Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO$_2$ measurements

Tianfeng Chai[1,2]**, Hyun-Cheol Kim**[1,2]**, Pius Lee**[1]**, Daniel Tong**[1,2,3]**, Li Pan**[1,2]**, Youhua Tang**[4,5]**, Jianping Huang**[4,5]**, Jeff McQueen**[4]**, Marina Tsidulko**[4,5]**, and Ivanka Stajner**[6]

[1]NOAA Air Resources Laboratory (ARL), NOAA Center for Weather and Climate Prediction, 5830 University Research Court College Park, MD 20740, USA
[2]Cooperative Institute for Climate and Satellites, University of Maryland, College Park, MD 20740, USA
[3]Center for Spatial Information Science and Systems (CSISS),George Mason University, Fairfax VA 22030, USA
[4]NOAA/NWS/NCEP/EMC, NOAA Center for Weather and Climate Prediction, 5830 University Research Court College Park, MD 20740, USA
[5]I.M. Systems Group, Rockville, MD 20852
[6]NOAA/NWS/OST, Silver Spring, MD 20910, USA

Correspondence to: Tianfeng Chai
(tianfeng.chai@noaa.gov)

Abstract

The National Air Quality Forecast Capability (NAQFC) project provides the US with operational and experimental real-time ozone predictions using two different versions of the three-dimensional Community Multi-scale Air Quality (CMAQ) Modeling System. Routine evaluation using near-real-time AIRNow ozone measurements through 2011 showed better performance of the operational ozone predictions. In this work, quality-controlled and -assured Air Quality System (AQS) ozone and nitrogen dioxide (NO_2) observations are used to evaluate the experimental predictions in 2010. It is found that both ozone and NO_2 are overestimated over the contiguous US (CONUS), with annual biases of $+5.6$ ppbv and $+5.1$ ppbv, respectively. The annual root mean square errors (RMSEs) are 15.4 ppbv for ozone and 13.4 ppbv for NO_2. For both species the over-predictions are most pronounced in the summer. The locations of the AQS monitoring sites are also utilized to stratify comparisons by the degree of urbanization. Comparisons for six predefined US regions show the highest annual biases for ozone predictions in Southeast ($+10.5$ ppbv) and for NO_2 in the Lower Middle ($+8.1$ ppbv) and Pacific Coast ($+7.1$ ppbv) regions. The spatial distributions of the NO_2 biases in August show distinctively high values in Los Angeles, Houston, and New Orleans areas. In addition to the standard statistics metrics, daily maximum eight-hour ozone categorical statistics are calculated using the current US ambient air quality standard (75 ppbv) and another lower threshold (70 ppbv). Using the 75 ppbv standard, the hit rate and proportion of correct over CONUS for the entire year are 0.64 and 0.96, respectively. Summertime biases show distinctive weekly patterns for ozone and NO_2. Diurnal comparisons show that ozone overestimation is most severe in the morning, from 07:00 to 10:00 local time. For NO_2, the morning predictions agree with the AQS observations reasonably well, but night-time concentrations are over-predicted by around 100 %.

1 Introduction

2

The US National Air Quality Forecast Capability (NAQFC) started as a joint effort between the National Oceanic and Atmospheric Administration (NOAA) and the US Environmental Protection Agency (EPA) to provide advance notice for future air pollution events with potential adverse health effects. By linking the National Centers for Environmental Prediction (NCEP) Eta Model with the Community Multi-scale Air Quality (CMAQ) modeling system, the NAQFC began providing next-day predictions of ground-level ozone concentrations at a 12 km horizontal grid resolution for the Northeast US in 2004 (Otte et al., 2005; Ryan et al., 2005). In 2005, the CMAQ coverage was expanded to include the states east of the Rocky Mountains (Pleim and Mathur, 2005; Davidson et al., 2008; Eder et al., 2009). The next NAQFC phase, operationally deployed in 2007, expanded coverage to the contiguous United States (CONUS) and replaced the hydrostatic Eta Model with the non-hydrostatic mesoscale model (NMM) within the the Weather Forecasting and Research framework (Eder et al., 2009). A prediction system that includes an aerosol module version 4 (AERO-4) and Carbon Bond version CB05 gas-phase chemical mechanism (Sarwar et al., 2008) was initially tested in 2006 (Gorline and Lee, 2009a) and it has been producing experimental ozone predictions for several years. Since 2007, both operational and experimental prediction systems have been continuously updated (Stajner et al., 2012).

The real-time operational NAQFC predictions, which rely on the Carbon Bond Mechanism version IV (CBMIV) gas-phase chemical mechanism (Gery et al., 1989), are accessible through NOAA's website at http://airquality.weather.gov/. These operational ozone predictions are used by state and local environmental agencies as a basis for air quality forecasts that they issue in terms of the Air Quality Index (AQI) to protect public health from impending poor air quality. Public also obtains operational hour-by-hour predictions directly from this web site. Vulnerable public uses NAQFC predictions to protect their health by adjusting their daily activities or medications.

The experimental NAQFC ozone predictions, accessible at http://airquality.weather.gov/expr/, are produced using the newer CB05 chemical mechanism. Due to higher ozone biases in the experimental predictions than those in the operational predictions through year 2011 (Saylor and Stein, 2012), these experimental predictions have not yet been transitioned to operations. Our study

3

provides a detailed evaluation of the experimental ozone predictions, and a precursor species nitrogen dioxide (NO_2), in order to understand and improve performance of the experimental predictions system, with a view towards its potential transition to operations.

A large amount of information created by continuous predictions is amenable to study of the chemical transport model (CTM) performance. A careful evaluation of the model predictions over CONUS may help researchers better understand, assess, and improve chemical mechanisms, coupling methods between the meteorological model and the CTM, and emission inventories along with the processing algorithms.

The NAQFC ozone predictions up to 2009 have been extensively evaluated. Eder et al. (2006) compared the daily maximum eight-hour ozone predictions for the Northeast US with AIRNow observations (http://www.epa.gov/airnow) from 1 June to 30 September 2004. They found that the NAQFC system over-predicted ozone with a domain-averaged mean bias (MB) of $+10.2$ ppbv and a root mean square error (RMSE) of 15.7 ppbv. The NAQFC predictions in the expanded eastern US domain during the warm season from 2004 to 2007 were evaluated using AIRNow observations (Eder et al., 2009). It was found that the operational NAQFC predictions steadily and gradually improved year after year as demonstrated by decreases in MB and RMSE. The four-month MBs in the eastern US are $+11.4$ ppbv, $+10.9$ ppbv, $+10.5$ ppbv, and $+7.9$ ppbv in 2004, 2005, 2006, and 2007, respectively. Correspondingly, the RMSEs are 16.8 ppbv, 16.3 ppbv, 15.6 ppbv, and 14.5 ppbv. They also showed that the MB and RMSE for the whole CONUS domain in the summer months (June, July, and August) of 2007 are $+4.3$ ppbv and 13.0 ppbv, respectively. The CONUS categorical statistical metrics for the same three-month period in 2007 were presented using both the 84 ppbv and the 75 ppbv daily maximum eight-hour ozone standards. With the 75 ppbv standard, the proportion of correct (POC), critical success index (CSI) or threat score (TS), hit rate (HIT), and false alarm rate (FAR) are 0.924, 0.232, 0.425, and 0.663, respectively. Recently, the NAQFC ozone predictions during the summers of 2007, 2008, and 2009 were compared with the AIRNow measurements by Gorline and Lee (2009b). In their study, the 2007 operational ozone predictions with the CB-MIV chemical mechanism were evaluated, while the 2008 and 2009 predictions were obtained from the experimental predictions using the CB05 chemical mechanism. They found that the

4

MB in August 2009 was about 2 ppbv higher than that in August 2008, and about 5 ppbv higher than the MB in August 2007. The unusually cool summer of 2009 was speculated as a contributing factor to the deteriorating predictions in 2009. Recently, Saylor and Stein (2012) presented the NAQFC predictions in 2009 from both operational and experimental versions. They showed that the use of CB05 in the experimental version systematically increased ground-level ozone over-predictions. The primary causes of the differences between the CBMIV and CB05 systems were identified as two sets of reactions in the CB05 mechanism that are absent from the CBMIV mechanism.

Many operational air quality forecasting systems using 3D CTMs exist worldwide. In Europe, atmospheric composition forecast products have been delivered under the Monitoring Atmospheric Composition and Climate-Interim Implementation project as part of the pre-operational GMES Atmosphere Service (http://www.gmes-atmosphere.eu, see also Menut and Bessagnet, 2010). Similar forecasts are also available in Japan (Maki, 2012), Taiwan(http://taqm.epa.gov.tw/taqm/) and Canada (Talbot et al., 2008). Zhang et al. (2012) summarized some recent real-time air quality forecasting evaluation results. Among all evaluation statistics for hourly ozone, the median positive MB, negative MB, and RMSE, are +4.5 ppbv, -8.1 ppbv, and 16.8 ppbv, respectively. For daily maximum eight-hour ozone categorical statistics, the median POC, CSI, TS, and HIT, are 0.92, 0.18, 0.32, and 0.65, respectively [1].

The main goal of the paper is to continue the NAQFC evaluations as a reference for real-time regional air quality forecasts and future model developments. All the previous NAQFC evaluations have utilized near-real-time AIRNow measurements instead of quality-controlled and -assured Air Quality System (AQS) data, which is the US EPA's repository of ambient air quality data and is available through the agency's Technology Transfer Network (http://www.epa.gov/ttn/airs/airsaqs). Rather than reporting in near-real-time as the AIRNow network requires, the AQS only mandates the monitoring stations to report quarterly. In addition to ozone and particulate matter ($PM_{2.5}$ and PM_{10}) observations available through AIRNow, a suite of other measurements such as nitrogen dioxide (NO_2), carbon monoxide (CO), and sulfur dioxide (SO_2) are also available. As pointed out by Sillman (1999), the model uncertainty can be greatly reduced if observations

---

[1]When a range is presented, the midrange value is used in calculating the median value.

of additional species besides ozone can be utilized in model evaluation and diagnosis. In this study, the AQS NO_2 measurements along with the AQS ozone observations are used for the NAQFC evaluations. NO_2 is not only an important ozone precursor, it is also one of the critical air pollutants regulated through the National Ambient Air Quality Standards in the US, with its annual and hourly limits set as 53 ppbv and 100 ppbv, respectively. In the current evaluation, the NAQFC model predictions are the original predictions without any post-processing.

The remainder of the paper is organized as follows. A brief description of the NAQFC model setup is given in Sect. 2. Section 3 presents the AQS observations, including a comparison between the AQS and AIRNow ozone measurements. Detailed comparisons between the model results and observations are provided in Sect. 4, followed by a summary and discussion in Sect. 5. A list of abbreviations and acronyms can be found in Appendix A.

2 Description of the NAQFC prediction system

The real-time NAQFC air quality prediction system during the year 2010 comprised the CMAQ modeling system (Byun and Schere, 2006) driven by the NCEP's North American Mesoscale (NAM) meteorological predictions with the WRF-NMM core (Janjic, 2003), similar to that described by Eder et al. (2009). A pre-processor to CMAQ, PREMAQ, prepares the CMAQ input files after taking WRF-NMM post-processor outputs (Otte et al., 2005).

Figure 1 shows the computational domain, which is covered by a grid with 442 columns and 265 rows in the longitudinal and latitudinal directions, respectively. The grid has a 12 km horizontal grid resolution and follows the Lambert conformal conic projection. There are 22 hybrid pressure/sigma layers extending from the surface to 100 hPa, which combine those of the WRF-NMM model (see Lee and Ngan, 2011, for details). At lateral boundaries, fixed profiles based on climatological averages are used at in-flow grid cells, and zero-flux-gradient is imposed at the outflow locations. However, Tang et al. (2009) showed that the NAQFC surface ozone predictions can be improved with the use of the MOZART global model predictions to better account long-range transport, especially over the US west coast. A zero-flux assumption at the top boundary is made in the CMAQ computation. Not considering the stratospheric ozone

intrusion may cause ozone underestimations at high-latitudes (Browell et al., 2003). Note that the real-time air quality predictions for the Alaska and Hawaii domains were tested and designated operational in September of 2010, but they are not included in the evaluation presented here. In 2010, real-time air quality predictions for CONUS were continuously provided with both the CBMIV and CB05 chemical mechanisms. In this study, only the experimental version based on CMAQ V4.6 with the CB05 chemical mechanism is evaluated. Each day, there are four different predicting cycles, initialized at 00Z, 06Z, 12Z, and 18Z, which use the newest meteorological fields available. The cycles starting at 06Z and 12Z produce predictions for next 48 h. In this study, only the first 24 h of the NAQFC experimental predictions initialized at 12Z are evaluated. Recently, Savage et al. (2013) demonstrated that there is little difference between day 1 and day 2 ozone forecasts for all metrics using Met Office Unified Model results over the period May 2010 to April 2011.

Gaseous and particulate emissions from anthropogenic and natural sources were divided into four sectors (area, mobile, point, and biogenic) and were processed using data provided by various agencies. Area emissions including off-road engine emissions are based on the US EPA 2005 National Emission Inventory version 1 (NEI05v1) for CONUS, the province-level 2000 Canadian Emissions Inventory for Canada, and the 1999 Mexico National Emission Inventories for Mexico (http://www.epa.gov/ttn/chief/eiinformation.html, also see U.S. EPA, 2011, for details). These inventory data were processed using Sparse Matrix Operator Kennel Emission (SMOKE) version 2.6 to represent monthly, weekly, daily, and holiday/non-holiday variations that are specific for each year (Houyoux et al., 2000). Emissions from wildfires, prescribed burning, and residential wood burning are based on a multi-year average inventory for the years from 1996 to 2002. Ignoring the temporal and spatial variability of the emission sources could cause large ozone and NO_x biases (McKeen et al., 2002; Duncan et al., 2003; Martin et al., 2006). The current operational NOAA Smoke Forecasting System (SFS) establishes the locations and extents of the fires by utilizing fire and smoke data from seven polar and geostationary satellites brought together by the Hazard Mapping System (Rolph et al., 2009; Ruminski et al., 2008). Incorporating the SFS to provide the CMAQ model with near-real-time emissions from large wildfire and agricultural burning is being explored. The EPA Office of Transportation and

Air Quality 2005 on-road emissions inventory was used to generate mobile emissions over the US Both the Electric Generating Unit (EGU) and the non-EGU point sources were based on the NEI05v1 data. Oxides of nitrogen (NO_x) and SO_2 emissions from the US EGU sources rely on 2008 Continuous Emission Monitoring data. Annual Energy Outlook (AEO) from the Department of Energy released in April 2010 (U.S. EIA, 2010) was used to project the EGU emissions to 2010 and was implemented on 6 July 2010. Before that date, a similar projection was made based on 2009 AEO data. Biogenic emissions were calculated dynamically using the Biogenic Emissions inventory System version 3.13 (Schwede et al., 2005), which considers variability in temperature and solar radiation to estimate NO_x and volatile organic compound (VOC) emissions from forests and grasslands.

3 Observations

3.1 AIRNow and AQS observations

Real-time ozone and $PM_{2.5}$ measurement data across the US, Canada, and parts of Mexico are provided by the US EPA through the AIRNow Gateway (http://www.airnowgateway.org). Because of their easy accessibility, AIRNow observations are widely used. Although the AIRNow data are only preliminary and not fully verified, they serve the purpose for real-time AQI reporting and forecasting. Observational data that have been subjected to additional quality control are available from the EPA's AQS, which is designed to meet the needs of regulatory, academic, and public health research communities. Without the requirement to disseminate data in real-time, the AQS system includes monitors from many other surface networks and its measured species extend from ozone and particulate matter ($PM_{2.5}$ and PM_{10}) to multiple atmospheric chemistry components, such as NO_2, CO, SO_2, and many VOC species. The AQS measurement data were downloaded from http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm.

Figure 2 displays the daily count of hourly observations in 2010 for both the AQS (version 5/16/12) and the AIRNow systems. For both systems, there are almost twice as many ozone measurements available in warm seasons as in cold seasons since some monitors do not operate

during the winter. The number of ozone measurements in both the AQS and the AIRNow data sets typically exceeds 10 000 per day. It should be noted that some AIRNow measurements are not available from the AQS system. This could be caused by delays in reporting to the AQS system or elimination of poor-quality data during the validation period. The data are considered to "overlap" if the measurements are reported from the same monitor at the same time, even if measurement values differ. The daily counts of "overlapped" measurement pairs are also plotted in Fig. 2. A snapshot of differences between "overlapped" data is displayed in Fig. 3a, which shows the paired data between AQS and AIRNow at the same sites and hours on 31 May 2010. While most data agree, some differences are seen, probably due to the quality control work carried out after AIRNow reporting.

3.2 Consistency check of ozone observations

Upon the examination of consistency between the AQS and AIRNow data sets, potential problems with reporting of the measurement time are suspected at several isolated sites. Additional quality control is applied to remove these questionable sites. In this process, hourly AQS and AIRNow ozone observations from each monitor are separated into daily files which run from 00:00 EDT to 23:00 EDT, or 00:00 EST to 23:00 EST following the US daylight saving time schedule. Consecutive hourly measurements at one location over one day form a 24-dimensional vector. At each location and for each day, the L2-norm $\Gamma(\Delta t)$ is calculated for the difference vector between AQS and AIRNow ozone observations at the matching hours ($\Delta t = 0$), as well as for the lower-dimensional difference vectors obtained by shifting AQS vector forward or backward by 1 or 2 h ($\Delta t = \pm 1, \pm 2$), as given in the following Eq. (1).

$$\Gamma(\Delta t) = \|O_3^{AQS}(t + \Delta t) - O_3^{AIRNow}(t)\|, \ \Delta t = 0, \pm 1, \pm 2 \text{h} \tag{1}$$

In addition to the shifting, missing data in either AQS and AIRNow ozone observations results in reduced dimensions of the difference vector. To account for variations in the dimension (N) of difference vectors, Ω is calculated in Eq. (2).

$$\Omega(\Delta t) = \frac{24}{N(\Delta t)} \Gamma(\Delta t)^2, \Delta t = 0, \pm 1, \pm 2 \text{h}, \quad \text{for} N(\Delta t) \geq 12 \tag{2}$$

Note that $\Omega(\Delta t)$ is calculated only when there are no less than 12 pairs of observations to form the difference vector. A monitor is flagged if $\Omega(\Delta t) < \frac{\Omega(0)}{10}$, for any $\Delta t = \pm 1, \pm 2\,\text{h}$. This condition indicates a closer match between AQS and AIRNow data sets after the measurement time is adjusted by Δt of $-2, -1, 1$, or $2\,\text{h}$ for this monitor on the particular day, implying

5 a possible inconsistency between measurement times reported in the two data sets. A total of 74 sites were flagged after checking the whole year. Observations from those flagged sites over the entire year were then removed. Figure 3b shows the comparison between AQS and AIRNow after removing the questionable sites. The agreement between AQS and AIRNow observations improves after eliminating the measurements from the flagged sites, with the coefficient of

10 determination R^2 increasing from 0.995 to 0.997. Ozone measurements from the monitor sites that are unique to AQS cannot be examined in this fashion and they are not included in the following evaluation either. Figure 2 shows the data counts after these two exclusion criteria are applied. Overall, more than 80 % of the AQS ozone data are retained for the evaluation. The observations are from 1124 AQS ozone monitors.

15 ## 3.3 AQS NO_2 observations

Without AIRNow NO_2 data to check against the There are no AIRNow NO_2 data available to perform the similar consistency examinination between AIRNow and AQS as what is done for ozone in section 3.2. All AQS NO_2 measurements are used in the following evaluation. It should be noted that most of the AQS NO_2 measurements were from chemiluminescence

20 monitors equipped with molybdenum converters, which systematically overestimate NO_2 concentrations (Dunlea et al., 2007; Steinbacher et al., 2007). Using Mexico City Metropolitan Area (MCMA) field campaign data during April of 2003, Dunlea et al. (2007) reported that the chemiluminescence monitor interference resulted in an aveage concentration up to 22 % greater than that from co-located spectroscopic measurements. In this study, the AQS NO_2 measure-

25 ments were used without any correction to account for this issue. NO_2 hourly measurements are also shown in Fig. 2. Unlike ozone monitoring that has a seasonal variation, the daily NO_2 measurement count from 408 sites is almost constant throughout the year.

4 NAQFC evaluation results

When comparing model predictions with AQS observations, model-predicted concentration counterparts are taken from the monitor-residing grid cells. With such direct matching, there is no interpolation applied and it is consistent with previous NAQFC evaluation studies (Eder et al.,
5 2006, 2009; Gorline and Lee, 2009b). However, there is a slight difference from what is described in Eder et al. (2009), where multiple observations inside a single grid cell are averaged as the representative measurement for the grid cell. In this paper, each measurement is compared against model prediction independently when there are two or more monitors located within one grid cell. In the following evaluations, the urbanization characteristics of each mon-
10 itor site are utilized to filter observations into urban, suburban, and rural categories. Among 1124 ozone sites, there are 200 urban, 455 suburban, 462 rural, and 7 unknown stations. The number of NO_2 sites at urban, suburban, rural, and unknown settings are 130, 148, 126, and 4, respectively.

In addition, separate evaluations in the six predefined regions shown in Fig. 1 are performed
15 to investigate regional variability in model performance.

4.1 Annual performance

Figure 4 shows the daily and domain-wide average ozone and NO_2 concentrations from AQS and CMAQ. Similar to the 2009 NAQFC prediction results (Saylor and Stein, 2012), the model significantly overestimates ozone during the summer. Until the end of May, there is very good
20 agreement between model predictions and AQS observations for ozone. As the NAQFC NO_2 predictions are compared with the AQS observations for the first time, it shows that the model overestimates NO_2 for all four seasons. The NO_2 overestimation is more severe in the summer than during the other seasons. The normalized monthly mean NO_2 biases are 74.6 %, 79.8 %, and 76.1 % for June, July, and August, respectively. January has the lowest normalized monthly
25 mean NO_2 bias of 34.6 %.

Figure 5 shows the annual performance in different local settings for both ozone and NO_2. The urban and suburban sites mostly resemble what is shown in Fig. 4. In rural areas, NO_2

11

concentrations are more than 50 % lower than those at urban sites, displayed by both the model and the observations. However, NAQFC still significantly overestimates NO_2 at the rural sites. For ozone, the model overestimation in rural areas during the summer is more pronounced than that in urban and suburban areas. As rural areas are mostly in NO_x-sensitive chemical regimes (Choi et al., 2012), the overestimated NO_2 in the area, especially in the forest-dominant Southeast region, can produce ozone much more efficiently than in the urban and suburban areas. Figure 5 also shows that the average ozone concentrations are slightly larger at rural sites than those at urban sites. The lower ozone concentrations in urban areas may be due to NO_x titration at night-time. This also indicates that due to its long lifetime ozone pollution has non-local impacts.

The time series of daily and regionally averaged ozone and NO_2 are shown in Figs. 6 and 7. The ozone overestimation in summer is seen in all the regions, but it is the most pronounced in the Southeast region. NO_2 is also overestimated in all the regions during the summer, ranging from the highest biases in the Pacific Coast and Lower Middle regions to minimal overestimation in the Rocky Mountain and Northeast regions.

The detailed monthly and annual average ozone biases and RMSEs in different regions are listed in Tables 1 and 2. Similar results for NO_2 are listed in Tables 3 and 4. Ozone biases in Lower Middle and Pacific Coast are the lowest, with the annual average being $+3.7$ ppbv and $+4.0$ ppbv, respectively. The most pronounced negative biases are seen in February in the Upper Middle and Northeast regions, with monthly average biases of -8.5 ppbv and -5.6 ppbv, respectively. The largest positive monthly average bias of $+17.6$ ppbv is seen in the Southeast Region in August. Table 2 also shows that the Southeast region has the largest annual RMSE of 17.6 ppbv. The highest monthly RMSE of 22.5 ppbv is seen in the Southeast in August. In agreement with Fig. 7, Table 3 also points to Lower Middle and Pacific Coast as the worst regions for NO_2 predictions, with their annual average biases $+8.1$ ppbv and $+7.1$ ppbv, respectively. When normalized by the observation mean, the relative biases show more than 100 % overestimation in Lower Middle from April to August, and in Pacific Coast in June and July. In July, the normalized monthly mean NO_2 bias reaches its peak in July (167.2 %). The Rocky Mountain region has the smallest annual NO_2 model bias of 0.4 ppbv (4.2 %) among all

regions and its monthly average biases range from -0.9 ppbv (-7.1 %) in January to 1.7 ppbv (33.0 %) in July. All other regions show consistent positive biases throughout the year. The CONUS RMSEs for NAQFC NO_2 predictions, listed in Table 4, range from 12.1 ppbv in May to 15.4 ppbv in September. In September, Pacific Coast and Lower Middle have the highest monthly NO_2 RMSEs of 19.6 ppbv and 19.1 ppbv, respectively.

4.2 Spatial patterns

The spatial distributions of the monthly average ozone and NO_2 AQS concentrations, model biases, and RMSEs at monitoring sites in August are shown in Fig. 8. They are similar to the other summer months such as July (not shown here). Higher monthly average ozone measurements are mostly located in the California, Rocky Mountain, and Mid-Atlantic (areas bordering Northeast and Southeast regions) areas. Multiple sites in Los Angeles and an isolated one in Denver, Colorado show very high NO_2 observations. The spatial distribution of ozone biases in Fig. 8 shows a broad spread of high positive ozone biases in the Southeast region. This is consistent with Fig. 6, which identifies Southeast as the region with the most severe ozone overestimation in summer. As this region is mostly covered with forest, the abundance of biogenic VOCs during the growing season helps to translate NO_2 overestimations into high ozone biases under the NO_x-sensitive regime.

Negative ozone biases are found around Los Angeles and New Orleans, where high positive NO_2 biases are shown in Fig. 8. It is possible that the emissions inventories do not fully account for the actual emissions reduction due to the long-lasting economic aftermath of hurricane Katrina on New Orleans, thus resulting in the overestimation in that area. The combination of high positive NO_2 biases with negative ozone biases suggests Los Angeles and New Orleans are probably under a VOC-sensitive regime, in which the increased NO_2 may lead to ozone reductions. Such model behavior in NO_x-rich urban regions is common. For instance, Tong et al. (2006) showed that increasing NO_x emissions actually reduced ozone in central Atlanta in their sensitivity studies to assess ozone impacts from NO_x emissions. Figure 8 shows that most of the higher ozone RMSEs are seen in the Southeast region and around Los Angeles. The Los Angeles and New Orleans areas also have the highest NO_2 RMSEs, as shown in Fig. 8.

13

4.3 Daily maximum eight-hour average ozone and its categorical statistics

Eight-hour running averages are calculated for both the model and the AQS hourly concentrations. A minimum of six hourly observations in any eight-hour time window is required for the calculation. Otherwise, the eight-hour ozone observation is flagged as missing. As the primary ozone standard in the US, the daily maximum eight-hour average concentration is currently set as 75 ppbv revised from its previous 0.08 ppm (effectively 84 ppbv due to rounding) in March 2008 (Environmental Protection Agency, 2008). Using the standard as a threshold for daily maximum eight-hour average ozone, there are four possible scenarios:

(a) prediction is above, but observation is below the threshold (false alarm);

(b) prediction and observation are above the threshold;

(c) prediction and observation are below the threshold;

(d) prediction is below, but observation is above the threshold.

In Fig. 9 a scatter plot of one day's observations in CONUS and collocated NAQFC predictions is presented and four quadrants are marked according to scenarios a–d that they correspond to. Hit rate (HIT), critical success index (CSI) or threat score (TS), false alarm rate (FAR), equitable threat score (ETS), and proportion of correct (POC) which is referred as Accuracy in Eder et al. (2006), are calculated for the NAQFC predictions for the entire year. The definitions are shown in Eqs. (3–7), where N_a, N_b, N_c, and N_d represent the number of incidences in each scenario a, b, c and d, respectively, as shown in Fig. 9.

$$\text{HIT} = \frac{N_b}{N_b + N_d} \tag{3}$$

$$\text{CSI} = \frac{N_b}{N_a + N_b + N_d} \tag{4}$$

$$\text{FAR} = \frac{N_a}{N_a + N_b} \tag{5}$$

14

$$\mathrm{ETS} = \frac{N_b - N_r}{N_a + N_b + N_d - N_r}, \quad \text{where} \quad N_r = \frac{(N_a + N_b) \times (N_b + N_d)}{N_a + N_b + N_c + N_d} \tag{6}$$

$$\mathrm{POC} = \frac{N_b + N_c}{N_a + N_b + N_c + N_d} \tag{7}$$

The HIT, CSI (or TS), FAR, and POC for the NAQFC predictions in previous years have been reported (Eder et al., 2006, 2009). HIT measures the fraction of observed above the threshold
5 events which are predicted correctly. It is also referred as probability of detection. FAR is the fraction of predicted above the threshold events that are wrong. CSI measures the fraction of correctly predicted above the threshold events after removing correctly predicted below the threshold incidences. ETS measures the prediction skill more critically by negating the correct predictions by chance. While "ETS = 1" means a perfect prediction, positive ETS values in-
10 dicate skillful predictions relative to a random forecast (Schaefer, 1990). ETS ≤ 0 denotes no skill for the forecast. POC is the fraction of predictions that match the above/below threshold with the observations.

Using the AQS observations and NAQFC predictions for the entire year and summer months (June–August), the categorical statistics for the daily maximum of eight-hour ozone exceeding
15 two thresholds are listed in Tables 5 and 6. Overall, the HIT values calculated for summer are better than those calculated for the entire year, but CSI, FAR, ETS, and POC values for summer are worse. The Rocky Mountain region is an exception in that the CSI and FAR values for summer are slightly better than those calculated for the entire year. Using the current 75 ppbv standard as the threshold, out of the total 4065 ($N_b + N_d = 2616 + 1449$) observed cases ex-
20 ceeding this threshold in AQS measurements, 2541 ($N_b + N_d = 1812 + 729$) cases happened during the summer months. HIT, CSI, FAR, ETS, and POC over CONUS for the entire year are 0.64, 0.17, 0.81, 0.16, and 0.96, respectively; while the same statistics calculated over CONUS for the summer are 0.71, 0.17, 0.82, 0.15, and 0.91. The summer HIT value is much better than HIT = 0.43 reported by Eder et al. (2009) for the 2007 summer months with the same standard.
25 However, the CSI, FAR, and POC values during the summer are worse, with the current 0.17, 0.82, and 0.91 compared with 0.23, 0.66, and 0.92. The ETS values of 0.15 and 0.16 indicate some skill in the NAQFC predictions. In all regions, the ETS scores are positive, showing that

15

the predictions are better than predictions by chance. The highest ETS scores are 0.24 and 0.23 for the Pacific Coast and Northeast regions. In the Rocky Mountain region, ETS = 0.06 reflects little skill of the model, mostly caused by the high FAR values (0.93 for summer and 0.92 for the entire year). The annual POC values are greater or equal to 0.95 in all regions, but the
5 summer values drops to as low as 0.87 in the Pacific Coast Region.

The categorical statistics are sensitive to the threshold used to define the exceedance events, as shown by Eder et al. (2009) using both the 85 ppbv and the 75 ppbv standards. Similar metrics calculated using a 70 ppbv threshold for daily maximum eight-hour ozone are also listed for CONUS in Tables 5 and 6. With the new threshold, the exceedances increase to
10 $7577(N_b + N_d = 5753 + 2824)$ from 4065 with the 75 ppbv standard for the year. The annual POC value drops from 0.96 to 0.93 and all other metrics improve for CONUS. It should be noted that the large model biases greatly affect the categorical statistics. By implementing a bias-adjustment technique, Kang et al. (2010) showed significant improvement in the categorical evaluation metrics, with increased HIT and decreased FAR at almost all locations during their
15 study period in 2008.

4.4 Weekly patterns of NAQFC performance

CTM predictions are highly sensitive to the model-ready emissions inputs, which are generated using a large number of month-of-year, day-of-week, and hour-of-day temporal profiles. Section 4.1 already showed that the NAQFC performance for ozone and NO_2 predictions varies
20 significantly by month. These monthly variations in model performance are influenced by differences in the meteorological conditions, specifically the temperature change from month to month. It is difficult to separate the emissions-induced effects caused by the month-of-year profile from the meteorological impacts. However, it has been well documented that the ozone concentrations in urban areas peak at weekends, while nitrogen oxides and VOC emissions are
25 generally lower at weekends than those on weekdays (Marr and Harley, 2002; Murphy et al., 2007; Pierce et al., 2010). Instead of focusing on the "weekend ozone effect", here we study the weekly patterns of NAQFC performance in order to investigate possible systematic errors in weekly profiles that are used in emissions processing.

16

In this section, the NAQFC predictions during the warm months, i.e. from June to September are grouped into days of the week. Strong weekly patterns are shown in the ozone biases for different days of the week listed in Table 7. Over CONUS and most regions, O_3 biases are higher on weekends than on weekdays. The RMSEs calculated for the different days of
5 the week do not show a clear weekly pattern. This indicates that the variability in prediction errors is influenced by interactions among the emissions, chemistry and meteorology, rather than stemming from the emissions alone.

Similarly, the day-of-week biases for NAQFC NO_2 predictions are listed in Table 8. Contrary to ozone, the NO_2 biases over CONUS are lower on weekends than on weekdays. The lowest
10 biases in NO_2 predictions occur on Saturdays in all regions except Northeast. The weekday-weekend contrast is especially evident in Pacific Coast, where the average model biases are no less than 9.1 ppbv on weekdays and no greater than 7.5 ppbv at weekends.

4.5 Diurnal cycles

Ozone and its precursors have distinctive diurnal cycles. Examination of corresponding cycles
15 in a CTM may help identify and correct shortcomings in the model and thus improve model predictions. van Loon et al. (2007) showed large diurnal cycle variations among seven different regional air quality models. The diurnal patterns of the NAQFC prediction biases are also studied here. Unlike the weekly patterns that mainly exhibit the emissions signals, the diurnal patterns of model performance are greatly affected by many diurnal characteristics coming from
20 the meteorological inputs. Diurnal profiles are obtained by averaging model-observation pairs by their local time (LT). Note that LT here is based on the official time zone of each AQS site and daylight saving regime is not considered. In order to remove the impact of monthly variations in meteorological conditions, the diurnal patterns are studied separately for each month.

The diurnal profiles of ozone and NO_2 for August, stratified by the degree of urbanization are
25 shown in Fig. 10. Ozone is overestimated for all hours, except at 19:00 LT for suburban sites and 18:00–20:00 for urban sites. The domain-averaged ozone predictions at rural sites have positive biases throughout the day. Ozone model biases peak in the early morning, from 07:00 to 10:00 LT in all three urbanization settings. NO_2 biases are positive for all hours at urban and

suburban sites, dipping to lowest levels between 08:00 and 13:00 LT. For the same time period, there are slight underestimations at rural sites. The NO_2 overestimation is most pronounced at night, from 18:00 to 06:00 LT, by around 100 % for all urbanization settings. The standard deviations of model predictions exceed those of the observations at almost all hours for NO_2. Meanwhile, the ozone variations in the model and observations are comparable.

Figures 11 and 12 show the regional diurnal profiles in August for ozone and NO_2, respectively. Ozone biases in the Southeast region are positive for all 24 h. The other regions display large positive ozone biases from morning until noon and minimal positive to slight negative biases between 18:00 and 20:00 LT, similar to the urban and suburban ozone diurnal profiles in Fig. 10. Note the close agreement between predicted and observed ozone with respect to the average values and the variability during nighttime in the Pacific Coast and Rocky Mountain regions. The regional diurnal profiles of NO_2 in Pacific Coast, Lower Middle, Southeast, and Upper Middle exhibit good agreement between the model and the observations from early morning until early afternoon, but show large biases at night-time, resembling the urban and suburban NO_2 diurnal profiles in Fig. 10. In Northeast, the diurnal profile is similar, but NO_2 biases at night are much smaller. Good agreement between average NAQFC NO_2 and AQS observations for most hours of the day is found in the Rocky Mountain region. However, NO_2 is still overestimated at 19:00 and 20:00 LT by more than 100 % in this region.

5 Summary and discussion

In this paper, the NAQFC experimental ozone predictions and real-time testing of prediction of precursor species NO_2 in 2010 are evaluated against quality-assured AQS observations of ozone and NO_2.

It is found that the CONUS- and daily-averaged predictions for both ozone and NO_2 are overestimated throughout the year, with peak overestimation in the summer. This seasonal pattern persists when sites are stratified by the degree of urbanization into urban, suburban, and rural sites. In August, over-prediction is more pronounced for rural than for urban and

18

suburban sites. The highest regional ozone biases were found in Southeast during the summer. NO_2 over-prediction is pronounced in the Pacific Coast and Lower Middle regions. The spatial distributions during the summer show the largest positive NO_2 biases in Los Angeles and New Orleans, where ozone levels were underestimated. This suggests that VOC-sensitive regimes prevailed during those months in 2010 for these two areas.

The ozone categorical statistics using the current US ambient air quality standard (75 ppbv) for daily maximum eight-hour average ozone show mixed results when comparing the 2010 experimental ozone predictions generated using CB05 mechanism with the operational ozone predictions for earlier years that rely on CBMIV mechanism. For a lower threshold of 70 ppbv, HIT, CSI, FAR, and ETS evaluated over CONUS for 2010 experimental predictions improve, but POC deteriorates in comparison to the same statistics evaluated for the 75 ppbv threshold.

The ozone and NO_2 biases show distinct weekly patterns in summer. While ozone biases are larger during the weekends than they are on weekdays, NO_2 biases show the opposite patterns in most regions. Diurnal patterns show that ozone overestimation is most severe in the morning, from 07:00 to 10:00 local time, lower overnight, and the lowest in the evening hours, around 19:00 local time. For NO_2, the morning predictions are in close agreement with the AQS observations, but night-time concentrations are over-predicted by around 100 %.

Comparisons on regional or domain-wide scales together with monthly or annual evaluations aim to eliminate influence of dynamical meteorological and chemical conditions, which vary significantly from site to site and from day to day. The averaging avoids large uncertainties associated with each individual site and time, thus exposes systematic model errors, which could be reduced in the future to improve NAQFC predictions. For example, NO_2 overestimation throughout the year in almost all regions may have contributed to the overall ozone estimation for the entire year. This is especially true during the growing season in the Southeast region where forests are predominant. Under the NO_x-sensitive chemical regime with abundant biogenic VOCs, the NO_2 overestimations likely caused the severe positive ozone biases from May to September. Higher NO_2 biases were found in the summer, and they are believed to contribute to the larger ozone overestimations seen in the summertime in all regions. The clear weekly signals shown by both ozone and NO_2 model biases suggest that weekly profiles resulting from

19

emissions processing may need adjustments. It should be noted that other factors, such as chemical mechanism, not considering long range transport at lateral boundaries or ozone intrusion from stratosphere at domain top, all contribute to the current model errors.

However, drawing conclusions on the exact causes for the current model problems requires further studies. There are several limitations in our current evaluation study. For instance, the AQS stations are quite sparse in some regions, especially for NO_2 monitoring. Uncertainties in emission rates, photochemical reaction rates, and meteorological inputs such as surface temperature, wind speed and direction, and cloud cover all contribute to uncertainties in NAQFC ozone and NO_2 predictions. Further analyses would benefit from meteorological measurements, observations of VOC species, and vertical profiles of most parameters in order to fully explain the evaluation results.

The type of analysis presented here has guided recent updates to the NAQFC system that produces experimental ozone predictions. Concurrently with the updates to the NCEP NAM model and the land use and land cover data for emissions in October 2011, three additional updates were made with the goal of reducing ozone biases discussed here. Previous constant lateral boundary condition profiles for most chemical species were replaced with monthly mean profiles from GEOS-CHEM global model simulations for 2006 that follow the methodology of Bey et al. (2001). Dry deposition was modified based on the Monin-Obukhov similarity theory (Wu et al., 2003) as well as by including canopy height and density based on recent Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System satellite observations (Lefsky, 2010). Planetary boundary layer (PBL) height was constrained to be at least 50 m. This mitigated the previous high ozone bias problem due to low PBLs at areas close to large water bodies. It also allows dilution of the mobile emissions near urban centers and lessened the severity of ozone titration at nighttime. Testing during the summer of 2011 has shown shown positive impacts of these changes and they were all incorporated into the experimental ozone predictions for 2012. The emission data sets have been updated in June 2012, with about 35% decrease in total mobile NO_x emissions. Preliminary evaluation of the latest experimental predictions shows improvements from this combination of updates. Examples of additional modifications that may prove beneficial for ozone predictions include:

20

assimilation of observed chemical composition data, increase of the model resolution, inclusion of newer versions of chemical and meteorological models, as well as a closer coupling among system components.

Appendix A   List of abbreviations and acronyms.

AEO	Annual Energy Outlook
AERO-4	Aerosol module version 4
AQI	Air Quality Index
AQS	Air Quality System
CB05	Carbon Bond Mechanism with 2005 updates
CBMIV	Carbon Bond Mechanism version IV
CMAQ	Community Multi-scale Air Quality modeling system
CONUS	Contiguous United States, Alaska and Hawaii not included
CSI	Critical Success Index
CTM	Chemical Transport Model
EDT	US Eastern Daylight saving Time
EGU	Electric Generating Unit
EPA	US Environmental Protection Agency
EST	US Eastern Standard Time
ETS	Equitable Threat Score
FAR	False Alarm Rate
HIT	Hit rate
LM	Lower Middle
LT	Local Time
MB	Mean bias
NAM	North American Mesoscale
NAQFC	National Air Quality Forecast Capability
NCEP	US National Centers for Environmental Prediction
NE	Northeast
NEI05v1	US EPA 2005 National Emission Inventory Version 1
NMM	Non-hydrostatic Mesoscale Model
NOAA	US National Oceanic and Atmospheric Administration
NOx	Oxides of nitrogen
PC	Pacific Coast
PBL	Planetary boundary layer
PM10	Particles that are 10 micrometers in diameter or smaller
PM2.5	Particles that are 2.5 micrometers in diameter or smaller
POC	Proportion Of Correct
PREMAQ	CMAQ pre-processor
RM	Rocky Mountain
RMSE	Root Mean Square Error
SE	Southeast
SFS	Smoke Forecasting System
SMOKE	Sparse Matrix Operator Kennel Emission
TS	Threat Score
UM	Upper Middle
VOC	Volatile Organic Compound
WRF	Weather Forecasting and Research model

References

Bey, I., Jacob, D., Yantosca, R., Logan, J., Field, B., Fiore, A., Li, Q., Liu, H., Mickley, L., and Schultz, M.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. of Geophys. Res., 106, 23 073–23 095, 2001.

Browell, E. V., Hair, J. W., Butler, C. F., Grant, W. B., DeYoung, R. J., Fenn, M. A., Brackett, V. G., Clayton, M. B., Brasseur, L. A., Harper, D. B., Ridley, B. A., Klonecki, A. A., Hess, P. G., Emmons, L. K., Tie, X., Atlas, E. L., Cantrell, C. A., Wimmers, A. J., Blake, D. R., Coffey, M. T., Hannigan, J. W., Dibb, J. E., Talbot, R. W., Flocke, F., Weinheimer, A. J., Fried, A., Wert, B., Snow, J. A., and Lefer, B. L.: Ozone, aerosol, potential vorticity, and trace gas trends observed at high-latitudes over North America from February to May 2000, J. of Geophys. Res., 108, doi:10.1029/2001JD001390, 2003.

Byun, D. and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, Appl. Mech. Rev., 59, 51–77, 2006.

Choi, Y., Kim, H., Tong, D., and Lee, P.: Summertime weekly cycles of observed and modeled NOx and O_3 concentrations as a function of satellite-derived ozone production sensitivity and land use types over the continental United States, Atmos. Chem. Phys., 12, 6291–6307, 2012.

Davidson, P., Schere, K., Draxler, R., Kondragunta, S., Wayland, R. A., Meagher, J. F., and Mathur, R.: Toward a US National Air Quality Forecast Capability: Current and Planned Capabilities, in: Air Pollution Modeling and Its Application XIX, edited by Borrego, C. and Miranda, A., pp. 226–234, Springer, Dordrecht, The Netherlands, 2008.

Duncan, B., Bey, I., Chin, M., Mickley, L., Fairlie, T., Martin, R., and Matsueda, H.: Indonesian wildfires

of 1997: Impact on tropospheric chemistry, J. of Geophys. Res., 108, doi:10.1029/2002JD003195, 2003.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., San Martini, F., Sheehy, P. M., Zahniser, M. S., Shorter, J. H., Wormhoudt, J. C., Lamb, B. K., Allwine, E. J., Gaffney, J. S., Marley, N. A., Grutter, M., Marquez, C., Blanco, S., Cardenas, B., Retama, A., Ramos Villegas, C. R., Kolb, C. E., Molina, L. T., and Molina, M. J.: Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment, Atmos. Chem. Phys., 7, 2691–2704, doi:10.5194/acp-7-2691-2007, 2007.

Eder, B., Kang, D., Mathur, R., Yu, S., and Schere, K.: An operational evaluation of the Eta-CMAQ air quality forecast model, Atmos. Environ., 40, 4894–4905, 2006.

Eder, B., Kang, D., Mathur, R., Pleim, J., Yu, S., Otte, T., and Pouliot, G.: A performance evaluation of the National Air Quality Forecast Capability for the summer of 2007, Atmos. Environ., 43, 2312–2320, 2009.

Environmental Protection Agency: National Ambient Air Quality Standards for Ozone, Federal Register, 73, 16 436–16 514, 2008.

Gery, M. W., Whitten, G. Z., Killus, J. P., and Dodge, M. C.: A photochemical kinetics mechanism for urban and regional scale computer modeling, J. of Geophys. Res., 94, 12 925–12 956, 1989.

Gorline, J. L. and Lee, P.: Performance evaluation of NOAA-EPA developmental aerosol forecasts, Environ. Fluid Mech., 9, 109–120, 2009a.

Gorline, J. L. and Lee, P.: Performance of NOAA-EPA air quality predictions, 2007 - 2009, in: the 8th Annual CMAS Conference, Chapel Hill, NC, October 19-21, 2009, 2009b.

Houyoux, M. R., Vukovich, J. M., Coats, C. J., Wheeler, N. W., and Kasibhatla, P. S.: Emission inventory development and processing for the seasonal model for regional air quality SMRAQ, J. of Geophys. Res., 105, 9079–9090, 2000.

Janjic, Z. I.: A nonhydrostatic model based on a new approach, Meteorology and Atmospheric Physics, 82, 271–285, doi:10.1007/s00703-001-0587-6, 2003.

Kang, D., Mathur, R., and Rao, S. T.: Real-time bias-adjusted O_3 and PM2.5 air quality index forecasts and their performance evaluations over the continental United States, 44, 2203–2212, doi:10.1016/j.atmosenv.2010.03.017, 2010.

Lee, P. and Ngan, F.: Coupling of important physical processes in the planetary boundary layer between meteorological and chemistry models for regional to continental scale air quality forecasting: an overview, Atmosphere, 2, 464–483, 2011.

Lefsky, M. A.: A global forest canopy height map from the Moderate Resolution Imaging Spectro-

24

radiometer and the Geoscience Laser Altimeter System, Geophysical Research Letters, 37, doi: 10.1029/2010GL043622, 2010.

Maki, T.: The outline of air quality operational and research activities in the Japan Meteorological Agency, in: 4th International Workshop on Air Quality Forecasting Research, Geneva, Switzerland, 12-14 December, 2012, 2012.

Marr, L. and Harley, R.: Modeling the effect of weekday-weekend differences in motor vehicle emissions on photochemical air pollution in central California, Environmental Science and Technology, 36, 4099–4106, 2002.

Martin, M. V., Honrath, R. E., Owen, R. C., Pfister, G., Fialho, P., and Barata, F.: Significant enhancements of nitrogen oxides, black carbon, and ozone in the North Atlantic lower free troposphere resulting from North American boreal wildfires, J. of Geophys. Res., 111, doi:10.1029/2006JD007530, 2006.

McKeen, S., Wotawa, G., Parrish, D., Holloway, J., Buhr, M., Hubler, G., Fehsenfeld, F., and Meagher, J.: Ozone production from Canadian wildfires during June and July of 1995, J. of Geophys. Res., 107, doi:10.1029/2001JD000697, 2002.

Menut, L. and Bessagnet, B.: Atmospheric composition forecasting in Europe, Annales Geophysicae, 28, 61–74, 2010.

Murphy, J. G., Day, D. A., Cleary, P. A., Wooldridge, P. J., Millet, D. B., Goldstein, A. H., and Cohen, R. C.: The weekend effect within and downwind of Sacramento - Part 1: Observations of ozone, nitrogen oxides, and VOC reactivity, Atmos. Chem. Phys., 7, 5327–5339, 2007.

Otte, T. L., Pouliot, G., Pleim, J. E., Young, J. O., Schere, K. L., Wong, D. C., Lee, P. C. S., Tsidulko, M., McQueen, J. T., Davidson, P., Mathur, R., Chuang, H.-Y., DiMego, G., and Seaman, N. L.: Linking the Eta model with the Community Multiscale Air Quality (CMAQ) modeling system to build a national air quality forecasting system, Weather and Forecasting, 20, 367–384, doi: 10.1175/WAF855.1, 2005.

Pierce, T., Hogrefe, C., Rao, S. T., Porter, P. S., and Ku, J.: Dynamic evaluation of a regional air quality model: Assessing the emissions-induced weekly ozone cycle, Atmos. Environ., 44, 3583 – 3596, doi:10.1016/j.atmosenv.2010.05.046, 2010.

Pleim, J. and Mathur, R.: Diagnostic evaluation, sensitivity analysis, and new developments in the Eta/CMAQ air quality forecast system, in: the 7th Conference on Atmospheric Chemistry, American Meteorology Society, San Diego, CA, Paper 4.7, 2005.

Rolph, G. D., Draxler, R. R., Stein, A. F., Taylor, A., Ruminski, M. G., Kondragunta, S., Zeng, J., Huang, H.-C., Manikin, G., Mcqueen, J. T., and Davidson, P. M.: Description and verification of the NOAA Smoke Forecasting System: the 2007 fire season, Weather and Forecasting, 24, 361–378,

doi:10.1175/2008WAF2222165.1, 2009.

Ruminski, M., Simko, J., Kibler, J., Kondragunta, S., Draxler, R., Davidson, P., and Li, P.: Use of multiple satellite sensors in NOAA's operational near real-time fire and smoke detection and characterization program, in: Remote sensing of fire: science and application, edited by Hao, WM, vol. 7089 of *Proceedings of the society of photo-optical instrumentation engineers (SPIE)*, SPIE, doi: 10.1117/12.807507, Conference on Remote Sensing of Fire - Science and Application, San Diego, CA, AUG 10, 2008, 2008.

Ryan, W., Davidson, P., and Stokols, P.: Evaluation of the National Air Quality Forecast System (NAQFS) developmental large domain model: summary of the air quality forecasters focus group workshop, in: the 7th Conference on Atmospheric Chemistry, American Meteorology Society, San Diego, CA, Paper 4.12, 2005.

Sarwar, G., Luecken, D., Yarwood, G., Whitten, G., and Carter, W. P.: Impact of an updated carbon bond mechanism on predictions from the CMAQ modeling system: preliminary assessment, J. Appl. Meteor. Climatol., 47, 3–14, doi:10.1175/2007JAMC1393.1, 2008.

Savage, N. H., Agnew, P., Davis, L. S., Ordóñez, C., Thorpe, R., Johnson, C. E., O'Connor, F. M., and Dalvi, M.: Air quality modelling using the Met Office Unified Model (AQUM OS24-26): model description and initial evaluation, Geosci. Model Dev., 6, 353–372, doi:10.5194/gmd-6-353-2013, 2013.

Saylor, R. D. and Stein, A. F.: Identifying the causes of differences in ozone production from the CB05 and CBMIV chemical mechanisms, Geosci. Model Dev., 5, 257–268, doi:10.5194/gmd-5-257-2012, 2012.

Schaefer, J. T.: The critical success index as an indicator of warning skill, Weather and Forecasting, 5, 570–575, 1990.

Schwede, D., Pouliot, G., and Pierce, T.: Changes to the biogenic emissions inventory system version 3 (BEIS3), in: 4th Annual CMAS Models-3 User's Conference, Chapel Hill, NC, September 26 - 28, 2005, 2005.

Sillman, S.: The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments, Atmos. Environ., 33, 1821–1845, 1999.

Stajner, I., Davidson, P., Byun, D., McQueen, J., Draxler, R., and amd J. Meagher, P. D.: US National Air Quality Forecast Capability: Expanding Coverage to Include Particulate Matter, in: Air Pollution Modeling and Its Application XXI, edited by Steyn, D. and Castelli, S., pp. 379–384, Springer, Dordrecht, The Netherlands, 2012.

Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevot,

A. S. H., and Hueglin, C.: Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, J. of Geophys. Res., 112, D11 307, doi:10.1029/2006JD007971, 2007.

Talbot, D., Moran, M. D., Bouchet, V., Crevier, L. P., Menard, S., Kallaur, A., and Team, G.-M.: Development of a new Canadian operational air quality forecast model, in: AIR POLLUTION MODELING AND ITS APPLICATION XIX, edited by Borrego, C. and Miranda, A., NATO Science for Peace and Security Series C-Environmental Security, pp. 470–478, NATO; CCMS; Univ Aveiro, doi: 10.1007/978-1-4020-8453-9_52, 29th NATO/CCMS International Technical Meeting on Air Pollution Modeling and Its Applications, Aveiro, PORTUGAL, SEP 24-28, 2007, 2008.

Tang, Y., Lee, P., Tsidulko, M., Huang, H.-C., McQueen, J., DiMego, G., Emmons, L., Pierce, R., Thompson, A., Lin, H.-M., Kang, D., Tong, D., Yu, S., Mathur, R., Pleim, J., Otte, T., Pouliot, G., Young, J., Schere, K., Davidson, P., and Stajner, I.: The impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the continental United States, Environ. Fluid Mech., 9, 43–58, doi:10.1007/s10652-008-9092-5, 2009.

Tong, D., Muller, N., Mauzerall, D., and Mendelsohn, R.: Integrated assessment of the spatial variability of ozone impacts from emissions of Nitrogen Oxides, Environmental Science and Technology, 40, 2006.

U.S. EIA: Annual Energy Outlook 2010, with projections to 2035, Tech. rep., United States Energy Information Administration, DOE/EIA-0383(2010), 2010.

U.S. EPA: Technical Support Document (TSD) for the final Transport Rule, Tech. rep., U.S. Environmental Protection Agency, Docket ID EPA-HQ-OAR-2009-0491, 2011.

van Loon, M., Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P., Christensen, J., Cuvelier, C., Graff, A., Jonson, J., Krol, M., Langner, J., Roberts, P., Rouil, L., Stern, R., Tarrasón, L., Thunis, P., Vignati, E., White, L., and Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, Atmos. Environ., 41, 2083 – 2097, doi:10.1016/j.atmosenv.2006.10.073, 2007.

Wu, Y., Brashers, B., Finkelstein, P. L., and Pleim, J. E.: A multilayer biochemical dry deposition model. 1. Model formulation, J. of Geophys. Res., 108, doi:10.1029/2002JD002293, 2003.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-time air quality forecasting, part I: History, techniques, and current status, 60, 632–655, doi:10.1016/j.atmosenv.2012.06.031, 2012.

Table 1. Monthly and annual average ozone biases in different regions and CONUS in 2010. Unit: ppbv.

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Pacific Coast	5.6	3.7	0.3	−2.6	0.4	3.9	6.3	5.2	7.1	5.2	7.1	4.8	4.0
Lower Middle	−0.3	−0.2	−1.6	−2.3	2.1	8.4	10.4	9.7	8.6	2.6	4.5	1.8	3.7
Southeast	4.7	2.9	3.3	3.7	9.8	14.0	15.4	17.6	13.9	10.4	9.8	6.2	10.5
Rocky Mountain	6.2	3.7	3.7	−1.1	−0.3	3.1	6.9	5.2	8.7	7.0	7.3	6.5	4.7
Upper Middle	−3.0	−8.1	−0.6	−1.8	0.6	6.6	8.2	9.3	8.0	3.7	3.6	−1.5	4.4
Northeast	−1.0	−5.6	−3.5	−1.1	1.5	7.0	10.5	10.4	11.0	6.9	5.2	0.2	5.1
CONUS	2.9	0.8	0.5	−0.6	2.6	7.5	9.8	9.9	9.7	6.3	6.4	3.7	5.6

Table 2. Monthly and annual average ozone RMSEs in different regions and CONUS in 2010. Unit: ppbv.

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Pacific Coast	14.7	14.4	15.3	14.0	12.8	15.1	17.8	17.4	18.7	15.9	15.9	13.8	15.7
Lower Middle	12.7	12.9	15.0	13.9	15.2	16.3	17.3	18.8	17.9	16.2	13.3	11.8	15.4
Southeast	11.5	12.0	13.2	14.0	16.3	19.6	20.3	22.5	19.6	17.0	15.3	12.5	17.6
Rocky Mountain	14.1	13.8	13.8	12.5	12.5	14.0	15.4	15.1	16.3	14.8	14.2	13.8	14.3
Upper Middle	8.8	12.9	13.4	12.1	13.1	14.7	16.2	17.3	14.3	12.6	10.9	8.6	14.3
Northeast	8.4	10.6	12.3	12.8	12.9	15.1	17.8	17.5	16.3	13.0	11.4	8.3	14.4
CONUS	12.7	13.1	14.0	13.2	13.9	16.0	17.6	18.4	17.1	15.1	14.0	12.3	15.4

Table 3. Monthly and annual averaged NO_2 Biases in different regions and CONUS in 2010. Unit: ppbv.

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Pacific Coast	6.2	6.3	6.2	4.9	5.6	8.3	8.7	9.2	8.9	7.0	5.9	8.1	7.1
Lower Middle	8.6	8.4	8.2	7.1	7.0	7.2	8.1	8.7	9.7	8.2	7.9	7.6	8.1
Southeast	4.3	4.6	3.4	2.9	3.7	4.0	3.7	5.0	4.1	3.5	4.3	4.8	4.0
Rocky Mountain	−0.9	−0.4	0.3	0.2	0.4	1.1	1.7	1.3	0.7	0.5	−0.2	0.0	0.4
Upper Middle	4.6	5.6	6.2	4.2	5.7	5.9	5.9	6.1	6.1	5.7	5.9	5.5	5.6
Northeast	4.0	4.9	3.4	2.9	2.9	2.4	2.2	3.0	3.7	3.9	4.3	3.6	3.4
CONUS	4.7	5.0	4.8	3.9	4.4	5.2	5.5	6.0	6.1	5.1	4.8	5.2	5.1

Table 4. Monthly and annual averaged NO_2 RMSEs in different regions and CONUS in 2010. Unit: ppbv.

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Pacific Coast	15.5	14.8	15.4	12.7	12.8	15.6	16.0	17.5	19.6	16.4	16.0	16.1	15.8
Lower Middle	16.0	16.0	17.3	15.7	15.6	14.2	15.0	17.8	19.1	18.6	15.8	15.2	16.4
Southeast	10.4	11.3	11.0	9.5	8.7	9.6	9.5	10.5	10.4	10.4	10.9	11.1	10.3
Rocky Mountain	12.0	11.3	11.0	8.5	8.1	8.0	8.5	8.6	9.6	9.6	9.8	10.5	9.7
Upper Middle	11.0	12.5	13.6	11.0	11.6	11.6	12.4	12.9	12.6	13.0	11.6	11.1	12.1
Northeast	11.1	12.1	12.4	11.4	9.5	9.4	9.6	10.1	10.3	10.3	11.3	10.2	10.7
CONUS	13.5	13.5	14.1	12.1	11.7	12.4	12.8	14.1	15.4	14.2	13.5	13.3	13.4

Table 5. Daily maximum eight-hour ozone categorical statistics for 2010, with the 75 ppbv and 70 ppbv thresholds. See text for details.

Region, standard (ppbv)	N_a	N_b	N_c	N_d	HIT	CSI	FAR	ETS	POC
Pacific Coast, 75	2047	977	47 796	681	0.59	0.26	0.68	0.24	0.95
Lower Middle, 75	1156	168	44 875	217	0.44	0.11	0.87	0.10	0.97
Southeast, 75	2774	345	54 107	132	0.72	0.11	0.89	0.10	0.95
Rocky Mountain, 75	939	70	42 667	88	0.44	0.06	0.93	0.06	0.98
Upper Middle, 75	1963	281	52 056	171	0.62	0.12	0.87	0.11	0.96
Northeast, 75	2232	770	42 637	157	0.83	0.24	0.74	0.23	0.95
CONUS, 75	11 119	2616	284 706	1449	0.64	0.17	0.81	0.16	0.96
CONUS, 70	18 192	5753	273 121	2824	0.67	0.21	0.76	0.19	0.93

Table 6. Daily maximum eight-hour ozone categorical statistics for summer months (June–August) in 2010, See text for details.

Region, standard (ppbv)	N_a	N_b	N_c	N_d	HIT	CSI	FAR	ETS	POC
Pacific Coast, 75	1507	632	12 427	424	0.60	0.25	0.70	0.20	0.87
Lower Middle, 75	763	85	11 138	49	0.63	0.09	0.90	0.09	0.93
Southeast, 75	1835	219	16 984	22	0.91	0.11	0.89	0.09	0.90
Rocky Mountain, 75	712	61	12 001	62	0.50	0.07	0.92	0.06	0.94
Upper Middle, 75	1599	226	20 356	77	0.75	0.12	0.88	0.11	0.92
Northeast, 75	1799	584	13 703	92	0.86	0.24	0.75	0.20	0.88
CONUS, 75	8223	1812	86 747	729	0.71	0.17	0.82	0.15	0.91
CONUS, 70	12 810	3764	79 758	1179	0.76	0.21	0.77	0.17	0.86

Table 7. Ozone biases for the different days of the week in the six predefined regions and CONUS. June–September 2010. Unit: ppbv.

Region	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Week
Pacific Coast	7.3	6.0	4.9	4.9	4.9	5.3	6.2	5.6
Lower Middle	10.1	9.7	9.3	8.7	8.3	9.1	9.9	9.3
Southeast	16.0	15.6	15.1	14.8	14.4	15.2	15.7	15.2
Rocky Mountain	6.6	5.7	4.8	5.9	6.6	6.1	6.3	6.0
Upper Middle	8.7	8.4	7.6	8.8	7.4	7.1	8.3	8.0
Northeast	10.8	9.4	9.2	10.8	10.2	8.5	9.1	9.7
CONUS	10.1	9.4	8.7	9.3	8.8	8.8	9.5	9.2

Table 8. NO_2 biases for the different days of the week in the six predefined regions and CONUS. June–September 2010. Unit: ppbv.

Region	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Week
Pacific Coast	7.5	9.3	9.4	9.1	9.4	9.2	7.2	8.8
Lower Middle	8.3	8.7	8.5	8.7	8.6	8.3	7.8	8.4
Southeast	4.2	4.6	4.8	4.6	4.1	3.9	3.1	4.2
Rocky Mountain	1.0	1.6	1.4	1.5	1.3	1.2	0.3	1.2
Upper Middle	5.5	6.5	7.0	6.4	6.2	5.4	5.1	6.0
Northeast	3.2	2.5	3.1	2.9	2.5	2.7	2.8	2.8
CONUS	5.3	6.0	6.1	6.0	5.9	5.7	4.7	5.7
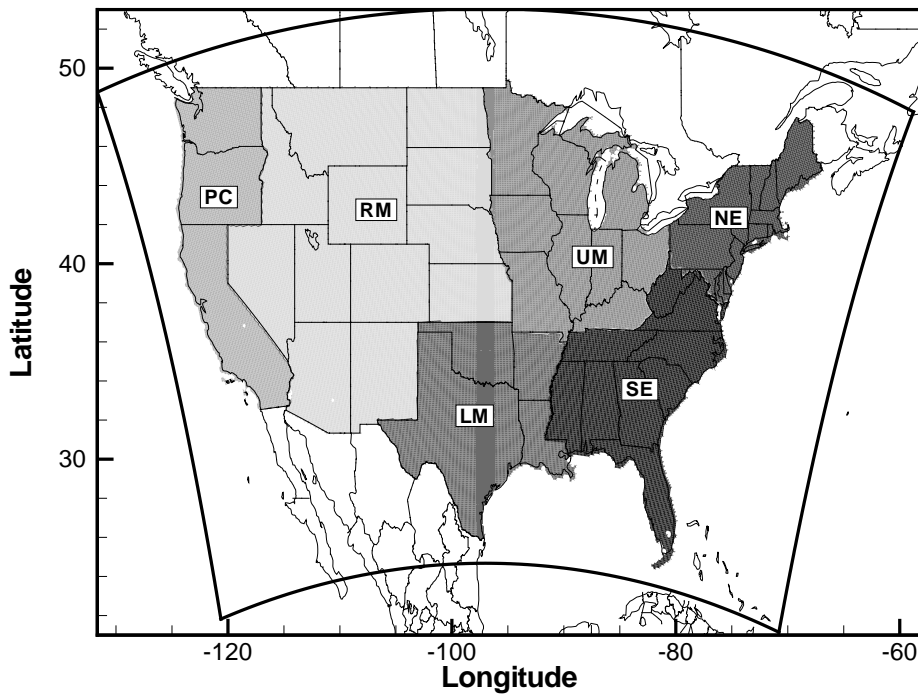
Fig. 1. NAQFC computational domain and six predefined US regions: Pacific Coast (PC), Rocky Mountain (RM), Southeast (SE), Lower Middle (LM), Upper Middle (UM), and Northeast (NE).
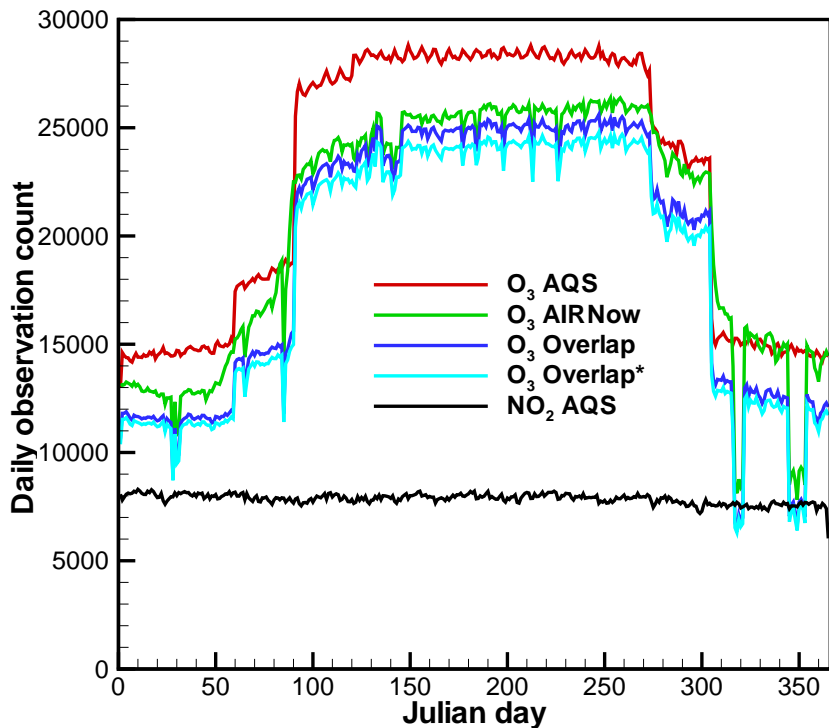
Fig. 2. Data counts of ozone and NO_2 hourly measurements for each day during 2010. "O_3 overlap" shows the number of overlapping ozone observation pairs and "O_3 overlap*" is the number of observation pairs after removing measurements from the 74 questionable sites (see text for details).

.

33
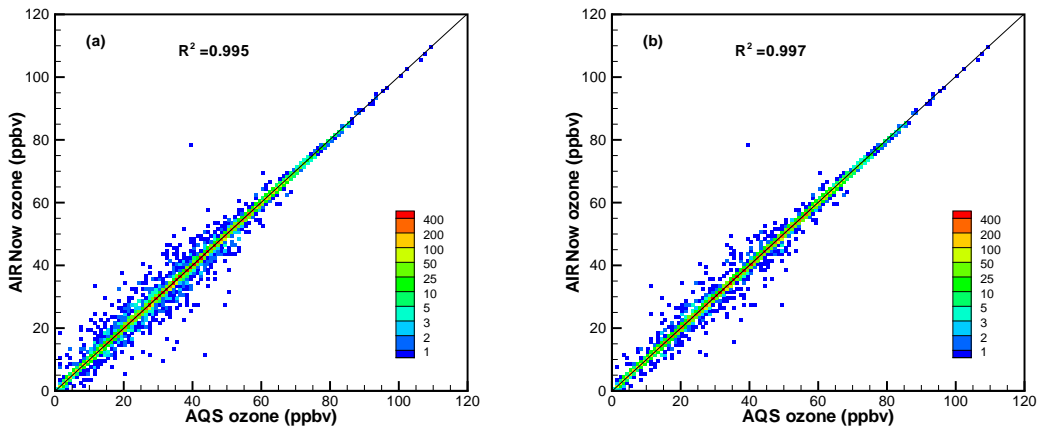
Fig. 3. Comparison between AIRNow and AQS data on 31 May 2010 before **(a)** and after **(b)** removing the 74 questionable sites using density plots, in which color represents the count of observation pairs at each pixel. The data between 00:00 EDT to 23:00 EDT are included here.
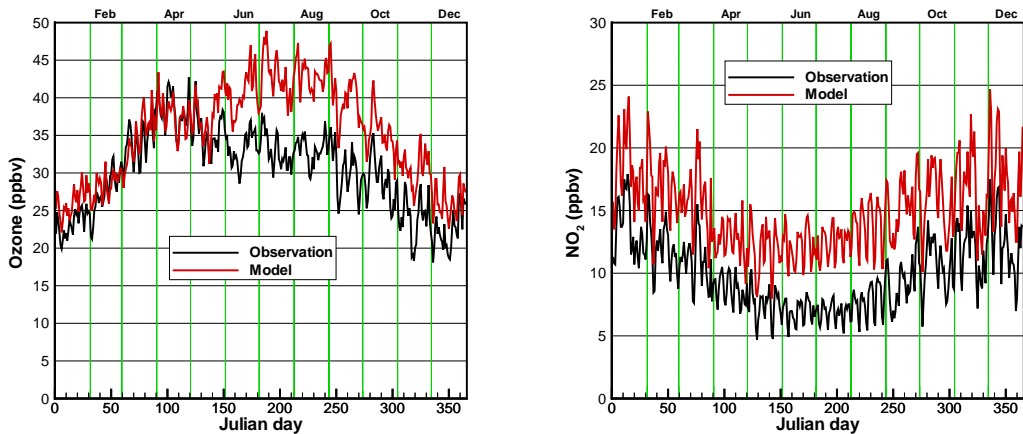
Fig. 4. Daily domain-wide average ozone and NO_2 concentrations in 2010.
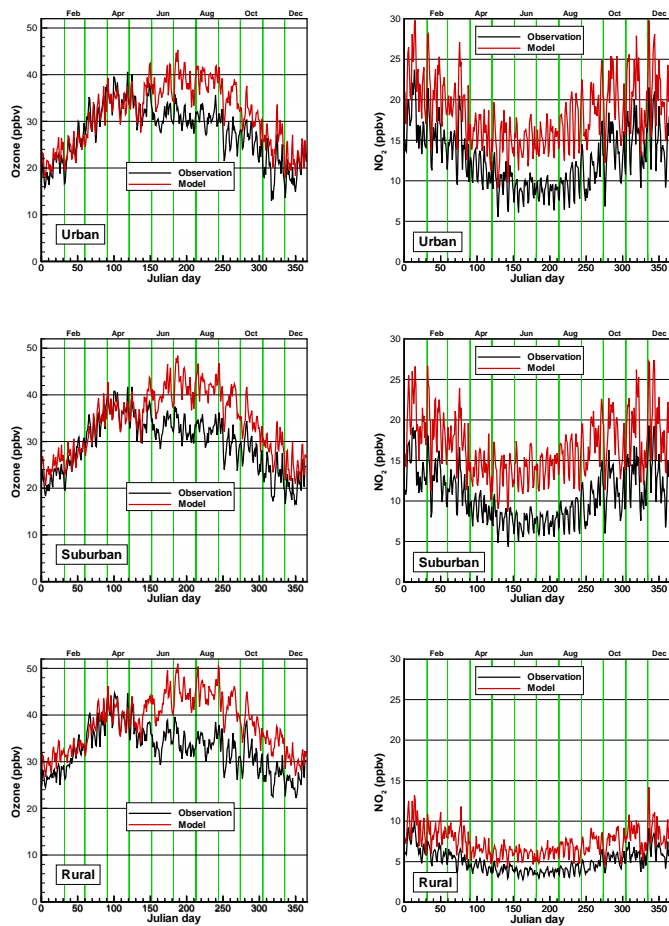
Fig. 5. Daily domain-wide average ozone and NO_2 concentrations at urban, suburban, and rural sites in 2010.
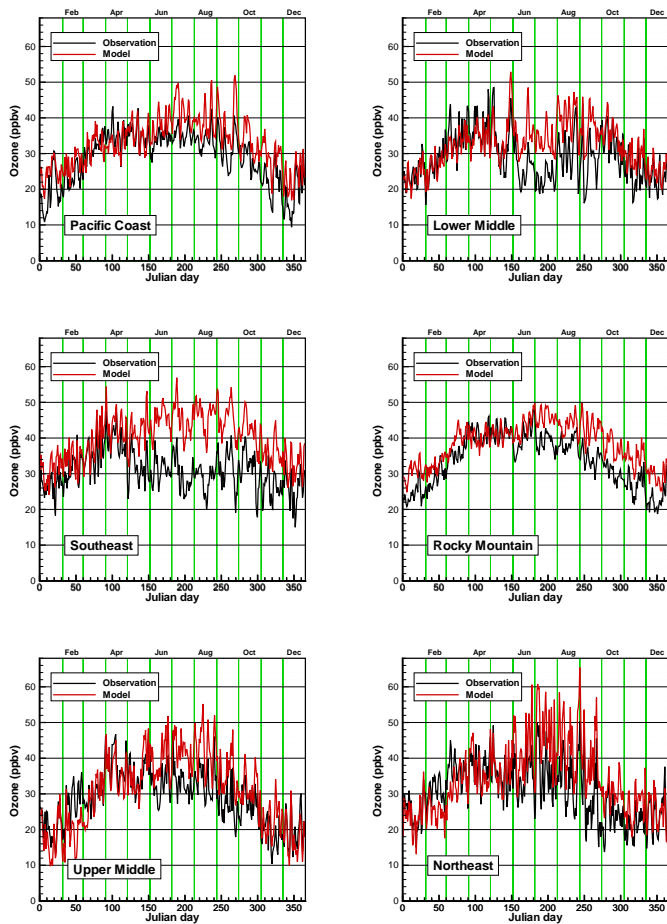
Fig. 6. Daily average ozone concentrations in 2010 for each of the six regions listed in Fig. 1.

37
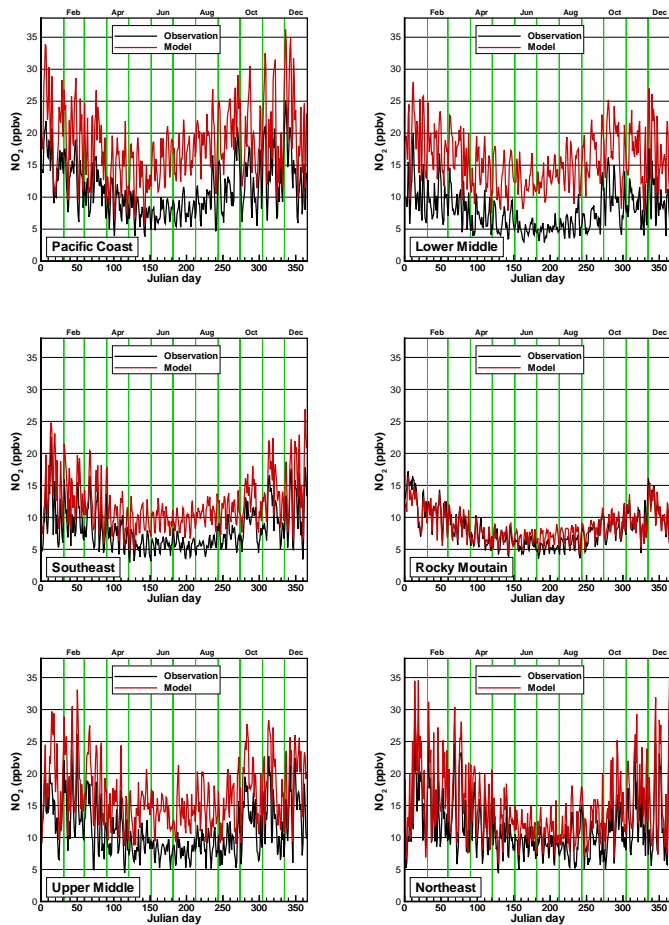
Fig. 7. Daily average AQS (black) and NAQFC NO_2 concentrations in 2010 for each of the six regions listed in Fig. 1.
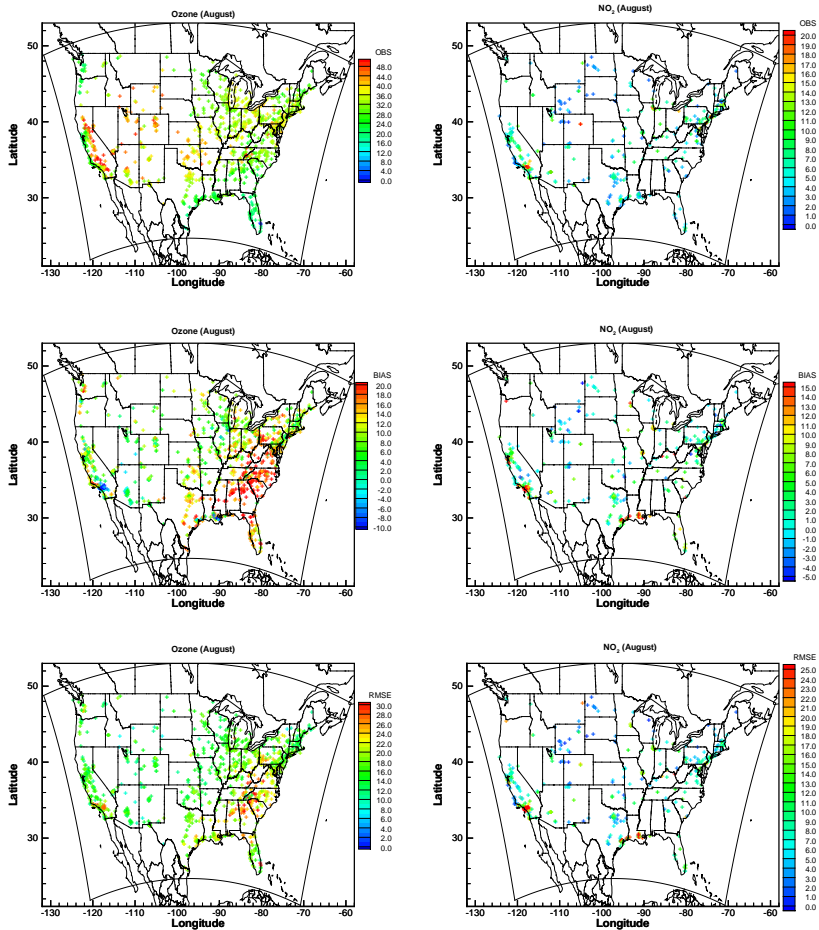
Fig. 8. Spatial distributions of ozone (left) and NO_2 (right) measurements (upper row), model biases (middle row), and RMSEs (lower row) in August 2010. Unit: ppbv.
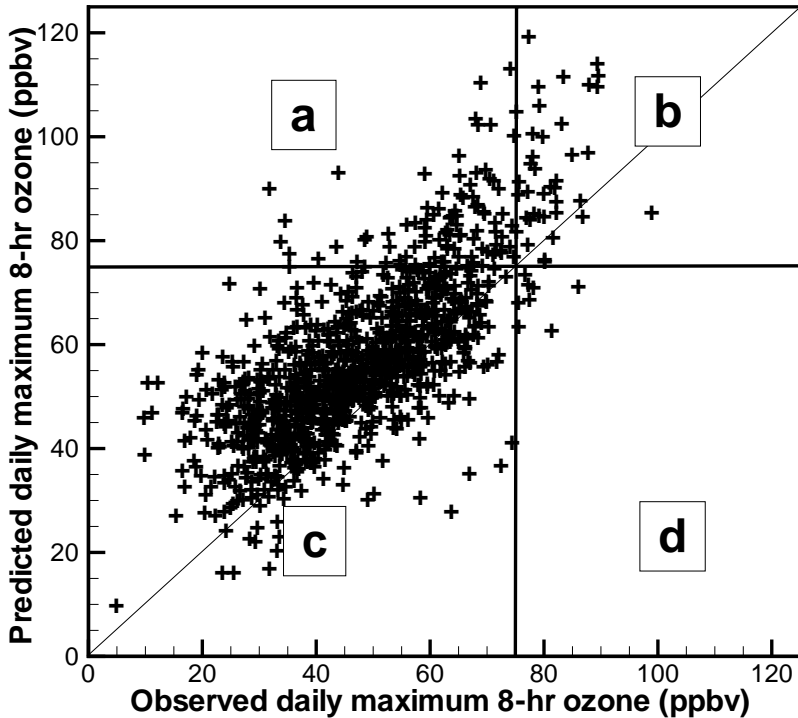
Fig. 9. Diagram of categorical statistics calculation. A scatter plot with AQS observed and NAQFC predicted daily maximum eight-hour average ozone on 17 August 2010 is shown as an example. The US standard for daily maximum eight-hour average ozone of 75 ppbv is used as the threshold to delimit the scatter plots into four regions, **(a)** prediction is above, but observation is below the threshold; **(b)** prediction and observation are above the threshold; **(c)** prediction and observation are below the threshold; **(d)** prediction is below, but observation is above the threshold.
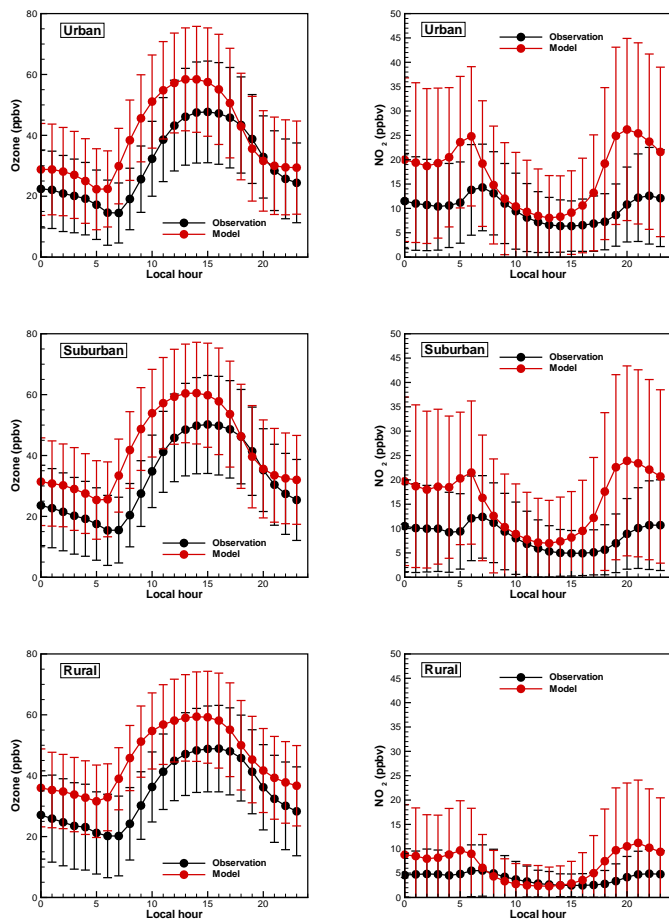
Fig. 10. Diurnal profiles of ozone and NO_2 at urban, suburban, and rural sites in August 2010. Average concentrations of the AQS observations and their NAQFC counterparts are shown with their standard deviations as a function of local hours.
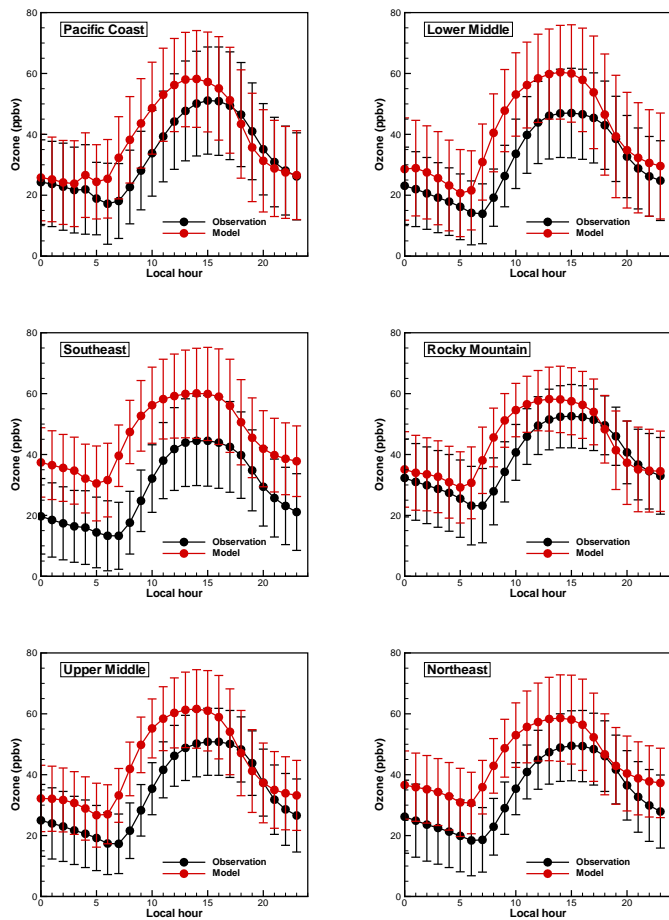
Fig. 11. Diurnal profiles of ozone in August 2010 for each of the six regions listed in Fig. 1. Average concentrations of the AQS observations and their NAQFC counterparts are shown with their standard deviations as a function of local hours.
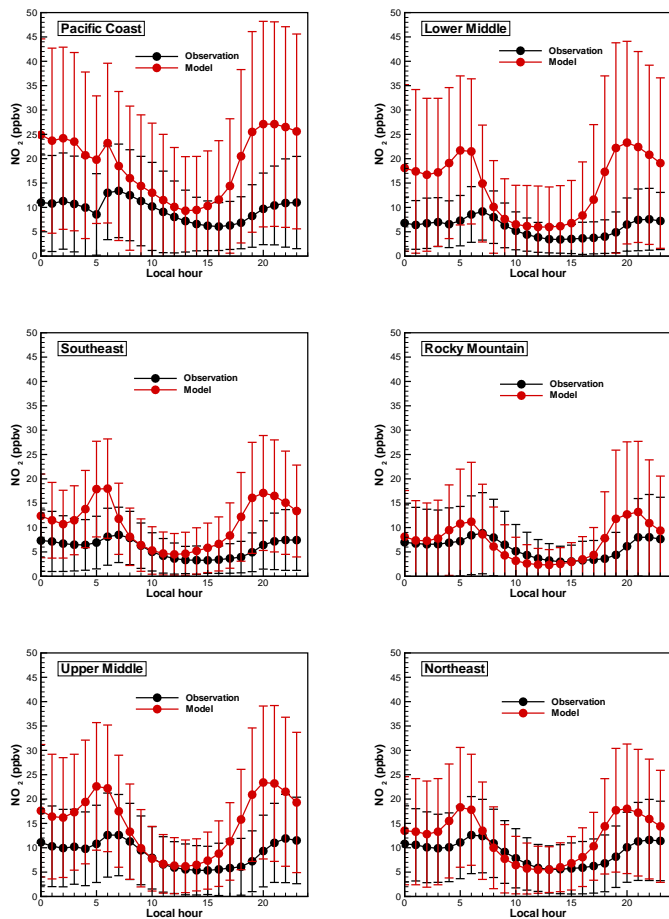
42

Fig. 12. Diurnal profiles of NO_2 in August 2010 for each of the six regions listed in Fig. 1. Average concentrations of the AQS observations and their NAQFC counterparts are shown with their standard deviations as a function of local hours.