Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 42 (2008) 5110–5123
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Dynamic evaluation of regional air quality models: Assessing changes in O_3 stemming from changes in emissions and meteorology

Alice B. Gilliland[a,*,1], Christian Hogrefe[b], Robert W. Pinder[a,1], James M. Godowitch[a,1], Kristen L. Foley[a,1], S.T. Rao[a,1]

[a]*Atmospheric Sciences Modeling Division, Air Resources Laboratory, National Oceanic and Atmospheric Administration, Research Triangle Park, NC 27711, USA*
[b]*Atmospheric Sciences Research Center, University at Albany, 251 Fuller Road, Albany, NY 12203, USA*

Received 20 August 2007; received in revised form 3 February 2008; accepted 4 February 2008

Abstract

Regional-scale air quality models are used to estimate the response of air pollutants to potential emission control strategies as part of the decision-making process. Traditionally, the model-predicted pollutant concentrations are evaluated for the "base case" to assess a model's ability to reproduce past observations. Dynamic evaluation approaches, which evaluate a model's ability to accurately simulate air quality changes from given changes in emissions, are critically important to regulatory applications. Here, we investigate approaches to evaluate the Community Multiscale Air Quality (CMAQ) model's predicted ozone (O_3) response to large NO_x emission reductions associated with the NO_x State Implementation Plan (SIP) Call and on-road mobile emissions. This case has the advantages that emission changes associated with the NO_x SIP Call can be well characterized and substantial changes are observed in O_3 levels. To consider the modeled response to emission changes in light of the strong meteorological influences on O_3, two time periods after the NO_x SIP Call are included with very different meteorological conditions. The sensitivity to chemical mechanisms is also considered by including simulations with the CB4, SAPRC, and CB05 chemical mechanisms. The evaluation results suggest that the air quality model predictions underestimate the O_3 reductions observed after the NO_x SIP Call was implemented. While the emission estimate uncertainties may also be a factor, the results suggest that the contribution of long-range transport of O_3 and precursors is underpredicted, especially when using the CB4 chemical mechanism. Further investigation of the chemical mechanisms' ability to characterize tropospheric chemistry aloft is recommended. Results based on the most recent CMAQ version 4.6 with CB05 and updated emission inventories show incremental improvements to the modeled O_3 response to NO_x emission reductions.

Keywords: Dynamic evaluation; NO_x emission reductions; CMAQ air quality model

1. Introduction

Regional-scale atmospheric chemical transport models that predict tropospheric and near-surface

*Corresponding author.
E-mail address: gilliland.alice@epa.gov (A.B. Gilliland).
[1]In partnership with the U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory.

1352-2310/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2008.02.018

concentrations of O_3 and aerosol species are generally referred to as air quality models. Emissions for air quality models are estimated from spatially heterogeneous and concentrated emission sources, such as electrical generating units (EGUs) and motor vehicles, and the physical and chemical processes are simulated within an Eulerian grid structure. For air quality management decisions, these models are used to simulate air quality impacts from hypothetical emission reduction scenarios. Costs and benefits are ultimately considered before final emission reductions are selected; however, the ability of an air quality model to accurately predict the changes in pollutant concentrations stemming from emission controls is important for developing meaningful and cost-effective emission control strategies. Since the real impact of potential emission control scenarios on air quality cannot be directly evaluated, the usual approach taken in regulatory modeling is to establish a model's credibility based on the model's ability to reproduce observed concentrations for the "base case". This base case reflects a simulation under recent historical meteorological conditions with the best emission inventory available. Evaluating the model's ability to reproduce past air quality conditions is very different from evaluating the model's ability to predict changes in air quality given changes in emissions. The latter issue referred to here as a "dynamic evaluation" is only possible if a retrospective case exists where (1) substantial emission reductions have resulted in discernable changes in air quality over time and (2) the change in emissions can be quantified accurately. An additional challenge is that the air quality changes over time are also driven by meteorological variability.

A strong candidate for a dynamic evaluation case study is the time period between 2002 and 2005 because large and abrupt NO_x emission changes occurred during this time as a result of the U.S. Environmental Protection Agency's (USEPA) NO_x State Implementation Plan (SIP) Call in addition to a more gradual decreasing trend in mobile on-road emissions (USEPA, 2005a). The NO_x SIP Call required substantial reductions in NO_x emissions from twenty-one states in the eastern United States by 31 May, 2004 to reduce interstate transport of O_3 and its precursors across the eastern United States. USEPA (2005a, 2006) reported large decreases in the observed daily maximum 8-h O_3 concentrations during 2004 and 2005 as compared to 2002 O_3 levels within the NO_x SIP Call region. Gégo et al. (2007) have shown that the observed O_3 reductions after the NO_x SIP Call implementation are most evident in the regions that are downwind of major EGUs in the Ohio River Valley. Therefore, this time period satisfies the first requirement for the dynamic evaluation, namely, that a discernable change in air quality is associated with known emission reductions. Satisfying the second requirement for the dynamic evaluation is the fact that the emission reductions for the NO_x SIP Call are directly measured. As required by the USEPA Acid Rain Program, major coal-fired EGUs are required to equip each stack with Continuous Emission Monitoring Systems (CEMS) that measure hourly SO_2 and NO_x emission rates and heat input (http://camddataandmaps.epa.gov/gdm/index.cfm?fuseaction=emissions.wizard). To meet the NO_x SIP Call requirements, all states involved chose to participate in the NO_x Trading Budget Program and implemented reductions for power industry EGUs. These emission reductions have also been detected in satellite-derived NO_2 columns in areas where major utility emission sources exist and not in the areas of heavy mobile emissions (Kim et al., 2006). Since these EGUs are instrumented with CEMS equipment, a large part of the NO_x SIP Call emission reductions are directly measured and can be incorporated into air quality model simulations.

Previous modeling studies have shown that emission inventories that do not account for NO_x emission reductions from the NO_x SIP Call introduce substantial biases in the O_3 predictions during the summer 2004 (Frost et al., 2006; Hudman et al., 2007; Kim et al., 2006), which illustrates the need for an accurate emission inventory for good model performance of the "base case." In the context of dynamic evaluation for this study, emission estimates and modeling simulations are developed for time periods before and after the NO_x SIP Call was implemented, and the modeled and observed changes in O_3 levels from summer 2002 to summers 2004 and 2005 are compared. It should be noted that these O_3 changes are influenced both by changes in emissions as well as meteorology, making it difficult to isolate the influence of either on the O_3 changes. Therefore, this dynamic evaluation approach entails simulating the year-to-year changes in O_3 that are influenced by variability in emissions and meteorological conditions to determine if insights can be gained about the model's response to both driving factors.

2. Approach for the dynamic modeling evaluation

Air quality model simulations before and after the NO_x SIP Call emission changes have been developed for this study to compare predicted and observed changes in O_3 levels. Summer (June, July, and August) 2002 was selected as the pre-NO_x SIP Call time period and summers 2004 and 2005 were selected as the post-NO_x SIP Call time periods. As discussed in the introduction, year-to-year changes in O_3 are driven both by changes in emissions and meteorology. Whereas both 2002 and 2005 summer conditions were very conducive to O_3 formation and accumulation, summer 2004 experienced much cooler, wetter conditions (USEPA, 2005a, 2006). USEPA (2006) illustrated that the O_3 reductions between 2002 and 2004 were much larger than those from 2002 to 2005 because of the differences in meteorological conditions. The comparison of modeled O_3 changes from 2002 to 2004 and 2005 is included here to consider how much of the O_3 change is attributable to changes in emissions.

Another approach to attribute the modeled O_3 changes to emission or meteorological influences could be to perform 'cross' simulations where the meteorology is held constant and emissions changed or vice versa. Godowitch et al., (2008a) demonstrate that these 'cross' simulations are quite useful for interpreting the model sensitivity to meteorology and emission changes; however, these hypothetical scenarios cannot be compared directly to observational data. Meteorologically-adjusted O_3 data (Camalier et al., 2007; USEPA, 2005a, 2006) account for about 70–80% of the interannual meteorological variability influence (Cox and Chu, 1993; Zheng et al., 2007; Camalier et al., 2007); therefore, it is not a direct measurement and would introduce additional uncertainty into the comparison.

For this application, we instead will use two different post-NO_x SIP Call summer periods in 2004 and 2005; the varying meteorology will help address the influence of meteorological changes on the O_3 response. In addition to the comparison of these post-NO_x SIP Call periods to the summer 2002 pre-NO_x SIP Call O_3 levels, predictions for each summer are compared to observed O_3 concentrations.

The O_3 metrics for comparisons will be the maximum daily 8-h O_3 (MDA8) values at the median and those equal to and above the 95th percentile of the cumulative frequency distribution (hereafter referred to as $\geqslant$95th%). The portion of the distribution $\geqslant$95th% is near the regulatory metric of the O_3 design value, which is the 3-year average of the observed 4th highest MDA8 O_3 values at a site (http://www.epa.gov/oar/oaqps/greenbook/40cfr50.html#appi). The observed and modeled O_3 metrics are calculated for each available observational site from the Air Quality System (AQS) network and Clean Air Status and Trends network (CASTNet) (Section 4).

3. Model description

The air quality simulations for this study were developed using the Community Multiscale Air Quality (CMAQ) model (Byun and Schere, 2006, www.cmaq-model.org). Simulations were performed for summers 2002, 2004, and 2005 with each summer period including June 1 through August 31 with a 10-day spin-up at the end of May. The modeling domain covers the eastern U.S. with 12 km × 12 km horizontal grid cells, and boundary conditions are specified from a 36 km × 36 km continental U.S. domain simulation for that time period. Boundary conditions for the 36 km × 36 km continental U.S. CMAQ domain simulation reflected average "background" profiles. For both 36 km and 12 km simulations, the model has 14 vertical layers extending up to 100 mb. Two series of CMAQ version 4.5 simulations were developed using both the Carbon Bond IV (Gery et al., 1989) and SAPRC99 (Carter, 2000) chemical kinetic mechanisms to assess the sensitivity of O_3 changes to the choice of chemical mechanism. These simulations are hereafter referred to as CMAQ-CB4 and CMAQ-SAPRC, respectively. Previous studies indicate that SAPRC tends to predict higher O_3 levels than CB4 (Jimenez et al., 2003) and shows a larger O_3 response to NO_x emission reductions (Arnold and Dennis, 2006). An additional set of CMAQ version 4.6 simulations for summers 2002 and 2005 was added during the course of this study that included an updated version of the carbon bond mechanism (CB05) (Sarwar et al., 2008) and new emission inventories for 2002 and 2005 (discussed below).

The meteorological fields for these CMAQ simulations were generated by the MM5 v3.6.3 model (Grell et al., 1994). Specific options used for these MM5 simulations include the Pleim-Xiu land surface scheme, Reisner 2 microphysics, Kain-Fritsch cumulus parameterization, and RRTM radiation parameterization. The MM5 analysis

nudging option (four-dimensional data assimilation) was exercised to assimilate observations into the simulation, and the MM5 simulation was reinitialized every four days. To create model-ready meteorology inputs for CMAQ, the MM5 results were then processed using the Meteorology-Chemistry Interface Program (MCIP) version 3. Further details on the model configuration can be found in Godowitch et al. (2008a).

To estimate criteria pollutant emissions for the CMAQ-CB4 and CMAQ-SAPRC simulations of summers 2002, 2004, and 2005, the USEPA National Emission Inventory (NEI) 2001 (http://www.epa.gov/ttn/chief/net/critsummary.html) was updated with year specific point source emissions from the observed CEMS data and estimated on-road mobile emissions from the Mobile6 emission model (http://www.epa.gov/otaq/m6.htm). Mobile6 was executed through the Sparse Matrix Operator Kernel Emissions (SMOKE) processor (http://www.smoke-model.org/version2.3.2/html/ch02s16.html). Within the NEI 2001 SMOKE-formatted Mobile6 dataset available for this study, databases for emission control programs (e.g., inspection and maintenance programs and reformulated gas) and vehicle miles traveled (VMT) are provided. The emission control program database provided information for reference counties within each state. Biogenic emissions were also estimated for each summer based on the meteorological conditions using BEIS 3.13 (Schwede et al., 2005). All other sectors (e.g., non-road and area emissions) were left unchanged from the 2001 NEI. If, in fact, there were reductions in emissions from the other sectors, these would not be reflected in the modeling scenarios studied here and could result in smaller predicted O_3 responses than observed.

In addition to the emission estimates projected from the 2001 NEI as described above, USEPA has recently released the 2002 and 2005 version 1 NEI inventories (http://www.epa.gov/ttn/chief/net/2002inventory.html; http://www.epa.gov/ttn/chief/net/2005inventory.html). The 2002 and 2005 NEI now include county specific control programs for Mobile6, and mobile source emissions were calculated by driving Mobile6 with the National Mobile Inventory Model (NMIM) (http://www.epa.gov/otaq/nmim.htm; http://www.epa.gov/otaq/models/nmim/420r05024.pdf). The 2002 and 2005 NEI estimates include the year specific EGU CEMs data and mobile source and non-road emissions calculated by driving Mobile6 and the NONROAD model through NMIM with county-specific control program information. The 2002 NEI also includes state, local, and tribal Consolidated Emissions Reporting Rule (CERR) submittals, but the 2005 NEI version 1 does not. Because of the uncertainties in the emission projections for the pre- and post-NO_x SIP Call and the fortuitous availability of SMOKE-processed emission files for the CMAQ v4.6 CB05 chemical mechanism, additional simulations are included in this study using these 2002 and 2005 NEI estimates (referred to as CMAQ-CB05 here). Including these new 2002 and 2005 NEI estimates offers the opportunity to consider the impact of the county-specific control program information on the on-road mobile NO_x emissions and include CB05 sensitivity tests in this study. These simulations were not part of the original design of this study but are being included because of these benefits; however, the differences in model predictions cannot be fully attributed to one factor since both emissions and the model version and chemical mechanism are changing.

Resulting NO_x emissions totals for summers 2002, 2004, and 2005 are shown in Table 1; these summary changes are based on states that fall entirely within the 12 km eastern U.S. modeling domain. The comparison of emission estimates as projected from the 2001 NEI are presented in the top half of Table 1. In summer 2004, a reduction of −41% occurred in EGU emissions and −12% in on-road mobile emissions as compared to the pre-NO_x SIP Call summer 2002 period. The large reductions in the EGU NO_x emissions are most evident in the Ohio River Valley area, where the majority of NO_x emissions come from the utility sector (Godowitch et al., 2008a). Since the reductions from the EGU emissions are based on the CEMS data, these changes have low uncertainty. Due to this major reduction in emissions from EGUs, other categories including on-road and non-road mobile emissions become relatively larger contributors to the total NO_x emission estimated inventory. In summer 2005, a total domain-wide reduction of −40% occurred in EGU emissions and −18% in on-road mobile emissions from the pre-NO_x SIP Call summer 2002 period. A comparison of the 2002 and 2005 NEI (Table 1) shows similar reductions from 2002 to 2005 in the EGU and mobile sectors, with little change in the non-EGU and non-road sectors. However, since 2005 version 1 NEI did not include CERR information, reductions in the non-EGU sector may actually

Table 1
Domain-wide NO_x emission estimate summaries for June, July, and August in 2002, 2004, and 2005

	NO_x emissions (ktons)				
	EGUs	Non-EGUs	Mobile	Non-road	Total[a]
Projections from 2001 National Emission Inventory (NEI)					
Summer 2002	858	466	1423	770	3517
Summer 2004	507	466	1257	770	3000
Difference (2004–2002)	−41%		−12%		−15%
Summer 2005	513	466	1172	770	2921
Difference (2005–2002)	−40%		−18%		−17%
2002 and 2005 NEI					
Summer 2002 NEI	853	373	1148	734	3108
Summer 2005 NEI	493	370	911	722	2496
Difference (2005–2002)	−42%		−21%		−20%

[a]These sums only include the major emission sectors listed above. Other sectors that are less than 5% of the total emissions are included in the emission inputs to CMAQ but not in the table.

have been higher. While the trends are quite similar to the projected trends from the 2001 NEI, the total and state-level mobile source NO_x emissions are about 20% higher for 2002 and 2005 from the 2001 NEI-derived Mobile6 emissions than from the 2002 NEI- and 2005 NEI-derived Mobile6 estimates. While the exact same version of Mobile6 is used in both cases, the availability of county specific information in the later inventories versus the reliance on reference county information in the 2001 NEI dataset is likely one cause for the difference. Since county-specific control program information is considered more accurate, the projected on-road mobile emissions from the 2001 NEI used for the CMAQ-CB4 and CMAQ-SAPRC simulations are likely biased too high even though the trend is consistent.

4. Observations: CASTNet and AQS data

The observed O_3 concentrations for this study were extracted from the Clean Air Status and Trends Network (CASTNet, http://www.epa.gov/castnet/) and AQS (AQS, http://www.epa.gov/air/data/aqsdb.html) databases. Following the regulatory metric for O_3 non-attainment, the maximum daily 8 h (MDA8) O_3 concentration is then calculated from the hourly CASTNet and AQS O_3 data. For the MDA8, a running 8-h average O_3 concentration is calculated for each hour, and then the maximum value for the day is selected. For the comparison of 2002 and 2004 summers, 719 AQS and 56 CASTNet sites are available within the 12 km modeling domain. A total of 649 AQS sites and 55 CASTNet are available for comparison of 2002 and 2005 summers.

5. Results and discussion

Comparison is shown in Fig. 1 of the model-predicted MDA8 O_3 changes from 2002 to 2004 and 2005 against the combined CASTNet and AQS data for the CMAQ-CB4, CMAQ-SAPRC, and CMAQ-CB05 (2005–2002 only) simulations. Note that while included in the figures, the CMAQ-CB05 results will not be discussed until later in this section. For the 2002–2004 comparison, the model-predicted decrease in MDA8 O_3 is less than the observed decrease; however, the model-predicted decreases are closer to observed with the SAPRC99 chemical mechanism than CB4. For the 2002–2005 comparison, the model-predicted decrease in MDA8 O_3 is also less than the observed decrease with CMAQ-CB4 and CMAQ-SAPRC results comparing similarly. In both comparisons of 2004 and 2005 versus 2002, the CMAQ-CB4 results do show less scatter in a paired comparison against observations than the CMAQ-SAPRC results (see Fig. 1 in Appendix A).

Table 2 summarizes the average change across all AQS and CASTNet sites in observed and model-predicted O_3 levels from summers 2002–2004 and 2002–2005 based on both changes in absolute O_3 levels (ppb) as well as relative percent changes in O_3 (e.g., (2004−2002)/2002). Looking at absolute changes in O_3, the average modeled change in O_3 that agrees best with the observed changes in O_3 at all sites is the CMAQ-SAPRC simulation for 2002 versus 2004, which predicts about 63% and 70% on

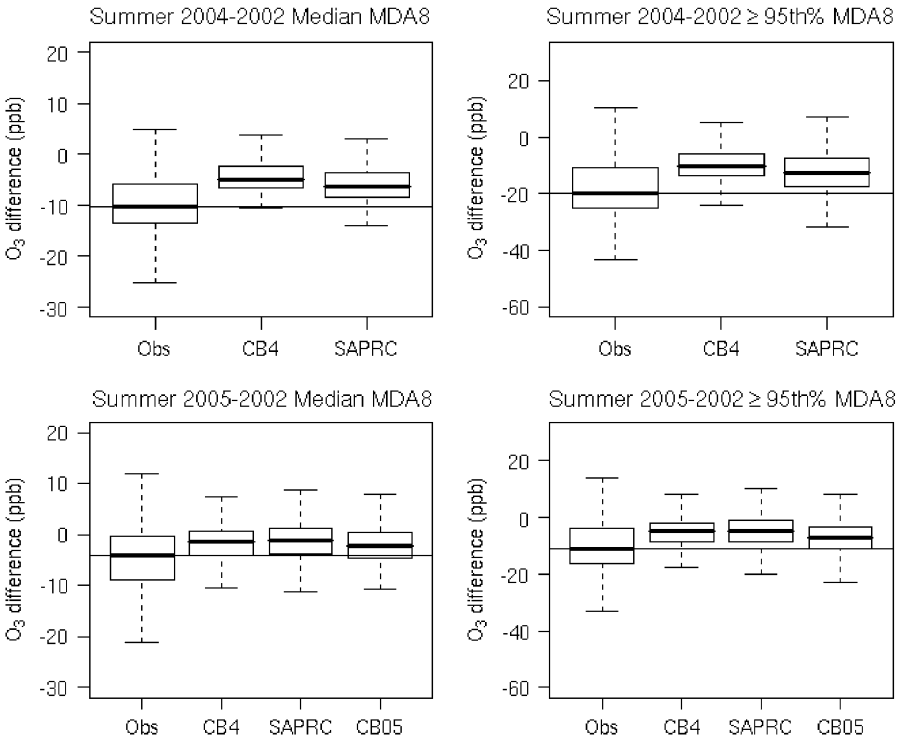

Fig. 1. Boxplot (median, inner, and outer quartile) comparison of the observed (CASTNet and AQS) and modeled change in daily maximum 8-h O_3 between summer 2002 and summer 2004 (2004–2002) and summer 2005 (2005–2002). Results are shown for both the CMAQ v4.5 CB4 and SAPRC and CMAQ v4.6 CB05 simulations. The range of O_3 values across the distribution in the boxplots represents the spatial variability in the observations and model simulations at the median or average above $\geqslant$95%.

average of the observed decrease at the median and the $\geqslant$95th%, respectively. The CMAQ-CB4 simulations for 2004 versus 2002 and both the CMAQ-CB4 and CMAQ-SAPRC simulations for 2005 versus 2002 simulate approximately 50% on average of the observed O_3 decreases. If the relative change with respect to 2002 is compared, results for most cases are similar (Table 2). The exception is that relative O_3 reductions at the $\geqslant$95th% for CMAQ-CB4 (2002 versus 2004 and 2002 versus 2005) is larger than if the absolute changes in ppb is compared.

In Fig. 1 and Table 2, an interesting point is that the CMAQ-SAPRC results compare more closely than CMAQ-CB4 with the observed decrease in O_3 for the 2004 versus 2002 results, but they do not when comparing 2005 versus 2002 comparison. Fig. 2 depicts the change from 2004 to 2005 in MDA8 O_3 for the $\geqslant$95th% for each site. The CMAQ-SAPRC simulations show similar changes in O_3 from 2004 to 2005 as observed, with increases reaching above 15 ppb under the warmer conditions of 2005. The CMAQ-CB4 simulations show smaller increases in O_3 that underestimate the change by approximately 50% or more at many sites. Since NO_x emissions are very similar for 2004 and 2005 (Table 1), this comparison suggests that the CB4 chemical mechanism does not capture O_3 variations associated with meteorological conditions as well as the SAPRC chemical mechanism. This would explain why the comparison of the CMAQ-SAPRC MDA8 O_3 change predicted from 2002 to 2004 agrees more closely with the observed changes in O_3 than the CB4-based simulations, since a substantial portion of that O_3 reduction is related to the cooler, wetter conditions in 2004.

The above analyses demonstrate how changes in O_3 across years include the influence from changes in both emissions and meteorology. For the dynamic evaluation approach used here, including both 2004 and 2005 summers for the post-NO_x SIP Call impacts provides some information about the influence of the meteorology. The O_3 decreases are clearly larger when comparing summers 2004 and 2002, as opposed to 2005 and 2002, because cooler, wetter conditions in 2004 also lead to lower O_3

Table 2
Average change (2004–2002) and (2005–2002) in maximum daily 8-h for the median and ⩾95th% of the distribution across sites

Observed change (AQS and CASTNet		Modeled change with CB4	Modeled change with SAPRC	Model change with CMAQ v4.6 CB05
(2004–2002)				
	N[a] = 656	$N = 623$	$N = 631$	
Median				
ppb				
Average ± σ	−10.6 ± 4.9 ppb	−5.1 ± 2.3 ppb	−6.7 ± 2.9 ppb	
Ratio[b]		47%	63%	
%				
Average ± σ	−18.3 ± 7.0%	−9.3 ± 4.9%	−10.8 ± 4.3%	
Ratio		51%	59%	
	$N = 647$	$N = 655$	$N = 646$	
⩾95th%				
ppb				
Average ± σ	−19.6 ± 8.7 ppb	−10.7 ± 5.0 ppb	−13.8 ± 6.5 ppb	
Ratio		55%	70%	
%				
Average ± σ	−20.7 ± 8.1%	−13.5 ± 5.7%	−14.9 ± 6.0%	
Ratio		65%	72%	
(2005–2002)				
	$N = 537$	$N = 476$	$N = 439$	$N = 495$
Median				
ppb				
Average ± σ	−7.0 ± 4.9 ppb	−3.5 ± 2.6 ppb	−3.7 ± 2.7 ppb	−3.8 ± 2.3 ppb
Ratio		50%	53%	54%
%				
Average ± σ	−11.9% ± 7.5%	−6.3 ± 4.5%	−5.9 ± 4.3%	−6.4 ± 3.7%
Ratio		53%	50%	54%
	$N = 591$	$N = 594$	$N = 551$	$N = 610$
⩾95th%				
ppb				
Average ± σ	−12.9 ± 7.2 ppb	−6.3 ± 4.0 ppb	−7.1 ± 4.6 ppb	−8.5 ± 4.7 ppb
Ratio		49%	55%	66%
%				
Average ± σ	−13.4 ± 6.9%	−8.0 ± 4.8%	−7.6 ± 4.7%	−9.8 ± 5.0%
Ratio		60%	57%	73%

[a]If O_3 increased at a site as compared to 2002, it was not included in the average shown here.
[b]The ratio of the mean modeled change to the mean observed change.

levels. Given similar meteorological conditions, the comparison between the 2002 and 2005 summer periods better isolates the emission influences on the lower O_3 concentrations. Weighing the results from all three summers, it is very unlikely that the modeled underestimation of O_3 changes between 2002, 2004, and 2005 can be attributed solely to a lack of response to changing meteorological conditions. Instead, it is anticipated that at least a portion of the underestimation of the O_3 change is related to the model's response to emission changes, but other factors must also be considered prior to that conclusion.

One potential factor that could affect the results presented above is model bias during the pre-NO_x SIP Call summer 2002 "base case." Previous evaluations (e.g., Appel et al., 2007) have shown that CMAQ-CB4 substantially underpredicts O_3 concentrations for observed concentrations exceeding about 90 ppb. At many sites, O_3 concentrations ⩾95th% are above 90 ppb in 2002, while summer 2004 and 2005 O_3 levels do not exceed 90 ppb at most sites (See Fig. 2 in Appendix A). It is, therefore, not surprising that the biases in CMAQ-CB4 at ⩾95th% are more severe during summer 2002 than the other two summers (Figs. 3a–c).

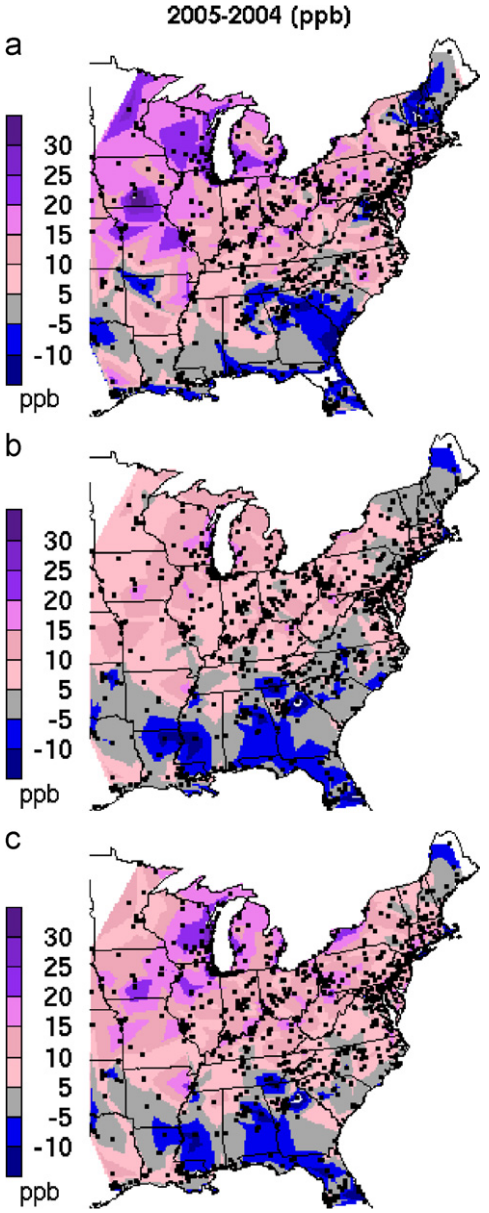

Fig. 2. Ozone changes in ⩾95th% MDA8 O_3 levels (ppb) between summers 2005 and 2004. Comparisons are shown for the O_3 change in the (a) observations, (b) CMAQ-CB4, and (c) CMAQ-SAPRC.The akima package included in the *R* statistical interp function (http://cran.r-project.org/doc/packages/akima.pdf; http://www.r-project.org/) was used to create the contour fields in these figures.

The more severe under-predictions in summer 2002 would contribute to a dampening of the O_3 reductions signal when comparing to summers 2004 and 2005. CMAQ-SAPRC results show a much better agreement with observations at the ⩾95th% with biases not larger than ±10 ppb at most sites within the domain for all summer periods presented (Fig. 3d–f). At the median of the O_3 levels (not shown), CMAQ-CB4 results have very small biases and better model performance than CMAQ-SAPRC, which has some overprediction at median and lower O_3 concentrations for all three summers.

Another factor that could contribute to the model underestimation of O_3 reductions is potential error in the estimated NO_x emission changes from 2002 to 2004 and 2005. Uncertainties should be quite small for EGUs since CEMS data were used, but mobile emission estimates have more uncertainty. EPA (2005) suggests a reduction in NO_x emissions from on-road vehicles from 2002 to 2004 of about 5% per year, which is consistent with the Mobile6 estimates shown in Table 1. While the trends are consistent, we have identified during the analysis that the on-road NO_x emission estimates using the information available in the USEPA NEI 2001 are likely over-estimated because of the lack of county-specific mobile fleet information. Comparisons with the newest USEPA NEI 2002 and 2005 (Table 1) suggest that the on-road NO_x emissions estimated using the NEI 2001 information are approximately 20% higher than the NEI 2002 and 2005 mobile emissions based on more detailed county-specific mobile fleet information. This could impact the model's sensitivity to the EGU emission reductions in areas of dense mobile emissions.

Having looked at the influence of model bias and estimates of NO_x emission changes, the model-predicted change in O_3 is revisited to identify other potential problems. In Figs. 4 and 5, the O_3 changes simulated by the model are spatially compared with observations at the ⩾95th% of the distribution for 2004 and 2005, respectively. Both absolute (ppb) and relative with respect to 2002 (%) O_3 changes are shown. While the change in O_3 from 2002 to 2004 is underestimated more with CMAQ-CB4 than with CMAQ-SAPRC based on the absolute O_3 change in ppb, the relative percentage change is more similar between the two chemical mechanisms for the difference between 2002 and 2004 (Figs. 4b and c versus Figs. 4e and f). This is because SAPRC O_3 predictions are higher than CB4, so that the relative change is more similar. Model underestimates of the O_3 response in 2004 in the Northeast and East coast are more substantial; in addition to the potential influence of over-estimates of mobile NO_x emissions, this could suggest that O_3 changes due to the reduction in transported O_3 and O_3 precursors may not be simulated as well by the model since this area

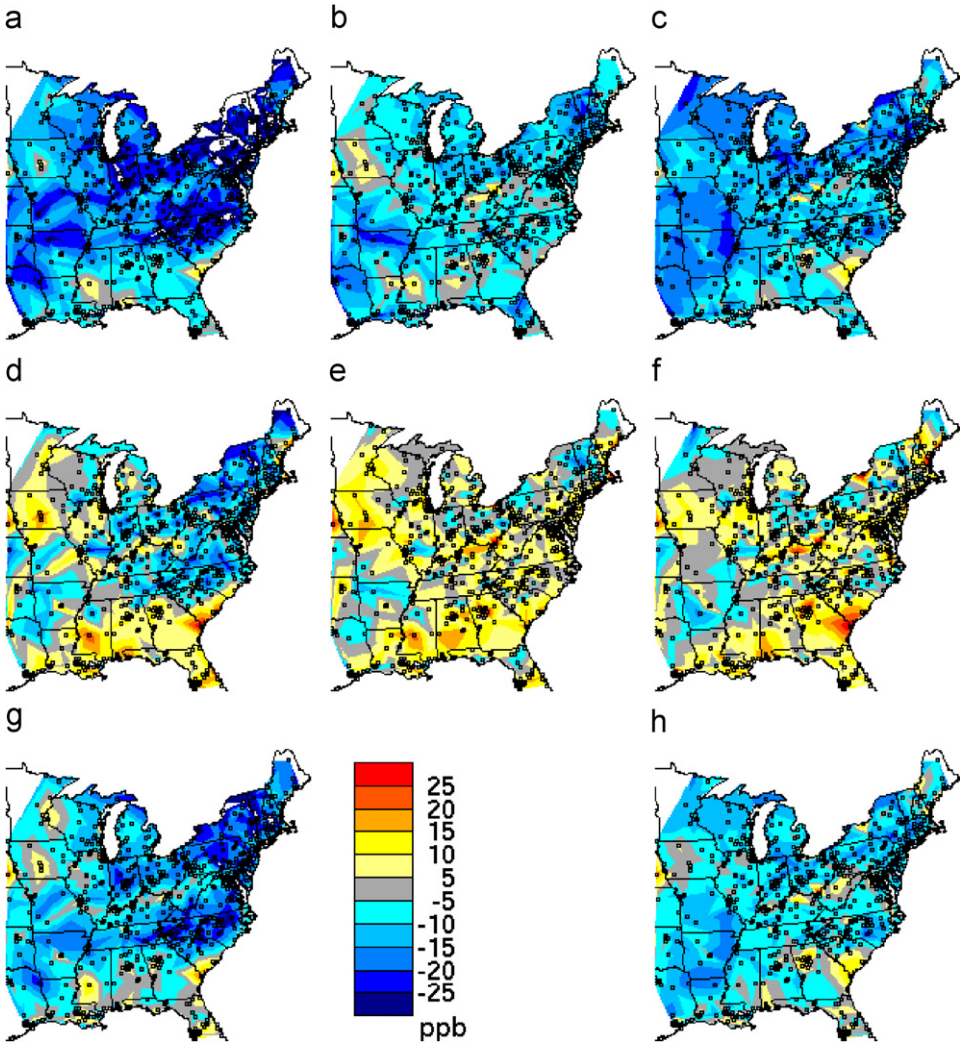

Fig. 3. Comparison of simulated and observed ozone changes at ⩾95th% O_3 mixing ratios for different configurations of the CMAQ model: (a–c) CMAQ v4.5 CB4, (d–f) CMAQ v4.5 SAPRC, (g–h) CMAQ v4.6 CB05. Comparisons are shown for summers 2002 (first column), 2004 (second column), and 2005 (third column).

is downwind of the Ohio River Valley where elevated NO_x sources were substantially reduced. These results are consistent with Kim et al. (2006) summer 2004 results using WRF-Chem, which also do not show O_3 decreases in the Northeast after EGU NO_x emissions were reduced to be in agreement with the NO_x SIP Call.

To more closely assess the model's response to transported NO_x emissions from elevated EGU sources, we estimate the spatial correlation structures of the synoptic forcing imbedded in the O_3 time series (i.e., weather-induced variations in O_3). As described in Godowitch et al. (2008b), the spatial correlation structures are estimated by determining the correlation coefficients between spectrally decomposed time series of hourly O_3 at various monitoring stations in the modeling domain. The e-folding distance is then defined as the distance at which the correlation between synoptic-scale O_3 time series drops below 0.37. In our analysis, the synoptic-scale component was estimated from hourly observed and modeled time series through the use of the Kolmogorov–Zurbenko (KZ) moving average filter (Rao et al., 1997; Hogrefe et al., 2000). Fig. 6a–c shows the e-folding distance for AQS observations and the corresponding CMAQ-CB4, CMAQ-SAPRC, and CMAQ-CB05 simulations for 2002, 2004 and 2005, grouped by the density of total non-EGU NO_x emissions. This grouping was chosen to identify whether the e-folding distance

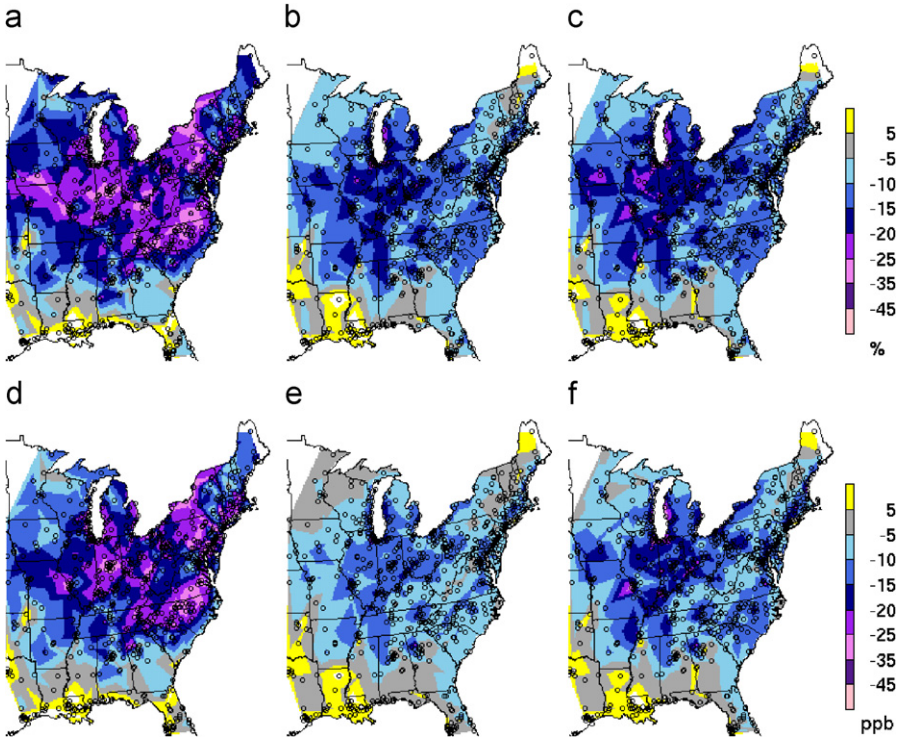

Fig. 4. Comparison of the relative change in O_3 at $\geqslant$95th% between 2004 and 2002 summers with respect to 2002 for (a) observations, (b) CMAQ v4.5 CB4, and (c) CMAQ v4.5 SAPRC. The absolute changes are shown in (d–f).

differs for areas with varying amounts of ground-level NO_x emissions from low-level sources. The e-folding distances in modeled O_3 concentrations are typically shorter than those observed, which implies that sources nearer the monitor are having more effect and transported emissions have less effect on modeled O_3 fluctuations than what observations suggest. This underprediction of spatial scales is more pronounced for CMAQ-CB4 than CMAQ-SAPRC and more pronounced for areas with highest (see TOP10 in Fig. 6) ground-level NO_x emissions. These results lend support to the discussion above that the contributions of precursor emissions and O_3 transport are underestimated.

Since these findings suggest that the influence of regional transport of O_3 and NO_x from aloft EGU emission sources could also play a role in the underestimated O_3 change, the model results are compared to aircraft observations available from the summer 2004 ICARTT field experiment (Neuman et al., 2006). In Fig. 7, we compare modeled concentrations of NO_x and O_3 to observations from daytime NOAA P-3 flights downwind of the Ohio River Valley during. Above approximately 2 km, both CMAQ-CB4 and CMAQ-SAPRC underpredict these concentrations as well as other relevant species (Yu et al., 2007). Possible explanations for the underpredictions from the P-3 flights include excessive NO_x termination, insufficient NO_x recycling from PAN and isoprene nitrates, excessive NO_x and O_3 scavenging and deposition, insufficient non-EGU sources aloft (i.e. lightning, aircraft), and insufficient contributions from boundary conditions. Preliminary sensitivity tests with lightning NO_x emissions do not improve this bias; hence, closer inspection of the chemistry in the mid- and upper-troposphere is needed to see if the chemical mechanisms are sufficiently accounting for the contributions of regional-scale transport of O_3 and its precursors to surface O_3 downwind.

The results presented for CMAQ-CB4 and CMAQ-SAPRC suggest that biases in the mobile on-road NO_x emissions and the long range transport of O_3 and O_3 precursors could both impact the underestimation of observed O_3 decreases after the NO_x SIP Call was implemented. To address the possibility that recent updates could have a positive impact on the results, we repeated our test with the most recent updates to CMAQ, the

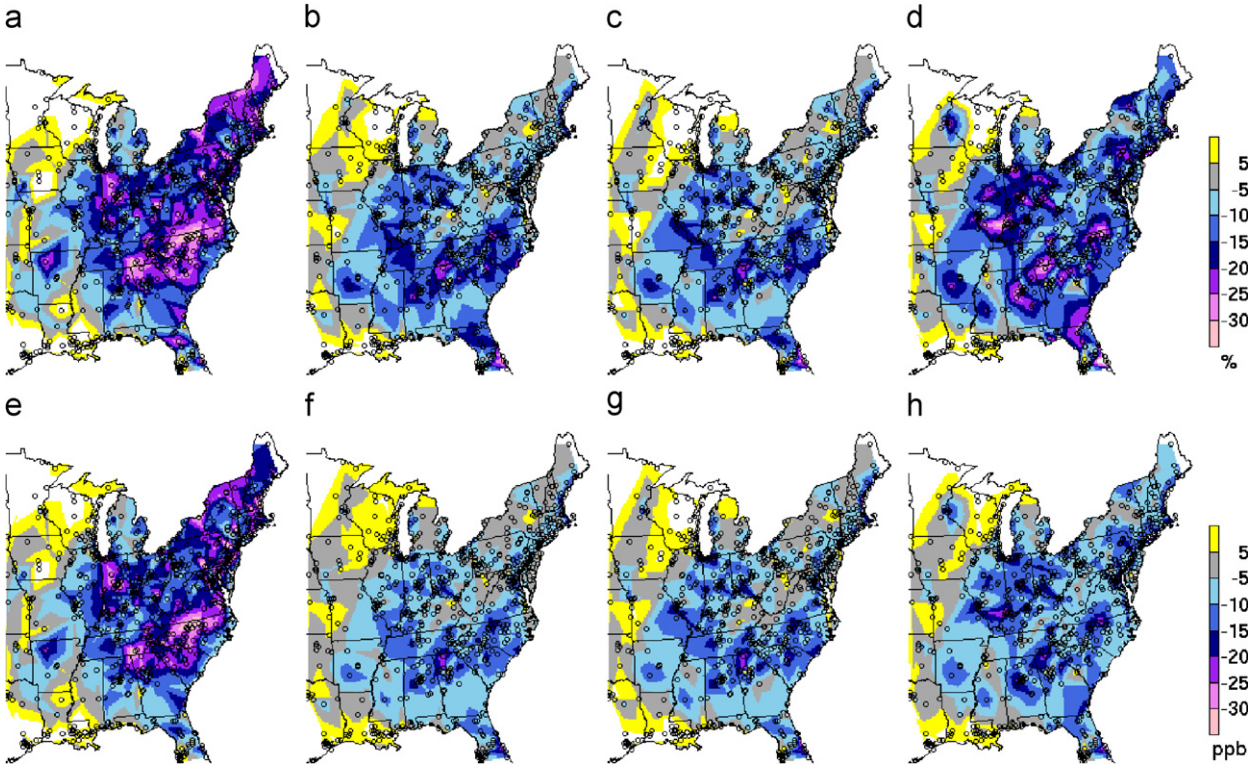

Fig. 5. Comparison of the relative change in O_3 at $\geqslant$95th% between 2005 and 2002 summers with respect to 2002 for (a) observations, (b) CMAQ v4.5 CB4, and (c) CMAQ v4.5 SAPRC, and (d) CMAQ v4.6 CB05 chemical mechanism with the updated 2002 and 2005 NEI. The absolute changes are shown in (e–h).

chemical mechanism CB05, and mobile emission inventories that are included in the NEI 2002 and 2005. The results do suggest incremental improvements in the ozone reductions from 2002 to 2005, especially at the $\geqslant$95th% of the ozone levels (Fig. 1, Table 2). The CMAQ-CB05 O_3 reduction from 2002 to 2005 is approximately 73% of observed relative (65% of the absolute) O_3 changes at the $\geqslant$95th%. Comparing Fig. 5b–d, CMAQ-CB05 predicted O_3 change from 2002 to 2005 could be showing better comparison to observations in the Northeastern corridor because of improved estimates of the dense mobile emissions for 2002 and 2005. The lower mobile NO_x emission estimates could also play a role here because the transported EGU emissions would then be a larger contribution to the overall NO_x emission budget. An additional reason could be that CB05 better accounts for the contribution of transported O_3 and precursors; however, Fig. 6 suggests similar results for SAPRC and CB05 regarding the e-folding distance and the estimated contribution of transported versus local emissions. Regardless, the newest available modeling and emissions platform does appear to provide more sensitivity to the emission changes at the $\geqslant$95th%.

6. Summary

The study presented here introduces a new challenge to model performance evaluation, where we illustrate the need to evaluate the model's O_3 response to emission changes. This approach, referred to here as the dynamic evaluation, is complex because events must be identified where emission changes can be quantified with reasonable confidence and where a change in air quality in response to these emission changes is evident. The time period surrounding the NO_x SIP Call is an ideal case study because the largest emission reductions were in the EGU utility sector that has direct measurement of emissions using CEMS. The model-predictioned sensitivity to chemical mechanisms and varying meteorology were assessed by using the CB4, SAPRC, and CB05 chemical mechanisms during the summers of 2002, 2004, and 2005.

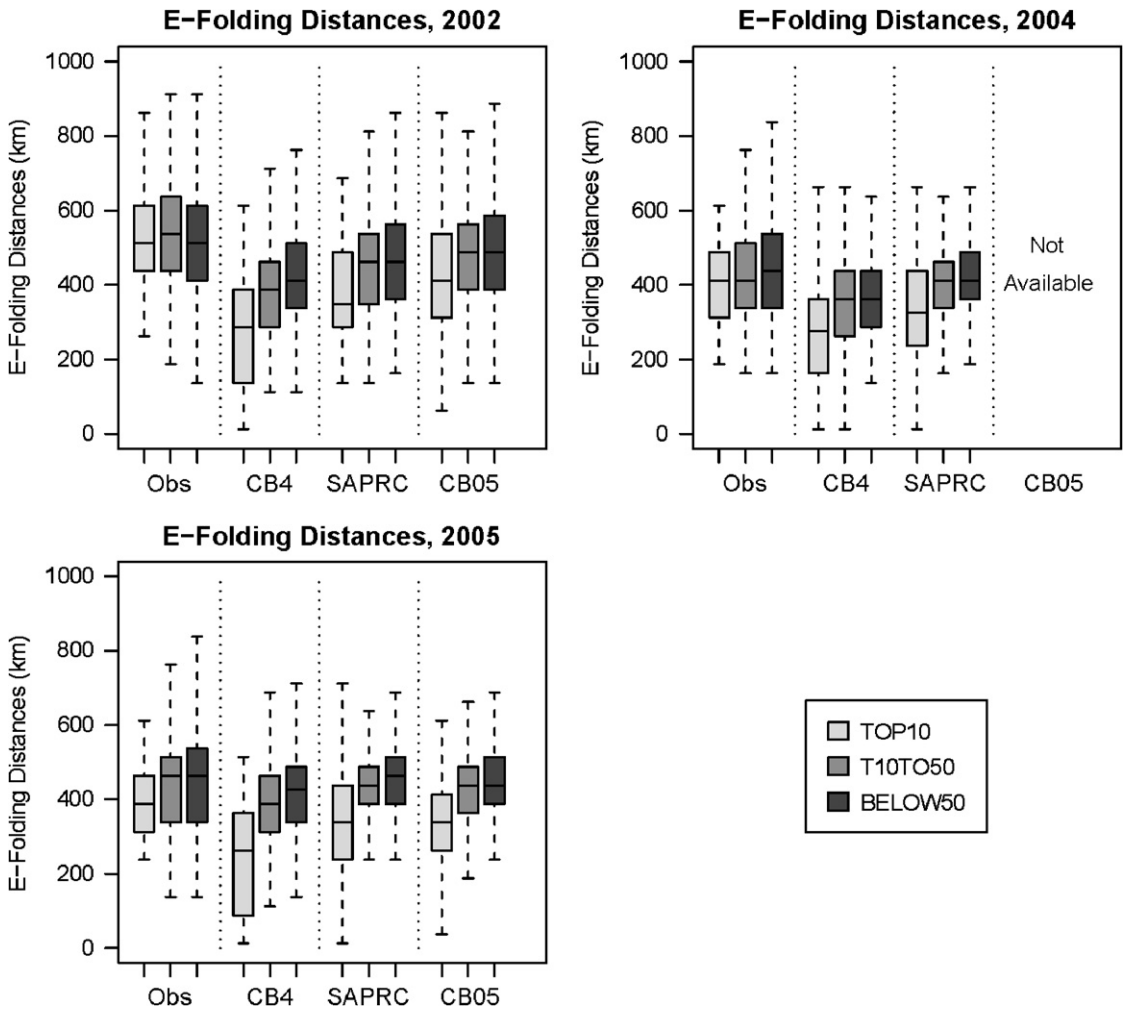

Fig. 6. e-Folding distance extracted from AQS observations and from CMAQ-CB4 and CMAQ-SAPRC simulations grouped by the density of total non-EGU NO_x emissions for summers 2002, 2004, and 2005. "TOP10" indicates locations where non-EGU total NOx emissions exceeded the 90th percentile of emissions at all locations. "T10TO50" indicates locations with non-EGU total NO_x emissions between the 50th and 90th percentiles, and "BELOW50" indicates locations with below-median non-EGU total NO_x emissions.

The evaluation results presented in this study suggest that the air quality model predictions underestimate the O_3 reductions observed after the NO_x SIP Call was implemented. The spatial correlation analysis and comparison with observations aloft suggest that the model underestimates the contribution of long-range transport of O_3 and precursors, especially when using the CB4 chemical mechanism. Further investigation of the chemical mechanisms' ability to characterize tropospheric chemistry aloft is recommended. The results also demonstrate that the SAPRC chemical mechanism better simulates O_3 changes associated with meteorology than CB4. The most recent CMAQ version 4.6, including the new CB05 chemical mechanism and updated emission inventories with improved mobile source estimates, collectively, shows incremental improvements to the modeled O_3 response to NO_x emission reductions from 2002 to 2005. Continued improvements in the 2005 NEI, such as the non-road mobile sector, may show additional incremental improvements.

This study demonstrates both the challenges and value in dynamic evaluation approaches. Results highlight the influence chemical mechanisms can have on a model's ability to correctly respond to these changes. Dynamic evaluation approaches critically depend on the best estimates of emissions and emission trends for air quality model evaluation. As new emission control programs are

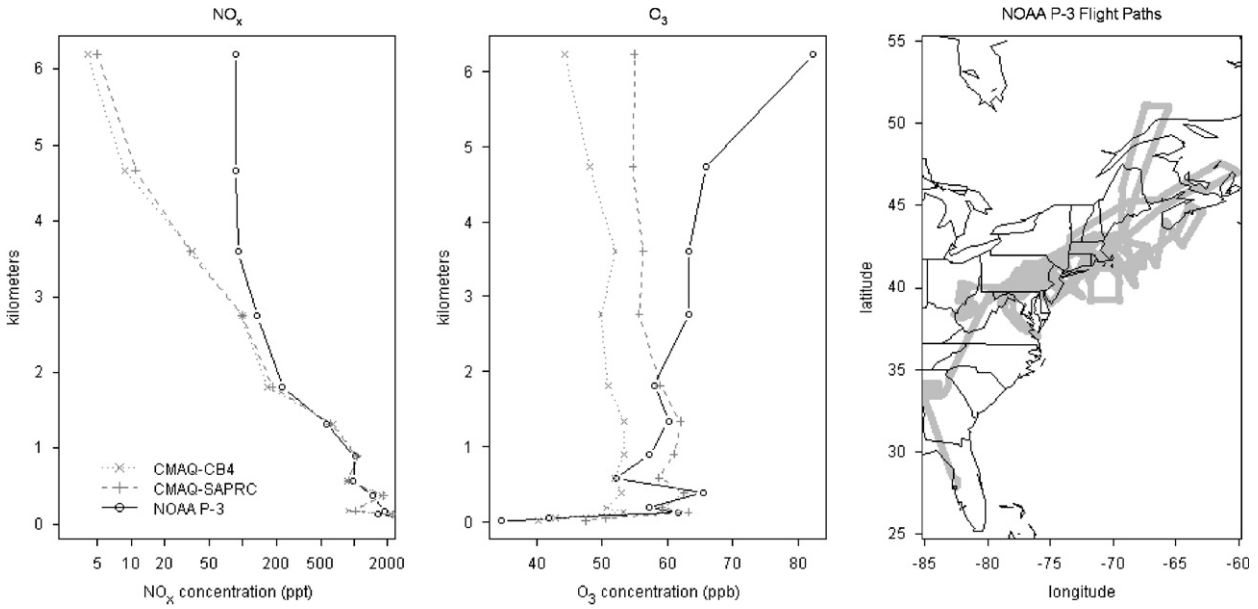

Fig. 7. For observations aloft averaged across all daytime NOAA P-3 flights, CMAQ-CB4 and CMAQ-SAPRC underpredict ozone and NO_x concentrations above 2 km. A more complete comparison is available in Yu et al. (2007).

implemented in the future, dynamic evaluation studies such as this should continue as a critical component of comprehensive evaluation of air quality models.

Acknowledgements

The authors would like to thank Charles Chang and Lucille Bender at Computer Sciences Corporation for their assistance in developing the emission scenarios and model simulations for this study. Additionally, our thanks go to Brian Timin and Norm Possiel from the U.S. Environmental Protection Agency (EPA) for comments and suggestions. We also thank Tom Ryerson and the ICARTT P-3 measurement team for providing data collected aloft during the summer of 2004. The research presented here was performed under the Memorandum of Understanding between the U.S. EPA and the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) and under agreement number DW13921548. This work constitutes a contribution to the NOAA Air Quality Program. Although it has been reviewed by EPA and NOAA and approved for publication, it does not necessarily reflect their policies or views. One of the authors (C. Hogrefe) gratefully acknowledges partial support for the work presented in this manuscript through a research fellowship from the Oak Ridge Institute for Science and Education (ORISE).

Appendix A. Supplementary materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.atmosenv.2008.02.018

References

Appel, K.W., Gilliland, A.B., Sarwar, G., Gilliam, R., 2007. Evaluation of the community multiscale air quality (CMAQ) model version 4.5: sensitivities impacting model performance; part I—ozone. Atmospheric Environment 41, 9603–9615.

Arnold, J.R., Dennis, R.L., 2006. Testing CMAQ chemistry sensitivities in base case and emissions control runs at SEARCH and SOS99 surface sites in the southeastern US. Atmospheric Environment 40, 5027–5040.

Byun, D., Schere, K.L., 2006. Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (CMAQ) modeling system. Applied Mechanics Reviews 55, 51–77.

Camalier, L., Cox, B., Dolwick, P., 2007. The effects of meteorology on O_3 in urban areas and their use in assessing O_3 trends. Atmospheric Environment 41, 7127–7137.

Carter, W.P.L., 2000. Documentation of the SAPRC-99 chemical mechanism for VOC reactivity assessment. Final Report to the California Air Resources Board under Contracts 92-329 and 95-308, May (Refereed). 〈http://www.arb.ca.gov/research/abstracts/92-329.htm#Pg〉.

Cox, W.M., Chu, S.H., 1993. Meteorologically adjusted ozone trends in urban areas: a probabilistic approach. Atmospheric Environment. Part B. Urban Atmosphere 27, 425–434.

Frost, G.J., et al., 2006. Effects of changing power plant NO_x emissions on O_3 in the eastern United States: proof of concept. Journal of Geophysical Research 111, D12306.

Gégo, E., Porter, P.S., Gilliland, A., Rao, S.T., 2007. Observation-based assessment of the impact of nitrogen oxides emissions reductions on O_3 air quality over the eastern United States. Journal of Applied Meteorology and Climatology 46, 994–1008.

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1989. A photochemical mechanism for urban and regional scale computer modeling. Journal of Geophysical Research 94, 12925.

Godowitch., J.M., Gilliland, A.B., Draxler, R.R., Rao, S.T., 2008a. Modeling assessment of point source NO_x emission reductions on ozone air quality in the eastern United States. Atmospheric Environment 42, 87–100.

Godowitch, J.M., Hogrefe, C., Rao, S.T., 2008b. Influence of point source NO_x emission reductions on modeled processes governing ozone concentrations and chemical/transport indicators. Journal of Geophysical Research in review.

Grell, G.A., Dudhia, J., Stauffer, D.R., 1994. A description of the 5th generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note, NCAR/TN−398+STR, Boulder, CO.

Hogrefe, C., Rao, S.T., Zurbenko, I.G., Porter, P.S., 2000. Interpreting the information in O_3 observations and model predictions relevant to regulatory policies in the eastern United States. Bulletin of the American Meteorological Society 81, 2083–2106.

Hudman, R.C., Jacob, D.J., Turquety, S., Leibensperger, E.M., Murray, L.T., Wu, S., Gilliland, A.B., Avery, M., Bertram, T.H., Brune, W., Cohen, R.C., Dibb, J.E., Flocke, F.M., Fried, A., Holloway, J., Neuman, J.A., Orville, R., Perring, A., Ren, X., Sachse, G.W., Singh, H.B., Swanson, A., Wooldridge, P.J., 2007. Surface and lightning sources of nitrogen oxides over the United States: magnitudes, chemical evolution, and outflow. Journal of Geophysical Research 112, doi:10.1029/2006JD007912.

Jimenez, P., Baldasano, J.M., Dabdub, D., 2003. Comparison of photochemical mechanisms for air quality modeling. Atmospheric Environment 37, 4179–4194.

Kim, S.-W., Heckel, A., McKeen, S.A., Frost, G.J., Hsie, E.-Y., Trainer, M.K., Richter, A., Burrows, J.P., Peckham, S.E., Grell, G.A., 2006. Satellite-observed U.S. power plant NO_x emission reductions and their impact on air quality. Geophysical Research Letters 33, doi:10.1029/2006GL027749.

Neuman, J.A., Parrish, D.D., Trainer, M., Ryerson, T.B., Holloway, J.S., Nowak, J.B., Swanson, A., Flocke, F., Roberts, J.M., Brown, S.S., Stark, H., Sommariva, R., Stohl, A., Peltier, R., Weber, R., Wollny, A.G., Sueper, D.T., Hubler, G., Fehsenfeld, F.C., 2006. Reactive nitrogen transport and photochemistry in urban plumes over the north Atlantic ocean. Journal of Geophysical Research 111, doi:10.1029/2005JD007010.

Rao, S.T., Zurbenko, I.G., Neagu, R., Porter, P.S., Ku, J.Y., Henry, R.F., 1997. Space and time scales in ambient O_3 data, 1997. Bulletin of the American Meteorological Society 78, 2153–2166.

Sarwar, G., Luecken, D.J., Yarwood, G., Whitten, G.D., Carter, W.P., 2008. Impact of an updated carbon bond mechanisms on predictions from the Community Multiscale Air Quality (CMAQ) Modeling System: preliminary assessment. Journal of Applied Meteorology and Climatology 47, doi:10.1175/2007JAMC1393.1.

Schwede, D., Pouliot, G., Pierce, T., 2005. Changes to the Biogenic Emissions Inventory System Version 3 (BEIS3). 4th CMAS Models-3 Users' Conference, Chapel Hill, NC, 26–28 September 2005. Available at 〈http://www.cmascenter.org/conference/2005/abstracts/2_7.pdf〉.

USEPA, 2005. Evaluating O_3 control programs in the Eastern United States: Focus on the NO_x Budget Trading Program, EPA-454-K-05-001. 〈http://www.epa.gov/airtrends/2005/O3nbp.pdf〉.

USEPA, 2006. NO_x Budget Trading Program: 2005 program compliance and environmental results, EPA430-R-06-013. 〈http://www.epa.gov/air/airtrends/2006/O3nbp/〉.

Yu, S., Mathur, R., Schere, K., Kang, D., Pleim, J., Otte, T., 2007. A detailed evaluation of the Eta-CMAQ forecast model performance for O_3, its related precursors, and meteorological parameters during the 2004 ICARTT study. Journal of Geophysical Research 112, doi:10.1029/2006JD007715.

Zheng, J., Swall, J.L., Cox, W.M., Davis, J.M., 2007. Interannual variation in meteorologically adjusted ozone levels in the eastern United States: a comparison of two approaches. Atmospheric Environment 41, 705–716.