Available online at www.sciencedirect.com

SCIENCE DIRECT

Atmospheric Environment 40 (2006) 4811–4824

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

A performance evaluation of the 2004 release of Models-3 CMAQ

Brian Eder*[,1], Shaocai Yu[2]

ASMD, ARL, NOAA, Mail Drop E243-01, Research Triangle Park, NC 27711, USA

Received 14 March 2005; accepted 5 August 2005

Abstract

This performance evaluation compares a full annual simulation (2001) of Community Multi-scale Air Quality (CMAQ) (Version 4.4) covering the contiguous United States against monitoring data from four nationwide networks. This effort, which represents one of the most spatially and temporally comprehensive performance evaluations of the model, reveals that CMAQ varies considerably in its ability to simulate ambient air concentrations of critical gas and particulate matter species. Simulations of the peak 1- and 8-h ozone (O_3) concentrations during the "O_3 season" (April–September, 2001) were relatively good (correlation (r) = 0.68, 0.69; normalized mean bias (NMB) = 4.0%, 8.1%; and normalized mean error (NME) = 18.3%, 19.6%, respectively). The annual simulation of sulfate (SO_4^{2-}) was also good ($0.77 \leqslant r \leqslant 0.92$, depending upon network) with relatively small error ($25.0\% \leqslant \text{NME} \leqslant 42.0\%$), though slightly negatively biased ($-2.0\% \leqslant \text{NMB} \leqslant -10.0\%$). The quality of ammonium (NH_4^+) simulations is similar to that of SO_4^{2-} ($0.56 \leqslant r \leqslant 0.79$; $-4.0\% \leqslant \text{NMB} \leqslant 14.0\%$; $35.0\% \leqslant \text{NME} \leqslant 63.0\%$). Simulations of nitrate (NO_3^-), elemental carbon (EC) and organic carbon (OC) are relatively poor, as compared to the simulations of the other species. For NO_3^-, the simulation resulted in: $0.37 \leqslant r \leqslant 0.62$; $-16.0\% \leqslant \text{NMB} \leqslant 4.0\%$; $80.0\% \leqslant \text{NME} \leqslant 94.0\%$. For the carbon species, the r ranged from 0.35 (OC) to 0.47 (EC), with fairly large amounts of error (NME = 68.0% for OC, 58.0% for EC) though small amounts of bias (NMB = −6.0% for EC and 12% for OC). The quality of the $PM_{2.5}$ simulations, like $PM_{2.5}$ itself, represented a compilation of the quality of all of the simulated particulate species ($0.51 \leqslant r \leqslant 0.70$, NMB = −3.0% and $45.0\% \leqslant \text{NME} \leqslant 46.0\%$).
Published by Elsevier Ltd.

Keywords: Community multiscale air quality (CMAQ) model; Performance evaluation; Ozone; $PM_{2.5}$; Sulfate; Nitrate; Ammonium; Organic and elemental carbon

1. Introduction

The Clean Air Act and its Amendments require that the US Environmental Protection Agency (EPA) establish National Ambient Air Quality Standards (NAAQS) for ozone (O_3) and fine particulate matter ($PM_{2.5}$) and to assess current and future air quality regulations designed to protect human health and welfare. Air quality models, such as EPA's Models-3 Community

*Corresponding author. Tel.: +1 919 541 3994; fax: +1 919 541 1379.
E-mail address: eder@hpcc.epa.gov (B. Eder).
[1]In partnership with the NERL, US Environmental Protection Agency, RTP, NC 27711, USA.
[2]On assignment from Science and Technology Corporation, Hampton, VA 23666, USA.

1352-2310/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.atmosenv.2005.08.045

Multi-scale Air Quality (CMAQ) model (Byun and Ching, 1999), provide one of the most reliable tools for performing such assessments. CMAQ simulates ambient air concentrations (as well as wet and dry deposition) of numerous pollutants on a myriad of spatial and temporal scales. These model simulations are designed to support both regulatory assessments by EPA program offices as well as scientific studies conducted by research institutions. In order to characterize its performance and to build confidence in its use by the air quality regulatory community, it is essential that CMAQ's ability to simulate concentrations over a wide range of meteorological conditions and geographical areas be tested. Recent CMAQ evaluation studies (Eder et al., 2001, 2002; Seigneur, 2003; Mebust et al., 2003; Zhang et al., 2004) have typically focused on a single geographical area of the country and/or a single episode. While informative, these studies have been unable to cover the breadth of meteorological conditions that occur throughout the year over the entire continental United States (US).

Accordingly, this study provides a comprehensive performance evaluation of the 2004 release of CMAQ (Version 4.4) using an annual simulation (2001) covering the contiguous US. In this evaluation, we compare simulated ambient air concentrations of O_3, $PM_{2.5}$ and various aerosol species including: sulfate (SO_4^{2-}), nitrate (NO_3^-), ammonium (NH_4^+) and elemental carbon (EC) and organic carbon (OC) with measurement data collected by four networks, including: the Interagency Monitoring of PROtected Visual Environments (IMPROVE) network, the Clean Air Status and Trends Network (CASTNet), the Speciated Trends Network (STN) and the Air Quality System (AQS).

2. CMAQ simulation attributes

This evaluation utilized a full, 1-year simulation (2001) using the 2004 release of CMAQ (Version 4.4). The modeling domain covered the contiguous US using a 36 km × 36 km horizontal grid resolution (resulting in 148 $(x) \times 112$ $(y) = 16{,}576$ grid cells) and a 14-layer logarithmic vertical structure (set on a terrain following σ coordinate). The depth of the first layer is ~38 m. The meteorological fields were derived from MM5, the Fifth-Generation Pennsylvania State University/National Center for Atmospheric Research (NCAR) Mesoscale Model (Grell et al., 1994) using the same vertical and horizontal resolution. These fields were then processed using Version 2.2 of Meteorology–Chemistry Interface Program (MCIP). The model simulation used the CB-IV gas-phase chemistry mechanism and the efficient Euler Backward Interactive (EBI) solver. Additional information concerning CMAQ, including updates associated with Version 4.4, can be found at the laboratory's website: http://www.epa.gov/asmdnerl/.

Emissions of the following gas-phase species: sulfur dioxide (SO_2), carbon monoxide (CO), nitrous oxides (NO_x), ammonia (NH_3) and volatile organic compounds (VOCs), were based on EPA's 2001 National Emissions Inventory (NEI) and processed by the Sparse Matrix Operator Kernel Emission (SMOKE) processor. Because current NEI NH_3 emissions are limited to annual estimates (with no intra-annual information), monthly factors developed by Gilliland et al. (2003) were utilized to provide better temporal representation. Primary anthropogenic $PM_{2.5}$ emissions were separated into different species including particle SO_4^{2-}, NO_3^-, OC, EC. Mobile emissions of VOC, CO, NO_x, and particulate matter from cars, trucks, and motorcycles were based on MOBILE6 (US EPA, 2003), while biogenic emissions were obtained from the Biogenic Emission Inventory System (BEIS) Version 3.12.

3. Observational data

A total of four monitoring networks were employed in this evaluation (IMPROVE, STN, CASTNet and AQS), each with its own, and often disparate sampling protocol and standard operating procedures. These differences result in species-specific differences in levels of accuracy, biases, and precision thereby hindering comparability across the networks. This is especially true for NO_3^- and the semi-volatile organic species, where measurement error associated with volatility, interference from gaseous organic species, limitations of analytical methods and lack of calibration standards vary across networks. Because of both the number and complexity of these sampling differences, and because it is beyond the scope of this work, no adjustments have been made to the observations obtained from the networks. The evaluation results will be segregated by network, each of which is briefly described below.

3.1. IMPROVE

The IMPROVE network, which began operations in 1985, represents a collaborative monitoring effort governed by a consortium of Federal, regional, and State organizations. The majority of IMPROVE monitors, which collect 24-h integrated samples every 3rd day (midnight to midnight LST), are located in the western US. For a detailed description of the network, including sampling protocol, the reader is referred to Malm et al. (2004) or the IMPROVE website: http://vista.cira.colostate.edu/improve/. A total of 115 days were available for this evaluation, given CMAQ's 1 year simulation and IMPROVE's 1-in-3 day sampling schedule. IMPROVE species used in this evaluation include $PM_{2.5}$, SO_4^{2-}, NO_3^-, EC and OC.

3.2. STN

The more recently established STN, developed by EPA, follows the protocol of the IMPROVE network (i.e. every 3rd day collection) with the exception that most of the sites are found in urban areas. The number of STN sites available during 2001 varied as the new network was being deployed. STN species used in this evaluation include: SO_4^{2-}, NO_3^-, NH_4^+ and $PM_{2.5}$. Additional network information can be obtained from http://www.epa.gov/air/data/aqsdb.html.

3.3. CASTNet

CASTNet evolved from EPA's National Dry Deposition Network (NDDN) in 1990. The concentration data are collected at predominately rural sites, the majority of which are in the eastern US, using filter packs that are exposed for 1-week intervals (i.e., Tuesday to Tuesday). Given CMAQ's 1 year simulation period and CASTNet's weekly sampling schedule, a total of 51 weekly observations were available from a total of 73 sites. CASTNet species used in this evaluation include: SO_4^{2-}, NO_3^-, and NH_4^+. Because of the length of CASTNet's sampling period, the volatility issues discussed in 3.0 become exacerbated. Accordingly, care must be used in interpretation of the NO_3^- and NH_4^+ data (Sickles II, 1999). Additional information about CASTNet can be obtained from the website http://www.epa.gov/castnet/.

3.4. AQS

The O_3 data employed in this evaluation were obtained from the US EPA's AQS (formerly the Aerometric Information Retrieval System (AIRS)). This data repository contains a multitude of hourly aerometric data, including O_3 concentrations, collected by Federal, State, and local agencies at thousands of locations nationwide. For this evaluation over 1000 AQS monitors were employed, the majority of which are in eastern locations. Additional information can be obtained from http://www.epa.gov/air/data/aqsdb.html.

4. Statistics

Because of the considerable differences in sampling protocols discussed in 3.0, evaluation statistics were calculated separately for each network. The CMAQ model output species were post processed in order to achieve compatibility with the observation species. The statistics were calculated for observations and model results that were paired in space (without interpolation) and time (daily or weekly, depending on the sampling period of each network). It must be noted that the measurements are made at specific locations, whereas CMAQ concentrations represent a volume-average value (corresponding to the volume of the grid cell). This discrepancy in spatial representativeness, referred to as incommensurability, is a fundamental source of uncertainty when evaluating models, especially in urban areas, where subgrid-scale gradients are more likely to be important (Seigneur, 2001).

The array of statistical metrics that have been developed for use in model evaluation continues to expand and can be overwhelming. Depending on the measure of interest, bias for example, numerous variations exist (i.e. mean bias (MB), mean normalized bias, normalized mean bias (NMB), fractional bias, etc.) each with their own advantages and disadvantages. For this evaluation, two standard and well-accepted measures of model bias were selected; the MB and NMB. Likewise, two accepted measures of model error; the root mean-square error (RMSE) and normalized mean error (NME), were selected. These metrics, which provide both actual (i.e. measured in either ppb or $\mu g\,m^{-3}$) and normalized (%) measures of performance, are defined below

$$\mathrm{MB} = \frac{1}{N}\sum_{1}^{N}(C_\mathrm{m} - C_\mathrm{o}), \tag{1}$$

$$\mathrm{NMB} = \frac{\sum_{1}^{N}(C_\mathrm{m} - C_\mathrm{o})}{\sum_{1}^{N} C_\mathrm{o}} 100\%, \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{1}^{N}(C_\mathrm{m} - C_\mathrm{o})^2}, \tag{3}$$

$$\mathrm{NME} = \frac{\sum_{1}^{N}|C_\mathrm{m} - C_\mathrm{o}|}{\sum_{1}^{N} C_\mathrm{o}} 100\%, \tag{4}$$

where C_m and C_o are modeled and observed concentrations, respectively.

As seen in Eqs. (2) and (4), the normalization is achieved by dividing by the sum of observed concentrations (as opposed to dividing by individual observations), thereby avoiding the inflation that other metrics (i.e. mean normalized bias or mean normalized error) are susceptible to when applied to small concentrations.

5. Evaluation results

Six month "seasonal" averages (April–September) of the four metrics (along with other summary statistics) for both the peak 1- and 8-h O_3 concentrations are provided in Table 1. Annual averages of the four metrics for each aerosol species and for each network are provided in Table 2. Scatter plots of monthly aggregated CMAQ simulation results (ordinate) versus observations (abscissa) are also provided for each specie, with factor of two reference lines (Fig. 1 (O_3) and Fig. 4 (aerosol species)). For the aerosol species, a distinction is made between locations east of 100°W (shown in color) and those west of 100°W (shown in black) in order to facilitate interpretation. For O_3, only 1 month is shown (June 2001, which is representative of the other months) in order to prevent the overwhelming number of simulated-observed pairs associated with the full O_3 season from rendering the figure ineffective. To facilitate interpretation of the results, values of the NMB and NME are also provided over space (Fig. 2 for O_3, Figs. 5, 6 for aerosol species) and over time (Fig. 7) for O_3 and the aerosol species.

Table 1
Summary statistics associated with CMAQ simulations of the peak 1- and 8-h ozone concentrations

Metric	Peak 1-H	Peak 8-H
Number	195,462	193,628
Mean modeled	59.3	54.4
Mean observed	57.0	50.3
r	0.68	0.69
MB (ppb)	2.3	4.1
NMB(%)	4.0	8.1
RMSE (ppb)	14.0	12.9
NME(%)	18.3	19.6

5.1. Ozone

Because of the tremendous amount of O_3 data available (over 4 million hourly observations) and for the sake of brevity, this evaluation focuses on the peak 1- and peak 8-h concentrations. Examination of Table 1 and Fig. 1 reveals that this latest release of CMAQ simulates O_3 concentrations quite well, as correlations approach 0.70 for both the peak 1- and peak 8-h concentrations. These values are consistent with previous evaluations (Eder et al., 2002). An overwhelming majority of simulated peak 8-h concentrations lie within the factor of two references lines (the other months and the peak 1-h scatter plots are very similar). A notable exception to this general agreement occurs when CMAQ over predicts low observed concentrations ($<$ 40 ppb).

The NMBs are both positive, but small (4.0% for 1-h, 8.1% for 8-h). When examined over space (Fig. 2, left panel (peak 8-h only)), the NMBs reveal several interesting features. Most notable is the tendency for CMAQ to overpredict (NMB generally > 15% and often > 30%) concentrations along coastal regions. Explanations for this tendency may be tied to poor representation of coastal boundary layers and their interaction with land/sea breezes by MM5, the meteorology model (Gilliam et al., 2005). Additional interesting spatial features include a small cluster of overprediction (NMB > 30%) found in/near the state Iowa, and a cluster of underpredictions (NMB < −15%) found in sections of southern California and Arizona.

The NMEs associated with simulations of O_3 are relatively small as well, with both the peak 1- and 8-h concentrations, producing NME < 20%. When examined over space (Fig. 2, right panel (peak 8-h only)), the NMEs mimic the NMBs in that poorer performance is found over coastal regions, especially from South Carolina to Texas. The poor performance seen in/near Iowa in terms of bias is also reflected in the error as the NME exceeds 45%.

Table 2
Summary statistics for each species/ network combination

Species	Metric		IMPROVE		CASTNET		STN	
$PM_{2.5}$	Number		13,217				6419	
	Mean modeled		5.64				12.55	
	Mean observed		5.81				12.89	
	r		0.70				0.51	
	MB ($\mu g\,m^{-3}$)	NMB (%)	−0.17	−3.0			−0.34	−3.0
	RMSE ($\mu g\,m^{-3}$)	NME (%)	4.11	45.0			8.83	46.0
SO_4	Number		13,447		3736		6970	
	Mean modeled		1.60		2.88		3.33	
	Mean observed		1.69		3.21		3.40	
	r		0.85		0.92		0.77	
	MB ($\mu g\,m^{-3}$)	NMB (%)	−0.09	−5.0	−0.32	−10.0	−0.07	−2.0
	RMSE ($\mu g\,m^{-3}$)	NME (%)	1.28	39.0	1.15	25.0	2.25	42.0
NH_4	Number				3736		6970	
	Mean modeled				1.12		1.44	
	Mean observed				1.16		1.26	
	r				0.79		0.56	
	MB ($\mu g\,m^{-3}$)	NMB (%)			−0.04	−4.0	0.17	14.0
	RMSE ($\mu g\,m^{-3}$)	NME (%)			0.58	35.0	1.27	63.0
NO_3	Number		13,398		3735		6130	
	Mean modeled		0.50		1.04		1.48	
	Mean observed		0.48		0.99		1.77	
	r		0.52		0.67		0.37	
	MB ($\mu g\,m^{-3}$)	NMB (%)	0.02	4.0	0.05	5.0	−0.29	−16.0
	RMSE ($\mu g\,m^{-3}$)	NME (%)	0.99	94.0	1.11	71.0	2.94	80.0
EC	Number		13,441					
	Mean modeled		0.22					
	Mean observed		0.24					
	r		0.47					
	MB ($\mu g\,m^{-3}$)	NMB (%)	−0.02	−6.0				
	RMSE ($\mu g\,m^{-3}$)	NME (%)	0.27	58.0				
OC	Number		13,427					
	Mean modeled		1.26					
	Mean observed		1.12					
	r		0.35					
	MB ($\mu g\,m^{-3}$)	NMB (%)	0.14	12.0				
	RMSE ($\mu g\,m^{-3}$)	NME (%)	1.59	68.0				

When examined over time, CMAQ's performance is fairly consistent, especially in terms of the NME (both the peak 1- and peak 8-h concentrations, right panel) as monthly averaged values range between 15% and 20%. The NMB does reflect some temporal variability, most notably in July, when CMAQ's general tendency to over predict is somewhat abated. Also, there is systematically more bias (~5%) and error (~2%) associated with the peak 8-h concentration as compared to the peak 1-h concentration. This is attributable to the fact that computation of the 8-h value generally includes evening hours and that CMAQ, like most models, has difficulty simulating the evolution of the nocturnal boundary layer and its subsequent impact on surface O_3 concentrations (Eder et al., 2005) (Fig. 3).

5.2. Sulfate

Sulfate is one of two species examined in this evaluation (NO_3^- being the other) that is measured by three networks, thereby affording a thorough evaluation, which for the most, part reveals relatively good performance. Correlation coefficients associated with each data set are high (Table 2), ranging from 0.77 (STN) to 0.92 (CASTNet) with the vast majority of the aggregated monthly simulations falling within a factor of two of the observations (Fig. 4). The biases, while all negative, are small, with the annual NMBs ranging from—10% (CASTNet) to −2% (STN). The errors are relatively small as well, with annual NMEs ranging from 25% (CASTNet) to 42% (STN). Examination of the NMB and NME over space (Fig. 5) reveals better performance over the eastern half of the domain, where the majority of NMBs lie within ±25% and the NMEs are less than 50%. Performance degrades somewhat in the west, especially in California where NME generally exceeds 50% and CMAQ significantly underpredicts concentrations (NMB < −25%). This underprediction was also evident in the scatter plot, where a small cluster of western sites fell outside the factor of two reference lines (Fig. 6).

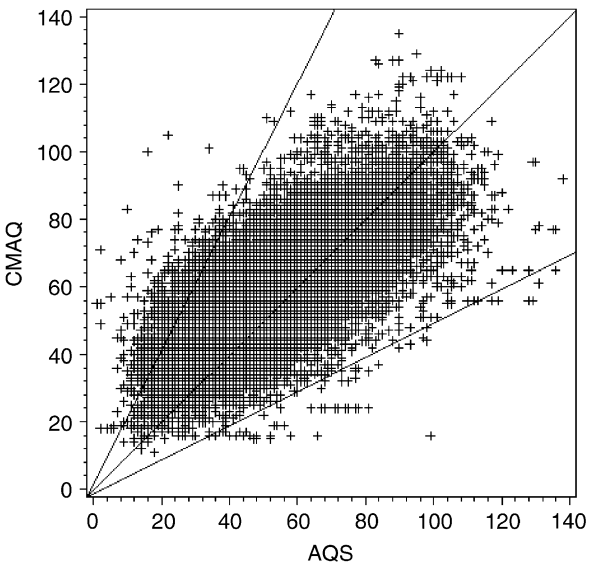

Fig. 1. Scatter plot of peak 8-h O_3 simulations versus AQS observations for the month of June 2001. 1:2, 1:1 and 2:1 reference lines are provided.

Examination of the domain-wide bias over time (Fig. 7) reveals a strong temporal trend in which performance is considerably poorer during the beginning of the year as NMBs (for each network) are large and negative. These negative biases decrease in magnitude throughout the spring and summer, briefly becoming positive during the autumn. Reasons for this trend may be tied to CMAQ's simulation of SO_4^{2-} wet deposition that has been found to be excessive during the winter months (personal communication, Robin Dennis, March, 2005). The error associated with CMAQ simulations of SO_4^{2-}, while considerable, are much more consistent over time. The only trend is found across networks, where errors calculated against CASTNet stations are considerably smaller (NME = 25%) than those associated with either STN or IMPROVE (NME ≈ 40%). This reduction in error is most likely attributable to the longer sampling time (weekly integrated averages versus 24 h averages) used by CASTNet and is not necessarily representative of improved performance.

The relative success of CMAQ in simulating SO_4^{2-} well, especially in eastern sections of the US, is not surprising, given CMAQ's derivation from the Regional Acid Deposition Model (Chang et al., 1990), which was developed to address problems associated with acid rain in the eastern US and Canada.

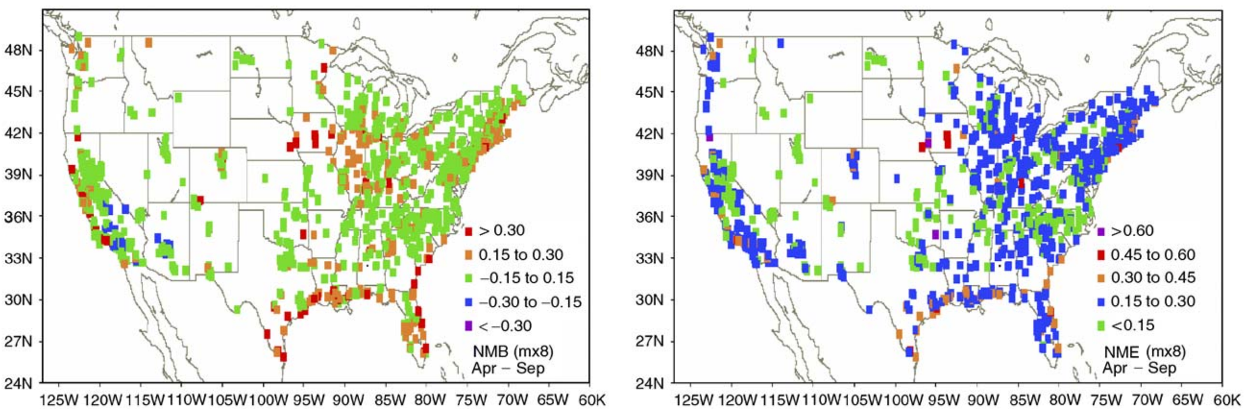

Fig. 2. Spatial plots of NMB (left panel) and NME (right) for the peak 8-h O_3 concentration for the 6 months period April–September, 2001.

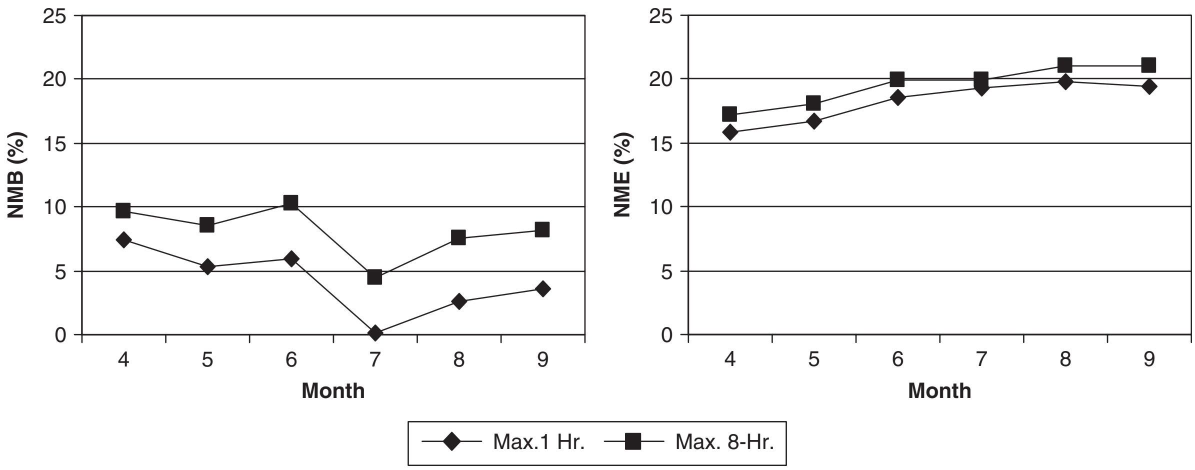

Fig. 3. Monthly plots of domain-wide NMB (%) (left panels) and NME (%) (right panels) for both the peak 1- and 8-h O_3 concentrations.

5.3. Nitrate

As with most deterministic models, CMAQ has historically had a difficult time simulating concentrations of NO_3^- accurately. This is due, in part, to volatility issues associated with NO_3^-, and their exacerbation due to uncertainties associated with SO_4^{2-} and total NH_4^+ simulations (which play a major role in determining the partitioning of NO_3^- (Yu et al., 2005)). The impact of these uncertainties is reflected in the scatter plot (Fig. 4), which depicts considerable scatter with many simulated values falling outside the factor of two reference lines. This is especially true of STN sites, which depict over-predicted in eastern sections of the domain and under-prediction in western section. The correlations reflect this scatter, ranging from 0.37 (STN) to 0.67 (CASTNet). The NMEs are considerably larger than for the other species as well, ranging from 71.0% (CASTNet) to 94.0% (STN). These values are however; considerably lower than previous CMAQ releases, where NMEs often exceeded 100% (depending on the simulation/evaluation configuration) (Eder et al., 2002, 2003; Mebust et al., 2003). When examined over space, the NMEs exhibit little, if any, spatial difference, with the possible exception of slightly better performance over the Ohio Valley and Great Lakes regions.

The 2004 release of CMAQ has also produced much smaller NO_3^- biases than previous releases, resulting in annual NMBs that range from −16% (STN) to 5% (CASTNet). This large reduction in bias is somewhat misleading, however; in that the NMB exhibits substantial and compensating spatial and temporal differences. For example, CMAQ tends to overpredict NO_3^- concentrations in the eastern domain, where NMBs often exceed 25% (and even 75%), while it tends to underpredict in most western locations, where NMBs generally range between −25% and −75%. This pattern of substantial and compensating performance is also evident when examining NO_3^- simulations over time, which is depicted in Fig. 6. The bias is positive and large during the transitional seasons, with peaks in NMB$\geqslant$40% during the months of April and October. During the summer and winter months, however; the bias becomes negative and large as NMBs approach and fall below −40%. This complex seasonal pattern is also evident in monthly plots of the error. NMEs exceed 50% for each month/network combination, peaking (as with the NMB) in the transitional months of April and October where they approach or exceed 100%.

Reasons for these large monthly biases and errors appear to be related to the NH_3 emission inputs used in this 2001 simulation. Currently, NEI NH_3 emission inventories are limited to annual estimates, with no intra-annual variability. In an attempt to alleviate this temporal limitation, Gilliland et al. (2003), developed an inverse modeling approach that provided estimates of monthly NH_3 emissions. More recent research (Gilliland et al., 2005) indicates that approach, while providing the much needed, temporally resolved estimates, may have inadvertently introduced over-compensating errors in the resulting emission adjustments. Such errors result in erroneous NH_4^+ simulations (discussed in Section 5.4), which can propagate into the NO_3^-

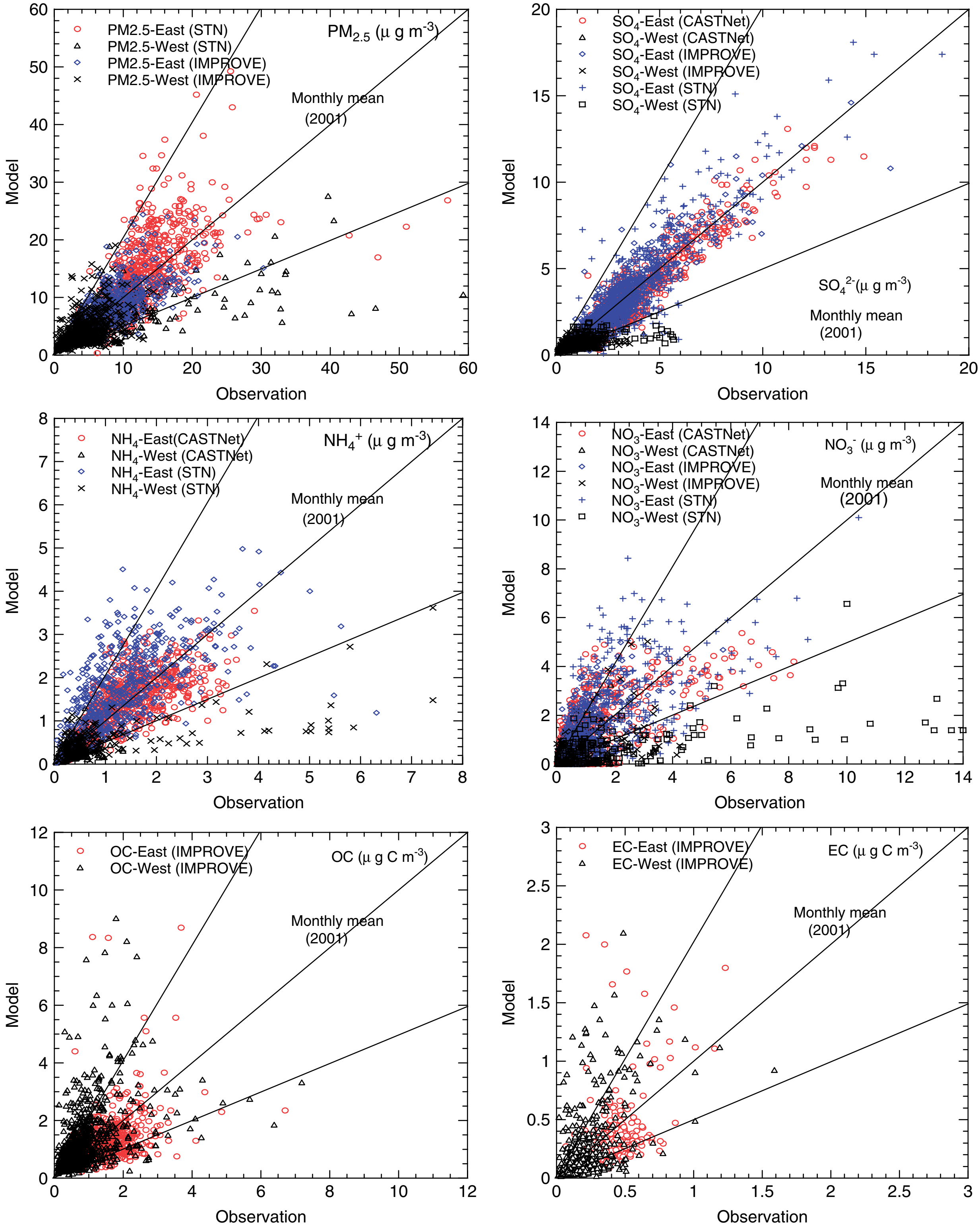

Fig. 4. Scatterplots of simulations (abscissa) versus observations (ordinate) for $PM_{2.5}$ (top left panel), SO_4 (top right), NH_4 (middle left) NO_3 (middle right), OC (bottom left) and EC (bottom right) for each available network. 1:1, 1:2 and 2:1 reference lines are provided. Black symbols denote western locations (Long > 100° W), color symbols eastern locations.

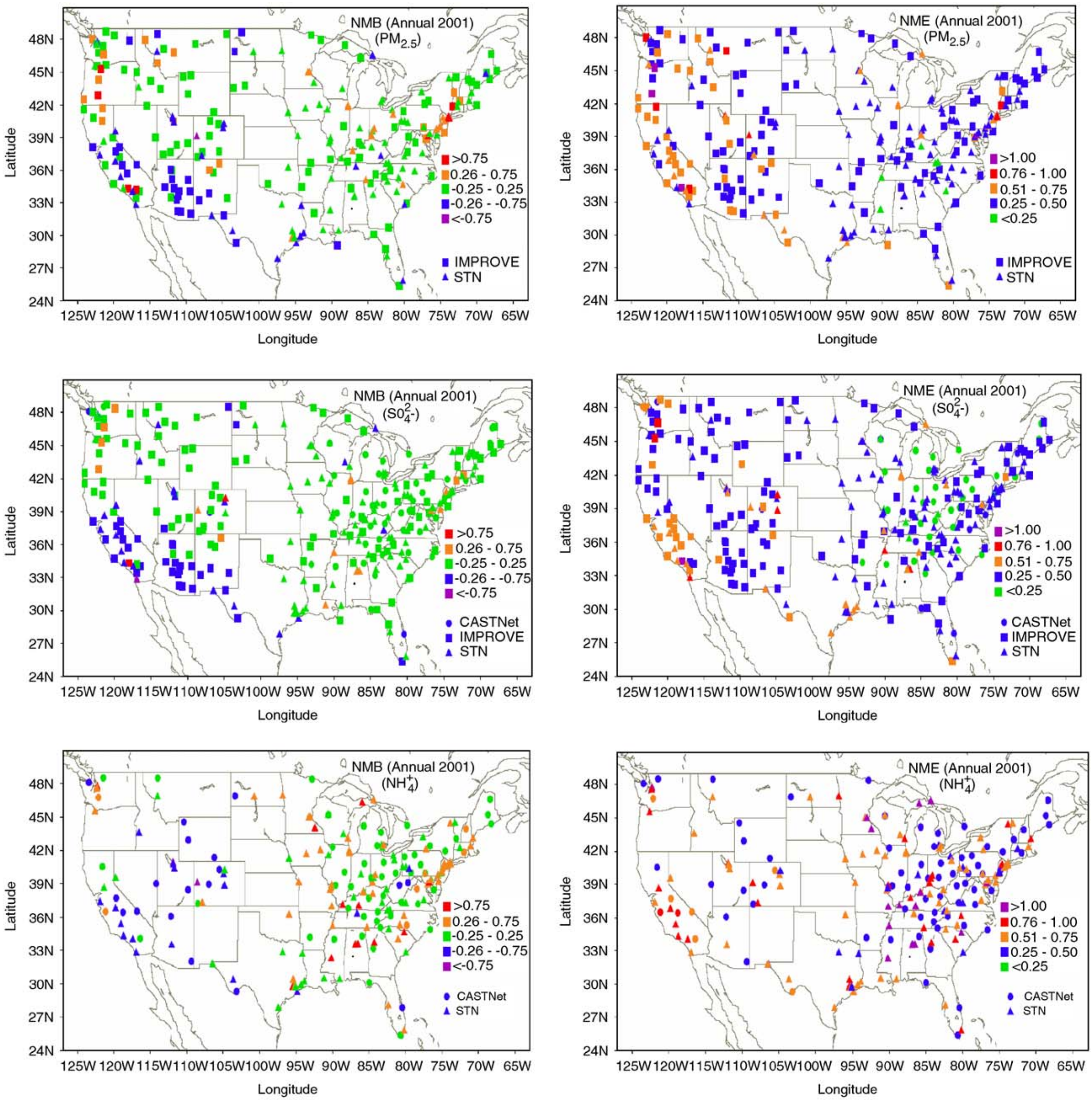

Fig. 5. Spatial plots of NMB (left panels) and NME (right panels) for $PM_{2.5}$ (top panels) SO_4 (middle panels) and NH_4 (bottom panels).

simulations. Gilliland et al. (2005) have developed a more recent series of monthly NH_3 emissions that seem to improve the simulations (not shown).

5.4. Ammonium

Examination of the scatter plot found in Fig. 4 reveals that the quality of NH_4^+ simulations is similar to, though somewhat better than that for NO_3^-, as a greater number of aggregated monthly simulated concentrations lie within a factor of two of the observations. The annual correlations reflect this improvement as well, ranging from 0.56 (STN) to 0.79 (CASTNet). As with NO_3^-, the annual NMBs associated with the NH_4^+ simulations are also quite small (−4% for CASTNet and 14% for STN). These annual values are also somewhat misleading, (just as those associated with NO_3^-), because the NH_4^+ simulations also exhibit substantial and compensating spatial and temporal

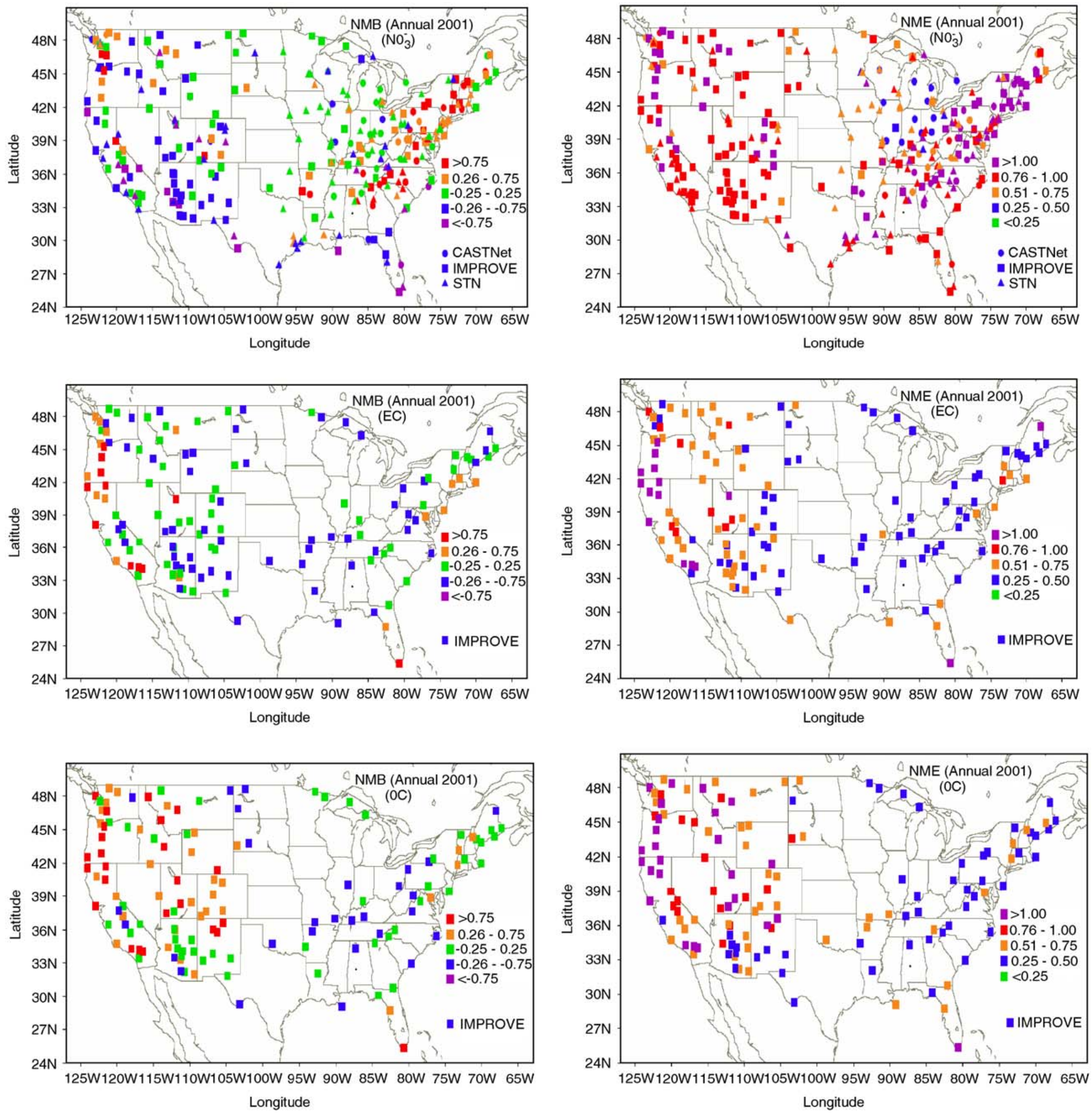

Fig. 6. Spatial plots of NMB (left panels) and NME (right panels) for NO_3 (top panels) OC (middle panels) and EC (bottom panels).

differences in their NMBs. Spatially, CMAQ tends to overpredict in eastern sections of the country, especially against STN sites where some sites have NMBs exceeding 75%. In the west, conversely, CMAQ tends to underpredict, with most sites recording biases between −25% and −75%. Temporally, the bias pattern is nearly identical to that of NO_3^- in that large positive NMBs are found in April, October and November (NMB$\geqslant$40%) with large negative (and compensating) NMBs in January and December (−20% to −40%).

These spatial and temporal differences in performance are also evident in plots of the NME, which annually, average between 35% (CASTNet) and 63.0% (STN). Spatially, the NME tends to be smaller, though still substantial in the east, where roughly half of the sites fall between 25% and 50%, the other half between 50% and 75%. In the west,

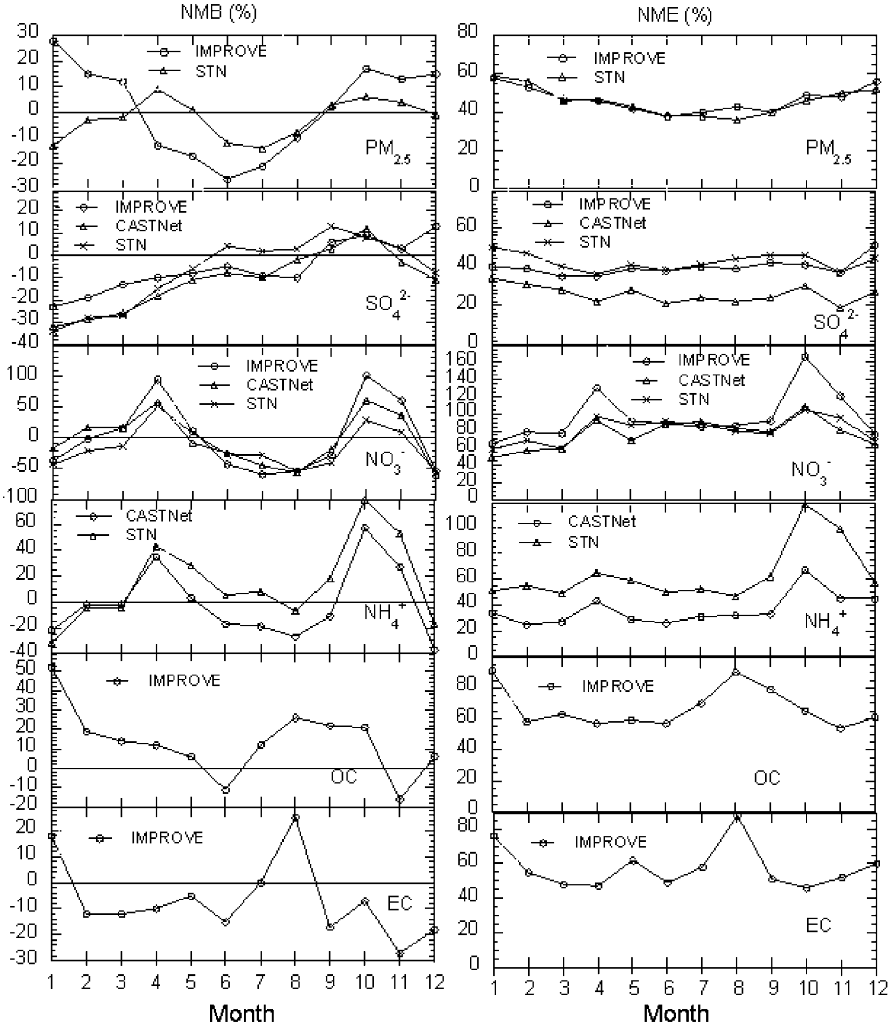

Fig. 7. Monthly plots of domain-wide NMB (%) (left panels) and NME (%) (right panels) for each of the aerosol species.

the majority of sites have NME that exceed 50%, with several (mostly along the west coast) exceeding 75%. Temporally, simulations of NH_4^+ result in the largest NME during the months of April, October and November. As with NO_3^-, these complex patterns are most likely related to the inverse modeling used in estimation of monthly NH_3 emissions discussed in Section 5.3.

5.5. Elemental and organic carbon

The quality of EC and OC simulation results, while similar, are both somewhat lacking. This result is not surprising, however; given the large uncertainties associated with (1) the emissions inventories and (2) the state-of-the-science concerning carbonaceous aerosols, especially OC, which is a complex mixture of hundreds of organic compound, each with different formation mechanisms. Correlations of the two species range from 0.35 for OC to 0.47 for EC (Table 2), with many monthly aggregated values falling outside the factor of two lines as seen in Fig. 4 (especially for OC and especially for western sites). This considerable scatter is reflected in the annual NMEs, which range from 58% for EC to 68% for OC (annually). Across space (Fig. 6), the NMEs associated with each specie are similar in that smaller errors are found in eastern locations (NME generally < 50%) when compared to western locations (NME > 50%). The poorest performance is found along the west coast where most NMEs exceed 75%. Similarities in the error associated with each carbonaceous species are also evident across time, as CMAQ simulations generally produce NMEs between 50% and 75%. The exception being the months January and August, when the NMEs are considerably larger (>75% for both OC and EC).

In terms of bias, there are more differences between the two species than there are similarities, as CMAQ generally overpredicts OC concentrations (annual NMB = 12%), while underpredicting EC (annual NMB = −6%). Temporally, the overprediction of OC is evident in all but 2 months (June and November) and is greatest in January and August (Fig. 7). Underprediction of EC occurs in all but 3 months, with two of these months (January and August) actually resulting in overprediction (NMBs near or above 20%).

An explanation as to why CMAQ performed more poorly for both EC and OC during the months of January and August appears to be tied to the flawed temporal allocation of wildfire emissions associated with several large fires that impacted the western US during 2001 (Roy et al., 2005). Wildfires, which are a major emissions source of both EC and OC, were observed in California during January and in California, Oregon, Montana and Nevada during August. The impact of these western wildfires is also evident when the NMBs are examined spatially as CMAQ's performance is generally better over the eastern sections of the domain for both carbonaceous species (many NMBs within ±25%; all within ±75%). Across most of the western states, biases associated with OC are positive and often large (many NMB⩾75%). Large positive biases associated with EC are generally confined to the West Coast States.

5.6. $PM_{2.5}$

The quality of CMAQ simulations of $PM_{2.5}$, like $PM_{2.5}$ itself, represents a compilation of the quality of all of the simulated particulate species. Overall, the performance is fairly good as the majority of the simulation results lie within a factor of two of the observations as seen in the scatterplot (Fig. 4). The correlations do vary with networks, however, ranging from 0.51 (STN) to 0.70 (IMPROVE). On an annual basis, the overall bias is very small and identical for each network (NMB = −3.0%). This annual metric is misleading, however, because when examined over time and space, the NMBs exhibit marked differences across months, the domain, as well as the two networks. Against IMPROVE sites (which are mostly remote sites located in the west), the model underpredicts concentrations during the period April through August, while overpredicting during the period October–March. Against STN sites (which again are mostly urban though distributed equally across the domain) the model underpredicts (though by smaller amounts) during the June–August period and is fairly unbiased the remaining months (with the exception of January, where the model underpredicts). Reasons for the discrepancy in performance during the month of January are intriguing though not understood at this time.

Spatially, the bias, though again generally small (most $PM_{2.5}$ simulation results across most of the domain are within ±25%), does exhibit considerable variability (Fig. 5). There are several areas of concentrated positive bias, most notably along the northeast coastline and in and near the state of Oregon. Conversely, a concentrated area of negative bias can be found stretching from southern Texas and New Mexico into Arizona and central California.

The error associated with CMAQ $PM_{2.5}$ simulations (annual NME = 45% and 46% for IMPROVE and STN, respectively) is considerably more consistent across time and the two networks (Fig. 7). Temporally, the NMEs are slightly smaller during the summer months (generally⩽40%) as compared to the winter months (NME⩾50%) for both networks. Examination of the error across space (Fig. 3) reveals somewhat better performance across the eastern two-thirds of the domain (most NME<50%), with some degradation of performance in the West Coast States, where most NMEs exceed 50%.

6. Summary

A performance evaluation of the 2004 release of the Models-3 CMAQ model (Version 4.4) has been presented that compares an annual simulation (2001) covering the contiguous United States against monitored data from four nationwide networks. This effort, which represents one of the most spatially and temporally comprehensive performance evaluations of the model, reveals that CMAQ varies in its ability to simulate ambient air concentrations of critical gas and particulate matter species. Simulations of the peak 1- and 8-h O_3 concentrations during the "O_3 season" (April–September) were good ($r = 0.68$, 0.69; normalized mean bias (NMB) = 4.0%, 8.1%; and normalized mean error (NME) = 18.3%, 19.6%) respectively. The annual simulations of SO_4^{2-} were also good ($0.77 \leqslant r \leqslant 0.92$, depending upon network) though slightly negatively biased ($-2.0\% \leqslant NMB \leqslant -10.0\%$) with relatively small error ($25.0\% \leqslant NME \leqslant 42.0\%$). Spatially, CMAQ's performance was better over the eastern half of the US, while temporally, the performance was somewhat degraded during the winter months. The performance of CMAQ's NO_3^- simulations, though still lagging that of O_3 and SO_4^{2-}, has shown marked improvement over previous releases. The correlations reflect this progress ($0.37 \leqslant r \leqslant 0.62$, depending upon network) as do the measures of error ($80.0\% \leqslant NME \leqslant 94.0\%$) and measures of bias ($-16.0\% \leqslant NMB \leqslant 4.0\%$), which have shown the largest, though somewhat misleading, improvement. Misleading in that when examined over space and time, the NO_3^- simulations exhibit large, though often compensating NMBs, which are thought to be attributable to an incomplete understanding of ammonia emissions. The quality of NH_4^+ simulations is similar to, though somewhat better than that for NO_3^-. Correlations range from 0.56 to 0.79, depending on the network. The model produces relatively modest amounts of error ($35.0\% \leqslant NME \leqslant 63.0\%$) and even less, though again somewhat misleading bias ($-4.0\% \leqslant NMB \leqslant 14.0\%$).

The quality of simulations of EC and OC, while similar, are both fairly poor, which is not surprising given the level of uncertainties associate with emissions and the current state-of-the-science. Correlations range from 0.35 (OC) to 0.47 (EC). The model produces fairly large, though not unreasonable amounts of error (NME = 68.0% for OC, 58.0% for EC) and encouragingly small amounts of bias (NMB = −6.0% for EC and 12% for OC). The quality of CMAQ simulations of $PM_{2.5}$, much like $PM_{2.5}$ itself, represents a compilation of the quality of all of the simulated particulate species. Overall, the performance of this release signifies a marked improvement over previous releases as correlations range from 0.51 to 0.70, depending on network. The annual bias is very small and identical for each network (NMB = −3.0%) and the error, though improved, is still considerable ($NME \approx 45\%$).

Acknowledgements

The authors would like to thank members of NERL's MEARB branch, especially Alfreida Torian and Steven Howard for the processing of the model and observation datasets.

Disclaimer: The research presented here was performed under the Memorandum of Understanding between the US Environmental Protection Agency (EPA) and the US 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.

References

Byun, D., Ching, J., 1999. Science algorithms of the EPA Models-3 Community Multi-scale Air Quality (CMAQ) modeling system, EPA-600/R-99/030. US EPA, US Government Printing Office, Washington, DC.

Chang, J., et al., 1990. The regional acid deposition model and engineering model. National Acid Precipitation Assessment Program, Acidic Deposition, State of Science and Technology, vol. 1, NAPAP SOS/T Rep. 4, Washington, DC.

Eder, B., Mebust, M., LeDuc, S., 2001. A Preliminary evaluation of Models-3 CMAQ using particulate matter data from the IMPROVE network. In: 25th NATO/CCMS International Technical Meeting on Air Pollution Modeling and Its Applications, 15-19 October 2001, Louvain-la-Neuve, Belgium, pp. 327–334 (Reprints).

Eder, B., Yu, S., Dennis, R., Pleim, J., Schere, K., 2002. Testing of the June 2002 version of EPA's CMAQ. In: US EPA-EMEP Workshop, 7–11 October, Bad Breisig, Germany.

Eder, B., Yu, S., Dennis, R., 2003. An Evaluation of the MODELS-3 CMAQ Aerosol Module 2003 AAAR PM Meeting Particulate Matter: Atmospheric Sciences, Exposure and the Fourth Colloquium on PM and Human Health, 31 March–4 April, 2003, Pittsburgh, PA.

Eder, B., Kang, D., Mathur, R., Yu, S., Schere, K., 2005. An operational evaluation of the Eta-CMAQ air quality forecast model. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2005.12.062.

Gilliam, R., Hogrefe, C., Rao, S., 2005. New methods for evaluating meteorological models used in air quality applications. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2006.01.023.

Gilliland, A., Dennis, R., Roselle, S., Pierce, T., 2003. Seasonal NH_3 emission estimates for the eastern United States based on ammonium wet concentrations and an inverse modeling method. Journal of Geophysical Research 108 (D15), 4477.

Gilliland, A., Roselle, S., Pinder, R., Dennis, R., 2005. Seasonal NH_3 emissions for an annual CMAQ simulation: inverse model estimation and evaluation. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2005.12.066.

Grell, G., Dudhia, J., Stauffer, D., 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.

Malm, W., Schichtel, B., Pitchford, M., Ashbaugh, L., Eldred, R., 2004. Spatial and monthly trends in speciated fine particle concentrations in the United States. Journal of Geophysical Research 109, D03306.

Mebust, M., Eder, B., Binkowski, B., Roselle, S., 2003. Model-3 CMAQ model aerosol component, 2. Model evaluation. Journal of Geophysical Research 108 (D6).

Roy, B., Gilliland, A., Pierce, T., Howard, S., Benjey, W., 2005. Using the MODIS rapid response fire products to refine wildland fire emission estimates. Atmospheric Environment, submitted for publication.

Seigneur, C., 2001. Current status of air quality models for particulate matter. Journal of Air and Waste Management Association 51, 1508–1521.

Seigneur, C., 2003. Review of CMAQ and REMSAD performance for regional PM modeling. Document Number CP163-03-01b; Report prepared for Utility Air Regulatory Group, Atlanta, GA, 2003.

Sickles II, J., 1999. A summary of airborne concentrations of sulfur and nitrogen containing pollutants in the Northeastern US. Journal of Air and Waste Management Association 49, 882–893.

US Environmental Protection Agency, 2003. User's guide to MOBILE6.1 and MOBILE6.2; EPA Office of Air and Radiation, EPA420-R-03-010, Assessment and Standards Division, Office of Transportation and Air Quality, US EPA, 262 pp.

Yu, S., Mathur, R., Kang, D., Schere, K., Eder, B., Pleim, J., 2005. Performance and diagnostic evaluations of a real-time ozone forecast by the Eta-CMAQ model suite during the 2002 New England Air Quality Study. Atmospheric Environment, in press.

Zhang, Y., Pun, B., Wu, S., Vijayaraghavan, K., Seigneur, C., 2004. Application and evaluation of two air quality models for particulate matter for a Southeastern US episode. Journal of Air and Waste Management Association 54, 1478–1493.