could still reach about 5 ppbv, and the derived CLBCs generally improved the ozone prediction over that region. It implies that long-lived wildfire pollutants, such as CO, could be transported to the farther downwind areas, and impact ozone concentrations. Throughout this period, the ozone differences between the NGAC-LBC and CMAQ_Base were very small, mainly caused by the aerosols' effect on the photolysis rates.

For $PM_{2.5}$ concentration, the CMAQ_Base run systematically underpredicted all the 4 EPA regions as shown in Figure 14, especially over the region 10, as the northwestern states encountered the major wildfire inflow. The AOT-NLBC and NGAC-LBC had similar performance over the northern states (i.e., regions 8, 10, and 5), while improving the predictions by reducing the mean bias up to 10 $\mu g/m^3$ over region 10 (Figure 14d). In region 9, however, they showed some differences in temporal variability (Figure 14h) as the AOT-NLBC only changed the north boundary. The AOT-NLBC overpredicted $PM_{2.5}$ during 21-23 August, 2018, and the NGAC-LBC yielded higher $PM_{2.5}$ after 08/25 over region 9. Even though the AOT-NLBC only changed the northern boundary conditions, that CLBC could influence the whole domain during the strong intrusion events. The domain-wide statistics of surface $PM_{2.5}$ predictions are R=0.39, 0.45, 0.50; MB=-7.53, -2.33, -2.70; RMSE=25.12, 24.04, 22.93 for the CMAQ_Base, NGAC-LBC, and AOT-NLBC runs, respectively. The AOT-NLBC had the best overall scores, except that the NGAC-LBC had a slightly better mean bias with its dynamic four boundaries.

These results demonstrate that the alternative CLBCs derived from VIIRS AOT may be useful for capturing the key intrusion signals in cases when the global model CLBCs are not available. This approach is useful in atmospheric composition forecasting as the satellite AOT retrievals can be obtained in near-real-time. The wildfire events of summer 2015 and 2018 are similar, which makes the quantitative derivation of CLBCs possible. However, this method may incur biases, which may be due to two reasons: 1) the relatively low correlation coefficient (Figure S3), and 2) the lack of detailed information on vertical distribution for the total column loading of pollutants. These factors depend on the chosen database, in this case summer 2015, where the major aerosol intrusion occurred below 3km (Figure 7). If other intrusion events have major elevated aerosol signals, the use of the AOT derived LBC may put too many aerosols in lower layers and cause surface $PM_{2.5}$ overpredictions.

5. Conclusion

In this study, we examined the influence of CLBCs on regional air quality prediction, and used surface ozone and $PM_{2.5}$ observations to verify the impacts. We developed a full-chemistry mapping table from the GEOS global model to CMAQ's CB05-Aero6 species. The simulations with the GEOS dynamic CLBCs performed the best compared with the surface observations in summer 2015 when the Saharan dust and Canadian wildfire intrusion events occurred. The base simulation (CMAQ_Base) had the worst model performance, as it did not account for these external influences. The NGAC-LBC only considered the GOCART aerosols (not full-chemistry). The simulation with the NGAC-LBC demonstrated good performance for capturing the dust storm intrusion but missed the ozone enhancements in the northern U.S. due to the Canadian fire events. The influences of CLBCs on the model performance depended on not only the distance from the inflow boundary, but also the specific species and their regional characteristics, exemplified by the difference distributions of CLBCs' impacts on ozone and

$PM_{2.5}$. During the studied events of summer 2015, The CLBCs affected both $PM_{2.5}$ mean background concentration and its spatiotemporal variability. The CLBCs' influences on $PM_{2.5}$'s correlation coefficient (R) mainly appeared near the inflow boundary, and decreased along with the distance from the boundary. The influence of the CLBCs on $PM_{2.5}$ background concentration, however, could be seen further inside the domain. The CLBCs' influence on ozone was more complex, and affected both by the boundary inflows of ozone and/or its precursors, as well as downward transport from the upper troposphere and stratosphere. In this study, only the aerosol dynamic CLBCs (GEOS-LBC or NGAC-LBC) showed the impacts on the model spatiotemporal variability, while the static CLBCs mainly impacted the background concentrations and mean bias. It should be noted that this study mainly focused on the CLBCs' influence on surface sites. For elevated observational platforms, such as airborne measurements, the spatiotemporal variability of the CLBCs may also affect the three-dimensional ozone model performance due to the relatively fast transport and weak local ozone production in the upper layers (Tang et al., 2007)

The AOT-derived CLBCs for the northern boundary (AOT-NLBC) demonstrated that it could be used as an alternative method to capture intrusion events when the dynamic CLBCs from global models are not available. Although the VIIRS AOT was updated only once per day and the CLBCs derived from it had a relatively noisy spatial distribution, this method still showed its value to replace the static CLBCs in a near-real-time air quality forecast. For the wildfire intrusion events of summer 2018, the AOT-NLBC showed generally better model performances than the NGAC-LBC. It should be cautioned that using this method may lead to biases stemming from the discrepancies in AOT regression, or inconsistent representations of the timing or vertical distributions of atmospheric pollutants between the actual events and the database events used in the derivation. It should be noted that other indicators, such as surface monitoring data, can be also used to derive the similar CLBCs if the historical CLBCs have a good correlation with these data, and there is a relatively dense number of stations available near the inflow boundary. Geostationary satellites can also achieve a near-real-time AOT retrieval with a high temporal resolution (on the order of minutes), which will likely provide a better solution for fast capturing the intrusions that vary significantly in space and time. Currently, the main issue for using geostationary AOT is their relatively poor retrieval quality in high latitudes or under high zenith angles. As such issues become alleviated, geostationary AOT retrievals may be used as an indicator to derive the CLBCs, or even replace the CLBCs provided by the global models.

Code and Data availability

The source code used in this study is available online at https://github.com/NOAA-EMC/EMC_aqfs (last access: 4 May 2020; NOAA-EMC, 2020). The VIIRS AOT data used here are in ftp://ftp.star.nesdis.noaa.gov/pub/smcd/VIIRS_Aerosol/npp.viirs.aerosol.data/epsaot550/. The surface AIRNow monitoring data can be obtained via https://airnow.gov.

Acknowledgements

This research was supported by National Oceanic and Atmospheric Administration (NOAA) under the Weather Program Office (funding no. NA16OAR4590118) and the NOAA National Air Quality Forecast Capability (grant # T8MWQAQ). We thank the NASA MAP program and NASA Center for Climate Simulation for support of the GEOS GMI model and NASA Health and Air Quality program for supporting NAQFC satellite applications. The surface AIRNow observations are provided by the US EPA, and satellite data are produced by NASA and NOAA.

Reference

Appel, K. W., S. J. Roselle, R. C. Gilliam, and J. E. Pleim, Sensitivity of the Community Multiscale Air Quality (CMAQ) model v4. 7 results for the eastern United States to MM5 and WRF meteorological drivers. Geosci. Model Dev., 3, 169–188, 2010.

Arlander, D.W., Brüning, D., Schmidt, U. and Ehhalt, D.H., 1995. The tropospheric distribution of formaldehyde during TROPOZ II. Journal of atmospheric chemistry, 22(3), pp.251-269.

Bey, I., D.J. Jacob, J.A. Logan, R.M. Yantosca, Asian chemical outflow to the Pacific in spring: origins, pathways, and budgets. *J. Geophys. Res.*, 106 (D19), pp. 23097-23113, 2001.

Bian, H., M. Chin, D. A. Hauglustaine, M. Schulz, G. Myhre, S. E. Bauer, M. T. Lund, V. A. Karydis, T. L. Kucsera, X. Pan, A. Pozzer, R. B. Skeie, S. D. Steenrod, K. Sudo, K. Tsigaridis, A. P. Tsimpidi, and S. G. Tsyro (2017), Investigation of global particulate nitrate from the AeroCom phase III experiment Atmos. Chem. Phys., 17, 12911–12940 (https://www.atmos-chem-phys.net/17/12911/2017/).

Carlton, A. G.; Bhave, P. V.; Napelenok, S. L.; Edney, E. D.; Sarwar, G.; Pinder, R. W.; Pouliot, G. A.; Houyoux, M., Model representation of secondary organic aerosol in CMAQv4.7. Environmental Science & Technology, 44 (22), 8553-8560, 2010.

Chin, M., Rood, R. B., Lin, S.-J., Muller, J.-F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties, *J. Geophys. Res.*, 105, 24671–24687, doi:10.1029/2000JD900384, 2000.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V. , Logan, J. A., and Higurashi, A.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, *J. Atmos. Sci.*, 59, 461–483, 2002.

Colarco, P., da Silva, A., Chin, M., Diehl, T., Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground- based aerosol optical depth. J. Geophys. Res. 115, D14207. https://doi.org/10.1029/2009JD012820. 2010.

Eastham, S.D., Weisenstein, D.K. and Barrett, S.R., Development and evaluation of the unified tropospheric–stratospheric chemistry extension (UCX) for the global chemistry-transport model GEOS-Chem. Atmospheric Environment, 89, pp.52-63. http://dx.doi.org/10.1016/j.atmosenv.2014.02.001, 2014.

Foley, K. M., Roselle, S. J., Appel, K. W., Bhave, P. V., Pleim, J. E., Otte, T. L., Mathur, R., Sarwar, G., Young, J. O., Gilliam, R. C., Nolte, C. G., Kelly, J. T., Gilliland, A. B., and Bash, J. O.: Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7, Geosci. Model Dev., 3, 205–226, https://doi.org/10.5194/gmd-3-205-2010, 2010.

Helmig, D., Petrenko, V., Martinerie, P., Witrant, E., Rockmann, T., Zuiderweg, A., Holzinger, R., Hueber, J., Thompson, C., White, J.W.C. and Sturges, W., Reconstruction of Northern Hemisphere 1950-2010 atmospheric non-methane hydrocarbons. Atmospheric chemistry and physics, 14(3), pp.1463-1483, doi:10.5194/acp-14-1463-2014, 2014.

Henderson, B. H., Akhtar, F., Pye, H. O. T., Napelenok, S. L., and Hutzell, W. T.: A database and tool for boundary conditions for regional air quality modeling: description and evaluation, Geosci. Model Dev., 7, 339–360, https://doi.org/10.5194/gmd-7-339-2014, 2014.

Janach, W.E., 1989. Surface ozone: trend details, seasonal variations, and interpretation. *Journal of Geophysical Research: Atmospheres*, 94(D15), pp.18289-18295.

Lee, P., J. McQueen, I. Stajner, J. Huang, L. Pan, D. Tong, H.-C. Kim, Y. Tang, S. Kondragunta, and M. Ruminski, NAQFC developmental forecast guidance for fine particulate matter (PM2. 5), Weather and Forecasting, 32: 343-60. doi:10.1175/waf-d-15-0163.1, 2017.

Lu, C.-H., da Silva, A., Wang, J., Moorthi, S., Chin, M., Colarco, P., Tang, Y., Bhattacharjee, P. S., Chen, S.-P., Chuang, H.-Y., Juang, H.-M. H., McQueen, J., and Iredell, M.: The implementation of NEMS GFS Aerosol Component (NGAC) Version 1.0 for global dust forecasting at NOAA/NCEP, *Geosci. Model Dev.*, 9, 1905-1919, https://doi.org/10.5194/gmd-9-1905-2016, 2016.

Molod, A., L. Takacs, M. Suarez, J. Bacmeister, I.-S. Song, and A. Eichmann, The GEOS Atmospheric General Circulation Model: Mean Climate and Development from MERRA to Fortuna. Technical Report Series on Global Modeling and Data Assimilation, 28, 2012.

Pan, L., Tong, D., Lee, P., Kim, H.C. and Chai, T.. Assessment of NOx and O_3 forecasting performances in the US National Air Quality Forecasting Capability before and after the 2012 major emissions updates. *Atmospheric Environment*, 95, pp.610-619, 2014.

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

Pierce, T., C. Geron, L. Bender , R. Dennis, G. Tonnesen, and A. Guenther A, Influence of increased isoprene emissions on regional ozone modeling. *J. Geophys. Res.* 103:25611–25629, 1998.

Mathur, R., Xing, J., Gilliam, R., Sarwar, G., Hogrefe, C., Pleim, J., Pouliot, G., Roselle, S., Spero, T.L., Wong, D.C. and Young, J., Extending the Community Multiscale Air Quality (CMAQ) modeling system to hemispheric scales: overview of process considerations and initial applications. Atmospheric chemistry and physics, 17, p.12449. doi: 10.5194/acp-17-12449-2017, 2017

Sarwar, G., Luecken, D., and Yarwood, G.: Chapter 2.9 Developing and implementing an updated chlorine chemistry into the community multiscale air quality model, in: Air Pollution Modeling and Its Application XVIII, edited by: Borrego, C. and Renner, E., vol. 6 of Developments in Environmental Science, 168– 176, Elsevier, Amsterdam, the Netherlands, doi:10.1016/S1474- 8177(07)06029-9, 2007.

Sillman, S., He, D., Cardelino, C. and Imhoff, R.E., The use of photochemical indicators to evaluate ozone-NOx-hydrocarbon sensitivity: Case studies from Atlanta, New York, and Los Angeles. *Journal of the Air & Waste Management Association*, 47(10), pp.1030-1040, 1997.

Sonntag, D. B., R. W. Baldauf, C. A. Yanca and C. R. Fulper, Particulate matter speciation profiles for light-duty gasoline vehicles in the United States, *Journal of the Air & Waste Management Association*, 64:5, 529-545, DOI:10.1080/10962247.2013.870096, 2014.

Strode S.A., J.R. Ziemke, L.D. Oman, L.N. Lamsal, M.A. Olsen, J. Liu, Global changes in the diurnal cycle of surface ozone, *Atmospheric Environment*, 199, 323-333, https://doi.org/10.1016/j.atmosenv.2018.11.028, 2019.

Tanaka, P. L., Allen, D. T., McDonald-Buller, E. C., Chang, S., Kimura, Y., Mullins, C. B., Yarwood, G., and Neece, J. D.: Development of a chlorine mechanism for use in the carbon bond IV chemistry model, *J. Geophys. Res.-Atmos.*, 108, 4145, doi:10.1029/2002JD002432, 2003.

Tang Y., Carmichael G. R., Thongboonchoo N., Chai T., Horowitz L.W., Pierce R. B., Al-Saadi J. A., Pfister G., Vukovich J. M., Avery M. A., Sachse G. W., Ryerson T. B., Holloway J. S., Atlas E. L., Flocke F. M., Weber R. J., Huey L. G., Dibb J. E., Streets D. G., and Brune W. H.: Influence of lateral and top boundary conditions on regional air quality prediction: a multiscale study coupling regional and global chemical transport models. *J. Geophys. Res.* 112:D10S18. doi:10.1029/2006JD007515, 2007.

Tang, Y., Lee, P., Tsidulko, M., Huang, H.C., McQueen, J.T., DiMego, G.J., Emmons, L.K., Pierce, R.B., Thompson, A.M., Lin, H.M. and Kang, D.: The impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the continental United States. *Environmental Fluid Mechanics*, 9(1), pp.43-58, DOI:10.1007/s10652-008-9092-5, 2009

Tyndall, G. S., Cox, R. A., Granier, C., Lesclaux, R., Moortgat, G. K., Pilling, M. J., Ravishankara, A.R. and Wallington, T. J.: Atmospheric chemistry of small organic peroxy radicals. Journal of Geophysical Research: Atmospheres, 106(D11), pp.12157-12182, 2001

Urbanski, S.P., Hao, W.M. and Baker, S., Chemical composition of wildland fire emissions. Developments in environmental science, 8, pp.79-107. DOI:10.1016/S1474-8177(08)00004-1. 2008

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

Whitby, E. R., and P. H. McMurry, Modal aerosol dynamics modeling. Aerosol Science and Technology 27: 673-688, 1997

Whitten, G. Z., Heo, G., Kimura, Y., McDonald-Buller, E., Allen, D. T., Carter, W. P., and Yarwood, G.: A new condensed toluene mechanism for Carbon Bond: CB05-TU, *Atmos. Environ.*, 44, 5346–5355, doi:10.1016/j.atmosenv.2009.12.029, 2010.

Yarwood, G., S. Rao, M. Yocke, and G. Whitten. Updates to the Carbon Bond Chemical Mechanism: CB05, Technical Report RT-0400675 ENVIRON International Corporation Novato, CA, USA. 2005.

Table 1. The runs with different lateral boundary conditions conducted in this study

Runs	Aerosol LBCs	Gaseous LBCs	Temporal Resolution
CMAQ_Base	static clean background	static GEOS-Chem 2006 with $O_3 \leq 100$ ppbV	static monthly mean
GEOS-LBC	dynamic full aerosol	dynamic full chemistry	3 hours
GLBC-Monthly	monthly mean full aerosol	monthly mean full chemistry	static monthly mean
NGAC-LBC	dynamic GOCART simple aerosol	Same as CMAQ_Base	3 hours
AOT-NLBC	daily AOT derived Northern LBC (NLBC) for EC and POC	daily AOT derived Northern LBC for CO, NOx, PAN, and HNO_3	24 hours for derived NLBC; static monthly mean for all others

Table 2. VOC species mapping table from GEOS to CMAQ CB05tucl

GEOS species (mole)	CMAQ Species (mole)
HCOOH	FACD
MO_2 (CH_3O_2)	XO_2
MP (methylhydroperoxide)	MEPX
A_3O_2 (primary RO_2 from C_3H_8: $CH_3CH_2CH_2OO$)	PAR + XO_2
ACTA (acetic acid)	AACD
ATO_2 (RO_2 from acetone: $CH_3C(O)CH_2O_2$)	2*PAR + XO_2
B_3O_2 (secondary RO_2 from C_3H_8: $CH_3CH(OO)CH_3$)	2*B_3O_2
ALK4 (C_4 or higher alkanes)	4*PAR
C_3H_8	1.5*PAR + NR
ETO_2 (ethylperoxy radical: CH_3CH_2OO)	MEO_2 + PAR
ETP (ethylhydroperoxide: CH_3CH_2OOH)	MEPX + PAR
GCO_3 (hydroxy peroxyacetyl radical: $HOCH_2C(O)OO$)	C_2O_3
GLYX (glyoxal)	FORM + PAR
GLYC (glycolaldehyde: $HOCH_2CHO$)	FORM + 2*PAR
GP (peroxide from GCO_3: $HOCH_2C(O)OOH$)	ROOH
GPAN (Peroxyacylnitrate: $HOCH_2C(O)OONO_2$)	PANX
HAC (hydroxyacetone: $HOCH_2C(O)CH_3$)	2*PAR
IALD (hydroxy carbonyl alkenes from isoprene)	ISOPX
IAO_2 (RO_2 from isoprene oxidation products)	$ISOPO_2$
IAP (peroxide from IAO_2)	ROOH
INO_2 (RO_2 from ISOP+NO_3)	0.2*ISPD + 0.8*NTR+ XO_2 + 0.8*HO_2 + 0.2*NO_2 + 0.8*ALDX + 2.4*PAR'
INPN (peroxide from INO_2)	0.2*ISPD + 0.8*NTR+ ROOH + 0.8*H_2O_2 + 0.2*PNA + 0.8*ALDX + 2.4*PAR
ISN1 (RO_2 from isoprene nitrate)	NTRI
ISNP (peroxide from ISN1)	$NTRIO_2$
KO_2 (RO_2 from C_3 or higher ketones)	XO_2 + PAR
MACR (methacrolein)	ISPD
MAN2 (RO_2 from MACR+NO_3)	0.925*HO_2 + 0.075*XO_2
MAO_3 (peroxyacyl from MVK and MACR)	$MACO_3$
MAOP (peroxide from MAO_3)	ISPD
MAP (peroxyacetic acid, $CH_3C(O)OOH$)	PACD
MCO_3 (peroxyacetyl radical)	C_2O_3
MEK (C_3 or higher ketones)	4*PAR
MRO_2 (RO_2 from MACR+OH)	0.713*XO_2 + 0.503*HO_2
MRP (Peroxide from MRO_2)	ROOH
MVK (methylvinylketone)	ISPD
MVN2 (RO_2 from MVK+NO_3)	0.925*HO_2 + 0.075*XO_2
PMN (peroxymethacryloyl nitrate)	OPEN

PO_2 (RO_2 from propene)	XO_2
PP (peroxide from PO_2: HOC_3H_6OOH)	ROOH
PPN (peroxypropionyl nitrate)	PANX
PRN1 (RO2 from propene+NO_3)	XO_2
PRPE (propene)	OLE + PAR
PRPN (peroxide from PRN1)	ROOH
R4N1 (RO_2 from C_4 and C_5 alkylnitrates)	ROOH + 2*PAR
R4O2 (RO_2 from C_4 alkane)	XO_2
R4P (peroxide from R4O2)	ROOH
RA3P (peroxide from A_3O_2)	ROOH
RB3P (Peroxide from B_3O_2)	ROOH
RCHO (C_3 or higher aldehydes)	ALDX
RCO3 (peroxypropionyl radical: $CH_3CH_2C(O)OO$)	XO_2
RCOOH (C_2 or higher organic acids)	AACD
RIO1 (RO_2 from isoprene oxidation products)	ISPD
RIO2 (RO_2 from isoprene)	$ISOPO_2$
RIP (Peroxide from RIO_2)	ISOPX
ROH (C_2 or higher alcohols)	3*PAR
RP (peroxide from RCO_3)	ROOH
VRO_2 (RO_2 from MVK+OH)	$ISOPO_2$
VRP (peroxide from VRO_2)	ROOH
ACET (acetone)	3*PAR

Table 3. Aerosol species mapping table from GEOS to CMAQ Aero6 ("D" represents the diameter of GEOS aerosol bin)

GEOS Aerosol (μg/m^3)	**CMAQ Aerosol Mass Concentration (μg/m^3)**	**CMAQ Aerosol Number Concentration (#/m^3)**
BCPHILIC	AECJ	2.72×10^7 (ACC)
BCPHOBIC	AECJ	2.72×10^7 (ACC)
OCPHILIC	APOCJ	2.72×10^7 (ACC)
OCPHOBIC	APOCJ	2.72×10^7 (ACC)
SO4	ASO4J	2.72×10^7 (ACC)
NH4a	ANH4J	2.72×10^7 (ACC)
NO3an1 (mean D=0.5μm)	ANO3J	2.72×10^7 (ACC)
NO3an2 (mean D=4.2μm)	0.8*ANO3J + 0.2 *ANO3K	5.4×10^6(ACC) + 1.2×10^4(COR)
NO3an2 (mean D=15μm)	ANO3K	6×10^3(COR)
DU001 (D: 0.2 – 2 μm)	AOTHRJ	2.72×10^7 (ACC)
DU002 (D: 2 – 3.6 μm)	0.45*AOTHRJ+0.55*ASOIL	3.3×10^5 (ACC)+5.1×10^4(COR)
DU003 (D: 3.6 – 6 μm)	ASOIL	1.15×10^4(COR)
DU004 (D: 6 – 12 μm)	0.75*ASOIL	1.4×10^3(COR)
SS001 (D: 0.06-0.2 μm)	0.39*ANAI+0.61*ACLI	7.4×10^8(ATKN)
SS002 (D: 0.2 - 1μm)	0.39*ANAJ+0.61*ACLJ	2.72×10^7 (ACC)
SS003 (D: 1- 3μm)	0.312*ANAJ+0.488*ACLJ +0.078*ASEACAT+0.122*ACLK	1.7×10^5(ACC)+1.26×10^4(COR)
SS004 (D: 3- 10μm)	0.39*ASEACAT+0.61*ACLK	1.36×10^4 (COR)

Table 4. Regional $PM_{2.5}$ statistics of the four simulations (CMAQ_Base, GEOS-LBC, GLBC-Monthly and NGAC-LBC) from 24 June to 8 July, 2015.

Regions	Simulations	Mean Bias ($\mu g/m^3$)	Root Mean Square Error ($\mu g/m^3$)	Correlation Coefficient, R	Index of Agreement
Contiguous U.S.	CMAQ_BASE	-6.74	13.69	0.18	0.37
	GEOS-LBC	-2.96	12.16	0.37	0.55
	GLBC-Monthly	-4.10	12.39	0.27	0.41
	NGAC-LBC	-3.30	12.09	0.30	0.44
Northeastern U.S.	CMAQ_BASE	-5.52	10.93	0.33	0.43
	GEOS-LBC	-3.81	9.89	0.40	0.50
	GLBC-Monthly	-4.25	10.31	0.34	0.45
	NGAC-LBC	-3.70	10.05	0.35	0.46
Pacific Coast	CMAQ_BASE	-3.96	10.63	0.16	0.31
	GEOS-LBC	-2.02	10.22	0.18	0.34
	GLBC-Monthly	-1.53	10.21	0.17	0.34
	NGAC-LBC	-0.79	10.33	0.16	0.34
Southeastern U.S.	CMAQ_BASE	-8.18	11.35	0.14	0.44
	GEOS-LBC	-3.07	8.39	0.37	0.58
	GLBC-Monthly	-4.78	9.08	0.27	0.49
	NGAC-LBC	-3.83	8.58	0.35	0.56
Rocky Mountain States	CMAQ_BASE	-7.62	17.57	0.02	0.31
	GEOS-LBC	-3.66	15.98	0.39	0.58
	GLBC-Monthly	-5.42	16.06	0.23	0.36
	NGAC-LBC	-4.65	15.78	0.24	0.36
North Central	CMAQ_BASE	-8.32	17.63	0.25	0.38
	GEOS-LBC	-2.95	16.47	0.33	0.52
	GLBC-Monthly	-5.25	16.41	0.27	0.40
	NGAC-LBC	-4.48	15.98	0.31	0.43
South Central	CMAQ_BASE	-9.65	13.12	0.07	0.42
	GEOS-LBC	-2.00	7.79	0.51	0.69
	GLBC-Monthly	-4.73	9.45	0.24	0.48
	NGAC-LBC	-3.52	8.31	0.46	0.63

Table 5. Same as Table 4 but for ozone

Regions	Simulations	Mean Bias (ppbV)	Root Mean Square Error (ppbV)	Correlation Coefficient, R	Index of Agreement
Contiguous U.S.	CMAQ_BASE	2.10	12.35	0.64	0.77
	GEOS-LBC	3.47	12.01	0.68	0.79
	GLBC-Monthly	4.84	12.52	0.68	0.78
	NGAC-LBC	1.88	12.29	0.64	0.77
Northeastern U.S.	CMAQ_BASE	1.87	10.68	0.66	0.78
	GEOS-LBC	4.88	11.54	0.68	0.78
	GLBC-Monthly	5.60	12.02	0.66	0.76
	NGAC-LBC	1.62	10.64	0.66	0.78
Pacific Coast	CMAQ_BASE	-2.58	12.04	0.78	0.86
	GEOS-LBC	-2.16	11.83	0.79	0.87
	GLBC-Monthly	0.46	11.79	0.78	0.87
	NGAC-LBC	-2.76	12.08	0.78	0.86
Southeastern U.S.	CMAQ_BASE	7.26	13.66	0.59	0.68
	GEOS-LBC	7.94	13.34	0.66	0.72
	GLBC-Monthly	9.06	14.20	0.65	0.70
	NGAC-LBC	7.04	13.50	0.60	0.69
Rocky Mountain States	CMAQ_BASE	-1.91	10.61	0.67	0.80
	GEOS-LBC	-0.17	10.45	0.67	0.80
	GLBC-Monthly	1.68	10.75	0.66	0.79
	NGAC-LBC	-2.08	10.63	0.67	0.80
North Central	CMAQ_BASE	-0.47	10.78	0.65	0.78
	GEOS-LBC	2.55	11.01	0.66	0.79
	GLBC-Monthly	3.00	11.22	0.65	0.78
	NGAC-LBC	-0.75	10.76	0.65	0.78
South Central	CMAQ_BASE	13.36	17.76	0.51	0.58
	GEOS-LBC	10.90	14.71	0.68	0.68
	GLBC-Monthly	12.66	16.24	0.66	0.64
	NGAC-LBC	13.12	17.56	0.51	0.58

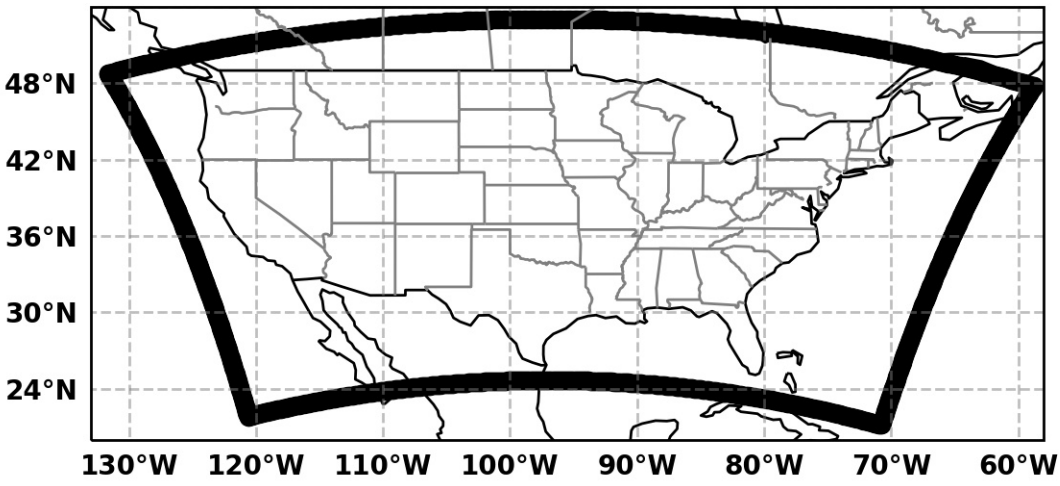

Figure 1, NAQFC contiguous U.S. domain (outlined in bolded black)

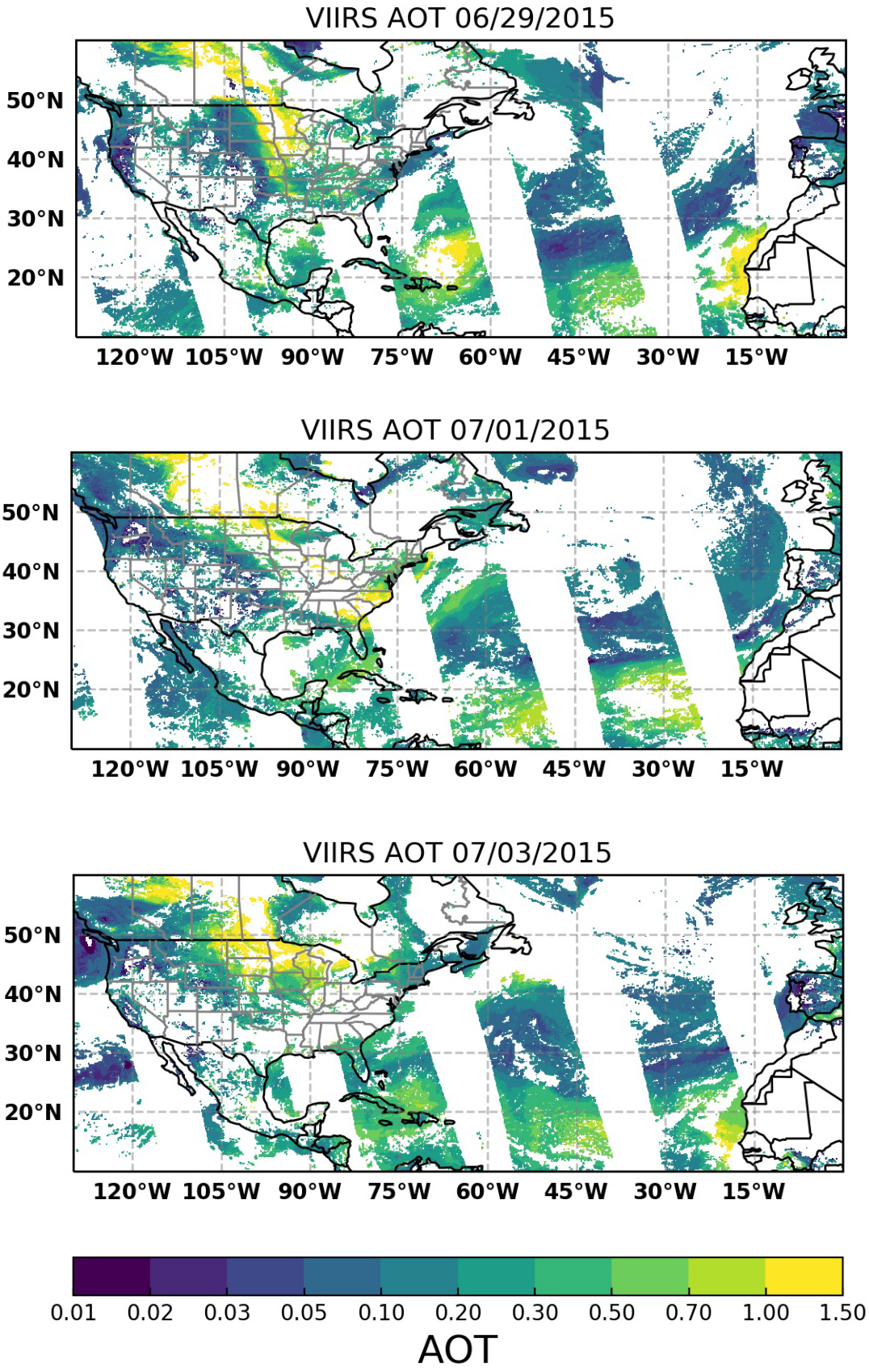

Figure 2. S-NPP VIIRS Aerosol Optical Thickness (AOT) on 29 June, 1 July, and 3 July of 2015.

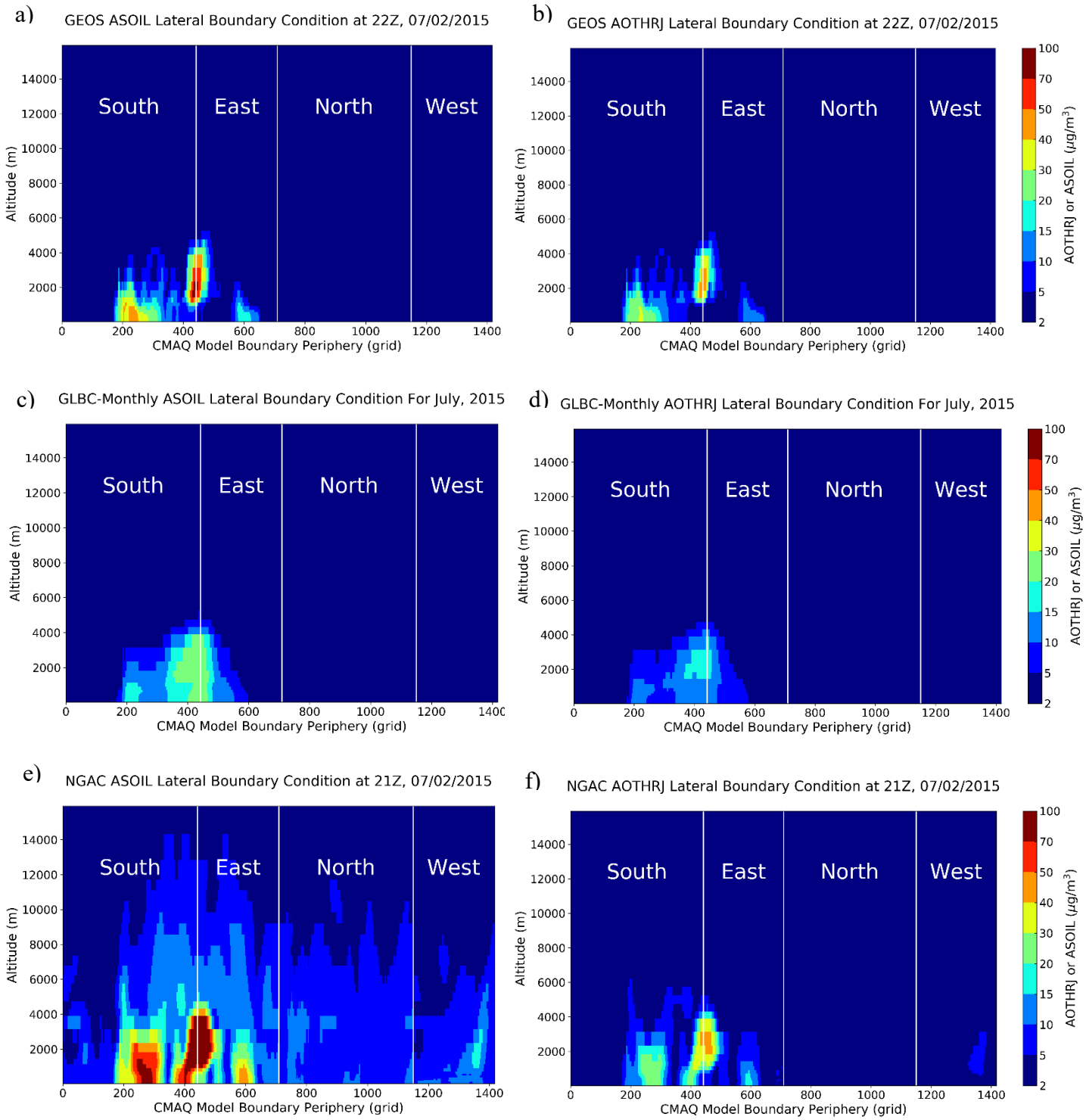

Figure 3. The lateral boundary conditions for ASOIL (left) and AOTHRJ (right) along the domain periphery for 2 July, 2015. The CMAQ LBC's grid index for each LBC segment is always from south to north and then from west to east, so the LBC index's start-points are reset instead of continuous for the north and west boundaries.

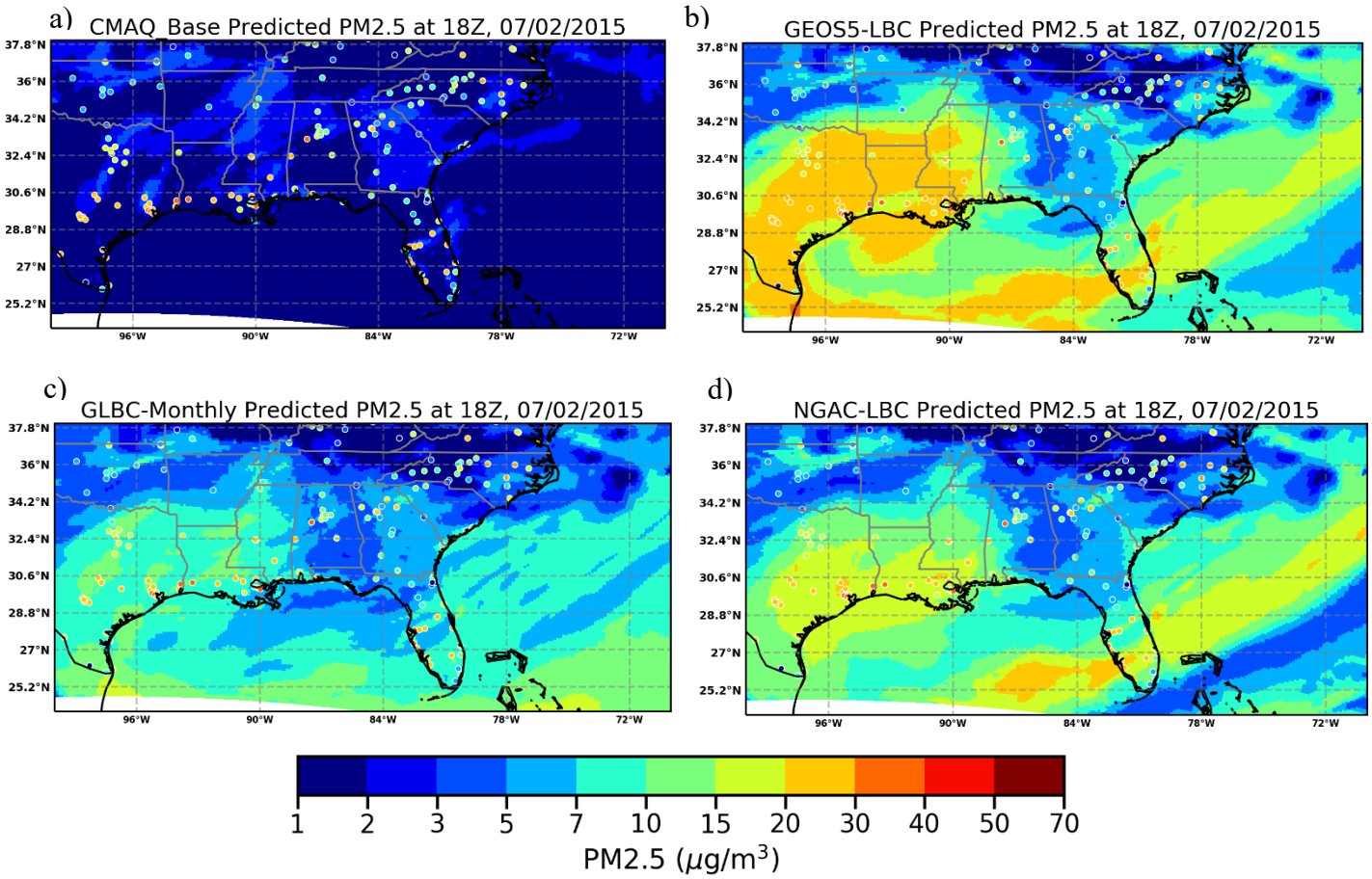

Figure 4. Model predicted surface $PM_{2.5}$ concentrations with the four LBCs on 2 July, 2015 (the colored circles showing the AIRNow observations)

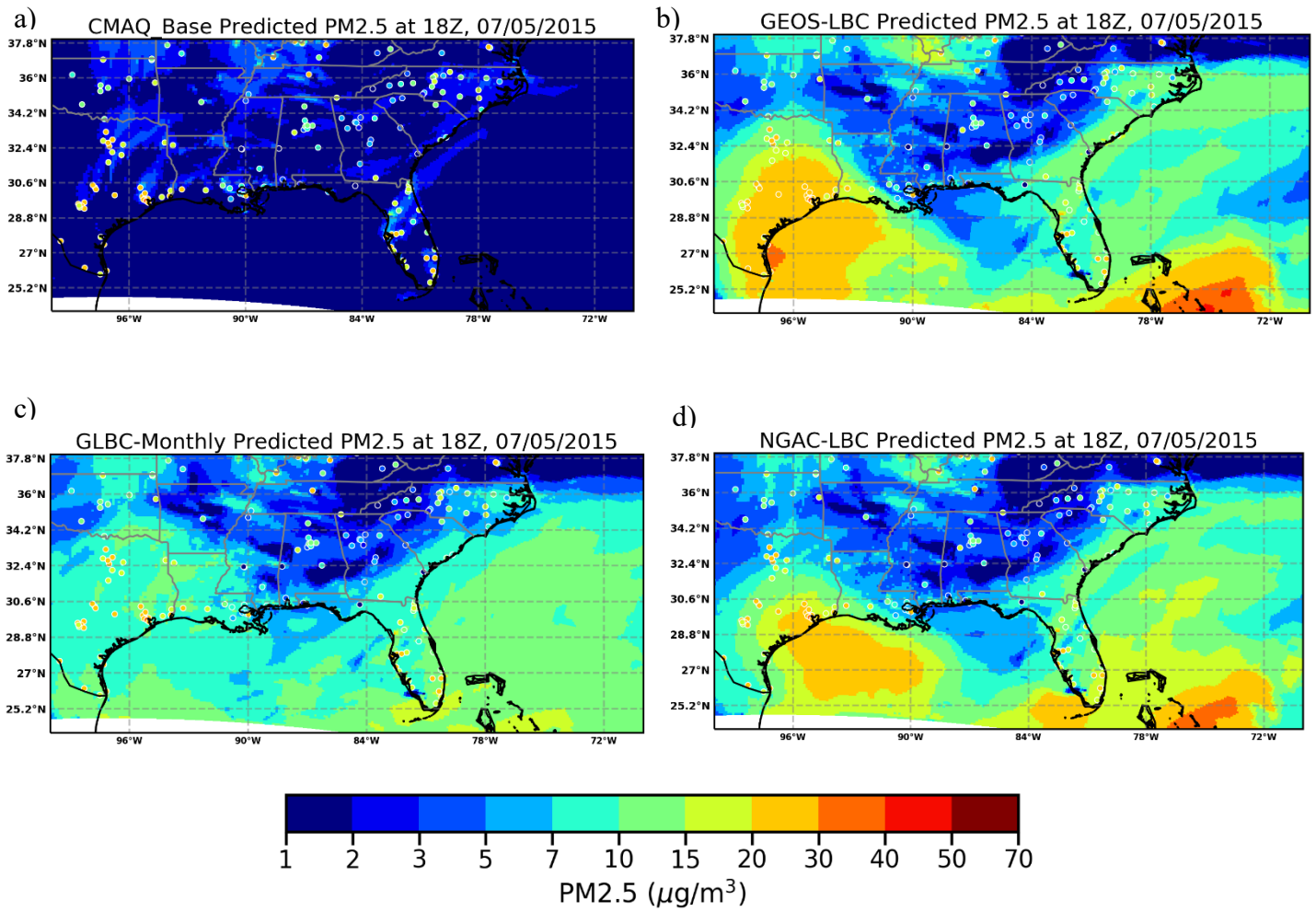

Figure 5. Same as figure 4 but for 5 July, 2015

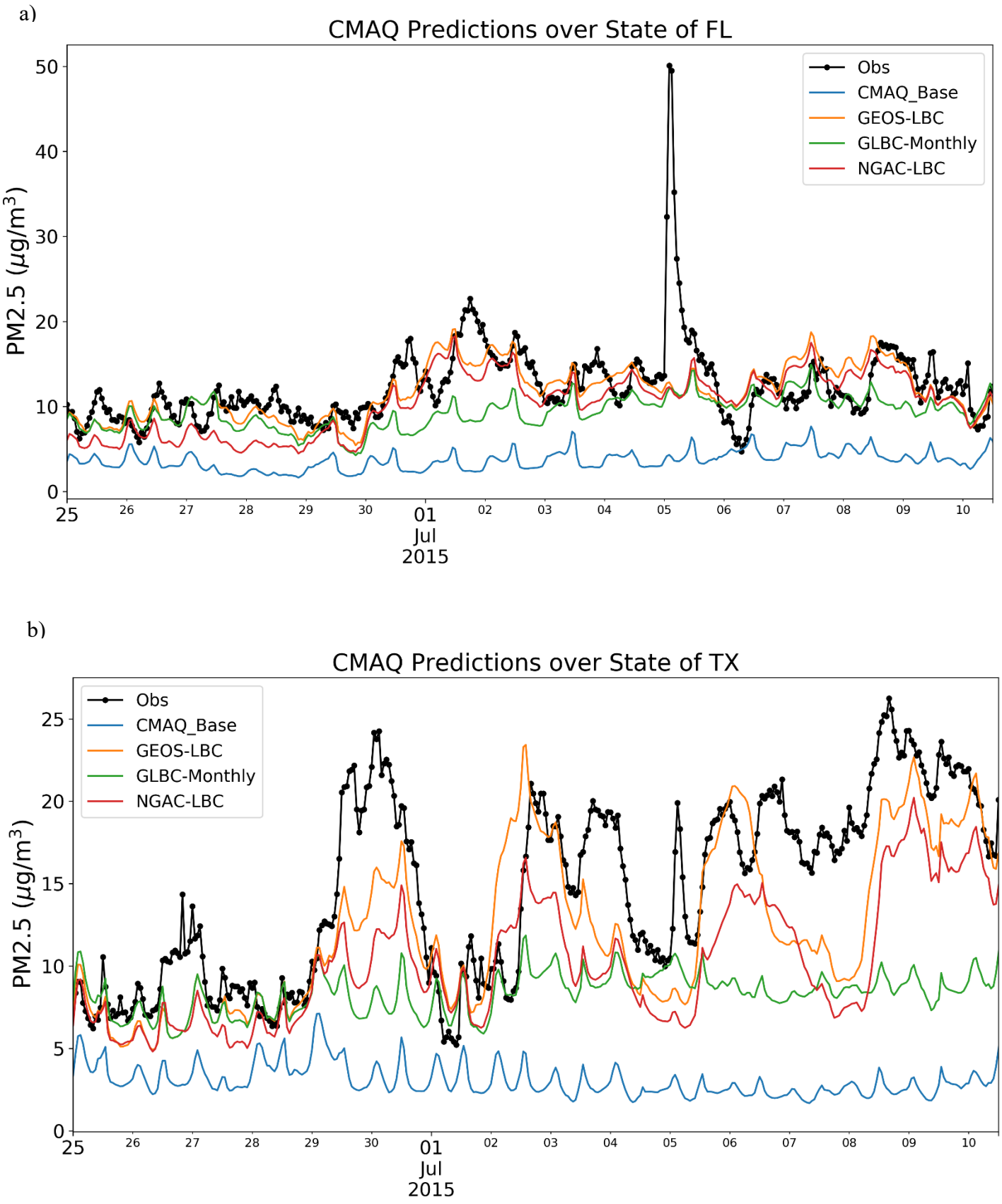

Figure 6. Time-series $PM_{2.5}$ comparisons over the states of Florida and Texas. All the times are in UTC.

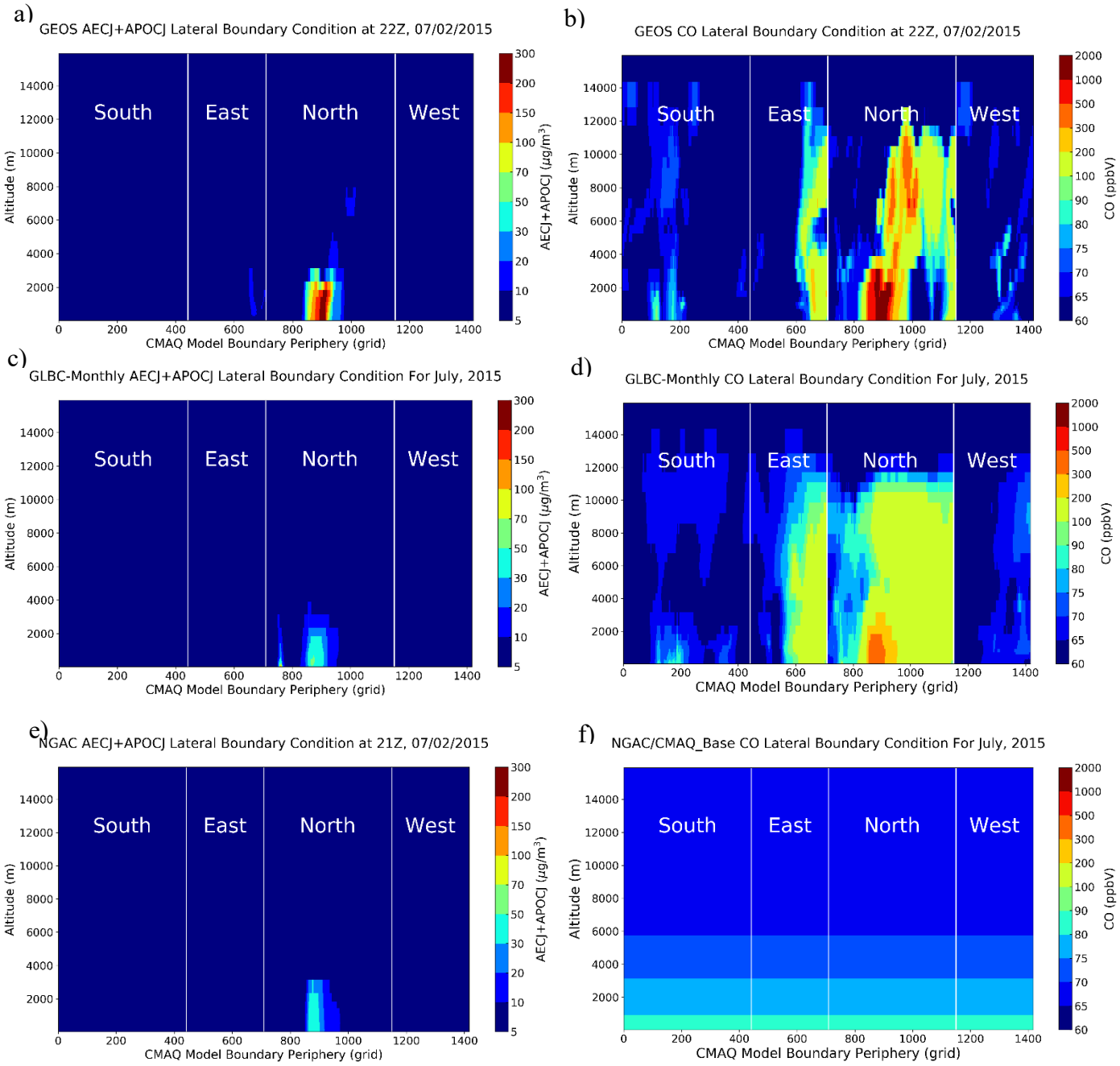

Figure 7, same as Figure 3 except for total EC and POC (AECJ+APOCJ) (left) and CO (right).

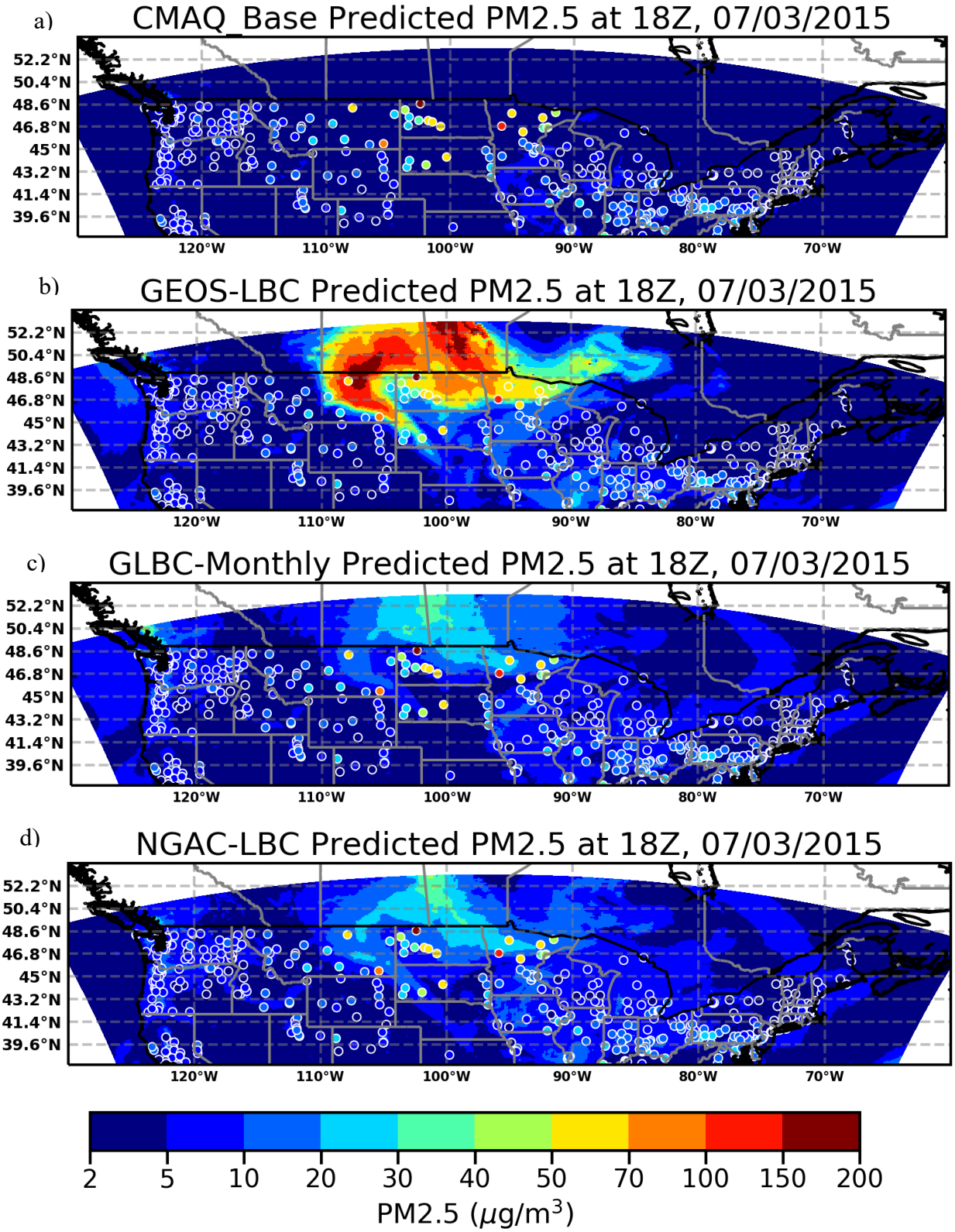

Figure 8, same as Figure 4, but for Northern USA on 3 July, 2015

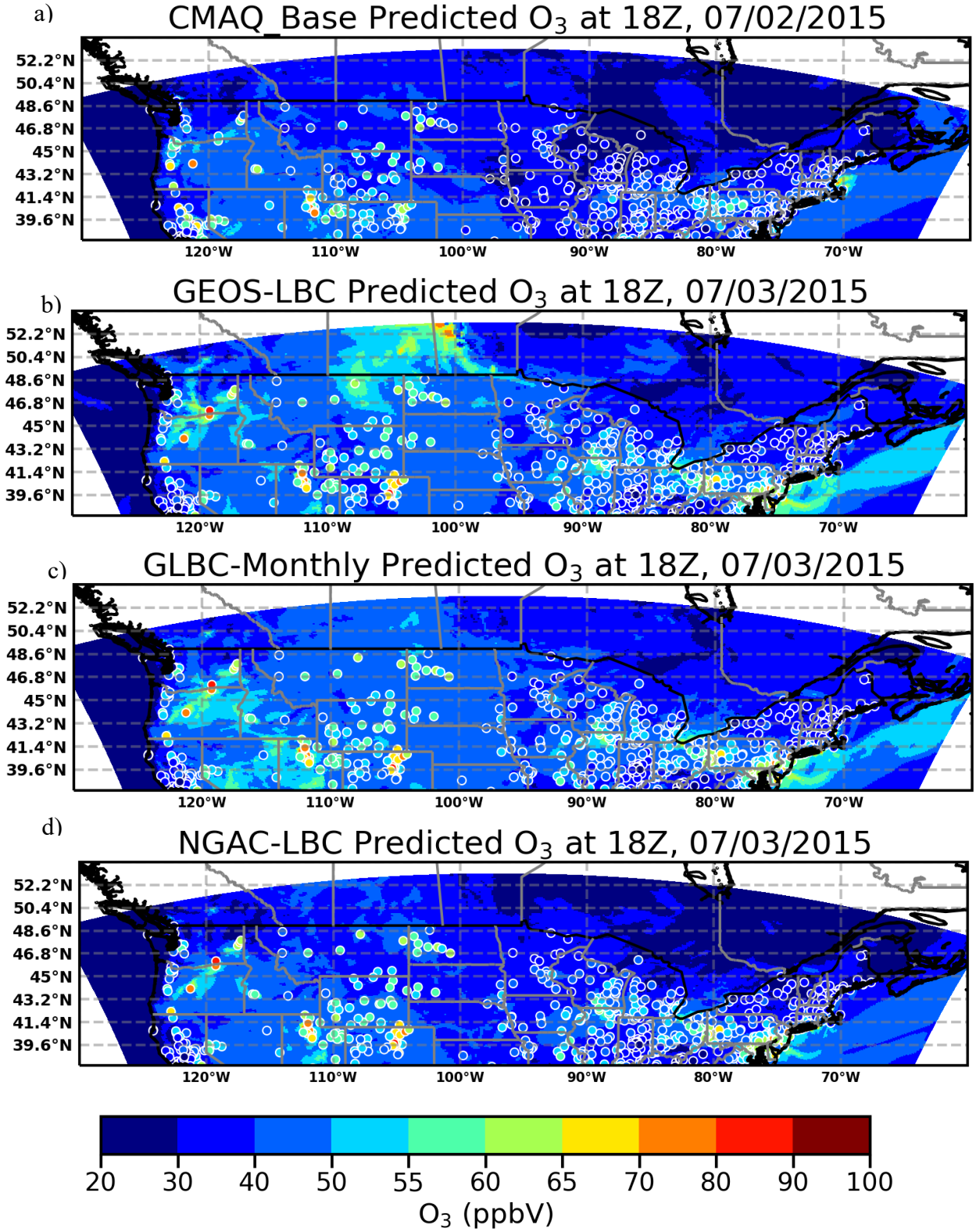

Figure 9, same as Figure 8, but for O_3.

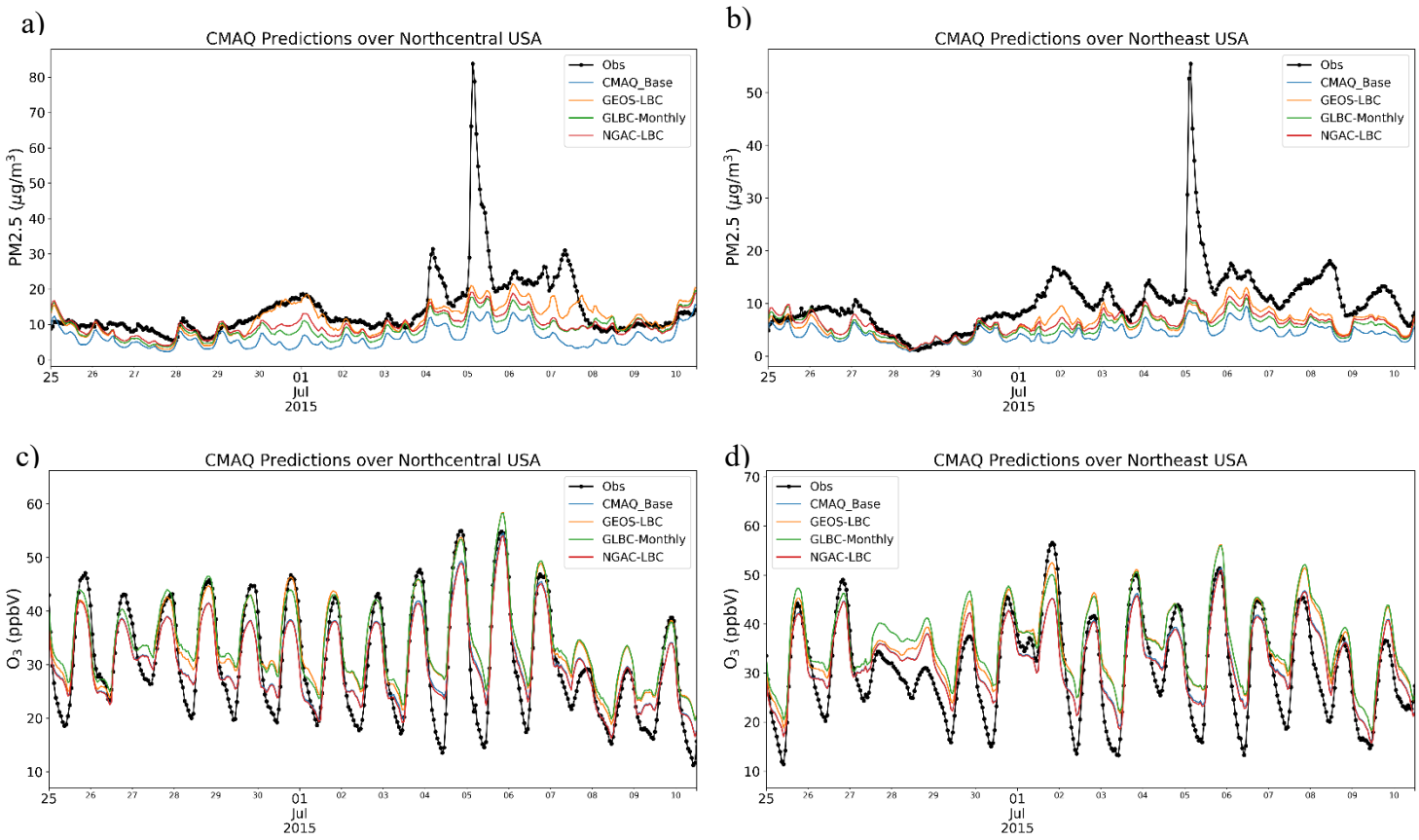

Figure 10. Time-series comparisons for $PM_{2.5}$ (top) and O_3 (bottom) over the northcentral region (left) (States of Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Ohio, and Wisconsin) and northeastern U.S. (right) (States of Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont and District of Columbia). All the time is in UTC.

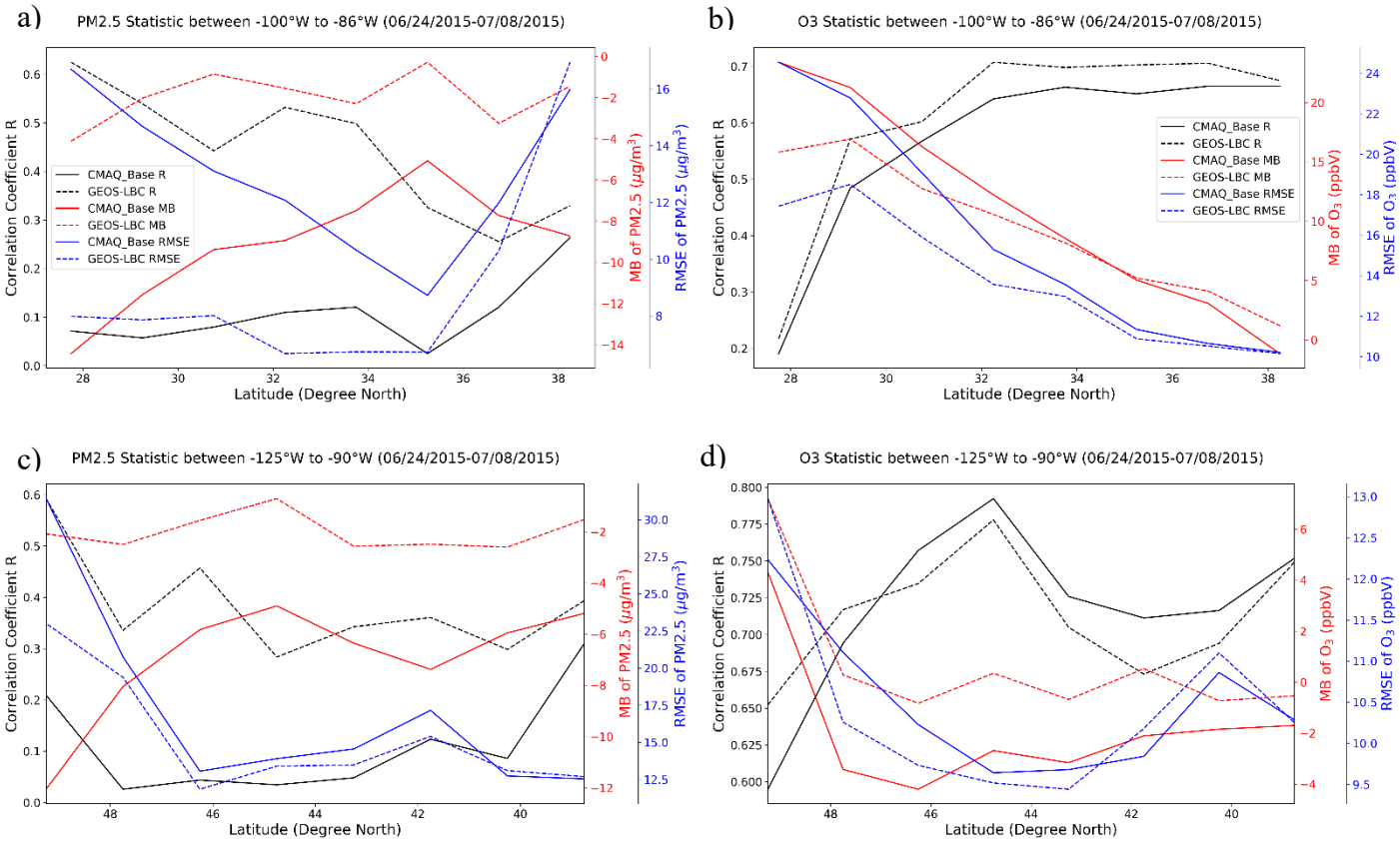

Figure 11, The latitudinal distributions of correlation coefficient R (black), mean bias (MB) (red), and root mean square error (RMSE) (blue) of $PM_{2.5}$ (left) and O_3 (right) concentrations from 24 June to 8 July, 2015 over the southern U.S. (top) and northern U.S. (bottom) for CMAQ_Base (solid line) and GEOS-LBC (dash line) runs.

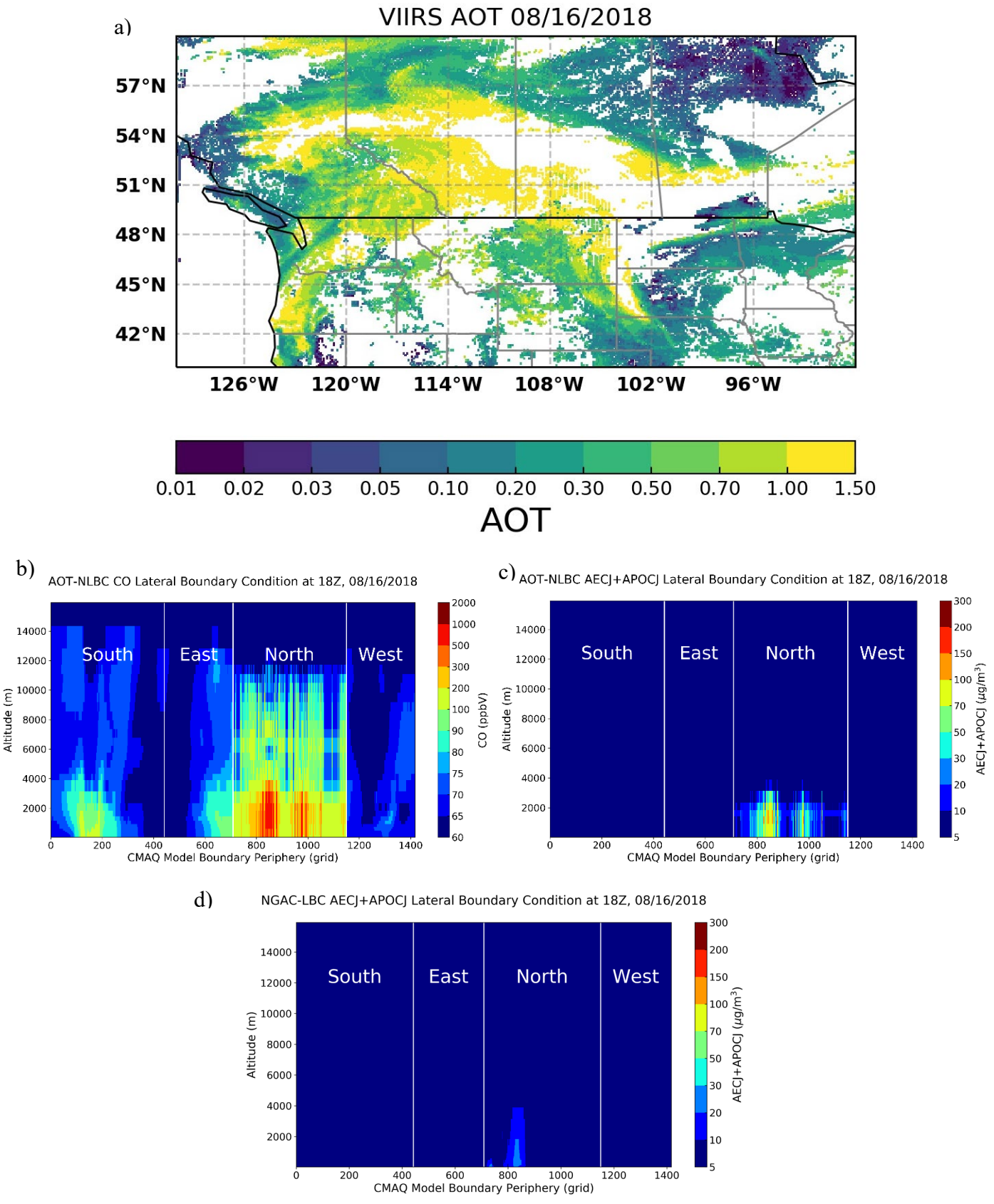

Figure 12. VIIRS-AOT (a) on 16 August, 2018, and the corresponding derived AOT-NLBC for CO (b) and AECJ+APOCJ (c). Plot d shows the NGAC-LBC's AEC+APOCJ at the same time.

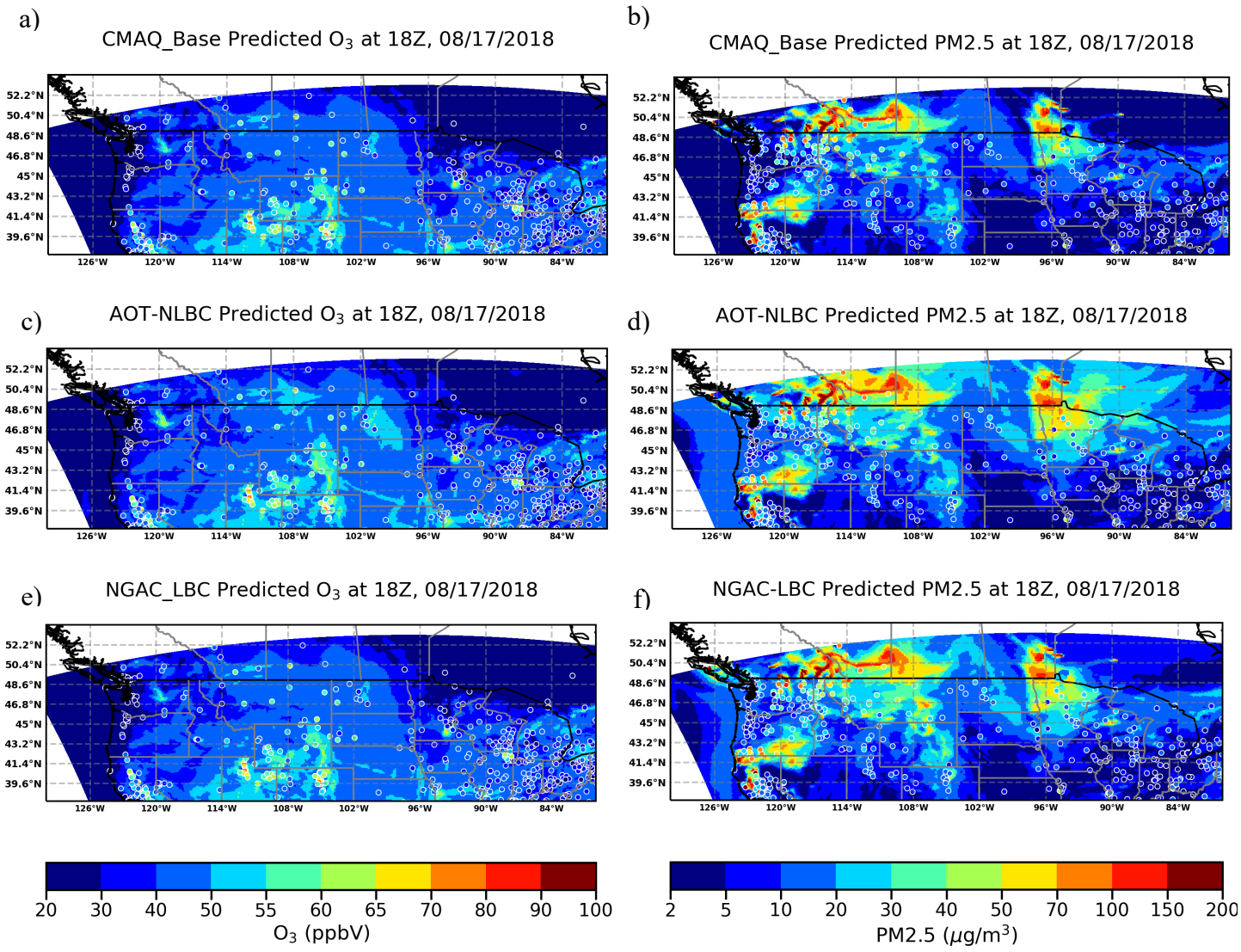

Figure 13. Model predicted surface ozone (left) and $PM_{2.5}$ (right) with the CMAQ_Base (a, b), AOT-NLBC (c, d) and NGAC-LBC (e, f) for 17 August, 2018 (the colored circles show the AIRNow observations)

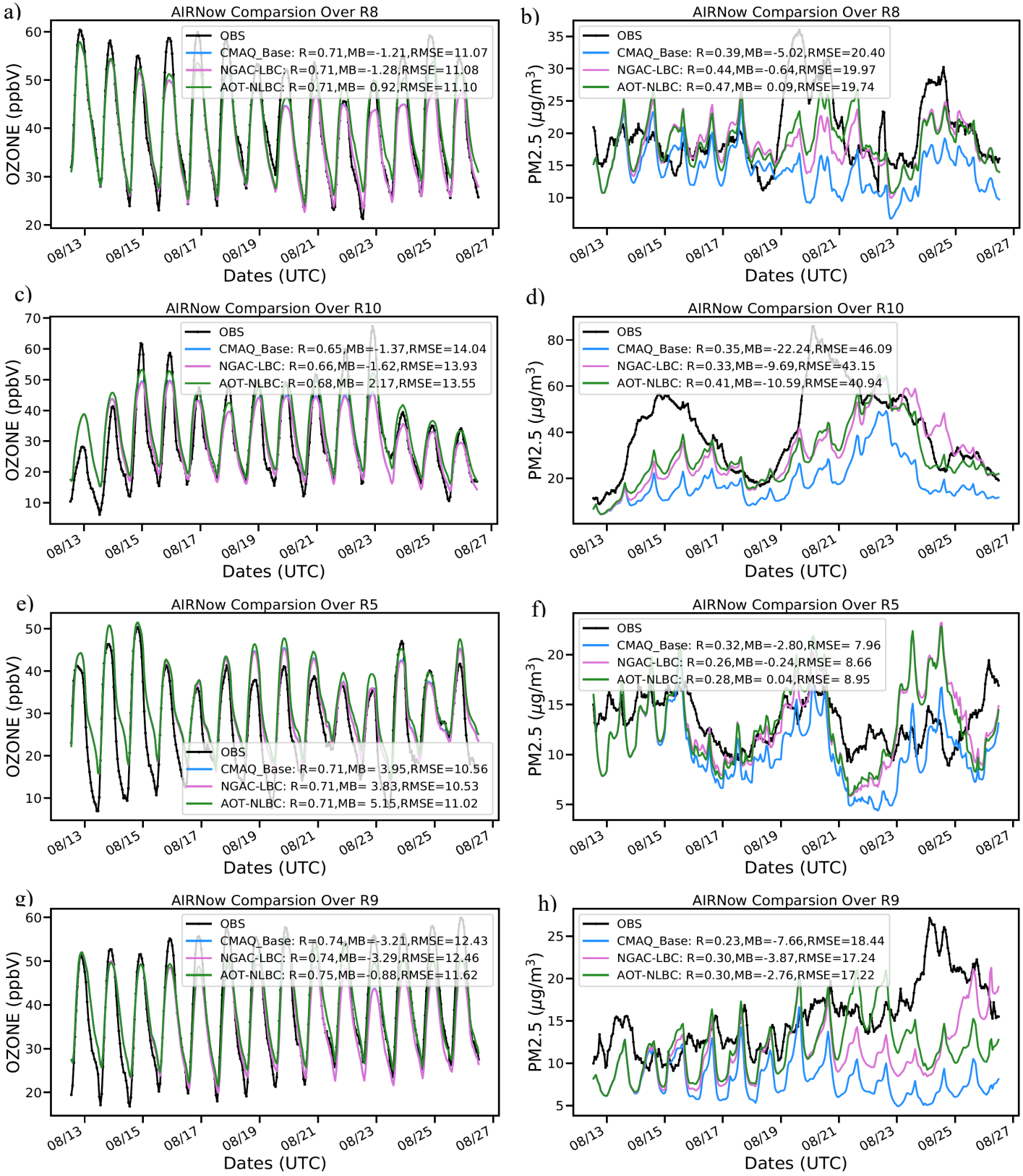

Figure 14, AIRNow time-series comparisons for surface ozone (left) and $PM_{2.5}$(right) over EPA Region 8 (R8, states of MT, ND, SD, WY, CO and UT), Region 10 (R10, states of WA, ID and OR), Region 5 (R5, states of MN, WI, IL, IN, MI, and OH) and Region 9 (R9, states of CA, NV, and AZ) predicted by CMAQ_Base, NGAC-LBC and AOT-NLBC in August, 2018