Available online at www.sciencedirect.com
SCIENCE DIRECT
Atmospheric Environment 40 (2006) 3041–3056
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Spatial variability of summertime tropospheric ozone over the continental United States: Implications of an evaluation of the CMAQ model

Daniel Q. Tong[a,*], Denise L. Mauzerall[a,b,**]

[a]*Science, Technology and Environmental Policy Program, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA*
[b]*Geosciences Department, Princeton University, USA*

Received 25 July 2005; received in revised form 23 November 2005; accepted 23 November 2005

Abstract

This study evaluates the ability of the Community Multiscale Air Quality (CMAQ) model to simulate the spatial variability of summertime ozone (O_3) at the surface and in the free troposphere over the continental United States. Simulated surface O_3 concentrations are compared with 987 Air Quality System (AQS) sites and 123 Clean Air Status and Trends Network (CASTNet) sites. CMAQ's ability to reproduce surface observations varies with O_3 concentration. The model best simulates observed O_3 for intermediate concentrations (40–60 ppbv), while over-(under-) predicting at lower (higher) levels. CMAQ reproduces surface O_3 for a wide range of conditions (30–80 ppbv) with a normalized mean error (NME) less than 35% and normalized mean bias (NMB) lying between $\pm 15\%$ for the whole domain. Although systematically over-predicting O_3 in the east and under-predicting it in the western United States, CMAQ is able to reproduce 1- and 8-h daily maxima with a cross-domain mean bias (MB) of 1 and 8 ppbv, or NMB of 8% and 25%, respectively. The model underestimates observed O_3 at rural sites (MB $= -5$ ppbv, NMB $= -5\%$ and NME $= 23\%$ with a 40 ppbv cut-off value) and over-predicts it at urban and suburban sites by a similar magnitude (MB $= 6$ ppbv, NMB $= 7\%$ and NME $= 25\%$). Apparent errors and biases decrease when data is averaged over longer periods, suggesting that most evaluation statistics are dependent on the time scale of data aggregation. Therefore, performance criteria should specify an averaging period (e.g., 1- or 8- h) and not be independent of averaging period as some current model evaluation studies imply. Comparisons of vertical profiles of simulated O_3 with ozonesonde data show both overestimation and underestimation by 10–20 ppbv in the lower troposphere and a consistent under-prediction in the upper troposphere. Vertical O_3 distributions are better simulated when lateral boundary conditions obtained from the global Model of Ozone and Related Tracers version 2 (MOZART-2) are used, but under-prediction remains. The assumption of zero-flux at the top boundary and the resulting exclusion of the contribution of stratosphere–troposphere exchange to tropospheric O_3 concentrations limit the ability of CMAQ to reproduce O_3 concentrations in the upper troposphere.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: CMAQ; Model evaluation; Ozone; Ozonesonde; Vertical profile; Air quality model

*Corresponding author. Tel.: +1 919 541 5150.
**Also corresponding author. Tel.: +1 609 258 2439.
E-mail addresses: tong.daniel@epa.gov (D.Q. Tong), mauzeral@princeton.edu (D.L. Mauzerall).

1352-2310/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2005.11.058

1. Introduction

Tropospheric ozone (O_3) continues to be a major concern for both the research and regulatory communities due to its importance in atmospheric chemistry (Logan et al., 1981) and climate change (Ramaswamy, 2001), as well as its harmful effects on public health (EPA, 1999; Bell et al., 2005; Ito et al., 2005; Levy et al., 2005), agriculture and ecosystems (Mauzerall and Wang, 2001; NRC, 2004). In an effort to protect public health and welfare, in 1971, the US Environmental Protection Agency (EPA) set the National Ambient Air Quality Standards (NAAQS) for O_3 at 0.12 ppmv for a 1-h daily maximum value. In 1997, EPA revised the O_3 NAAQS to require that a 3-year average of the annual fourth highest daily 8-h average concentration not exceed 0.08 ppmv. Despite these regulatory measures, there has been little progress in reducing nationwide O_3 levels in recent years (EPA, 2000; Lin et al., 2001). Between 1989 and 1998, there was only a 4% decrease in the national second highest daily maximum 1-h O_3 value and "no change" in the fourth highest daily maximum 8-h average (EPA, 2000). With the exception of the southwest, exceedances of both 1- and 8-h NAAQS have not significantly changed in the past decade, suggesting that air quality gains from the 1980s to 1990s have not continued (Lin et al., 2001). Compliance with the O_3 NAAQS within any state is complicated by inter-state transport. Long-range transport of O_3 and its precursors from upwind areas may contribute significantly to persistent nonattainment of the O_3 standards at many locations. The need to address inter-state transport of O_3 and its precursors has led to the formation of multi-state organizations to mitigate air pollution, including the Ozone Transport Commission (OTC) and the Ozone Transport Assessment Group (OTAG) (NRC, 2004), and most recently the Clean Air Interstate Rule (CAIR) in 2005 which is designed to reduce NO_x and SO_2 emissions in the eastern United States by capping regional emissions.

Air quality models are valuable tools for simulating the spatial distribution of O_3 concentrations, resulting from both local emissions and long-range transport of O_3 precursors, in locations where measurements are not available. In addition, they allow an exploration of emission control strategies to mitigate elevated concentrations of O_3, which can help states attain the NAAQS. The EPA's Models-3 Community Multi-scale Air Quality (CMAQ) model (Byun and Ching, 1999) has increasingly been used by both the regulatory and scientific communities to investigate these issues. However, evaluation of CMAQ's ability to reproduce the spatial distribution of surface O_3 is needed to assist interpretation of simulation results.

This paper evaluates the ability of CMAQ to reproduce O_3 measurements at the surface and in the troposphere over the continental United States. Most previous evaluations of CMAQ performance have focused on a single geographic area, including the eastern US (Hogrefe et al., 2001, 2004), the mid-Atlantic states (Bell et al., 2003), and the southeastern US (Kang et al., 2003). While informative, these studies have not analyzed the spatial variability of O_3 across the entire country. This paper, together with other emerging papers (e.g., Eder and Yu, 2005), evaluates model performance over the entire continental US in a consistent fashion. Our study focuses on CMAQ's ability to: (1) capture the spatial variability in surface O_3 concentrations, (2) accurately simulate O_3 concentrations in both rural and urban locations, and (3) reproduce high O_3 concentrations during pollution episodes as well as low concentrations in clean air. Furthermore, we (4) examine the influence that averaging time periods have on apparent model performance. Finally, we (5) evaluate the vertical distribution of tropospheric O_3 simulated by CMAQ and assess the impacts of lateral boundary conditions (LBC) on simulated O_3 concentrations at the surface and in the free troposphere using both a predefined LBC and an LBC obtained from the global Model of Ozone and Related Tracers version 2 (MOZART-2).

We describe the CMAQ model, measurements and methods used for model evaluation in Section 2. Evaluation results are presented in Section 3 and their implications for CMAQ applications are discussed in Section 4. We conclude in Section 5.

2. Model, observations and methods

2.1. CMAQ simulation

We use the CMAQ model (Dennis et al., 1996; Byun and Ching, 1999) version 4.2 to simulate the transport and chemical transformation of tropospheric O_3 and its precursors for July 1996 over the continental United States. CMAQ is a comprehensive Eulerian grid model that simulates the complex interactions among multiple atmospheric pollutants

and between regional and urban scales (Dennis et al., 1996). In this study the model is configured to include detailed implementation of horizontal and vertical advection, turbulent diffusion based on K-theory, chemistry in the gas, liquid, and particulate phases using a modified version of the CBM-IV chemical mechanism (Gery et al., 1994), dry deposition, and cloud physics and chemistry (Byun and Ching, 1999). Although at the time of this evaluation, a newer version of CMAQ was available, it was not suitable for our intended applications. Notable numerical instabilities were found under certain conditions in the newer version (Tong, 2004), possibly due to discontinuous changes in the partitioning of gaseous HNO_3 and NO_3 aerosol that affects the concentrations of OH and NO_2, which in turn affects simulated O_3 concentrations (Tong and Mauzerall, 2005).

Meteorological parameters, including wind speeds and directions, temperature, humidity, pressure, and solar radiation, are obtained from the fifth generation Mesoscale Model (MM5) developed by Pennsylvania State University (PSU) and the National Center for Atmospheric Research (NCAR) (Grell et al., 1994). Hourly meteorological data from MM5 is converted into CMAQ compatible form using version 2.2 of the Meteorology-Chemistry Interface Processor (MCIP). Anthropogenic emissions of nitrogen oxides (NO_x), volatile organic compounds (VOCs), sulfur dioxide (SO_2), carbon monoxide (CO) and ammonia (NH_3) are based on the 1996 county level US EPA National Emissions Trends inventory and processed by the Sparse Matrix Operator Kernel Emissions (SMOKE) model (Carolina Environmental Program, 2003). Mobile emissions of NO_x, VOCs, CO and primary particulate matter (PM) are prepared using the vehicle emission model MOBILE5 (EPA, 2003). Biogenic emissions, including NO_x and VOCs, are obtained using the Biogenic Emissions Inventory System, version 3 (BEIS3) (Pierce, 2001). BEIS3 accounts for variability in temperature and solar radiation when estimating biogenic emissions. NO_x formation by lightning and emissions from biomass burning are not included in the modeling system.

The model domain covers the 48 contiguous US states and parts of southern Canada and northern Mexico, with western and eastern borders over the Pacific and Atlantic Oceans. The horizontal domain has a grid resolution of $36 \times 36\,km^2$ (132 columns by 90 rows) with 12 vertical layers extending from the surface to approximately 15 km. Vertical layers are unevenly distributed with six layers in the lowest kilometer, and a surface layer of approximately 38 m. Our CMAQ simulations are conducted from 1 July to 31 July 1996. To minimize the effect of initial conditions, the first 3 days of the simulation are discarded. A 3-day spin-up period has been shown to virtually eliminate the effect of initial conditions on O_3 distributions (Berge et al., 2001; Hogrefe et al., 2004).

Chemical concentrations are specified at the lateral boundaries of the domain using mean July concentrations from a simulation of the global Model of Ozone and Related Chemical Tracers, Version 2 (MOZART-2) (Horowitz et al., 2003). MOZART-2 simulates the distribution of 63 chemical species in the troposphere with a horizontal resolution of 2.8° latitude $\times$ 2.8° longitude and 34 hybrid vertical levels extending from the surface to 4 mbar, using a time step of 20 min for all chemical and transport processes. Surface emissions of chemical species in MOZART-2 include those from fossil fuel burning and other industrial activities, biomass burning, biogenic emissions from vegetation and soils, and oceanic emissions, and are intended to be representative of emissions in the early 1990s. Meteorological variables are obtained from the middle atmosphere version of the Community Climate Model (MACCM3). A full description and evaluation of the version of MOZART-2 used here can be found in Horowitz et al. (2003). The chemical species for which we extracted concentrations from MOZART-2 include both gas species (i.e. O_3, NO_2, NO, NO_3, OH, N_2O_5, HNO_3, H_2O_2, CO, SO_2, PAN, formaldehyde, high molecular weight aldehydes, isoprene, ethene, and NH_3) and particulate sulfate and organic carbon.

CMAQ assumes zero-flux at the top of the model domain (approximately 200 mb). Therefore, the contribution of stratosphere–troposphere exchange (STE) to tropospheric O_3 distributions is not included in this CMAQ simulation.

2.2. Observations

We compare simulated O_3 concentrations with hourly surface O_3 measurements obtained from 987 Air Quality System (AQS) (previously AIRS) and 123 Clean Air Status and Trends Network (CASTNet) monitors. The AQS network collects ambient air pollution data from monitoring stations located in urban, suburban and rural areas. CASTNet sites

are primarily in rural locations and within national parks and monuments, and are intended to provide insight into background levels of pollutants where urban influences are minimal.

Observations of O_3 vertical distributions are obtained from multi-year ozonesonde measurements at Boulder, Colorado and Wallops Island, Virginia, as compiled by Logan et al. (1999). Measurements of vertical O_3 profiles are sparse and these are the only two stations active in 1996 in the continental United States. In July 1996, five and six launches were made at the Boulder and Wallops Island sites, respectively. We compare both the July 1996 monthly mean and a multi-year (1985–2000) mean July profile of sonde measurements with CMAQ predictions.

2.3. Statistical analysis method

A variety of procedures have been used to evaluate the performance of air quality models (EPA, 1991; Russell and Dennis, 2000; Hogrefe et al., 2001, 2004; Fiore et al., 2003; Fuentes, 2003). Our evaluation focuses on CMAQ's ability to reproduce spatial variability in O_3 at the surface and in the free troposphere. We therefore first compare observed and simulated surface O_3 concentrations following the statistical approach of Russell and Dennis (2000). Although no single set of evaluation techniques is universally recommended, the statistical measures listed in Table 1 have been widely used in recent regional air quality model evaluations (Hogrefe et al., 2001, 2004; Kang et al., 2003; Bell et al., 2003; Eder and Yu, 2005).

Table 1
Definitions of statistical measures (EPA, 1991)

Measures	Definition
Mean bias (MB)	$\frac{1}{N}\sum_{i=1}^{N}(C_{\mathrm{mod}}(i)-C_{\mathrm{obs}}(i))$
Mean error (ME)	$\frac{1}{N}\sum_{i=1}^{N}\lvert C_{\mathrm{mod}}(i)-C_{\mathrm{obs}}(i)\rvert$
Normalized mean bias (NMB)	$\frac{1}{N}\frac{\sum_{i=1}^{N}C_{\mathrm{mod}}(i)-C_{\mathrm{obs}}(i)}{\sum_{i=1}^{N}C_{\mathrm{obs}}(i)}\times 100\%$
Normalized mean error (NME)	$\frac{1}{N}\frac{\sum_{i=1}^{N}\lvert C_{\mathrm{mod}}(i)-C_{\mathrm{obs}}(i)\rvert}{\sum_{i=1}^{N}C_{\mathrm{obs}}(i)\rvert}\times 100\%$
Unpaired peak prediction accuracy (UPA)	$\frac{C_{\mathrm{mod}}(i)_{\max}-C_{\mathrm{obs}}(j)_{\max}}{C_{\mathrm{obs}}(j)_{\max}}\times 100\%$

N, the number of existing observation–prediction pairs at a monitoring site; C_{mod} and C_{obs}, the modeled and observed O_3 concentrations; i, the ith model–observation pair; j, the jth model–observation pair; and $C(i)_{\max}$, the maximum simulated or observed O_3 concentration.

hours in July 1996 and for at least 75% of possible measurements over each examined averaging period.

Model performance for the entire continental US is obtained by calculating aggregate statistics from individual site statistics:

$$D_{\mathrm{total}}=\sum_{i=1}^{M}(N_i * D_i)/\sum_{i=1}^{M}N_i, \tag{1}$$

$$\mathrm{UPA}_{\mathrm{total}}=\frac{\max(C_{\mathrm{mod}}(1)_{\max},\ldots C_{\mathrm{mod}}(M)_{\max})-\max(C_{\mathrm{obs}}(1)_{\max},\ldots,C_{\mathrm{obs}}(M)_{\max})}{\max(C_{\mathrm{obs}}(1)_{\max},\ldots,C_{\mathrm{obs}}(M)_{\max})} \tag{2}$$

We calculate these statistics differently than previous analyses in order to examine spatial variability in model performance. Instead of obtaining a single value across the entire modeling domain, we calculate the statistics in Table 1 for both individual monitoring sites and individual model grid boxes (where we aggregate monitors within a grid box). Examining statistical metrics at different locations provides a useful measure of spatial variability in model performance. For inclusion in our analysis, a monitor must have reported measurements for at least 75% of the

Here D is a statistical measure, such as mean bias (MB), mean error (ME), normalized mean bias (NMB), or normalized mean error (NME) (see Table 1 for definitions), weighted by the number of observation–prediction pairs; M is the number of monitoring sites, and N_i is the number of data pairs at site i; UPA is unpaired peak prediction accuracy as defined in Table 1; C_{mod} and C_{obs} are O_3 concentrations calculated by CMAQ and observed at a monitoring site, respectively; and max indicates the maximum value at a monitoring site for the period.

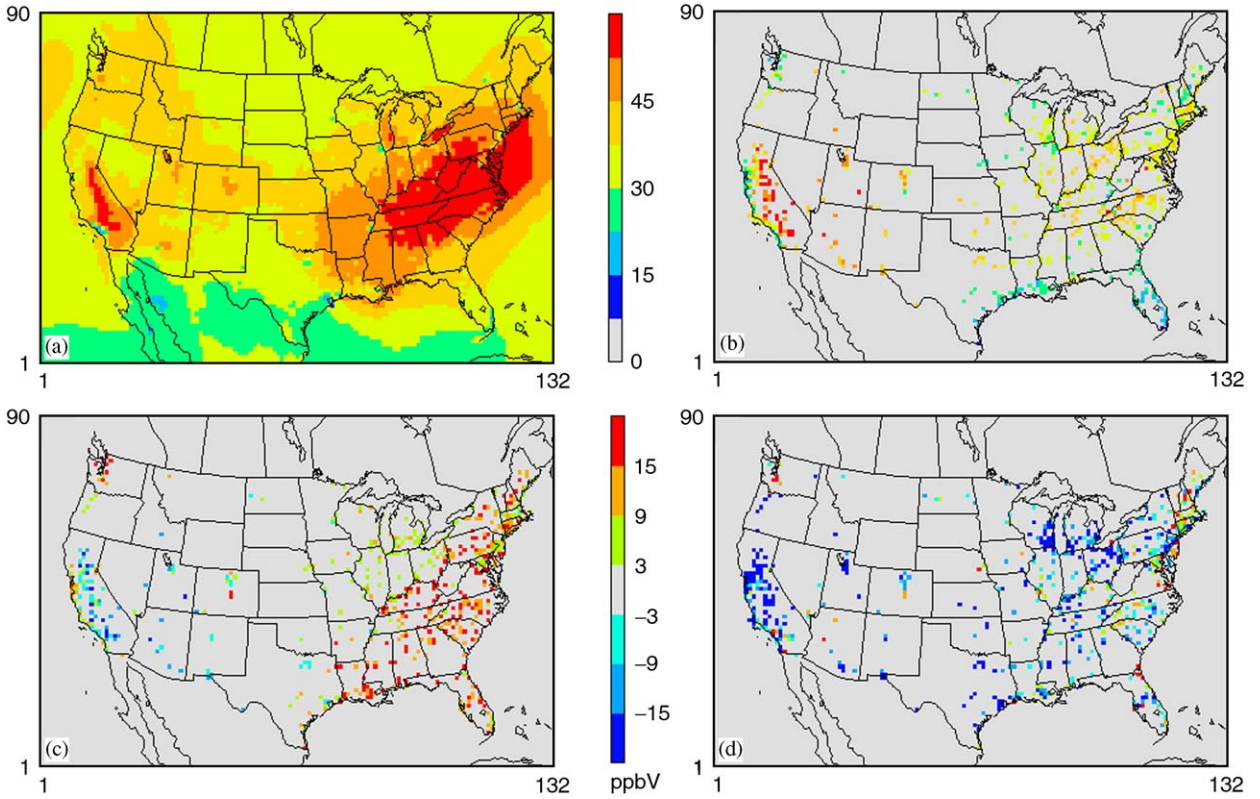

Fig. 1. Monthly average surface O_3 concentrations (a) predicted by CMAQ and (b) observed within the AQS network in July 1996. Differences (model minus observation) are shown for July 1996 (c) monthly average and (d) 1-h peak value. There are 987 monitors distributed among 550 grid cells. If several monitors are located inside a single cell, simultaneous hourly measurements are averaged together to represent the hourly concentration in the cell.

3. Results of model evaluation

3.1. Spatial variability of model performance

Fig. 1 presents simulated and observed monthly mean surface O_3 concentrations (from AQS) and the difference between them (model minus observations) in July 1996. Comparison of Figs. 1a and b shows that CMAQ generally captures the spatial pattern of surface O_3, including high O_3 (>45 ppbv monthly mean) in the California Central Valley and in the eastern US extending from northern Georgia to the mid-Atlantic seaboard.

High O_3 concentrations, however, were simulated with varying accuracies in different regions. Fig. 1c shows that CMAQ systematically overestimates 24-h monthly mean surface O_3 at most locations in the east by 15–20 ppbv, and underestimates it in the western US by 10–15 ppbv, particularly in the California Central Valley. Over-prediction in the east is likely due to the assumed immediate dilution of NO_x emissions from large point sources across a grid box, leading to an increase in O_3 production efficiency (Gillani and Pleim, 1996). Under-prediction in the California Central Valley and Rocky Mountains is likely linked to uncertainties in precursors emissions (Marr et al., 2002; Dreyfus et al., 2002) and the inability of MM5, which provides meteorology to drive CMAQ, to accurately represent wind and stability fields over complex topography (Deng et al., 2004; Gilliam et al., 2006). Fig. 1c also reveals that CMAQ frequently overestimates surface O_3 in coastal regions. Overestimation at coastal sites has been linked to poor representation of the vertical structure of the mixed layer between land and ocean, and its interaction with land/sea breezes (Colle et al., 2003). Differences between model and measurements in 90% of the data pairs in Fig. 1c are lying within ± 18 ppbv, and the MB is 6 ppbv. Fig. 1d compares simulated

and observed 1-h O_3 peak concentrations in July 1996. Although overestimating peak O_3 at a few locations along the mid- and northern Atlantic, the model primarily underestimates monthly 1-h peak O_3 concentrations across the country. Ninety percent of the differences between simulated and observed peak O_3 are less than 35 ppbv, with a MB of −17 ppbv. The higher average model values in Fig. 1c and lower maximum concentrations in Fig. 1d partly result from the $36 \times 36\,km^2$ horizontal resolution in the CMAQ simulation. Jang et al. (1995) studied the sensitivity of NO_x emissions and O_3 production to grid resolution. They found that lower model resolution leads to higher average and lower peak concentrations as dilution across the grid cell increases O_3 production efficiency resulting in higher average concentrations while local maxima can not be adequately resolved.

Next, we evaluate the ability of CMAQ to accurately reproduce O_3 concentrations observed at individual monitoring stations. Unlike Fig. 1, Figs. 2 and 3 directly compare measurements at individual sites with matching predictions instead of averaging measurements inside a grid cell. Fig. 2 compares observed 1-h average surface O_3 concentrations at individual AQS measurement sites with CMAQ simulated O_3 concentrations in co-located grid boxes. Here we use a cut-off value of 40 ppbv for all data pairs (i.e., data pairs containing measured concentrations below 40 ppbv were not used in the analysis). The cut-off was applied to examine model performance at levels above background concentrations in the continental US (Russell and Dennis, 2000; Hogrefe et al., 2001). Due to the cut-off value, the MB calculated in Fig. 2b is smaller (96% varying within 10 ppbv) than that in Fig. 1c. Along the Pacific coast, a large fraction of O_3 concentrations are below 40 ppbv and thus the cut-off eliminates many data pairs in which over-prediction occurs hence reducing the MB. The distinction suggests that a large portion of MB is due to difficulty CMAQ has in reproducing low concentrations of O_3. This issue will be further examined in Section 3.4.

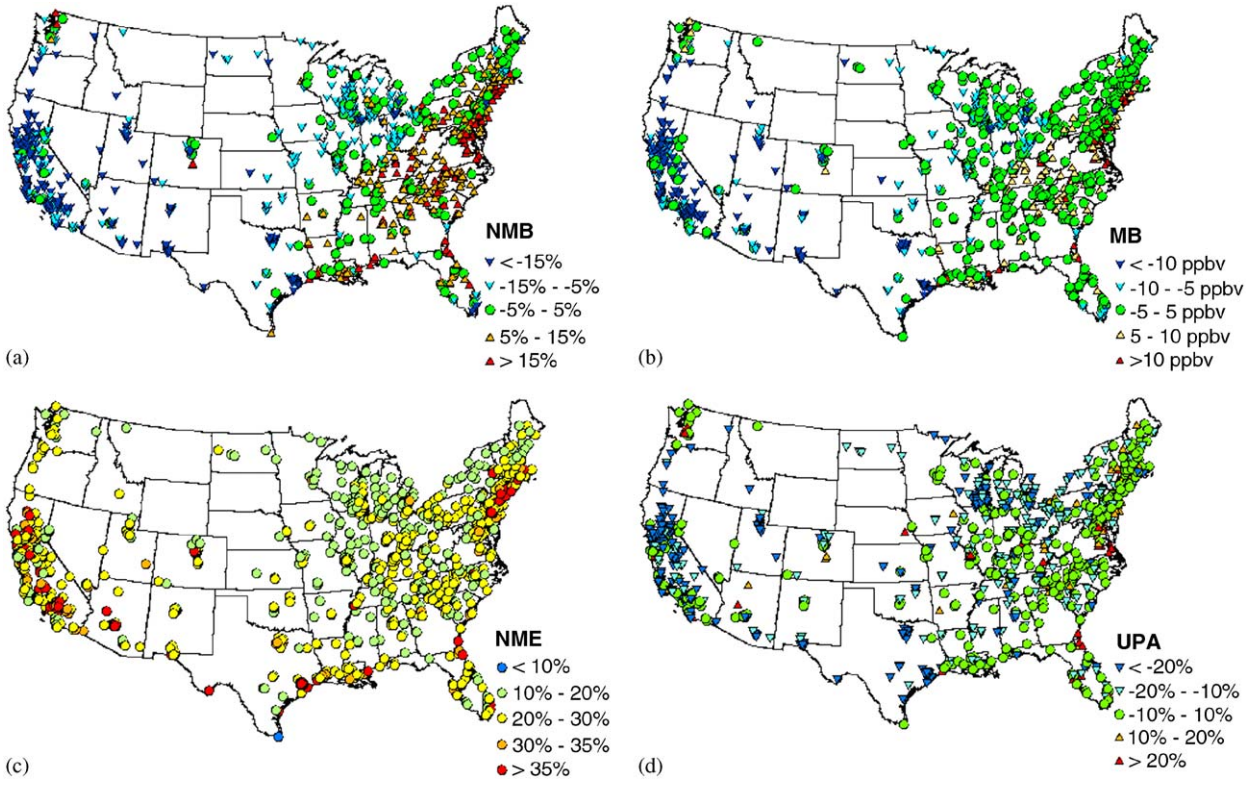

Fig. 2. A statistical comparison of 1-h average surface O_3 concentrations between individual AQS measurement sites and CMAQ predictions: (a) normalized mean bias (NMB); (b) mean bias (MB); (c) normalized mean error (NME), and (d) unpaired peak prediction accuracy (UPA) (see Table 1 for definitions). Only data pairs with observations above 40 ppbv are included here.

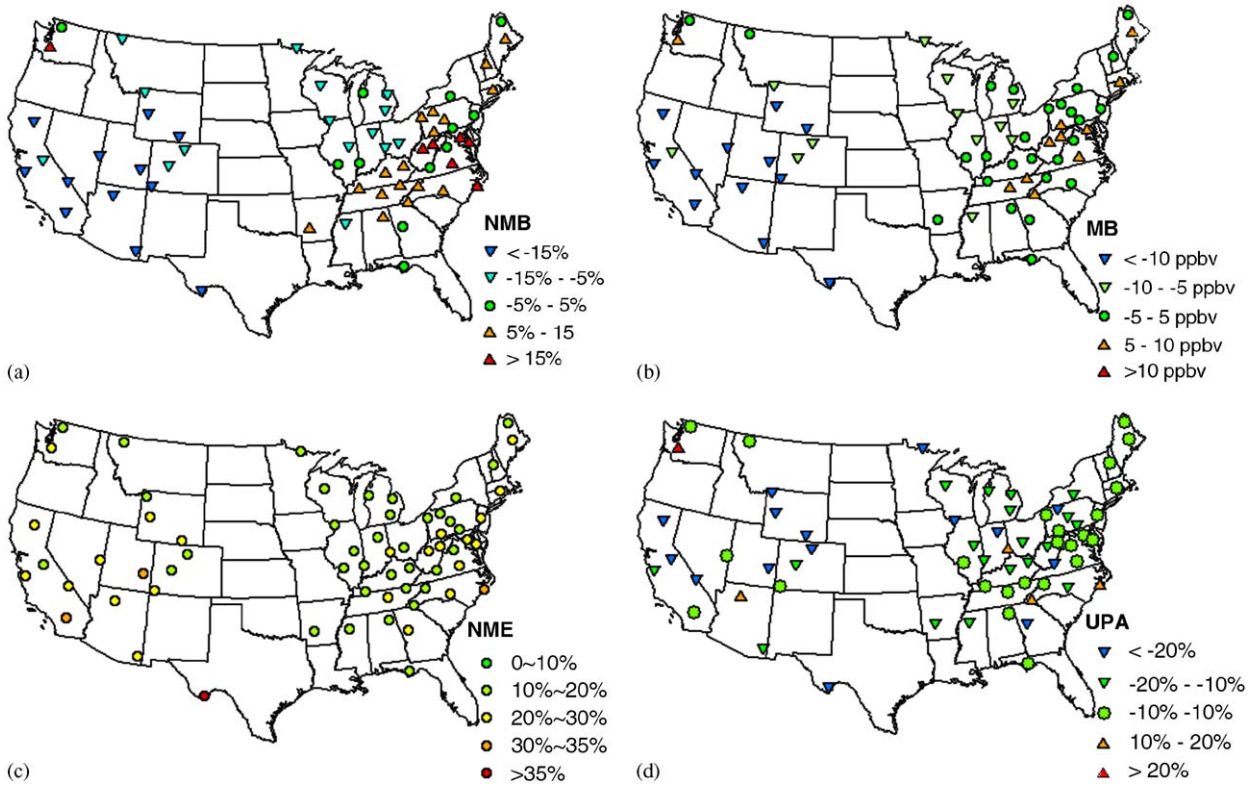

Fig. 3. Same as Fig. 2, but using measurements from CASTNet sites.

Fig. 2a shows a spatial pattern of over-prediction to under-prediction from the eastern to western US. The model tends to overestimate surface O_3 mostly in the east (NMB > 15%) with over-predictions decreasing (5% < NMB < 15%) over the Appalachian Mountains, being well predicted in the mid-west (−5% < NMB < 5%), and underestimated in California (NMB < −15%). MB values (Fig. 2b) show a similar pattern with greatest over-prediction along the east coast (MB > 10 ppbv), modest over-prediction in the Ohio Valley (5 ppbv > MB > 10 ppbv), the center of the country being well predicted (−5 ppbv > MB > 5 ppbv) and the California Central Valley being under-predicted by more than 10 ppbv. NME is generally < 30% (Fig. 2c), except for higher values found in the California Central Valley, along the northeast corridor between Washington DC and Boston, and in locations along the Gulf of Mexico. UPA (Fig. 2d) varies from −68% to 90%, with 75% of values falling within ±20%. The spatial distribution of UPA suggests that simulated O_3 peaks are underestimated in the Great Lakes region, Texas, the California Central Valley and Oregon, and are over-predicted at a few sites primarily in the eastern US.

Fig. 3 uses CASTNet data to present a comparison similar to Fig. 2. While the AQS network consists of a mixture of urban, suburban and rural sites, CASTNet data are mainly from rural and national park sites. Many rural areas are not able to attain the new 8-h NAAQS for O_3 (Chameides et al., 1997), making a rural evaluation important. Distribution of NMB (Fig. 3a) shows over-predictions of 5% to 15% or more along the Appalachian Mountains stretching from Tennessee to Pennsylvania, and at a few northeastern sites. Negative values of NMB are found in the west indicating CMAQ under-predicts observed concentrations frequently by more than 15%. Differences between simulated and observed O_3 are all 20 ppbv or less. Eastern sites have positive MB (indicating CMAQ is over-predicting observations) of up to 10 ppbv while western sites have a negative MB from −5 to lower than −10 ppbv (Fig. 3b). Values of NME (Fig. 3c) are between 10% and 30%, with only one site in Texas exceeding 35%. O_3 peaks at rural locations are frequently under-predicted by CMAQ,

Table 2
Model evaluation statistics of hourly ozone (O_3) concentrations for July 1996

Sites (cutoff value)	MB (ppbv)	ME (ppbv)	NMB (%) (±5–15)	NME (%) (30–35)	UPA (%) (±15–20)
AQS (>20 ppbv)	2	14	12	35	−12
AQS (>40 ppbv)	−5	14	−6	25	−12
AQS (>60 ppbv)	−14	19	−17	24	−12
CASTNet (>20 ppbv)	1	12	10	29	−17
CASTNet (>40 ppbv)	−4	12	−6	21	−17
CASTNet (>60 ppbv)	−12	16	−16	22	−17
AQS Rural (>40 ppbv)	−5	14	−5	23	−19
AQS Suburban (>40 ppbv)	6	16	7	26	11
AQS Urban (>40 ppbv)	6	15	7	25	4

Simulated values are compared to measurements from rural, suburban and urban locations within the air quality system (AQS) and clean air status and trends network (CASTNet) sites for different cut-off values. Acceptable performance criteria recommended by Russell and Dennis (2000) are shown in parenthesis.

particularly in the western and mid-western US (Fig. 3d).

3.2. Comparisons between rural and urban/suburban sites

After examining spatial variability, we now evaluate the overall performance across the US domain. Results calculated from Eqs. (1) and (2) (with the same simulated and observed data as in Section 3.1) are shown in Table 2, using different cut-off values. Comparisons between AQS data and predictions show that MB and NMB are positive for a cut-off value of 20 ppbv, and negative for cut-off values of 40 and 60 ppbv. This indicates that CMAQ generally overestimates (underestimates) surface O_3 below (above) 20 (40) ppbv. Underestimations are larger when 60 ppb rather than 40 ppb is used as the cutoff. Increasing cutoff values from 40 to 60 ppbv results in MB decreasing from −5 ppbv to −14 ppbv, and NMB from −6% to −17%, respectively. Approximately 250,000 prediction–observation pairs lie both above and below 40 ppbv. On average, CMAQ over-predicts the lower half of the AQS observations by 7 ppbv and under-predicts the upper half of AQS observations by −5 ppbv.

A comparison between predictions and rural CASTNet data shows similar bias to that of AQS data. At rural sites, CMAQ overestimates observations at low O_3 levels, and underestimates with a larger bias at higher O_3 levels. Across-domain peak CASTNet O_3 is underestimated by 17%, 5% more than aggregated AQS data. ME increases with increasing cut-off values (from 12% to 16%), but values of NME are larger with a 20 ppbv cutoff than with a 40 or 60 ppbv cutoff. Comparisons of both CASTNet and AQS observations with CMAQ predictions show similar tendencies, indicating that the model simulates surface O_3 with similar biases between the two observational networks.

To further differentiate model performance between urban, suburban and rural settings, we recalculate the AQS statistics for each category using a 40 ppbv cut-off (Table 2). Statistics at AQS rural sites are similar to that at CASTNet sites, with the model under-predicting observations (MB = −5 ppbv; NMB = −5%; NME = 23%; and UPA = −19%). Statistics at urban and suburban sites, however, show a positive bias (MB = 6 ppb; NMB = 7%; NME = 25%; and UPA = 7.5%), suggesting that the model systemically over-predicts observations at these urban/suburban locations by a similar magnitude to the under-predictions at rural sites. This type of bias has implications for model applications to quantify health impacts as it has a tendency to under-predict O_3 concentrations in urban high-population areas.

US EPA has suggested informal criteria for regulatory modeling practices of ±5 to ±15% for NMB, 30 to 35% for NME, and ±15 to ±20 for UPA (EPA, 1991; Russell and Dennis, 2000). With a 40 ppbv cut-off value, our evaluation statistics lie within the suggested ranges. The criteria are also met when the cut-off value is lowered to 20 ppbv or raised to 60 ppbv, except that NMB is slightly (1–2%) out of range for the 60 ppbv cut-off value. However, we see with 20, 40 and 60 ppbv cut-off

values, the MB is 2, −5, and −14 ppbv, respectively, indicating a shift in model performance from over-prediction to under-prediction. Thus, the choice of cut-off value can influence the user's perception of model performance. We analyze this further below.

3.3. CMAQ performance as a function of O_3 concentrations

In this section, we examine the ability of CMAQ to reproduce O_3 observations at different levels. Results in Sections 3.1 and 3.2 show that CMAQ systematically over-predicts (under-predicts) surface O_3 at low (high) concentrations, a conclusion in agreement with the findings of Hogrefe et al. (2004). Here we group the observation–model pairs into 10 ppbv bins based on observed O_3 concentrations. Statistics are calculated for each bin and tabulated in Table 3. For the lowest O_3 concentrations (<10 ppbv), CMAQ over-predicts observations with a MB of 21 ppbv and an extreme UPA value (940%). MB and NMB decrease with increasing observed concentrations and become negative above 50 ppbv. ME and NME decrease from 0 to 50 ppbv and increase thereafter. For very high observed O_3 concentrations (>90 ppbv), CMAQ under-predicts observations with a MB of 32 ppbv and an UPA of −12%.

Although epidemiological studies that have examined relationships between O_3 and premature mortality have not found evidence of a threshold below which no adverse health effects exist (Bell et al., 2005; Levy et al., 2005), the NAAQS establish a threshold above which O_3 concentrations are not allowed and below which they are deemed acceptable. We therefore analyze the ability of CMAQ to successfully predict both 1- and 8-h daily maximum O_3 concentrations, two measures previously and currently used to trigger regulatory enforcement. Comparisons are made between 1-h daily maximum O_3 predictions by CMAQ and measurements from AQS sites (see figure in online Supplementary Material). Similar to Figs. 1 and 2, CMAQ overestimates 1-h daily maximum O_3 along the east coast of the US, and underestimates it in the California Central Valley, the Great Lakes area, and Texas. Despite the spatial discrepancies, the overall performance for 1-h daily maxima (without a cut-off value) when aggregated across the US is only 1 ppbv (MB), 15 ppbv (ME), 8% (NMB), 25% (NME), and −12% (UPA). All statistics lie within the US EPA suggested ranges. The daily 1-h maximum results are also comparable to the best performance CMAQ displays when simulating 1-h average O_3 concentrations as shown in Table 2, suggesting that although CMAQ under-predicts extremely high O_3 events, it simulates 1-h daily maxima well.

In 1999, the 1-h average O_3 NAAQS was replaced by the 8-h average standard. Fig. 4 presents the comparison between 8-h daily maximum O_3 concentrations predicted by CMAQ and observed at AQS sites. CMAQ considerably overestimates the 8-h daily maxima (NMB > 15% and NME > 35%) in the eastern US and western coastal states, particularly in the southeast US, New England and the state of Washington. The model under-predicts 8-h daily peaks in the California Central Valley and a few western locations. Note that larger biases are also co-located with dense human population. Despite the spatial discrepancies the overall performance for 8-h daily maxima across the domain is 7 ppbv (MB), 14 ppbv (ME), 26% (NMB), 36% (NME), and 0% (UPA).

Table 3
Model performance statistics for hourly O_3 concentration intervals observed at AQS sites in July 1996

OBS O_3 levels (ppbv)	Number of pairs	OBS_{max} (ppbv)	MOD_{max} (ppbv)	MB (ppbv)	ME (ppbv)	NMB (%) (±5–15)	NME (%) (30–35)	UPA (%) (±15–20)
0–10	87,400	10	104	20.8	21.6	626.3	640.9	940
10–20	108,292	20	126	15.7	17.4	104.8	115.9	531
20–30	128,215	30	139	10.4	13.9	41.9	55.4	363
30–40	115,508	40	137	6.0	12.4	17.3	35.3	243
40–50	91,901	50	148	1.7	12.0	3.9	26.6	196
50–60	66,156	60	169	−2.9	12.5	−5.1	22.6	181
60–70	41,387	70	152	−7.7	14.3	−11.8	22.0	118
70–80	23,091	80	184	−13.2	17.5	−17.5	23.3	131
80–90	12,423	90	171	−18.7	21.6	−21.9	25.4	90
90+	10,857	215	190	−32.4	33.8	−30.7	32.1	−12

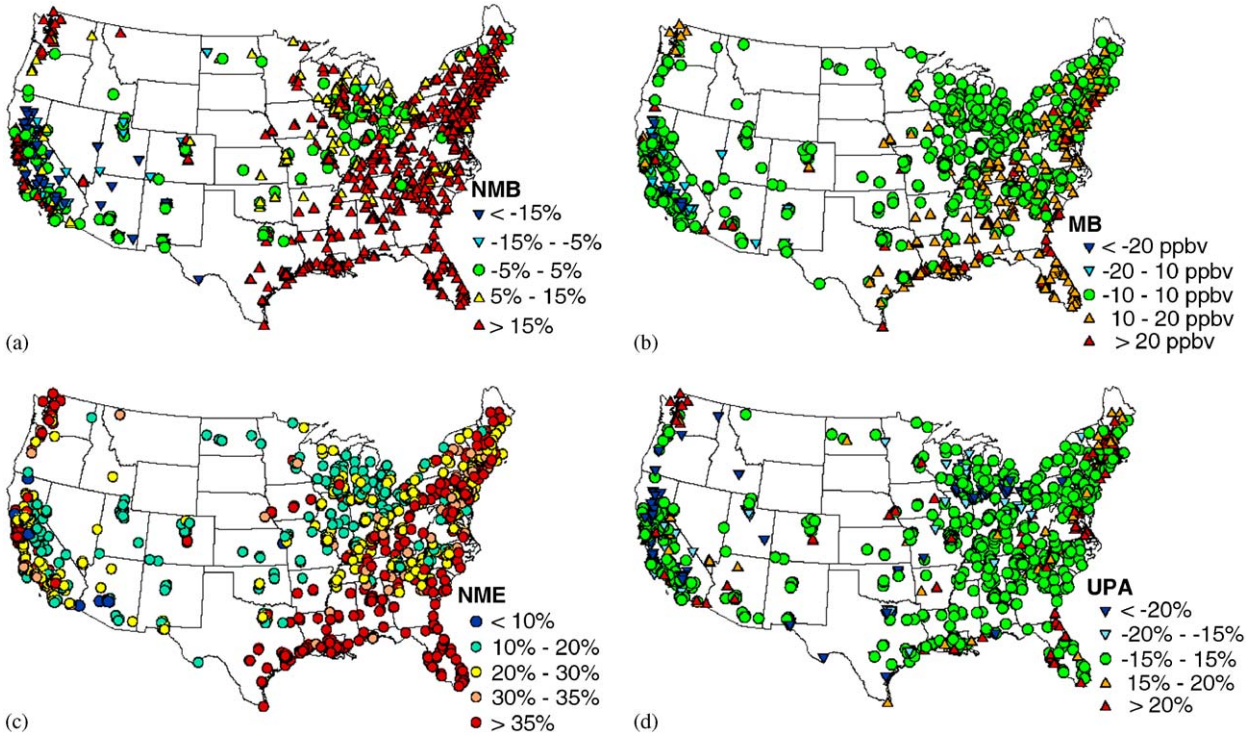

Fig. 4. A statistical comparison of 8-h average daily maximum surface O_3 concentrations between CMAQ and AQS measurement sites. Statistical metrics are the same as in Fig. 2.

In sum, our results show that CMAQ best simulates observed O_3 at intermediate concentrations (40–60 ppbv), while over-(under-) predicting at lower (higher) levels. The model is able to reproduce surface O_3 for a wide range of conditions (30–80 ppbv) with a cross-domain NME less than 35% and NMB lying between $\pm$15%. At over 80 ppbv under-predictions exceed 20 ppbv and below 30 ppbv over-predictions are between 10 and 20 ppbv. Although systematically over-predicting O_3 in the east and under-predicting it in the west, CMAQ is capable of reproducing 1- and 8-h daily maxima with cross-domain NMBs of 8% and 25%, and NMEs of 25% and 36%, respectively. Although our results are calculated based on only July data, it has been reported that CMAQ's ability to accurately simulate O_3 varies less temporally than spatially (Eder and Yu, 2006). The evaluation by Eder and Yu (2006), using a new version of CMAQ to simulate year 2001 surface O_3 concentrations, found similar spatial patterns of model performance across the continental US as described here.

3.4. Influence of data averaging time period on performance statistics

In this section we examine the relationship between performance statistics and time scales of data aggregation in order to determine if CMAQ is able to capture short-term variability in O_3 concentrations resulting from local processes. Bouchet et al. (1999) showed that current modeling systems perform better when simulating longer-term events rather than transient, smaller-scale features. Hogrefe et al. (2004) showed that CMAQ better represents regional ozone climatology and fluctuations of a week or more, rather than small-scale fluctuations on time scales of 24 h or less. These studies have focused on a single geographic area. To evaluate model performance across the entire continental United States, we aggregate raw observation–model data pairs into 1-, 8-h, daily, weekly, and monthly averages, using a moving average. No cut-off values are applied and only data pairs that include at least 75% of possible measurement data are included. Statistics of data pairs over

Table 4
Model evaluation statistics for different averaging periods of hourly ozone (O_3) concentrations in July 1996

Aggregating period (h)	Number of pairs	OBS_{max} (ppbv)	MOD_{max} (ppbv)	MB (ppbv)	ME (ppbv)	NMB (%) (±5–15)	NME (%) (30–35)	UPA (%) (±15–20)
1 (1 h)	683,973	215	190	6.3	15.4	105.2	125.4	−11.6
8 (8 h)	86,401	154	154	6.2	13.7	50.3	66.9	−0.1
24 (daily)	28,927	114	107	6.2	11.8	30.2	41.9	−6.5
168 (weekly)	3776	86	74	6.0	10.1	24.5	33.3	−14.1
748 (monthly)	953	79	65	6.3	9.6	23.1	30.4	−17.9

Data pairs of measurements and predictions are obtained from the AQS network and from grid boxes in CMAQ, which include the measurement sites. Maximum values of measurements and model simulated O_3 concentrations over the domain are shown as OBS_{max} and MOD_{max}, respectively. Definitions of MB, ME, NMB, NME and UPA are given in Table 1.

different averaging time periods are calculated using the same procedures as described in Section 2.3 and are shown in Table 4.

Both observed and predicted maximum O_3 values decrease with increasing averaging time. UPA values, however, indicate the closest correspondence between observed and simulated O_3 concentrations when an 8-h averaging period is used. MB remains nearly the same when different averaging periods are used. The time scale of aggregation does not affect the simple sum of individual differences; the slight variations in MB are attributed to changes in the number of data pairs excluded from the calculations because of the completeness criteria we use (>75%). Values of other statistical metrics, including ME, NMB, and NME, consistently decrease with longer averaging periods, with greatest decreases for NME and NMB when the averaging period increases from 1 to 8 h. It is a common practice in model evaluation to average model and observations over a variety of time periods, rather than directly comparing simulated concentrations with raw observations (e.g., Mebust et al., 2003; Hogrefe et al., 2001, 2004; Eder and Yu, 2006). Our results show that performance criteria for these metrics should be specific to a time period (e.g., 1- or 8-h criteria), and not independent of averaging time as some current model evaluation studies imply.

3.5. Influence of LBC on simulated vertical O_3 profiles

In this section we compare the vertical profiles of simulated O_3 with ozonesonde observations from the Boulder, CO and Wallops Island, VA stations. In addition, we present the impacts of LBC on simulated vertical O_3 concentrations.

Fig. 5 compares two simulated vertical profiles of O_3 using different LBC. In one case the boundary conditions are extracted from a multi-year simulation of MOZART-2 as described in Section 2.1; in the other case the boundary conditions are established from predefined CMAQ vertical concentration profiles as given by Byun and Ching (1999). O_3 concentrations in the two simulations are similar in the lower troposphere, but diverge with increasing altitude with O_3 concentrations simulated with the MOZART-2 LBC consistently higher in the upper troposphere than those obtained with the predefined CMAQ LBC. Differences in monthly mean O_3 concentrations between the two simulations range from 1 ppbv (3 ppbv) at the surface to 15 ppbv (11 ppbv) at 200 mbar at Boulder (Wallops Island). The results reveal that LBC have a larger impact on simulated O_3 in the upper than lower troposphere.

Fig. 5 also compares the simulated vertical O_3 profiles with two ozonesonde profiles, the average of July 1996 measurements and the 15-year (1985–2000) average of July measurements. At Wallops Island, the CMAQ simulation with the predefined LBC over-predicts July 1996 O_3 concentrations in the boundary layer with a maximum overestimation at the surface, and under-predicts the observations in the free troposphere by about 25 ppbv between 700 and 200 mbar. Although the use of MOZART-2 boundary conditions improves model performance in the upper troposphere, the discrepancy remains large (an average underestimate by CMAQ of 15 ppbv above 700 mbar). The discrepancy in the upper troposphere is much larger when the multi-year average rather than July 1996 sonde data is used, suggesting that the discrepancy between model and the 1996 data is unlikely attributed to unusually high O_3 concentrations at upper altitudes that year. At the high-elevation

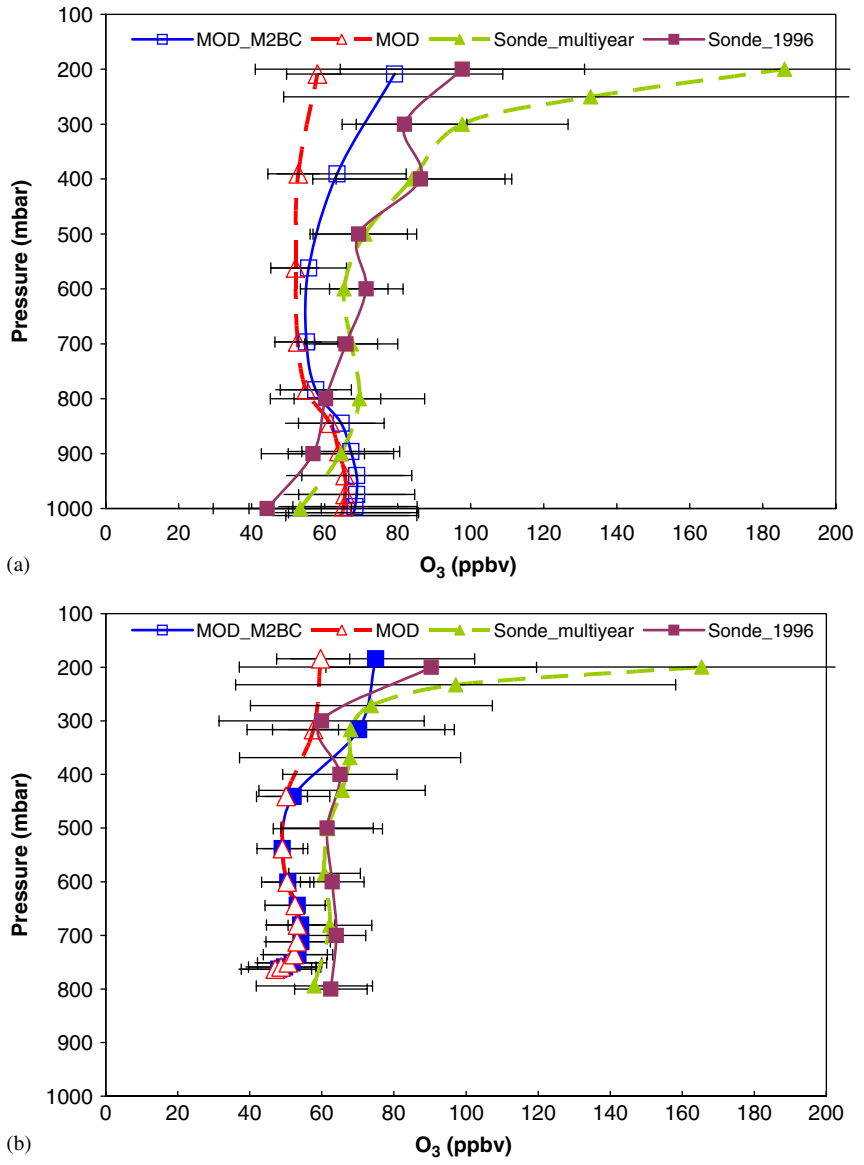

Fig. 5. Comparisons of vertical profiles of monthly averaged ozone (O_3) concentrations between July 1996 CMAQ simulations using predefined LBC (MOD) and using LBC from a multi-year 1990s simulation of MOZART-2 (MOD_M2BC), and ozonesonde measurements for July 1985–2000 (Sonde_multiyear) and for July 1996 (sonde_96) at (a) Wallops Island, VA (37.90°N, 75.50°W, 13 m above sea level); and (b) Boulder, CO (40.02°N, 105.27°W, 1743 m above sea level).

Boulder station, both simulations consistently underestimate observations at the surface (about 800 mbar) and above, except at 300 mbar. MOZART-2 boundary conditions again improve model performance in the upper troposphere while having a negligible effect near the surface. At 200 mbar, the normalized bias is −33% using the predefined LBC and −29% using the MOZART-2 LBC. Similar to Wallops Island, the July 1996 sonde data above 400 mbar at Boulder are considerably lower than the multi-year average.

4. Implications for air quality applications

The above evaluation results have important implications for various CMAQ applications. CMAQ has increasingly been employed to predict regional high O_3 episodes and NAAQS nonattainment. Our

results show that the model overestimates 1- and 8-h daily maxima in the eastern and coastal areas and underestimates the same at many western US locations with a cross-domain NME of 25% and 36%, respectively. Although the work by Eder and Yu (2006) reported an improvement in model performance, largely due to an improved emission inventory, values of NME remain notable (18% and 20% for 1- and 8-h daily peaks). This presents a concern when using CMAQ to determine whether emission controls will actually result in required O_3 attainment.

Recent studies have begun to use CMAQ to examine health impacts of air pollution (e.g., Tong et al., 2005). Pollution damages on public health are calculated from concentration–response functions derived from epidemiological studies, and population exposure to air pollutants such as O_3. In such applications, accurate prediction of O_3 concentrations in densely populated areas is particularly important because spatial variations in population density, combined with an overprediction of O_3 concentrations, can result in an over-prediction of calculated exposures. Statistics at AQS sites (Table 2) show that the model over-predicts observations at urban/suburban locations by a MB of 6 ppbv and an NME of 26%. Fig. 4 and the figure in the Supplementary Material also show that over-prediction in the eastern US is larger and more widespread when 8-h rather than 1-h average O_3 is used. The fact that larger discrepancies between predictions and observations co-locate with dense populations, particularly in the eastern US and coastal locations, indicates an area for further model development effort.

Compared to ozonesonde data, CMAQ both overestimates and underestimates O_3 concentration by 10–20 ppbv in the lower troposphere, and consistently underestimates it in the upper troposphere by up to 50 ppbv. CMAQ was originally designed to capture high surface O_3 events that are typically dominated by strong photochemical production of O_3 in the boundary layer. Recent air quality studies have extended the CMAQ domain to investigate pollution transport on inter-continental and hemispheric scales (e.g., EPA, 2004). Long-distance transport of air pollution at mid-latitudes is generally associated with lifting of surface emissions to the free troposphere by frontal systems or convection followed by rapid advection in the free troposphere westerlies and subsidence at downwind locations (Park et al., 2004). O_3 is the primary precursor of the hydroxyl radical which in turn determines the lifetime of most atmospheric constituents. Therefore, it is important for CMAQ to capture O_3 concentrations accurately in the free troposphere. The large and persistent discrepancies in the upper troposphere raise concern when applying CMAQ to long-distance transport issues.

Improving LBC in CMAQ appears to be insufficient to correct under-prediction of O_3 in the upper troposphere. The MOZART-2 simulation from which we extracted LBC slightly over-predicts O_3 in the upper troposphere in the northern extratropics (Horowitz et al., 2003); using MOZART-2 LBC, however, still results in an under-prediction of O_3 in the upper troposphere. The remaining discrepancy can be attributed to the use of a zero-flux upper boundary condition in CMAQ. While photochemistry largely controls O_3 concentrations in the polluted boundary layer, influx of O_3 from the stratosphere is important in the upper troposphere (Mauzerall et al., 1996). Excluding the contribution of STE through the use of either a concentration or flux upper boundary condition limits the ability of CMAQ to resolve the vertical profile of O_3.

5. Summary

This paper evaluates the ability of the Models-3 CMAQ model to reproduce surface O_3 measurements obtained from 987 AQS and 123 CASTNet sites and vertically at two ozonesonde stations. We evaluated model performance to determine how well CMAQ: (1) captured spatial variability in surface O_3 concentrations across the continental US, (2) simulated surface O_3 at rural compared to urban locations, (3) simulated O_3 across the spectrum of observed concentrations, (4) performed using a variety of data averaging times, and (5) simulated O_3 vertically as indicated by comparison with ozonesonde data.

Our results show that CMAQ overestimates surface O_3 in the eastern US and coastal locations, and underestimates it in the California Central Valley and many western locations. The model underestimates observed O_3 at rural sites (MB = −5 ppbv, NMB = −5% and NME = 23% with a 40 ppbv cut-off value) and over-predicts it at urban and suburban sites by a similar magnitude (MB = 6 ppbv, NMB = 7% and NME = 25%). Values of most statistics, except MB, decrease with longer averaging periods, indicating that performance criteria for these metrics should be specific to a time period (e.g., 1- or 8-h criteria), not

independent of time scale as some current model evaluation studies imply. CMAQ's ability to reproduce surface observations varies with O_3 concentrations. The model best simulates observed O_3 for intermediate concentrations (40–60 ppbv), while over-(under-) predicting at lower (higher) levels. For extremely high observations (>90 ppbv), CMAQ under-predicts observations with an average bias of 32 ppbv. The model reproduces surface O_3 observations for a wide range of conditions (30–80 ppbv) with NMEs less than 35% and NMBs lying between $\pm$15%. Although systematically over-predicting O_3 in the east and under-predicting it in the west, CMAQ is able to reproduce 1- and 8-h daily maxima with a cross-domain MB of 1 and 8 ppbv, or NMB of 8% and 25%, respectively. However, larger discrepancies between predictions and observations co-locate with dense populations, particularly in the eastern US and coastal locations, indicating areas for further model development effort. Although we focus here on O_3 concentrations in July 1996, model performance may vary with seasonality, grid resolution, and concentrations of other pollutants. Further evaluation is necessary to establish model performance for other time periods and species.

Comparisons of vertical profiles of simulated O_3 with ozonesonde data show both an over- (under-) prediction at the surface at Wallops Island, VA (Boulder, CO). Between 800 and 400 mbar, CMAQ under-predicts ozonesonde data by 10–20 ppbv at both stations. The under-prediction persists above 400 mbar when predefined LBC are used but decreases when LBC extracted from a MOZART-2 simulation are used, providing near agreement between measurements and model simulated values at 300 mbar at Boulder, CO. Above 300 mbar CMAQ significantly under-predicts measured O_3 concentrations regardless of the LBC used. The use of MOZART-2 LBC is not sufficient to correct under-predictions in the upper troposphere. CMAQ assumes zero-flux at the top boundary, and the exclusion of the contribution of STE limits the ability of CMAQ to resolve the vertical profile of O_3. Although this shortcoming has a relatively small effect on surface O_3 concentrations in polluted areas it may hinder analysis of long-distance transport of O_3.

Acknowledgements

This research was funded by the Glaser Progress Foundation. We are grateful to Namsoug Kim for processing observation data, Jennifer Logan and Inna Megretskaia for providing ozonesonde data, and Carey Jang and Binyu Wang for providing emissions and meteorological data for use in CMAQ. We thank Michelle Bell, Robin Dennis, Arlene Fiore, Daiwen Kang, Jenise Swall and Shaocai Yu for helpful comments on earlier versions of the manuscript.

Appendix A. Supplementary Material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.atmosenv.2005.11.058.

References

Bell, M.L., Ellis, H., 2003. Sensitivity analysis of tropospheric ozone to modified emissions for the Mid-Atlantic Region. Atmospheric Environment 38, 1879.

Bell, M.L., Dominici, F., Samet, J.M., 2005. A meta-analysis of time-series studies of ozone and mortality with comparison to the National Morbidity, Mortality, and Air Pollution Study. Epidemiology 16 (4).

Berge, E., Huang, H.-C., Chang, J.T.-H., 2001. A study of the importance of initial conditions for photochemical oxidant modeling. Journal of Geophysical Research 106, 1347–1363.

Bouchet, V.S., Laprise, R., Torlaschi, E., McConnell, J.C., Plummer, D.A., 1999. Studying ozone climatology with a regional climate model 2. Climatology. Journal of Geophysical Research 104, D23, 30,373–330,385.

Byun, D.W., Ching, J.K.S. (Eds.), 1999. Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. EPA Report No. EPA-600/R-99/030.

Carolina Environmental Program, 2003. Sparse Matrix Operator Kernel Emission (SMOKE) modeling system. University of North Carolina, Carolina Environmental Programs, Chapel Hill, NC.

Chameides, W.L., Saylor, R.D., Cowling, E.B., 1997. Ozone pollution in the rural United States and the new NAAQS. Science 276, 916.

Colle, B.A., Olson, J.B., Tongue, J.S., 2003. Multiseason verification of the MM5. Part I: Comparison with the Eta Model over the Central and Eastern United States and impact of MM5 resolution. American Meteorological Society, 431–457.

Deng, A., Seaman, N.L., Hunter, G.K., Stauffer, D.R., 2004. Evaluation of interregional transport using the MM5-SCIPUFF system. Journal of Applied Meteorology 43, 1864–1886.

Dennis, R.L., Byun, D.W., Novak, J.H., Galluppi, K.J., Coats, C.J., Vouk, M.A., 1996. The next generation of integrated air quality models: EPA's Models-3. Atmospheric Environment 30, 1925–1938.

Dreyfus, G.B., Schade, G.W., Goldstein, A.H., 2002. Observational constraints on the contribution of isoprene oxidation to ozone production on the western slope of the Sierra Nevada, CA. Journal of Geophysical Research 107 (D19), 4365.

Eder, B., Yu, S., 2005. A performance evaluation of the 2004 release of Models-3 CMAQ. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2005.08.045.

Fiore, A.M., Jacob, D.J., Mathur, R., Martin, R.V., 2003. Application of empirical orthogonal functions to evaluate ozone simulations with regional and global models. Journal of Geophysical Research 108, 4431.

Fuentes, M., 2003. Statistical assessment of geographic areas of compliance with air quality standards. J. Geophys. Res. 108 (D24), 9002.

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1994. A photochemical kinetic mechanism for urban and regional scale computer modeling. Journal of Geophysical Research 94, 12,925–12,956.

Gillani, N.V., Pleim, J.E., 1996. Subgrid scale features of anthropogenic emissions of NO_x and VOC in the context of regional eulerian models. Atmospheric Environment 30, 2043–2059.

Gilliam, R.C., Hogrefe, C., Rao, S.T., 2006. New methods for evaluating meteorological models used in air quality applications. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2006.01.023.

Grell, G.A., Dudhia, J., Stauffer, D.R., 1994. A description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note, NCAR/TN-398 + STR, National Center for Atmospheric Research, Boulder, CO.

Hogrefe, C., Rao, S.T., Kasibhatla, P., Hao, W., Sistla, G., Mathur, R., McHenry, J., 2001. Evaluating the performance of regional-scale photochemical modeling systems: Part II—ozone predictions. Atmospheric Environment 35, 4175–4188.

Hogrefe, C., Biswas, J., Lynn, B., Civerolo, K., Ku, J.Y., Rosenthal, J., Rosenzweig, C., Goldberg, R., Kinney, P.L., 2004. Simulating regional-scale ozone climatology over the eastern United States: model evaluation results. Atmospheric Environment 38, 2627–2638.

Horowitz, L.W., et al., 2003. A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2. J. Geophys. Res. 108 (D24), 4784.

Ito, K., De Leon, S.F., Lippman, M., 2005. Associations between ozone and daily mortality: analysis and meta-analysis. Epidemiology 16, 446.

Jang, J.C.C., Jeffries, H.E., Tonnesen, G.S., 1995. Sensitivity of ozone to model grid resolution. II: Detailed process analysis for ozone chemistry. Atmos. Environ. 29, 3101–3114.

Kang, D., Aneja, V.P., Mathur, R., Ray, J., 2003. Nonmethane hydrocarbons and ozone in the rural southeast United States national parks: a model sensitivity analysis and its comparison with measurement. J. Geophys. Res. 108 (D10), 4604.

Levy, J.I., Chemerynski, S.M., Sarnat, J.A., 2005. Ozone exposure and mortality: an empiric Bayes metaregression analysis. Epidemiology 16, 458.

Lin, C.C., Jacob, D.J., Fiore, A.M., 2001. Trends in exceedances of the ozone air quality standard in the continental United States, 1980–1998. Atmos. Environ. 35, 3217–3228.

Logan, J.A., Prather, M.J., Wofsy, S.C., McErloy, M.B., 1981. Tropospheric chemistry: a global perspective. Journal of Geophysical Research 86, 7210–7254.

Logan, J.A., Megretskaia, I.A., Miller, A.J., Tiao, G.C., Choi, D., Zhang, L., Stolarski, R.S., Labow, G.J., Hollandsworth, S.M., Bodeker, G.E., Claude, H., Muer, D.D., Kerr, J.B., Tarasick, D.W., Oltmans, S.J., Johnson, B., Schmidlin, F., Staehelin, J., Viatte, P., Uchino, O., 1999. Trends in the vertical distribution of ozone: A comparison of two analyses of ozonesonde data. Journal of Geophysical Research 104 (D21), 26,373–26,400.

Marr, L.C., Black, D.R., Harley, R.A., 2002. Formation of photochemical air pollution in central California. 1: development of a revised motor vehicle emission inventory. Journal of Geophysical Research 107(D6), paper no. 10.1029/2001JD000689.

Mauzerall, D.L., Wang, X., 2001. Protecting Agricultural crops from the effects of tropospheric ozone exposure: reconciling science and standard setting in the United States, Europe, and Asia. Annual Review of Energy and the Environment 26, 237–268.

Mauzerall, D.L., Jacob, D.J., Fan, S.M., Bradshaw, J.D., Gregory, G.L., Sachse, G.W., Blake, D.R., 1996. Origin of tropospheric ozone at remote high northern latitudes in summer. Journal of Geophysical Research 101, 4175–4188.

Mebust, M.R., Eder, B., Binkowski, F.S., Roselle, S.J., 2003. Models-3 community multiscale air quality (CMAQ) model aerosol component 2. Model evaluation. Journal of Geophysical Research 108, 4184.

National Research Council (NRC), 2004. Air Quality Management in the United States. National Research Council of the National Academies, Washington, DC.

Park, R.J., Jacob, D.J., Field, B.D., Yantosca, R.M., Chin, M., 2004. Natural and transboundary pollution influences on sulfate–itrate–mmonium aerosols in the United States: implications for policy. Journal of Geophysical Research 109, D15204.

Pierce, T.E., 2001. BEIS3 Version 0.9, Environmental Protection Agency, Research Triangle Park, NC (available via anonymous ftp at ftp://ftp.epa.gov/amd/asmd/beis3v09/).

Ramaswamy, V., 2001. Radiative Forcing of Climate Change, in Climate Change 2001, The Scientific Basis, Intergovernmental Panel on Climate Change. Cambridge University Press, USA.

Russell, A., Dennis, R., 2000. NARSTO critical review of photochemical models and modeling. Atmospheric Environment 34, 2284–2324.

Tong, D.Q., 2004. CMAQ in policy applications and remaining questions, Models-3 User's Workshop, October 18th–20th, 2004. Chapel Hill, NC.

Tong, D.Q., Mauzerall, D.L., 2005. Numerical instability in the Community Multiscale Air Quality model and its impacts on aerosol and ozone simulations. Atmospheric Environment, submitted.

Tong, D.Q., Muller, N.Z., Mauzerall, D.L., Mendelsohn, R.O., 2005. An integrated assessment of the spatial variability of ozone impacts from emissions of nitrogen oxides. Environmental Science & Technology 40 (5), 1395–1400.

US Environmental Protection Agency (US EPA), 1991. Guidance for regulatory application of the Urban Airshed Model. EPA-450/4-91-013, July 1991, United States Environmental Protection Agency, Research Triangle Park, NC 27711.

US Environmental Protection Agency (US EPA), 1999. The benefits and costs of the clean air act, 1990 to 2010, EPA Report to Congress, EPA-410-R-99-001, Washington DC, June, 1999.

US Environmental Protection Agency (US EPA), 2000. National air quality and emissions trends report: 1998, EPA-454-/F-00-002. US Office of Air Quality Planning and Standards, Research Triangle Park, NC.

US Environmental Protection Agency (US EPA), 2003. User's guide to MOBILE6.1 and MOBILE6.2. Office of Air and Radiation, EPA420-R-03-010, Assessment and Standards Division, Office of Transportation and Air Quality, US Environmental Protection Agency.

US Environmental Protection Agency (US EPA), 2004. Intercontinental Transport and Climatic Effects of Air Pollutants (ICAP): Final Report, EPA Task Order ID 4TCG68024006, available online (http://www.cep.unc.edu/empd/projects/ICAP/PDF/ICAP_PhaseII_final_JRE.pdf).