

FINAL REPORT CRC Project A-44

PERFORMANCE EVALUATION OF FOUR AIR QUALITY MODELS

# APPLIED FOR AN ANNUAL SIMULATION OF PM OVER THE WESTERN UNITED STATES

Prepared by

Christian Seigneur, Betty Pun, Shu-Yun Chen and Kristen Lohman

Atmospheric & Environmental Research, Inc. 2682 Bishop Drive, Suite 120 San Ramon, CA 94583

Prepared for

ENVIRON

101 Rowland Way Novato, CA 94945

Document CP152-04-01

October 2004

# **TABLE OF CONTENTS**

Exe	cutive Su	ımmary	E-1
	E.1	Overview	E-1
	E.2	Measurement Databases	E-2
	E.3	Summary of Model Performance	E-3
	E.4	Comparison of Performance among CAMx, CAMQ and REMSAD	E-3
	E.5	CAMx M4 versus CAMx 4Sec	E-4
	E.6	Recommendations	E-5
1.	Intro	duction	1-1
2.	Over	view of the Measurement Databases	2-1
	2.1	AQS	2-1
	2.2	IMPROVE	2-2
	2.3	CASTNet	2-8
	2.4	NADP	2-10
	2.5	Summary of Measurement Uncertainties	2-10
3.	Perfo	ormance Evaluation Measures	3-1
4.	Resu	Its of the Model Performance Evaluation	4-1
	4.1	Ozone	4-1
	4.2	Sulfate	4-7
	4.3	Nitrate	4-18
	4.4	Organic Matter	4-29
	4.5	Black Carbon	4-40
	4.6	PM <sub>2.5</sub>	4-47
	4.7	Coarse Mass	4-56
	4.8	PM <sub>10</sub>	4-61

	4.9	Wet Deposition	. 4-66
5.	Asse	ssment of Model Performance	5-1
	5.1	Overall Assessment of Model Performance	5-1
	5.2	CMAx M4 versus CAMx 4Sec	5-2
	5.3	Comparison of Performance among CMAx, CMAQ and REMSAD	5-5
	5.4	Recommendations	5-6
6.	Conc	lusion	6-1
7.	Refe	rences	7-1

# LIST OF TABLES

Table 2-1.	Uncertainties associated with ambient concentration measurements	2-11
Table 3-1.	Statistical metrics for model performance	3-2
Table 4-1.	Annual performance statistics for ozone concentrations (AQS)	4-3
Table 4-2.	Annual performance statistics for fine sulfate concentrations (IMPROV	/E) 4-9
Table 4-3.	Annual performance statistics for sulfate concentrations (CASTNet)	4-14
Table 4-4.	Annual performance statistics for fine nitrate concentrations (IMPROV	'E)4-20
Table 4-5.	Annual performance statistics for total nitrate concentrations (CASTNe	et).4-25
Table 4-6.	Annual performance statistics for fine organic matter concentrations	
	(IMPROVE, scaling factor of 1.4).	4-31
Table 4-7.	Annual performance statistics for fine organic matter concentrations	
	(IMPROVE, scaling factor of 2.1).	4-32
Table 4-8.	Annual performance statistics for fine black carbon concentrations	
	(IMPROVE).	4-42
Table 4-9.	Annual performance statistics for measured PM <sub>2.5</sub> concentrations	
	(IMPROVE)	4-48
Table 4-10.	Annual performance statistics for reconstructed PM <sub>2.5</sub> concentrations	
	(IMPROVE).	4-49
Table 4-11.	Annual performance statistics for coarse PM mass concentrations	
	(IMPROVE).	4-57
Table 4-12.	Annual performance statistics for PM <sub>10</sub> mass concentrations	
	(IMPROVE).	4-62
Table 4-13.	Annual performance statistics for sulfate wet deposition (NADP)	4-67
Table 4-14.	Annual performance statistics for nitrate wet deposition (NADP)	4-68
Table 4-15.	Annual performance statistics for ammonium wet deposition (NADP).	4-69

# LIST OF FIGURES

Figure 2-1.	Typical sampling efficiency for a PM <sub>2.5</sub> sampler with size segregation based
	on inertial impaction
Figure 2-2.	Scatter diagram of measured and reconstructed PM <sub>2.5</sub> concentrations from
	IMPROVE (55 sites in the western United States during 1996)2-7
Figure 2-3.	Schematic representation of CASTNet data and their use in model
	performance evaluation
Figure 4-1.	Scatter diagrams of observed and simulated ozone concentrations (AQS) 4-4
Figure 4-2.	Seasonal performance statistics for ozone concentrations (AQS) 4-5
Figure 4-3.	Scatter diagrams of observed and simulated fine sulfate concentrations
	(IMPROVE)
Figure 4-4.	Spatial display of the mean normalized error for sulfate (IMPROVE) 4-11
Figure 4-5.	Seasonal performance statistics for fine sulfate concentrations
	(IMPROVE)
Figure 4-6.	Scatter diagrams of observed and simulated sulfate concentrations
	(CASTNet)
Figure 4-7.	Seasonal performance statistics for sulfate concentrations (CASTNet) 4-16
Figure 4-8.	Scatter diagrams of observed and simulated fine nitrate concentrations
	(IMPROVE)
Figure 4-9.	Spatial display of the mean normalized error for nitrate (IMPROVE) 4-22
Figure 4-10	Seasonal performance statistics for fine nitrate concentrations
	(IMPROVE)
Figure 4-11	Scatter diagrams of observed and simulated total nitrate concentrations
	(CASTNet)
Figure 4-12	2.Seasonal performance statistics for total nitrate concentrations
	(CASTNet)
Figure 4-13	3. Scatter diagrams of observed and simulated fine organic matter
	concentrations (IMPROVE, scaling factor of 1.4)
Figure 4-14	. Scatter diagrams of observed and simulated fine organic matter
	concentrations (IMPROVE, scaling factor of 2.1)

Figure 4-15. Spatial display of the mean normalized error for organic matter (IMPROVE	Ì.,
OC x 2.1)	5
Figure 4-16.Seasonal performance statistics for fine organic matter concentrations	
(IMPROVE, scaling factor of 1.4)	6
Figure 4-17.Seasonal performance statistics for fine organic matter concentrations	
(IMPROVE, scaling factor of 2.1)	8
Figure 4-18. Scatter diagrams of observed and simulated fine black carbon concentration	IS
(IMPROVE)	3
Figure 4-19. Spatial display of the mean normalized error for black carbon	
(IMPROVE)	4
Figure 4-20.Seasonal performance statistics for fine black carbon concentrations	
(IMPROVE)	5
Figure 4-21. Scatter diagrams of observed and simulated PM <sub>2.5</sub> concentrations	
(IMPROVE)	0
Figure 4-22.Seasonal performance statistics for PM <sub>2.5</sub> concentrations (IMPROVE,	
measured)	1
Figure 4-23. Scatter diagrams of reconstructed and simulated PM <sub>2.5</sub> concentrations	
(IMPROVE)	3
Figure 4-24.Seasonal performance statistics for PM <sub>2.5</sub> concentrations (IMPROVE,	
reconstructed)	4
Figure 4-25. Scatter diagrams of observed and simulated coarse mass concentrations	
(IMPROVE)	8
Figure 4-26.Seasonal performance statistics for coarse mass concentrations	
(IMPROVE)	9
Figure 4-27. Scatter diagrams of observed and simulated $PM_{10}$ concentrations	
(IMPROVE)	3
Figure 4-28.Seasonal performance statistics for PM <sub>10</sub> concentrations (IMPROVE) 4-6	4
Figure 4-29. Scatter diagrams of sulfate wet deposition (NADP)	0
Figure 4-30.Seasonal performance statistics for sulfate wet deposition (NADP) 4-7	1
Figure 4-31. Scatter diagrams of nitrate wet deposition (NADP)	3
Figure 4-32.Seasonal performance statistics for nitrate wet deposition (NADP)	4

Figure 4-33	. Scatter diagrams of ammonium wet deposition (NADP) 4	-76
Figure 4-34	Seasonal performance statistics for ammonium wet deposition (NADP)4	-77
Figure 5-1.	Comparison of annual sulfate concentrations simulated by CAMx M4 and	
	CAMx 4Sec at the IMPROVE sites	5-3

#### **EXECUTIVE SUMMARY**

## E-1 Overview

We present in this report the performance evaluation of four air quality models that were applied to simulate particulate matter (PM) for an entire year (1996) over the western United States. The air quality models considered here are the following:

- The Comprehensive Air Quality Model with Extensions with Mechanism 4 and the default two particle size section option (CAMx M4) (version 4+, similar to version 4.1 to be released by ENVIRON in June 2004)
- The Comprehensive Air Quality Model with Extensions with Mechanism 4 and the four particle size section option (CAMx 4Sec) (version 4+)
- The Community Multiscale Air Quality model (CMAQ) (version 4.3 released by EPA in September 2003)
- The Regional Modeling System for Aerosols and Deposition (REMSAD) (version 7.06 released by ICF Consulting in July 2002)

CAMx M4 and REMSAD represent PM in two size fractions, fine and coarse, with a cut-off Stokes diameter of 2.5  $\mu$ m. CAMx 4Sec uses four size sections with three in the fine mode (from 0.039 to 2.5  $\mu$ m in Stokes diameter) and one in the coarse mode (from 2.5 to 10  $\mu$ m in Stokes diameter). CMAQ uses three modes; two modes (nuclei and accumulation) are assumed to represent fine PM and the third mode is assumed to represent coarse PM. All models use a single droplet size for aqueous-phase chemistry.

We evaluated the performance of these models by comparing the model simulation results with measurements of ozone concentrations, PM concentrations (including several PM species and two PM size fractions) and wet deposition of sulfate, nitrate and ammonium. The performance evaluation was conducted using the Model Performance Evaluation (MPE) package developed by AER for the Central Regional Air Planning Association (CENRAP). This MPE package includes a database of atmospheric data (ambient concentrations and wet deposition fluxes) and a software that calculates several performance metrics.

## E-2 Measurement Databases

The performance of the models was evaluated using non-urban ozone concentrations from the Air Quality System (AQS), PM concentrations from the Interagency Monitoring of Protected Visual Environments (IMPROVE), sulfate and total nitrate concentrations from the Clean Air Status & Trends Network (CASTNet) and wet deposition of sulfate, nitrate and ammonium from the National Atmospheric Deposition Program (NADP). We reviewed the uncertainties that are associated with those measurements because they affect the model performance evaluation.

Measurements of sulfate and nitrate (or total nitrate in the case of CASTNet) concentrations are likely to be reliable. Black carbon (BC) measurements have some uncertainty associated with the analytical technique (on the order of 30% on average). Organic matter (OM) concentrations are estimated from organic carbon (OC) measurements and are likely to be underestimates because of negative artifacts during warm months and an underestimation of the 1.4 scaling factor to convert OC to OM. Accordingly, alternative OM concentrations were estimated using a more realistic 2.1 scaling factor. Coarse mass should be a reliable estimate because coarse material is mostly non-volatile and the artifacts associated with the semi-volatile components of the fine fraction should be similar in the  $PM_{2.5}$  and the  $PM_{10}$  measurements and should cancel out. The soil component does not exactly correspond to the "other" component of the models; the comparison is, therefore, not meaningful and is not presented here.

The  $PM_{2.5}$  measurement is likely to be an underestimate during warm months because of the volatilization of ammonium nitrate and some organic compounds. An IMPROVE methodology can be used to calculate a reconstructed  $PM_{2.5}$  concentration from the  $PM_{2.5}$  component concentrations; however, this reconstructed  $PM_{2.5}$ concentration also has uncertainties associated with the calculation of some of its components (ammonium, OM and other primary components). Therefore, there are significant uncertainties associated with both measured and reconstructed  $PM_{2.5}$  concentrations and our assessment focuses on PM components rather  $PM_{2.5}$  and  $PM_{10}$  concentrations.

### E.3 Summary of Model Performance

Performance for non-urban ozone concentrations shows acceptable errors but significant biases (8 to 10 ppb underprediction on average for concentrations above 40 ppb). Such biases suggest possible uncertainties with the oxidants associated with secondary PM formation ( $O_3$ , OH and  $H_2O_2$ ) that may affect model performance for PM.

For PM, the models typically explain only between 15 and 30% of the variance in the observed BC and sulfate concentrations, between 20 and 25% of the variance in the observed OM concentrations, about 20% of the variance in the nitrate concentrations, and about 40% of the variance in the total inorganic nitrate concentrations (i.e.,  $HNO_3$  + particulate nitrate). All normalized errors (i.e., mean normalized error, normalized mean error, fractional error and normalized root mean square error) are greater than 50% except for CMAQ performance with CASTNet sulfate where the errors are slightly under 50%. In the case of nitrate, all normalized errors exceed 100%. All models show significant biases for nitrate (large overpredictions) and OM (large underpredictions).

Coarse PM concentrations (measured as the difference between  $PM_{10}$  and  $PM_{2.5}$  concentrations) are significantly underpredicted by all models (by about one order of magnitude). Less than 5% of the variance in the measured coarse PM concentrations is explained by the models.

Wet deposition of sulfate, nitrate and ammonium is significantly underestimated by all models.

# E.4 Comparison of Performance among CAMx, CMAQ and REMSAD

For sulfate, CAMx 4Sec and CMAQ show comparable performance. The performance of CAMx M4 is slightly poorer than that of CAMx 4Sec (see below). REMSAD showed the worst performance (only 7% of the variance in the IMPROVE data explained and a mean normalized error nearing 100%). The large differences obtained among the four models (factor of two for average annual concentrations)

suggest that improvements in the model inputs (boundary conditions, cloud and precipitation fields) and model formulation (sulfate formation in clouds, sulfate removal) are warranted.

For nitrate, CAMx overestimates more than CMAQ and REMSAD but shows similar coefficients of correlation. These results suggest that there is a systematic bias in CAMx that should be investigated. Uncertainties in the model inputs (ammonia inventory, boundary conditions, ambient temperature and relative humidity) are likely to introduce large errors in the modeling results.

CAMx underestimates OM more than CMAQ and REMSAD. We anticipate that OM formulations will evolve as our understanding of SOA formation continues to improve. Also, efforts should be directed toward improving the emission inventories of primary OM and biogenic SOA precursors.

Model performance for BC is comparable among the models. It is strongly affected by uncertainties in the emission inventory and the ambient measurements. It also reflects uncertainties associated with the formulation of transport and dispersion in all models.

# E.5 CAMx M4 versus CAMx 4Sec

CAMx M4 (two size sections for PM, fine and coarse) and CAMx 4Sec (four size sections for PM) show similar results except for sulfate, with CAMx 4Sec showing better performance than CAMx M4. The difference is due to the presence of some sulfate in the coarse mode in CAMx 4Sec and a likely faster removal of sulfate by dry deposition in CAMx 4Sec than in CAMx M4. However, it is not clear whether four sections are sufficient to correctly simulate the evolution of the particle size distribution and additional testing of CAMx 4Sec is recommended. First, simulations should be conducted with various size resolutions (e.g., 4 vs. 8 sections) to determine an optimal number of size sections that provides both numerical accuracy and computational efficiency. Next, this optimal model configuration should be evaluated with size distributed field data.

### E.6 Recommendations

The results of this model performance evaluation point out the need to use the output of a global model for the boundary conditions and to improve various aspects of the emission inventory (ammonia, SOA biogenic precursors, primary PM including BC and coarse PM). Data assimilation of clouds and precipitation is also recommended for meteorological modeling as clouds affect sulfate and nitrate formation (and possibly SOA formation) and precipitation is a major removal mechanism for PM that is significantly underestimated by the models in this 1996 application.

Because of significant uncertainties in the model inputs and the limited amount of data available for diagnostic evaluation, it is not feasible to provide definitive conclusions regarding the adequacy of the model formulations. Nevertheless, areas that deserve further investigation include sulfate formation (e.g., use of multiple droplet sizes), nitrate formation (e.g., heterogeneous reactions), SOA formation, the size distribution of secondary PM, and PM removal rates (mainly wet deposition rates but also dry deposition rates).

#### 1. INTRODUCTION

We present in this report the performance evaluation of four air quality models that were applied to simulate particulate matter (PM) for an entire year (1996) over the western United States. The air quality models considered are the following:

- The Comprehensive Air Quality Model with Extensions with Mechanism 4 and the default two particle size section option (CAMx M4) (version 4+, similar to version 4.1 to be released by ENVIRON in June 2004)
- The Comprehensive Air Quality Model with Extensions with Mechanism 4 and the four particle size section option (CAMx 4Sec) (version 4+)
- The Community Multiscale Air Quality model (CMAQ) (version 4.3 released by EPA in September 2003)
- The Regional Modeling System for Aerosols and Deposition (REMSAD) (version 7.06 released by ICF Consulting in July 2002)

CAMx M4 and REMSAD represent PM in two size fractions, fine and coarse, with a cut-off Stokes diameter of 2.5  $\mu$ m. CAMx 4Sec uses four size sections with three in the fine mode (from 0.039 to 2.5  $\mu$ m in Stokes diameter) and one in the coarse mode (from 2.5 to 10  $\mu$ m in Stokes diameter). CMAQ uses three modes; two modes (nuclei and accumulation) are assumed to represent fine PM and the third mode is assumed to represent coarse PM. All models use a single droplet size for aqueous-phase chemistry.

The two versions of CAMx (with two and four particle size sections, respectively) were applied by ENVIRON under CRC Project A-44. CMAQ and REMSAD were applied by the University of California at Riverside (UCR) and ENVIRON, respectively, under the Western Regional Air Partnership (WRAP) program to address regional haze. The output files of the CAMx, CMAQ and REMSAD simulations were obtained from ENVIRON except for the wet deposition outputs files of the CMAQ simulation which were obtained from UCR. These files were 1-hour average values except for the CMAQ concentration file that provided instantaneous (end of the hour) concentrations (1-hour average values were not available). The effect of using 1-hour average values versus

instantaneous values on performance measures is small (less than 1%; G. Tonnesen, communication to WRAP, May 2004) and, therefore, does not affect the performance evaluation presented here noticeably.

The performance evaluation was conducted using the Model Performance Evaluation (MPE) package developed by AER for the Central Regional Air Planning Association (CENRAP). This MPE package includes a database of atmospheric data (ambient concentrations and wet deposition fluxes) and a software that calculates several performance metrics (Pun et al., 2004a).

We present in Section 2 an overview of the databases that were used to evaluate the models. The uncertainties associated with the measurements are discussed because they can affect model performance. We discuss the statistical metrics that are used to assess model performance in Section 3. The results of the model performance evaluation are presented in Section 4 (the complete set of performance statistics is available as an electronic file). An assessment of model performance and some recommendations are provided in Section 5. Concluding remarks are presented in Section 6.

#### 2. OVERVIEW OF THE MEASUREMENT DATABASES

Four measurement databases were used in this performance evaluation. Those include the Air Quality System (AQS) for ozone concentrations, the Interagency Monitoring of Protected Visual Environments (IMPROVE) for PM concentrations, the Clean Air Status & Trends Network (CASTNet) for sulfate and total nitrate concentrations and the National Atmospheric Deposition Program (NADP) for wet deposition of sulfate, nitrate and ammonium. We discuss the measurements from these networks and how they are used in this model performance evaluation. We conclude with a summary of the uncertainties associated with those measurements and how they may affect model performance evaluation.

#### 2.1 AQS

To the extent possible, PM air quality models should be evaluated for gaseous precursors and oxidants in addition of PM species (e.g., Seigneur et al., 2000; EPA, 2001). For the year 1996, few data are available for gases. The AQS provides ozone  $(O_3)$ , nitrogen oxides  $(NO_X)$ , sulfur dioxide  $(SO_2)$ , carbon monoxide (CO) and some volatile organic compounds (VOC) measurements. Because the air quality models were applied with a coarse spatial resolution (about 36 km), only non-urban ambient concentrations should be used for model performance evaluation. Non-urban concentrations of  $NO_X$ ,  $SO_2$ , and CO are low in the western United States and, consequently, they are not very suitable for a performance evaluation. Because there are few VOC concentrations available in non-urban locations, those were not used for model performance evaluation. Therefore, only the O<sub>3</sub> 1-hour average concentrations at non-urban locations were used here.

#### 2.2 IMPROVE

IMPROVE provides PM concentrations in Class I areas (such as National Parks and National Wilderness Areas) twice a week (on Wednesdays and Saturdays). (This monitoring schedule was changed to every third day in 2000.) The measurements include  $PM_{10}$ ,  $PM_{2.5}$ ,  $PM_{2.5}$  components (sulfate, nitrate, metals, organic carbon and black carbon), light scattering and, in some locations, light absorption. It should be noted that there are some significant uncertainties with some of those measurements.

The  $PM_{10}$  and  $PM_{2.5}$  measurements are obtained by selectively sampling particles that have aerodynamic diameters less than 10 and 2.5 µm, respectively. It must be noted that those cut-off diameters are approximate because some particles with larger diameters will be sampled and some particles with smaller diameters will not be sampled (see Figure 2-1). Moreover, the models use Stokes diameters to simulate the dynamics of atmospheric PM. The Stokes and aerodynamic diameters are related by the square root of the particle density. Therefore, if the particle density differs from 1 g/cm<sup>3</sup>, a Stokes diameter of 2.5 µm will not correspond to an aerodynamic diameter of 2.5 µm. Also, models that use a modal representation (here, CMAQ) may not match the fine and coarse PM measurements exactly because the tails of the modes may overlap with the next size section (for example, the upper tail of the accumulation mode may extend into the coarse mode). These points have been discussed in detail by Bhave (2004) and Seigneur (2004). The uncertainties associated with the size fractions are likely to be small compared to those associated with the chemical mass measurements, as discussed below.

PM<sub>10</sub> and PM<sub>2.5</sub> mass measurements are made from a gravimetric analysis of PM collected on a Teflon filter. Although such measurements are likely to be accurate for non-volatile PM components, there are possible artifacts for semi-volatile PM components. Some negative artifacts may occur as semi-volatile compounds such as ammonium nitrate and some organic compounds volatilize due to a change in temperature, pressure or gas-phase composition. Such negative artifacts have been documented for ammonium nitrate (e.g., Hering and Cass, 1999; Ashbaugh and Eldred, 2004) and organic compounds (e.g., Pang et al., 2002). Positive artifacts may occur as some gases adsorb to the filter. Such positive artifacts are more likely to occur for



Aerodynamic diameter µm

Figure 2-1. Typical sampling efficiency for a  $PM_{2.5}$  sampler with size segregation based on inertial impaction.

organic compounds (e.g., Turpin et al., 2000). The  $PM_{10}$  and  $PM_{2.5}$  mass measurements are more likely to be inaccurate when the ambient temperature is high (i.e., during spring and summer) because the volatilization of some semi-volatile compounds leads to underestimation of PM mass. In addition to the problems associated with positive and negative artifacts, another source of discrepancy between measured and simulated PM concentrations results from the amount of water associated with PM. The ravimetric measurements are conducted in the laboratory at a temperature and a relative humidity that are within specified ranges. Some amount of water is typically retained with PM due to the hysteresis of hygroscopic particles. On the other hand, the simulated PM concentrations are typically reported without any water.

Sulfate concentrations can be obtained in two ways from the IMPROVE measurements. Because  $SO_2$  is in the gas phase, the sulfur measured on the filter corresponds solely to sulfate and the sulfur measurement multiplied by three (to account for the four associated oxygen atoms) provides a sulfate measurement. On the other hand, sulfate is also measured from a Nylon filter sample. The two values typically agree well with each other. Here, we used the sulfur measurement multiplied by three as the sulfate concentration.

Nitrate measurements are made from PM collected on a Nylon filter. A denuder is placed in front of the filter to collect HNO<sub>3</sub> and positive artifacts due to HNO<sub>3</sub> adsorption are, therefore, minimized. The Nylon filter is slightly alkaline so that volatilization of nitrate is also minimized. Therefore, nitrate measurements should be reasonably accurate since both positive and negative artifacts are minimized.

Ammonium measurements are not made routinely in the IMPROVE network and model performance was, therefore, not evaluated for ammonium.

Carbon is sampled on Quartz filters and a thermal analysis is conducted in the laboratory to differentiate between organic and black (also referred to as elemental or light-absorbing) carbon. The differentiation between organic carbon (OC) and black carbon (BC) is, therefore, operational because it depends on the laboratory analytical protocol. Chow et al. (2001) estimated that uncertainties due to different protocols may be on the order of a factor of two for BC. A recent comparison conducted by EPA (Solomon et al., 2004) suggests that differences may be less (on the order of 30% for

BC). Because OC concentrations are typically greater than BC concentrations in remote areas, the relative uncertainty due to the analytical protocol is typically less for OC.

In addition, there are two other sources of uncertainties for organic PM. First, there may be positive and negative artifacts (e.g., Turpin et al., 2000; Eatough et al., 2003). In the IMPROVE sampling, there is no denuder to remove organic vapors prior to the Quartz filter, therefore, some semi-volatile organic compounds could adsorb to the Quartz filter. The IMPROVE protocol attempts to compensate for this positive artifact by subtracting the organic PM collected on a back-up filter. However, these back-up filters are not specific to the sampling site and period. Instead, an average value obtained from a limited number of back-up filters (i.e., field blanks) is subtracted from the value obtained from the Quartz filter. This approach leads to two problems: (1) using an average value may not be representative of a given site and period and (2) there may also be negative artifacts (volatilization of organic PM from the Quartz filter) that are not accounted for by this method. As a result, it is possible that OC measurements are underestimates of actual OC values.

The second source of uncertainty arises from the conversion of the OC concentration to an organic mass (OM) concentration that includes oxygen, nitrogen and hydrogen associated with the measured carbon. Since the non-carbon atoms are not measured, the IMPROVE protocol recommends using a factor of 1.4 to convert OC to OM. Turpin and Lim (2001) have argued that a factor of 1.4 is too low for remote sites (such as the IMPROVE sites) and that a factor of 2.1 would be more representative of the chemical composition of OM in remote areas. Here, we used a factor of 1.4 in the base performance evaluation and a factor of 2.1 in an alternative assessment to investigate the effect of this major source of uncertainty.

Other components of PM include crustal material, sea salt and primary anthropogenic emissions. The IMPROVE protocol estimates a  $PM_{2.5}$  component classified as "soil" by summing several metal oxides (the metal concentrations are measured and an oxidation state is assumed for each metal). Air quality models typically have a PM component called "other PM". CAMx differentiates "other" PM among crustal material, primary anthropogenic emissions and sodium chloride. However, the attribution of PM primary emissions to the crustal and anthropogenic categories is

arbitrary because the emission model, SMOKE, does not provide sufficient information to distribute primary PM properly. Here, we compared the IMPROVE "soil" concentration with the "other" PM calculated by the models. Clearly, there are large uncertainties associated with this comparison and, consequently, we do not attempt to interpret the discrepancies between estimates derived from the measurements and the modeling results.

As mentioned above, PM<sub>2.5</sub> measurements may have large uncertainties due to positive and negative artifacts, as well as the presence of some water. Another  $PM_{2.5}$ concentration is calculated according to the IMPROVE methodology from the concentrations of sulfate, nitrate, OC, BC and "soil". This calculated concentration is referred to as "reconstructed" PM<sub>2.5</sub>. There are, however, uncertainties associated with this methodology. First, ammonium is not measured and it is assumed to be present to completely neutralize sulfate and nitrate. Although this may be a plausible assumption when nitrate is present (sulfate is generally neutralized before nitrate), it does not hold in ammonia-poor environments where ammonium bisulfate or letovicite may be present. Second, a factor of 1.4 is used to convert OC to OM (as discussed above, a factor of 2.1 Third, the "other" components are estimated from may be more appropriate). measurements of elements and assumptions on their oxidation states. A scatter diagram of the measured and reconstructed PM2.5 concentrations in the western IMPROVE network during 1996 is presented in Figure 2-2a. The reconstructed PM<sub>2.5</sub> concentrations tend to be lower than the measured  $PM_{2.5}$  concentrations. Although there are several sources of discrepancies between the two PM<sub>2.5</sub> concentrations, the major source of discrepancy is likely due to the use of a 1.4 OM/OC scaling factor for the reconstructed  $PM_{2.5}$  concentrations. Using a more realistic scaling factor (e.g., 2.1) reduces the differences to some extent, as depicted in Figure 2-2b. We used both the measured and the reconstructed (with a 1.4 OM/OC scaling factor) PM<sub>2.5</sub> concentrations for this model performance evaluation. However, one must keep in mind that both PM2.5 concentrations have associated uncertainties.

Coarse mass is estimated as the difference between the  $PM_{10}$  and  $PM_{2.5}$  concentrations. One would expect that the same artifacts apply to the fine fraction of



Figure 2.2. Scatter diagram of measured and reconstructed  $PM_{2.5}$  concentrations from IMPROVE (55 sites in the western United States during 1996), using OM/OC = 1.4 (top) and OM/OC = 2.1 (bottom).

 $PM_{10}$  and to  $PM_{2.5}$ . Therefore, the coarse mass estimate (which should contain little semi-volatile material) should be reasonably accurate.

#### 2.3 CASTNet

CASTNet provides 1-week averaged measurements of PM species (sulfate, nitrate and ammonium) and of some gases (nitric acid, HNO<sub>3</sub>, and SO<sub>2</sub>). In addition, estimates of dry deposition are also provided for the species measured by combining the measured concentrations with estimates of dry deposition velocities.

The fact that the measurements are obtained from 1-week averaged filter samples casts some doubt on the accuracy of those measurements for the semi-volatile species; 88

namely, nitrate, ammonium and HNO<sub>3</sub> (see discussion of uncertainties associated with ammonium nitrate above). Therefore, we did not use those measurements in this model performance evaluation. However, the sum of nitrate and HNO<sub>3</sub> should be representative of the actual ambient concentration of total inorganic nitrate (i.e., particulate + gas), because only the partitioning between the gas and particulate phases should be affected by the 1-week duration. Therefore, we used the sulfate and total inorganic nitrate measurements in the model performance evaluation. Note that the PM measurements are made with open-face filters and do not include any size segregation. Therefore, we compare total (i.e., fine and coarse) sulfate PM concentrations to the CASTNet measurements.

CASTNet measurements are reported at standard temperature and pressure (STN). The model outputs are converted to standard temperature and pressure by the MPE software using the modeled temperature and pressure.

Because dry deposition is not measured but estimated using a model to estimate the deposition velocities, it was not used in the model performance evaluation. Figure 2-3 depicts this rationale by deconstructing the CASTNet deposition estimates into its components.



Figure 2-3. Schematic representation of CASTNet data and their use in model performance evaluation.

#### 2.4 NADP

NADP provides wet deposition fluxes for sulfate, nitrate, ammonium, hydrogen and several other cations. These data were used here to evaluate the wet deposition components of the PM models. Although wet deposition is not necessarily an output of a PM air quality model, it affects PM concentrations by removing PM from the atmosphere. Therefore, this evaluation adds a diagnostic component to the overall evaluation process by providing a direct evaluation of an important removal pathway for PM.

We evaluated the models for sulfate, nitrate and ammonium wet deposition (hydrogen ions were not available from the models).

#### 2.5 Summary of Measurement Uncertainties

Table 2-1 summarizes the uncertainties associated with the ambient concentration measurements used for the model performance evaluation. This summary shows that the sulfate and nitrate (or total sulfate and nitrate in the case of CASTNet) components of  $PM_{2.5}$  are likely to be reliable measurements. BC measurements have some uncertainty associated with the analytical technique (on the order of 30% on average according to Solomon et al., 2004). OM estimates are likely to be underestimates because of negative artifacts during warm months and an underestimation of the scaling factor to convert OC to OM. Coarse mass should be a reliable estimate because coarse material is mostly nonvolatile and the artifacts associated with the semi-volatile components of the fine fraction should be similar in the  $PM_{2.5}$  and the  $PM_{10}$  measurements and should cancel out. The soil component does not exactly correspond to the "other" component of the models and the comparison is, therefore, not meaningful.

The  $PM_{2.5}$  measurement is likely to be an underestimate during warm months because of the volatilization of ammonium nitrate and some organic compounds. The reconstructed  $PM_{2.5}$  has also uncertainties associated with the calculation of some of its components (ammonium, OM and other primary components). Therefore, there are

Table 2-1. Uncertainties associated with ambient concentration measurements.

Monitoring network	Measurement	Major uncertainty	Consequence for performance evaluation
AQS	O <sub>3</sub>	Some interferences	Negligible
IMPROVE	Sulfate (as sulfur)	None	Negligible
	Nitrate	Negligible volatilization	Negligible
	OM (as OC)	Positive and negative artifacts; scaling factor	Likely to underestimate actual OM
	BC	Operational uncertainty of analytical method	30% to factor of 2 uncertainty
	PM <sub>2.5</sub> (measured)	Negative artifacts for ammonium nitrate; positive and negative artifacts for OM; presence of some water	More likely to underestimate during warm periods due to negative artifacts (volatilization)
	PM <sub>2.5</sub> (reconstructed)	Ammonium assumed to neutralize sulfate and nitrate; 1.4 scaling factor OM/OC; soil oxides assumptions	Ammonium may be overestimated (particularly if no nitrate); OM likely to be underestimated, "other" components uncertain
	PM <sub>10</sub> (measured)	Same as PM <sub>2.5</sub>	Same as PM <sub>2.5</sub>
	Coarse mass	PM <sub>2.5</sub> and fine PM <sub>10</sub> uncertainties are likely to cancel out; negligible uncertainties for the coarse fraction	Negligible
CASTNet	Sulfate	None	Negligible
	Total nitrate	Negligible (major uncertainties apply to the gas/particle partitioning)	Negligible

significant uncertainties associated with measured (or reconstructed)  $PM_{2.5}$  concentrations and the results of the model performance evaluation should be interpreted with caution.

A similar conclusion applies to the  $PM_{10}$  measurement, although the artifacts are likely to be less for the coarse fraction than for the fine fraction and the  $PM_{10}$  measurement should, therefore, be more accurate than the  $PM_{2.5}$  measurement.

There are no particular artifacts associated with the wet deposition measurements.

#### **3. PERFORMANCE MEASURES**

Most of the performance measures that are currently available in the MPE software (version 1.03) were used. Table 3-1 presents those measures as well as their mathematical expressions. When calculating the performance measures, a threshold value of 40 ppb was used for the ozone concentrations and a threshold value of 0.01  $\mu$ g/m<sup>3</sup> was used for all PM concentrations (data pairs were included in the statistics only if the observed value exceeded the threshold value). A threshold of 0.01 kg/ha was used for the wet deposition data. We briefly discuss below the meaning of the measures as well as their advantages and disadvantages.

Accuracy of peak: This measure is useful to assess the ability of a model to reproduce the peak concentration in an airshed (i.e., the concentration that is relevant for the attainment of the National Ambient Air Quality Standards). It is less relevant for regional studies and we do not use it here in our discussion of model performance.

Correlation coefficient (r): It is calculated here for all sites and all time periods. Therefore, it measures the ability of the model to reproduce the spatial and temporal variations in the observed variables. It ranges from -1 (perfect anti-correlation) to +1 (perfect correlation).

Coefficient of determination  $(r^2)$ : It represents the fraction of the variance in the observations that is reproduced by the model. It ranges from 0 to 1. A value of 0.5 means that the model explains half of the variance in the observations.

Mean error and bias: The arithmetic means of the error (absolute difference between the simulated and observed values) and bias (signed difference between the simulated and observed values) provide an absolute measure of the error and bias in the model simulations.

Mean normalized error and bias: These metrics have typically been used to evaluate the performance of air quality models for ozone with benchmarks of 35% and 15%, respectively. They give more weight to overpredictions than to underpredictions because there is no limit on the overpredictions whereas the underpredictions are

Metric	Formula	
Accuracy of peak unpaired in space and time	$A_{pu} = \frac{P_{peak}^u - O_{peak}}{O_{peak}}$	
Accuracy of peak paired in space but not in time	$A_{pm} = \frac{P_{peak} - O_{peak}}{O_{peak}}$	
Mean observed value	$\overline{O} = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left( O_{i}^{j} \right)$	
Mean modeled value	$\overline{P} = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left( P_{i}^{j} \right)$	
Mean error	$E = \frac{1}{N \bullet M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left  P_i^j - O_i^j \right $	
Mean bias	$B = \frac{1}{N \cdot M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left( P_i^j - O_i^j \right)$	
Mean normalized error	$E_N = \frac{1}{N \bullet M} \sum_{i=1}^N \sum_{j=1}^M \left  \frac{P_i^j - O_i^j}{O_i^j} \right $	
Mean normalized bias	$B_N = \frac{1}{N \bullet M} \sum_{i=1}^N \sum_{j=1}^M \left( \frac{P_i^j - O_i^j}{O_i^j} \right)$	
Normalized mean error	$NME = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \left  P_{i}^{j} - O_{i}^{j} \right }{\sum_{i=1}^{N} \sum_{j=1}^{M} O_{i}^{j}}$	
Normalized mean bias	$NMB = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (P_i^{j} - O_i^{j})}{\sum_{i=1}^{N} \sum_{j=1}^{M} O_i^{j}}$	

Table 3-1. Statistical metrics used for model performance.

Metric	Formula
Fractional error	$E_F = \frac{1}{N \bullet M} \sum_{i=1}^{N} \sum_{d=1}^{M} \left  2 \bullet \frac{P_i^j - O_i^j}{P_i^j + O_i^j} \right $
Fractional bias	$B_F = \frac{1}{N \bullet M} \sum_{i=1}^{N} \sum_{j=1}^{M} \left( 2 \bullet \frac{P_i^j - O_i^j}{P_i^j + O_i^j} \right)$
Correlation coefficient (r)	$\sum_{i=1}^{N} \sum_{i=1}^{M} \left( P_{i}^{\ j} - \overline{P}  ight) ullet \left( O_{i}^{\ j} - \overline{O}  ight)$
Coefficient of determination (r <sup>2</sup> )	$r = \frac{1}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M} (P_{i}^{j} - \overline{P})^{2}}} \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M} (O_{i}^{j} - \overline{O})^{2}}$
Index of agreement	$IA = 1 - \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (P_{i}^{j} - O_{i}^{j})^{2}}{\sum_{i=1}^{N} \sum_{j=1}^{M} (P_{i}^{j} - \overline{O} +  O_{i}^{j} - \overline{O} )^{2}}$
Normalized RMSE	$NRMSE = \frac{1}{N} \sum_{i=1}^{N} \frac{RMSE_i}{\overline{O_i}}$
	$=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sqrt{\sum_{j=1}^{M}\left(P_{j}-O_{j}\right)^{2}}}{M}\right)$ $\frac{\frac{M}{\sum_{j=1}^{M}O_{j}}}{M}$

Table 3-1. Statistical metrics for model performance (continued).

Notations:

 $P_{peak}^{u}$ : unpaired peak prediction anywhere in the modeling domain

 $O_{peak}$ : peak observation anywhere in the domain

 $P_{peak}$ : peak prediction at the location of the peak measurement

 $\dot{N}$ : total number of measurement sites

*M*: total number of averaging periods (e.g., M=1 for daily statistics of 24-hour concentrations)

 $P_i^{j}$ : Predicted value at site 'i' during averaging period 'j'  $O_i^{j}$ : Observed value at site 'i' during averaging period 'j'  $P_{i, component}^{j}$ : Predicted value at site 'i' during averaging period 'j' of a PM component

O<sub>i</sub>, <sub>component</sub>; Observed value at site 'i' during averaging period 'j' of a PM component

P<sub>i</sub>, <sup>j</sup><sub>total</sub>: Predicted value at site 'i' during averaging period 'j' of total PM

O<sub>i</sub>, total: Observed value at site 'i' during averaging period 'j' of total P

constrained to -100%. Also, errors and biases associated with small observations are given more weight than those associated with large observations. The error may range from 0 to infinity and the bias from -100% to + infinity.

Normalized mean error and bias: The excessive weight given to errors and biases associated with small observations is avoided with those metrics because the normalization is performed after summing up all errors (or biases) with respect to the mean observed value. However, overpredictions still carry more weight than underpredictions. The error may range from 0 to infinity and the bias from -100% to + infinity.

Fractional error and bias: These metrics place the same weight on overpredictions and underpredictions because the normalization is done with respect to the arithmetic mean of the observed and simulated values. Also, the error and biases associated with small observations are not given any excessive weight because the normalization is not done with respect to the observation alone. Seigneur et al. (2000) recommended these metrics for error and bias for these reasons. One potential disadvantage, however, is that the observations are not used as the reference. Instead, these metrics measure the discrepancy between observation and simulation without placing any a priori judgment on which one is the reference. It can be seen as an objective measure of the discrepancy between two data sets rather than a measure of the discrepancy of a data set with respect to a reference data set. The fractional error and bias range between 0 and 2 and between -2 and +2, respectively.

Root mean square error: It is a standard measure used in statistics to measure the average error between two data sets. It can be normalized (here, with respect to the observed value).

Index of agreement: It has been used to evaluate meteorological models and is provided here for completeness.

#### 4. **RESULTS OF THE MODEL PERFORMANCE EVALUATION**

The results of the performance evaluation of the four models (CAMx M4, CAMx 4Sec, CMAQ and REMSAD) are presented by species. We first present the results for ozone using the AQS non-urban data. Next, we present the performance evaluation for the following components of  $PM_{2.5}$ : sulfate, nitrate, organic matter (OM) and black carbon (BC). For sulfate, results are presented first for the IMPROVE data and next for the CASTNet data. For nitrate, results are presented for  $PM_{2.5}$  nitrate for the IMPROVE data and for total nitrate for the CASTNet data. No results are presented for ammonium since data are not available for the period of interest. No results are presented for the "other" category because no direct measurements are available and the "soil" category of the IMPROVE network does not correspond exactly to the "other" category that may also include primary anthropogenic PM and sea salt. Then, we present the results for  $PM_{2.5}$ , the PM coarse fraction and  $PM_{10}$ . The results are presented for both the measured and reconstructed values of  $PM_{2.5}$  concentrations. Finally, the evaluation against the wet deposition data of NADP is presented for sulfate, nitrate and ammonium.

For each species or PM size fraction, we present first the results of the performance evaluation for the annual simulation results. Next, we present seasonal performance statistics (correlation coefficient, normalized errors and normalized biases).

#### 4.1 Ozone

Table 4-1 presents the annual performance statistics and Figure 4-1 presents scatter plots for the non-urban ozone concentrations. A threshold value of 40 ppb was used. The models underpredicted the ozone concentrations by 8 to 10 ppb on average. CAMx shows slightly less bias then CMAQ. However, CAMx predicts zero ozone concentration values at times when the observations and the other two models show non-zero concentrations. This result may be due to a shallower mixing height at times in CAMx that leads to ozone titration by NO. The mean normalized errors would be

considered satisfactory for an ozone modeling study but the mean normalized biases would be considered slightly too large (underpredictions by more than 15%).

Figure 4-2 presents seasonal statistics for ozone. Because a 40 ppb threshold was used, few data pairs (27) are available in winter. Better correlations between simulated and measured values are obtained in spring and winter. The errors tend to be larger in spring and summer. Overall, the underpredictions displayed by the models suggest that the concentrations of oxidants involved in secondary PM formation ( $O_3$ , OH, and  $H_2O_2$ ) may have significant biases that may affect model performance for PM.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD	
Mean observed value (ppm)	46.8				
Mean simulated value (ppm)	38.6	38.6	36.9	39.0	
Mean error (ppm)	11.0	11.0	11.6	12.3	
Mean bias (ppm)	-8.1	-8.1	-9.8	-7.7	
Mean normalized error	0.233	0.233	0.243	0.261	
Mean normalized bias	-0.166	-0.166	-0.202	-0.156	
Normalized mean error	0.236	0.236	0.249	0.264	
Normalized mean bias	-0.174	-0.174	-0.211	-0.165	
Fractional error	0.289	0.289	0.292	0.305	
Fractional bias	-0.229	-0.229	-0.255	-0.215	
Correlation coefficient	0.158	0.159	0.144	0.115	

Table 4-1. Annual performance statistics for ozone concentrations (AQS).



Figure 4-1. Scatter diagrams of observed and simulated ozone concentrations (AQS)





Figure 4-2. Seasonal performance statistics for ozone concentrations (AQS).





Figure 4-2 (continued). Seasonal performance statistics for ozone concentrations (AQS).
### 4.2 Sulfate

Table 4-2 presents the annual performance statistics for sulfate and Figure 4-3 a scatter plot for the IMPROVE data set. On average, sulfate is the second most important PM<sub>2.5</sub> component after OM in the western United States with an annual average concentration of about 0.8  $\mu$ g/m<sup>3</sup>. All models underestimate this average concentration with a range of biases from REMSAD's significant underprediction of -0.4  $\mu$ g/m<sup>3</sup> to CAMx M4's negligible underprediction of  $-0.01 \ \mu g/m^3$ . The mean error ranges from 0.4  $\mu g/m^3$  (CMAQ and CAMx 4Sec) to 0.6  $\mu g/m^3$  (REMSAD). The mean normalized biases are positive because their formulation gives more weight to overpredictions and those tend to dominate (see Figure 4-2). On the other hand, the fractional biases that provide a balanced treatment for over- and underpredictions show negative values ranging from -0.59 (REMSAD) to -0.004 (CAMx M4). REMSAD tends to show the largest normalized and fractional errors. CAMx M4 also shows a large normalized mean error but its fractional error is similar to those obtained for CAMx 4Sec and CMAO. The correlation coefficients are positive. CAMx M4 and CAMx 4Sec show the largest correlation coefficients and REMSAD the smallest; however, less than 30% of the variance in the observations is explained by the models.

Figure 4-4 presents a spatial display of the mean normalized errors for the four model simulations. The largest errors occur along the Pacific coast (possibly due to uncertainty in the western boundary condition) and in the Wyoming/Colorado area.

For current PM models, one may expect the normalized error to be 50% or less for sulfate, a well-studied chemical species. Here, all models show large mean normalized errors that exceed 60%, normalized mean errors that exceed 50% and fractional errors that exceed 50%. The relatively low sulfate concentrations that are observed in the western United States (compared to the eastern United States) may explain in part the poor performance obtained here. However, there may be other causes that should be investigated further. For example, fixed boundary conditions were used for all the simulations. Irwin et al. (2004) have shown that there is some indirect evidence that there is a strong seasonal variation for sulfate concentrations at the upwind boundary over the Pacific Ocean Figure 4-5 presents seasonal statistics for sulfate from IMPROVE. The smallest normalized errors and biases tend to occur in summer and the largest ones in winter. These seasonal variations reflect to some extent the fact that the sulfate concentrations are lower in winter compared to summer; lower observed concentrations lead to larger normalized errors for a given absolute error. On the other hand, the correlation coefficients are lower in summer and larger in winter.

Performance statistics for CASTNet sulfate concentrations are summarized in Table 4-3. Figure 4-6 presents a scatter diagram of those observed and simulated concentrations. The mean observed sulfate concentration is underpredicted by all models, with REMSAD showing the largest underprediction (factor of 2.7). CMAQ tends to show the best performance, explaining slightly more than 30% of the variance in the observations and having normalized errors slightly below 50%. REMSAD shows the worst performance with a low coefficient of determination (6%) and normalized errors exceeding 60%. CAMx 4Sec shows typically lower concentrations than CAMx M4, probably because of a greater rate of removal of sulfate via dry deposition.

Seasonal statistics are presented in Figure 4-7. The correlation coefficient increases from spring to winter. There is no clear seasonal variation of error with season.

The CASTNet results are qualitatively similar to those obtained with the IMPROVE data. The models tend to underestimate sulfate concentrations significantly over both monitoring networks.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	0.819			
Mean simulated value ( $\mu g/m^3$ )	0.809	0.622	0.682	0.420
Mean error (µg/m <sup>3</sup> )	0.504	0.429	0.424	0.607
Mean bias (µg/m <sup>3</sup> )	-0.010	-0.196	-0.136	-0.399
Mean normalized error	0.947	0.656	0.645	0.992
Mean normalized bias	0.532	0.102	0.174	0.044
Normalized mean error	0.616	0.524	0519	0.742
Normalized mean bias	-0.012	-0.240	-0.167	-0.487
Fractional error	0.575	0.567	0.547	0.964
Fractional bias	-0.004	-0.218	-0.115	-0.588
Correlation coefficient	0.440	0.518	0.531	0.273

Table 4-2. Annual performance statistics for fine sulfate concentrations (IMPROVE).



Figure 4-3. Scatter diagrams of observed and simulated fine sulfate concentrations (IMPROVE).



Figure 4-4. Spatial display of the mean normalized error for sulfate (IMPROVE).





## Figure 4-5. Seasonal performance statistics for fine sulfate concentrations (IMPROVE).





Figure 4-5 (continued). Seasonal performance statistics for fine sulfate concentrations (IMPROVE).

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	1.603			
Mean simulated value ( $\mu g/m^3$ )	1.289	1.116	1.120	0.591
Mean error (µg/m <sup>3</sup> )	0.912	0.863	0.794	1.139
Mean bias (µg/m <sup>3</sup> )	-0.314	-0.487	-0.483	-1.013
Mean normalized error	0.646	0.539	0.498	0.651
Mean normalized bias	0.238	0.045	0.052	-0.329
Normalized mean error	0.569	0.538	0.495	0.711
Normalized mean bias	-0.196	-0.304	-0.301	-0.632
Fractional error	0.540	0.529	0.479	0.869
Fractional bias	-0.058	-0.188	-0.138	-0.685
Correlation coefficient	0.385	0.429	0.559	0.241

Table 4-3. Annual performance statistics for sulfate concentrations (CASTNet).



Figure 4-6. Scatter diagrams of observed and simulated sulfate concentrations (CASTNet).











Figure 4-7 (continued). Seasonal performance statistics for sulfate concentrations (CASTNet).

### 4.3 Nitrate

Table 4-4 presents the annual performance statistics and Figure 4-8 a scatter plot for the IMPROVE data set. On average, nitrate is significantly less important than OM and sulfate as a PM<sub>2.5</sub> component in the western United States with an annual average concentration of only 0.34  $\mu$ g/m<sup>3</sup>. However, there are some areas of the domain where nitrate is an important component (e.g., California). All models overestimate this average concentration; overpredictions range from CMAQ and REMSAD's 0.2  $\mu$ g/m<sup>3</sup> to CAMx 4Sec's 0.4  $\mu$ g/m<sup>3</sup>. The mean error ranges from 0.5  $\mu$ g/m<sup>3</sup> (CMAQ) to 0.7  $\mu$ g/m<sup>3</sup> (CAMx 4Sec). All normalized errors exceed 100%. The mean normalized biases are positive and exceed 200%. On the other hand, the fractional biases are negative for three of the models and a small positive value for the fourth model. These results are due to the fact that the overpredictions of nitrate occur for small observed values (see Figure 4-8). CAMx M4 (with both two and four sections) tends to show the largest normalized and fractional errors. The correlation coefficients are commensurate with those obtained for sulfate (slightly lower than for sulfate for CAMx M4 and CMAx 4Sec, comparable for CMAQ and better than for sulfate for REMSAD). However, less than 20% of the variance in the nitrate observations is explained by the models.

Figure 4-9 presents a spatial display of the mean normalized errors for the four model simulations. The largest errors occur at the Bridger Wilderness Area, WY and the Jarbidge Wilderness Area, NV. No other particular spatial pattern is apparent. Errors are not greater near the domain boundaries suggesting that boundary conditions are not a dominating source of error.

One may not currently expect the normalized error for nitrate to be less than 50%. However, all models here show large normalized errors that exceed 300% and fractional errors that exceed 100%. The relatively low nitrate concentrations that are observed during the warm months may explain in part the poor performance obtained here. However, there may be other causes that should be investigated further.

Figure 4-10 presents some seasonal statistics for nitrate measurements from IMPROVE. Except for the fractional error, the normalized errors and biases tend to be

the smallest in summer and the largest in fall or winter. The correlation coefficients are larger in spring and smaller in fall.

Table 4-5 presents the performance statistics for the evaluation of total nitrate (i.e., particulate nitrate and gaseous HNO<sub>3</sub>). Figure 4-11 presents scatter diagrams of the measured and simulated values. The results show that all models overpredict total nitrate. The largest overpredictions are obtained with CAMx and the smallest underpredictions are obtained with REMSAD. The coefficients of determination ( $r^2$ , i.e., the fraction of the variance in the measurements explained by the models) are in the range of 37 to 40%. Figure 4-12 presents seasonal statistics for CASTNet total nitrate. The models tend to show better performance in summer when the concentrations are higher.

These results are consistent with the overprediction of particulate nitrate measured in IMPROVE. The better  $r^2$  values suggest that the models predict total nitrate better than particulate nitrate because of uncertainties in the partitioning of total nitrate between the gas phase (HNO<sub>3</sub>) and the particulate phase (nitrate). Also, the longer averaging time used in CASTNet compared to IMPROVE may help in improving model performance.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	0.345			
Mean simulated value ( $\mu g/m^3$ )	0.759	0.797	0.549	0.544
Mean error (µg/m <sup>3</sup> )	0.681	0.699	0.480	0.505
Mean bias (µg/m <sup>3</sup> )	0.414	0.452	0.204	0.199
Mean normalized error	4.542	4.984	3.108	3.365
Mean normalized bias	3.686	4.281	2.387	2.514
Normalized mean error	1.976	2.030	1.393	1.465
Normalized mean bias	1.202	1.313	0.594	0.579
Fractional error	1.388	1.319	1.126	1.241
Fractional bias	-0.234	0.014	-0.136	-0.293
Correlation coefficient	0.434	0.431	0.437	0.433

Table 4-4. Annual performance statistics for fine nitrate concentrations (IMPROVE).



Figure 4-8. Scatter diagrams of observed and simulated fine nitrate concentrations (IMPROVE).



Figure 4-9. Spatial display of the mean normalized error for nitrate (IMPROVE).





Figure 4-10. Seasonal performance statistics for fine nitrate concentrations (IMPROVE).





Figure 4-10 (continued). Seasonal performance statistics for fine nitrate concentrations (IMPROVE).

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	1.630			
Mean simulated value ( $\mu g/m^3$ )	2.875	2.893	2.263	1.870
Mean error ( $\mu g/m^3$ )	1.645	1.665	1.139	1.037
Mean bias (µg/m <sup>3</sup> )	1.245	1.264	0.633	0.240
Mean normalized error	1.943	1.988	1.260	1.014
Mean normalized bias	1.784	1.830	1.074	0.682
Normalized mean error	1.009	1.022	0.699	0.636
Normalized mean bias	0.764	0.775	0.388	0.148
Fractional error	0.788	0.799	0.625	0.632
Fractional bias	0.536	0.546	0.372	0.150
Correlation coefficient	0.615	0.610	0.629	0.615

Table 4-5. Annual performance statistics for total (gas + particulate) nitrate concentrations (CASTNet).



Figure 4-11.Scatter diagrams of observed and simulated total nitrate concentrations (CASTNet).





# Figure 4-12. Seasonal performance statistics for total nitrate concentrations (CASTNet).





### Figure 4-12 (continued). Seasonal performance statistics for total nitrate concentrations (CASTNet).

### 4.4 Organic Matter

Table 4-6 presents the annual performance statistics and Figure 4-13 presents a scatter plot for the IMPROVE data set with an OM/OC scaling factor of 1.4. Similar results are presented in Table 4-7 and Figure 4-14 for an OM/OC scaling factor of 2.1. On average, organic matter (OM) is the most important  $PM_{2.5}$  component in the western United States with an annual average concentration of 1.5 µg/m<sup>3</sup> using an OM/OC scaling factor of 2.1. However, there are some time periods and some areas of the domain where OM is not the most important component (e.g., California where nitrate dominates).

All models underestimate these average concentrations. CAMx M4 shows the largest underestimations with an average OM concentration of about 0.7  $\mu$ g/m<sup>3</sup>. CMAQ and REMSAD show average OM concentrations of 1.1 and 1.2  $\mu$ g/m<sup>3</sup>, respectively. These average underpredictions are significant, particularly when the OM/OC factor of 2.1 is used (factors of about 3.1, 2.1 and 1.9 for CAMx M4, CMAQ and REMSAD, respectively). These underpredictions are reflected in all the different bias metrics. CAMx M4 shows the largest biases; REMSAD shows the smallest (yet, significant) biases.

The mean error is in the range of 0.9 to 1.0  $\mu$ g/m<sup>3</sup> when using the 1.4 scaling factor and in the range of 1.4 to 1.7  $\mu$ g/m<sup>3</sup> when using the 2.1 scaling factor. The mean normalized errors exceed 60% and the fractional errors exceed 70% with the 1.4 scaling factor. Typically, one does not currently expect the normalized error for OM to be less than 50%. However, there is a clear bias here toward underprediction of OM by all four models. This bias would likely be even larger if we took into account the sampling artifacts (which are more likely to be negative artifacts than positive artifacts). This suggests that primary OM emissions (and, possibly, boundary conditions) are underestimated and/or that secondary organic aerosol (SOA) formation is underestimated. Primary OM emissions originating, for example, from forest fires are difficult to estimate and improvements are needed for such emission inventories. As demonstrated by Pun et al. (2003a), there are still large uncertainties associated with SOA formation. Those uncertainties are present in the emissions of VOC precursors (particularly biogenic

precursors), the yields of formation of semi-volatile compounds, and the partitioning of SOA between the gas and particulate phases.

The correlation coefficients are in the range of 0.4 to 0.5, i.e., the models can explain between 20 and 25% of the variance in the measured OM concentrations.

Figure 4-15 presents a spatial display of the mean normalized errors for the four model simulations (using a 2.1 OM/OC scaling factor). No particular spatial pattern appears. The San Francisco Bay Area shows the largest error in the domain.

Figures 4-16 and 4-17 present seasonal statistics for OM concentrations derived from IMPROVE using the default 1.4 OM/OC scaling factor and a 2.1 scaling factor, respectively. The normalized biases tend to be larger in winter than in spring or summer. Seasonal trends are not as apparent for the normalized errors. On the other hand, better correlations are obtained in winter than in spring.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	1.528			
Mean simulated value ( $\mu g/m^3$ )	0.740	0.733	1.089	1.228
Mean error ( $\mu g/m^3$ )	0.963	0.968	0.900	0.974
Mean bias (µg/m <sup>3</sup> )	-0.788	-0.795	-0.439	-0.300
Mean normalized error	0.600	0.602	0.651	0.687
Mean normalized bias	-0.378	-0.384	-0.098	-0.078
Normalized mean error	0.631	0.634	0.589	0.638
Normalized mean bias	-0.516	-0.521	-0.287	-0.196
Fractional error	0.809	0.815	0.709	0.742
Fractional bias	-0.689	-0.697	-0.429	-0.434
Correlation coefficient	0.444	0.443	0.481	0.488

Table 4-6. Annual performance statistics for fine organic matter concentrations (IMPROVE, scaling factor of 1.4).

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	2.291			
Mean simulated value ( $\mu g/m^3$ )	0.740	0.733	1.089	1.228
Mean error ( $\mu g/m^3$ )	1.652	1.658	1.457	1.459
Mean bias (µg/m <sup>3</sup> )	-1.551	-1.558	-1.202	-1.062
Mean normalized error	0.691	0.693	0.652	0.696
Mean normalized bias	-0.565	-0.570	-0.376	-0,336
Normalized mean error	0.721	0.724	0.636	0.637
Normalized mean bias	-0.677	-0.680	-0.525	-0.464
Fractional error	1.034	1.041	0.872	0.893
Fractional bias	-0.964	-0.991	-0.744	-0.745
Correlation coefficient	0.444	0.443	0.481	0.488

Table 4-7. Annual performance statistics for fine organic matter concentrations (IMPROVE, scaling factor of 2.1).



Figure 4-13. Scatter diagrams of observed and simulated fine organic matter concentrations (IMPROVE, scaling factor of 1.4).



Figure 4-14. Scatter diagrams of observed and simulated fine organic matter concentrations (IMPROVE, scaling factor of 2.1).



Figure 4-15. Spatial display of the mean normalized error for organic matter (IMPROVE, 2.1xOC).





Figure 4-16. Seasonal performance statistics for fine organic matter concentrations (IMPROVE, scaling factor of 1.4).





Figure 4-16 (continued). Seasonal performance statistics for fine organic matter concentrations (IMPROVE, scaling factor of 1.4).





# Figure 4-17. Seasonal performance statistics for fine organic matter concentrations (IMPROVE, scaling factor of 2.1).





Figure 4-17 (continued). Seasonal performance statistics for fine organic matter concentrations (IMPROVE, scaling factor of 2.1).

#### 4.5 Black Carbon

Table 4-8 presents the annual performance statistics for black carbon and Figure 4-18 presents a scatter plot for the IMPROVE data set. On average, black carbon (BC) is the least important  $PM_{2.5}$  component (among those explicitly measured in IMPROVE) in the western United States with an annual average concentration of only 0.27 µg/m<sup>3</sup>.

Three models (CAMx M4 with two and four sections, and CMAQ) underestimate and one model (REMSAD) overestimates this average concentration. The underestimation is about 20% and the overestimation about 10%. BC is a primary component of  $PM_{2.5}$ ; i.e., it is directly emitted into the atmosphere. Therefore, the errors and biases reflect uncertainties in emissions, transport and deposition processes. Because emissions are the same for all four models, the difference that appears between REMSAD and the other three models reflects a difference in the formulation of the transport and deposition processes. Also, the REMSAD overestimation of BC explains in part the fact that REMSAD simulates greater OM than the other three models (see above) because some primary OM is associated with BC.

The mean normalized errors are in the range of 60 to 70% and the fractional errors are in the range of 70 to 80%. Such errors are not unexpected if one considers that there are uncertainties in the BC emissions (e.g., from forest fires) as well as in the measurements (on the order of 30%, see above).

The correlation coefficients are about 0.5. They are among the best for all PM components which suggests that the variance in the observations is best reproduced for primary PM components.

Figure 4-19 presents a spatial display of the mean normalized errors for the four model simulations. No synoptic spatial patterns appear. However, the Class I areas that exhibit the largest errors are located near large urban areas: San Gorgonio Wilderness Area near Los Angeles, CA, Tonto National Monument near Phoenix, AZ, Lone Peak Wilderness Area near Salt Lake City, UT, and Point Reyes National Seashore near San Francisco, CA. These large errors correspond to overestimations and suggest that BC emissions in urban areas are overestimated.

Figure 4-20 depicts seasonal statistics for BC from IMPROVE. CMAQ and CAMx show larger normalized biases in summer than winter but REMSAD shows larger biases in summer than winter. The correlation coefficients are larger in fall and smaller in spring.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	0.268			
Mean simulated value ( $\mu g/m^3$ )	0.216	0.216	0.217	0.293
Mean error ( $\mu g/m^3$ )	0.170	0.170	0.191	0.216
Mean bias (µg/m <sup>3</sup> )	-0.051	-0.052	-0.051	-0.025
Mean normalized error	0.681	0.682	0.715	0.899
Mean normalized bias	0.016	0.017	-0.095	0.343
Normalized mean error	0.635	0.635	0.713	0.805
Normalized mean bias	-0.192	-0.193	-0.189	0.095
Fractional error	0.669	0.669	0.772	0.667
Fractional bias	-0.317	-0.317	-0.474	-0.152
Correlation coefficient	0.534	0.534	0.529	0.494

Table 4-8. Annual performance statistics for fine black carbon concentrations (IMPROVE).


Figure 4-18. Scatter diagrams of observed and simulated fine black carbon concentrations (IMPROVE).



Figure 4-19. Spatial display of the mean normalized error for black carbon (IMPROVE).





## Figure 4-20. Seasonal performance statistics for fine black carbon concentrations (IMPROVE).





Figure 4-20 (continued). Seasonal performance statistics for fine black carbon concentrations (IMPROVE).

#### 4.6 PM<sub>2.5</sub>

Table 4-9 presents the annual performance statistics and Figure 4-21 presents a scatter plot for the  $PM_{2.5}$  concentrations measured by IMPROVE. Seasonal statistics are presented in Figure 4-22. Table 4-10 presents the annual performance statistics and Figure 4-23 presents a scatter plot for the  $PM_{2.5}$  concentrations reconstructed according to the IMPROVE methodology from measurements of individual  $PM_{2.5}$  components. Seasonal statistics are presented in Figure 4-24.

All models underestimate the  $PM_{2.5}$  average concentrations. The average underestimation ranges from about 0.6  $\mu$ g/m<sup>3</sup> (CAMx M4) to about 1  $\mu$ g/m<sup>3</sup> (CMAQ). However, the measured PM<sub>2.5</sub> concentrations are likely to underestimate actual concentrations during the warm months (due to volatilization of ammonium nitrate and some organics) and could overestimate actual concentrations during cold months because of the presence of water. The reconstructed PM<sub>2.5</sub> concentrations may overestimate (because of the assumption that sulfate and nitrate are fully neutralized by ammonia) or underestimate (because of the 1.4 scaling factor to convert OC to OM) the actual concentrations. Therefore, it is not feasible to reach definitive conclusions on the performance of the models based on the IMPROVE PM<sub>2.5</sub> estimates. Nevertheless, it is likely that the models underestimate the actual PM<sub>2.5</sub> concentrations for the following resasons. First, the measured and reconstructed PM<sub>2.5</sub> concentrations are more likely to be underestimated than overestimated so that the biases reported in the performance assessment would be lower limits. Second, the two major components of PM2.5 (sulfate and OM) were underestimated by the models.

Normalized errors and biases are larger in winter (except for the fractional errors that are larger in spring) and smaller in summer (except for the fractional errors that tend to be smaller in fall). The correlation coefficients increase significantly from spring to summer, fall and winter.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	4.641			
Mean simulated value ( $\mu g/m^3$ )	4.072	3.932	3.666	3.838
Mean error ( $\mu g/m^3$ )	2.949	2.900	2.746	3.000
Mean bias (µg/m <sup>3</sup> )	-0.569	-0.709	-0.983	-0.804
Mean normalized error	0.754	0.729	0.644	0.725
Mean normalized bias	0.203	0.162	0.020	0.101
Normalized mean error	0.635	0.625	0.592	0.646
Normalized mean bias	-0.123	-0.153	-0.212	-0.173
Fractional error	0.638	0.641	0.634	0.675
Fractional bias	-0.163	-0.189	-0.280	-0.257
Correlation coefficient	0.391	0.394	0.433	0.400

Table 4-9. Annual performance statistics for measured  $PM_{2.5}$  concentrations (IMPROVE).

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	4.031			
Mean simulated value ( $\mu g/m^3$ )	4.218	4.065	3.776	3.942
Mean error ( $\mu g/m^3$ )	2.640	2.563	2.380	2.642
Mean bias (µg/m <sup>3</sup> )	0.188	0.035	-0.255	-0.088
Mean normalized error	0.790	0.752	0.654	0.759
Mean normalized bias	0.340	0.285	0.136	0.234
Normalized mean error	0.655	0.636	0.590	0.655
Normalized mean bias	0.047	0.009	-0.063	-0.022
Fractional error	0.593	0.594	0.581	0.629
Fractional bias	-0.044	-0.072	-0.161	-0.139
Correlation coefficient	0.398	0.405	0.454	0.419

Table 4-10. Annual performance statistics for reconstructed PM<sub>2.5</sub> concentrations (IMPROVE).



Figure 4-21. Scatter diagrams of observed and simulated  $PM_{2.5}$  concentrations (IMPROVE).





# Figure 4-22. Seasonal performance statistics for $PM_{2.5}$ concentrations (IMPROVE, measured).





## Figure 4-22 (continued). Seasonal performance statistics for PM<sub>2.5</sub> concentrations (IMPROVE, measured).



Figure 4-23. Scatter diagrams of reconstructed and simulated  $PM_{2.5}$  concentrations (IMPROVE).





## Figure 4-24. Seasonal performance statistics for $PM_{2.5}$ concentrations (IMPROVE, reconstructed).





## Figure 4-24 (continued). Seasonal performance statistics for PM<sub>2.5</sub> concentrations (IMPROVE, reconstructed).

#### 4.7 Coarse Mass

Table 4-11 presents the annual performance statistics and Figure 4-25 presents scatter plots for the PM coarse concentrations measured by IMPROVE as the difference between  $PM_{10}$  and  $PM_{2.5}$  concentrations. The average coarse PM concentration is commensurate with that of  $PM_{2.5}$ ; i.e., about 5 µg/m<sup>3</sup>.

All models significantly underestimate the average coarse PM concentrations by factors of 6 (REMSAD) to 18 (CMAQ). The average underestimations are on the order of 5  $\mu$ g/m<sup>3</sup>. The normalized errors and biases also reflect the poor performance of the models for coarse PM. The correlation coefficients are lower than those obtained for the PM<sub>2.5</sub> components discussed above. Less than 5% of the variance in the coarse PM observations can be explained by the models.

Figure 4-26 presents seasonal statistics for the IMPROVE coarse PM concentrations. Little seasonal variation appears for those statistics that reflect large underpredictions by all four models.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	5.578			
Mean simulated value ( $\mu g/m^3$ )	0.315	0.815	0.772	0.854
Mean error (µg/m <sup>3</sup> )	5.269	4.841	4.846	4.819
Mean bias (µg/m <sup>3</sup> )	-5.264	-4.764	-4.807	-4.725
Mean normalized error	0.915	0.850	0.842	0.852
Mean normalized bias	-0.870	-0.649	-0.698	-0.649
Normalized mean error	0.945	0.868	0.869	0.864
Normalized mean bias	-0.944	-0.854	-0.862	-0.847
Fractional error	1.672	1.350	1.378	1.354
Fractional bias	-1.657	-1.286	-1.334	-1.287
Correlation coefficient	0.180	0.124	0.205	0.134

Table 4-11. Annual performance statistics for coarse mass concentrations (IMPROVE).



Figure 4-25. Scatter diagrams of observed and simulated coarse mass concentrations (IMPROVE).





## Figure 4-26. Seasonal performance statistics for coarse mass concentrations (IMPROVE).





### Figure 4-26 (continued). Seasonal performance statistics for coarse mass concentrations (IMPROVE).

#### 4.8 PM<sub>10</sub>

Table 4-12 presents the annual performance statistics and Figure 4-27 presents a scatter plot for the  $PM_{10}$  concentrations measured by IMPROVE. The average annual  $PM_{10}$  concentration is about 10  $\mu$ g/m<sup>3</sup>.

All models significantly underestimate the average  $PM_{10}$  concentration as a result of the underestimations in the coarse PM concentration. The average underestimations are on the order of 5 to 6 µg/m<sup>3</sup>; i.e., similar to the underpredictions in coarse mass. The normalized errors and biases also reflect the poor performance of the models for  $PM_{10}$ . The correlation coefficients range between 0.25 and 0.3; i.e., between the values obtained for the fine and coarse PM fractions. Less than 10% of the variance in the coarse PM observations can be explained by the models.

Figure 4-28 presents the seasonal statistics for  $PM_{10}$  from IMPROVE. The normalized biases are larger in winter and smaller in summer. However, the normalized errors tend to be smaller in winter (or fall) and larger in spring. The correlation coefficients follow a similar trend as those for  $PM_{2.5}$ , increasing from spring through winter.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value ( $\mu g/m^3$ )	10.165			
Mean simulated value ( $\mu g/m^3$ )	4.210	4.569	4.249	4.481
Mean error (µg/m <sup>3</sup> )	6.893	6.721	6.699	6.856
Mean bias (µg/m <sup>3</sup> )	-5.955	-5.595	-5.916	-5.684
Mean normalized error	0.637	0.628	0.615	0.649
Mean normalized bias	-0.400	-0.348	-0.427	-0.367
Normalized mean error	0.678	0.661	0.659	0.675
Normalized mean bias	-0.586	-0.551	-0.582	-0.559
Fractional error	0.889	0.847	0.877	0.885
Fractional bias	-0.748	-0.682	-0.762	-0.728
Correlation coefficient	0.247	0.250	0.303	0.253

Table 4-12. Annual performance statistics for measured  $PM_{10}$  concentrations (IMPROVE).



Figure 4-27. Scatter diagrams of observed and simulated  $PM_{10}$  concentrations (IMPROVE).











## Figure 4-28 (continued). Seasonal performance statistics for PM<sub>10</sub> concentrations (IMPROVE).

#### 4.9 Wet Deposition

Performance statistics for wet deposition of sulfate, nitrate and ammonium are presented in Tables 4-13, 4-14 and 4-15, respectively. (A complete set of simulation results from CMAQ were not available.) Scatter diagrams of the measured and simulated values are presented in Figures 4-29, 4-31 and 4-32, respectively. Seasonal statistics are presented in Figures 4-30, 4-32 and 4-34, respectively.

The mean simulated values are significantly lower than the measured values for all three chemical species and for all models (CAMx M4, CAMx 4Sec and REMSAD). Sulfate deposition is underpredicted by a factor of 5 to 7, nitrate by a factor of ~3 and ammonium by a factor of 6 to 9. Normalized errors exceed 80% and normalized biases are significant in most cases. There is little correlation between the simulated and measured wet deposition fluxes; less than 5 % of the variance on the measurements is explained by the models and, in the case of ammonium, there is a negative correlation between the simulated and measured values.

These results indicate that a major removal of PM from the atmosphere is underpredicted by the models.

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value (kg/ha)	0.246			
Mean simulated value (kg/ha)	0.050	0.033		0.034
Mean error (kg/ha)	0.218	0.222		0.219
Mean bias (kg/ha)	-0.196	-0.213		-0.212
Mean normalized error	0.935	0.887		0.876
Mean normalized bias	-0.555	-0.713		-0.734
Normalized mean error	0.887	0.901		0.892
Normalized mean bias	-0.795	-0.865		-0.734
Fractional error	1.479	1.544		1.520
Fractional bias	-1.325	-1.466		-1.460
Correlation coefficient	0.144	0.149		0.204

Table 4-13. Annual performance statistics for sulfate wet deposition (NADP).

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value (kg/ha)	0.246			
Mean simulated value (kg/ha)	0.083	0.078		0.085
Mean error (kg/ha)	0.213	0.212		0.205
Mean bias (kg/ha)	-0.163	-0.167		-0.161
Mean normalized error	1.030	1.013		0.992
Mean normalized bias	-0.231	-0.266		-0.261
Normalized mean error	0.866	0.864		0.834
Normalized mean bias	-0.664	-0.681		-0.654
Fractional error	1.295	1.303		1.281
Fractional bias	-1.007	-1.032		-1.008
Correlation coefficient	0.107	0.109		0.177

Table 4-14. Annual performance statistics for nitrate wet deposition (NADP).

Performance metric	CAMx M4	CAMx 4Sec	CMAQ	REMSAD
Mean observed value (kg/ha)	0.167			
Mean simulated value (kg/ha)	0.026	0.027		0.019
Mean error (kg/ha)	0.155	0.155		0.154
Mean bias (kg/ha)	-0.141	-0.139		-0.147
Mean normalized error	0.928	0.941		0.862
Mean normalized bias	-0.491	-0.454		-0.664
Normalized mean error	0.929	0.928		0.923
Normalized mean bias	-0.847	-0.835		-0.884
Fractional error	1.416	1.401		1.452
Fractional bias	-1.230	-1.195		-1.359
Correlation coefficient	-0.037	-0.036		-0.017

Table 4-15. Annual performance statistics for ammonium wet deposition (NADP).



Figure 4-29. Scatter diagrams of sulfate wet deposition (NADP).





Figure 4-30. Seasonal performance statistics for sulfate wet deposition (NADP).



## Figure 4-30 (continued). Seasonal performance statistics for sulfate wet deposition (NADP).



Figure 4-31. Scatter diagrams of nitrate wet deposition (NADP).



Figure 4-32. Seasonal performance statistics for nitrate wet deposition (NADP).



## Figure 4-32 (continued). Seasonal performance statistics for nitrate wet deposition (NADP).













Figure 4-34 (continued). Seasonal performance statistics for ammonium wet deposition (NADP).
#### 5. ASSESSMENT OF MODEL PERFORMANCE

#### 5.1 Overall Assessment of Model Performance

Performance for ozone non-urban concentrations shows acceptable errors but significant biases (8 to 10 ppb underprediction for concentrations above 40 ppb on average). Such biases suggest possible uncertainties with the oxidants associated with secondary PM formation ( $O_3$ , OH and  $H_2O_2$ ) that may affect model performance for PM.

Because of significant uncertainties associated with the measurements of total  $PM_{2.5}$  concentrations, our assessment focuses on the available measurements of  $PM_{2.5}$  components; i.e., sulfate, nitrate, organic matter (OM) and black carbon (BC).

The models typically explain only between 15 and 30% of the variance in the observed BC and sulfate concentrations, between 20 and 25% of the variance in the observed OM concentrations, about 20% of the variance in the nitrate concentrations and about 40% of the variance in the total inorganic nitrate concentrations (i.e.,  $HNO_3$  + particulate nitrate). (One exception is the low, 6 to 7%, variance explained for IMPROVE and CASTNet sulfate by REMSAD.) All normalized errors (i.e., mean normalized error, normalized mean error, fractional error and normalized root mean square error) are greater than 50%, except for CMAQ performance for CASTNet sulfate where the normalized errors are slightly less than 50%. In the case of nitrate, all normalized errors exceed 100%. All models show significant biases for nitrate (large overpredictions) and OM (large underpredictions).

Coarse PM concentrations (measured as the difference between  $PM_{10}$  and  $PM_{2.5}$  concentrations) are underpredicted by about one order of magnitude by all models. Less than 5% of the variance in the measured coarse PM concentrations is explained by the models.

Wet deposition fluxes of sulfate, nitrate and ammonium are significantly underpredicted by all models.

# 5.2 CAMx M4 versus CAMx 4Sec

CAMx M4 uses two PM size sections and sulfate, nitrate, ammonium, OM and BC are assumed to be present only in the fine section. CAMx 4Sec uses four PM sections (three for fine PM and one for coarse PM) and sulfate, nitrate, ammonium and SOA can be present in any section. The formation of secondary PM via gas-to-particle conversion occurs predominantly in the fine size sections because those sections typically have the largest surface area. A large fraction of sulfate is formed in cloud droplets via aqueous-phase chemistry. As those droplets evaporate, they may lead to the formation of some sulfate in the coarse size section. Thus, all sulfate in CAMx M4 is present in the fine section (i.e., PM<sub>2.5</sub>) whereas in CAMx 4Sec sulfate is distributed between the fine and coarse PM fractions.

There is no significant change in model performance associated with the use of a more detailed representation of PM in CAMx (i.e., 4 PM size sections instead of 2) except for sulfate. The change in the coefficient of determination  $(r^2)$  between the two CAMx versions is less than 0.2% for nitrate, OM and BC; however, it improves for sulfate from 19% for CAMx M4 to 27% for CAMx 4Sec. The mean normalized error for sulfate decreases from 95% (CAMx M4) to 66% (CAMx 4Sec) and the mean normalized bias decreases from 53% (CAMx M4) to 10% (CAMx 4Sec).

Figure 5-1 presents a comparison of the fine sulfate concentrations simulated by CAMx M4 and CAMx 4Sec as well as a comparison of the total (i.e., fine and coarse) sulfate concentrations simulated by these two models (all sulfate is in the fine fraction in CAMx M4). These comparisons are for concentrations at the IMPROVE sites. It appears that the lower fine sulfate concentrations simulated by CMAx 4Sec compared to CAMx M4 are not only due to the presence of some sulfate in the coarse PM fraction but also to lower total sulfate concentrations. The lower total sulfate concentrations are likely due to larger dry deposition rates for coarse sulfate particles than for fine sulfate particles (Seinfeld and Pandis, 1998), thereby leading to a more rapid removal of sulfate in CAMx 4Sec than in CAMx M4.



Figure 5-1. Comparison of annual sulfate concentrations simulated by CAMx M4 and CAMx 4Sec at the IMPROVE sites: (a) fine sulfate concentrations, (b) total sulfate concentrations.

The difference between CAMx M4 and CAMx 4Sec sulfate concentrations are significant enough that one should investigate whether a more detailed treatment of the sulfate size distribution is warranted.

First, one should assess whether four size sections, as used in CAMx 4Sec, are sufficient to provide an accurate representation of the sulfate size distribution (PM-CAMx, for example, used twice as many size sections over the same size range). Seigneur et al. (1986) showed that the ability of an aerosol dynamics model to simulate the PM size distribution deteriorates (in particular for condensational growth) as the size resolution decreases. The ability to correctly simulate the evolution of the particle size distribution also depends on the condensational growth algorithm used in the model (Zhang et al., 1999). It is not clear whether any tests were conducted with CAMx 4Sec over a wide range of conditions to select an optimal size representation.

Zhang et al. (2004) investigated the effect of the particle size resolution on PM concentrations with a 3-D air quality model. They found that the use of a finer particle size resolution (8 size sections versus 2 size sections) leads to lower sulfate concentrations in the  $PM_{2.5}$  size range. That result is qualitatively similar to that obtained with CAMx in this study; however, the differences obtained by Zhang et al. (2004) are significantly less than those obtained in this application of CAMx. Clearly, there are differences in the applications (five-day episode in the Los Angeles basin versus annual simulation over the western United States) and the models (e.g., different condensational growth algorithms). Nevertheless, a fundamental difference between these two modeling studies is the number of size sections (8 versus 4) used to represent the PM size distribution. One would expect that a coarser size resolution may lead to more numerical diffusion and, therefore, to more growth of sulfate into the coarse mode (see Seigneur et al., 1986). Also, a coarse size resolution will affect dry and possibly wet deposition rates for sulfate. Using a single size section for the coarse mode results in a representative particle diameter of 5 µm. Using two sections for the coarse mode would result in a representative particle diameter of 3.5 µm for the lower size section (where most coarse sulfate is likely to be present). Therefore, the use of a single size section for the coarse PM fraction may lead to overestimates of coarse sulfate wet and dry deposition. Such hypotheses will need to be tested in a systematic manner to determine whether 4 size sections are sufficient to provide a detailed representation of the particle size distribution.

Once an appropriate representation of the PM size distribution has been selected for CAMx, CAMx 4Sec (or CAMx Nsec if a different number of size sections is warranted) should be evaluated with field data that include sulfate size distribution data to evaluate whether the sulfate size distributions simulated by CAMx 4Sec (or CAMx Nsec) are realistic.

If a significant fraction of sulfate is indeed present in the coarse mode, the origin of that sulfate should be elucidated and possible parameterizations using a two-size representation should be investigated.

### 5.3 Comparison of Performance among CAMx, CMAQ and REMSAD

No model shows consistently better performance than the other models. There are, however, differences among the models for specific PM species that suggest possible areas of improvements.

For sulfate, CMAQ and CAMx 4Sec show comparable performance. REMSAD shows the worst performance with a low correlation coefficient (0.27) and large mean normalized error (nearly 100%). The fact that CAMx M4 shows a slightly worse performance than CMAQ indicates that there are several factors that affect sulfate model performance. Both CAMx M4 and CMAQ assume that all sulfate is present in the fine PM fraction whereas CAMx 4Sec allows sulfate to be distributed among the various PM size sections including coarse PM. Therefore, sulfate formation and/or removal must differ between CMAQ and CAMx because the version of CAMx that treats the sulfate size distribution differently than CMAQ is in better agreement with CMAQ than the version of CAMx that treats the sulfate size distribution similarly to CMAQ. The chemistry of sulfate formation is similar in CMAQ and CAMx (CBM-IV for gas-phase chemistry and RADM for aqueous-phase chemistry). Therefore, differences in dry deposition may explain this discrepancy between CMAQ and CAMx.

For nitrate, CAMx (M4 and 4Sec) overestimates observations significantly more than CMAQ and REMSAD. For example, the mean normalized errors exceed 450% for

CAMx whereas they are in the 300 to 350% range for CMAQ and REMSAD. However, the correlation coefficients are similar among the four models. This suggests that there is a systematic bias in the nitrate calculation of CAMx that should be investigated.

For OM, CAMx (M4 and 4Sec) tends to underestimate observations more than CMAQ and REMSAD. Although the errors are commensurate among the models, CAMx shows larger negative biases than CMAQ and REMSAD. On average, CAMx underestimates the OM concentration by a factor of 3.8 (using an OM/OC scaling factor of 2.1), compared to factors of 2.6 and 2.3 for CMAQ and REMSAD, respectively. There are currently many uncertainties associated with model performance for OM (including modeling uncertainties arising from emission inventories, chemical kinetics and gas/particle partitioning, measurement uncertainties). Nevertheless, these results suggest that, if the large uncertainties that appear in this particular performance evaluation are confirmed in other performance evaluations, the formation of SOA in CAMx may need to be improved.

For BC, REMSAD tends to overestimate whereas CAMx (M4 and 4Sec) and CMAQ tend to underestimate. Model performance for BC reflects uncertainties in the BC emission inventory, the model formulation for transport and dispersion (including meteorological inputs), as well as measurement uncertainties. It is, therefore, not feasible to draw any conclusions regarding the adequacy of model formulation based on the BC performance evaluation results. An evaluation of the transport and dispersion modules of 3-D air quality models should be conducted using tracer data (e.g., Moran and Pielke, 1996; Pun et al., 2003b).

## 5.4 Recommendations

The limited amount of data that is available to conduct diagnostic performance evaluations prevents us to develop specific recommendations to address model performance problems. Nevertheless, some general areas of investigation can be proposed to improve the model performance of future applications.

#### 5.4.1 Sulfate

Sulfate concentrations vary significantly among the models (difference in average concentrations of a factor of nearly 2 between CAMx M4 and REMSAD and a factor of 1.3 between CAMx M4 and CAMx 4Sec). Sulfate is considered to be the PM component that is the best understood and should be modeled with the best accuracy (e.g., Seigneur and Moran, 2004). Therefore, these results suggest that our current state of knowledge for sulfate needs to be reviewed. Sources of uncertainties include model inputs and model formulation (sulfate measurements are considered to be reliable).

Among the model inputs, emissions of  $SO_2$  and sulfate are reasonably well characterized and most of the uncertainties are likely to result from the boundary conditions and the meteorological inputs. Therefore, we recommend that boundary conditions be obtained from a global model as was done in the BRAVO study (Pun et al., 2004b) and that meteorological inputs for clouds and precipitation (that affect sulfate formation and removal) be improved. To that end, data assimilation of cloud and precipitation data during meteorological modeling is a possible approach.

The sensitivity of sulfate simulation results to the model formulation appears clearly in the four model simulations analyzed here. The difference between CAMx M4 and CAMx 4Sec is due to different size representations of sulfate (in the fine mode in CAMx M4 and in both the fine and coarse modes in CAMx 4Sec) that are likely to affect atmospheric removal rates. In addition, although all models applied here used a single cloud droplet size, there is some evidence that a size representation of cloud droplets will affect the sulfate formation rate, typically leading to more sulfate formation (10 to 50% more in a recent Los Angeles simulation, Fahey and Pandis, 2003).

One must note that a better model formulation may not necessarily lead to better model performance in cases where model performance is mostly governed by uncertainties in the model inputs. Also, a database such as the one available here for the western United States in 1996 is inadequate to evaluate detailed model formulations. For example, sulfate size distribution data are not available to evaluate the multi-size representation of CAMx 4Sec for sulfate.

#### 5.4.2 Nitrate

Nitrate was significantly overestimated by all models, with CAMx showing the largest overpredictions. Total inorganic nitrate (i.e.,  $HNO_3$  + particulate nitrate) was also significantly overpredicted by all models. Therefore, the gas/particle partitioning of inorganic nitrate is not the main cause for the overprediction of particulate nitrate.

Sources of uncertainties include model inputs and model formulation (nitrate measurements on Nylon filters are considered to be reliable).

Among the model inputs, emissions of  $NO_x$  are reasonably well characterized (although  $NO_x$  emissions from fires are uncertain) and most of the uncertainties are likely to result from the emissions of ammonia (NH<sub>3</sub>), the boundary conditions and the meteorological inputs. There are various ongoing efforts to improve NH<sub>3</sub> emission inventories. However, we will not be able to evaluate such inventories until routine measurements of total ammonium (i.e., particulate ammonium and gas-phase NH<sub>3</sub>) become available. At the moment, IMPROVE does not measure ammonium and the Speciation Trends Network (STN) only measures particulate ammonium. There is a dire need for routine measurements of gas-phase NH<sub>3</sub>

Although boundary conditions do not appear to be as critical for nitrate as they are for sulfate, they should be obtained from a global model along with those for sulfate for consistency. Meteorological inputs for clouds and precipitation affect nitrate formation and removal. However, their effect will be less than they are for sulfate because most sulfate is formed in the aqueous phase whereas a significant amount of nitrate is formed in the gas phase, and sulfate dry deposition is low whereas a significant nitrate fraction is removed as HNO<sub>3</sub> dry deposition. Nevertheless, the data assimilation of cloud and precipitation data during meteorological modeling that was recommended for improving sulfate simulations will also benefit nitrate simulations.

The ammonium nitrate gas-particulate equilibrium is very sensitive to temperature and relative humidity. The models use the temperature and relative humidity provided by the meteorological models. There are uncertainties associated with those variables that are not negligible (e.g., a few degrees for temperature). Using actual ambient temperature and relative humidity to calculate the gas/particle partitioning will provide a more representative evaluation of the air quality model formulation (e.g., Bhave, 2004).

Considering the current uncertainties with the model inputs, the CAMx formulation may not need any improvements. Nevertheless, a review of that formulation is recommended in light of the large overpredictions of nitrate concentrations obtained here.

### 5.4.3 Organic Matter

The significant underestimation of OM concentrations by CAMx and the other two models raises some doubts about the inputs relevant to OM and the formulation of the secondary organic areosol (SOA) module. The major sources of input uncertainties include emissions of primary OM, boundary conditions for OM concentrations and emissions of VOC that are precursors of SOA. Primary OM emissions are uncertain, particularly for non-routine emissions such as forest fires and biogenic emissions such as plant wax. Boundary conditions should be obtained from a global PM model. Emissions of anthropogenic VOC precursors are likely to be better known than those of biogenic VOC precursors (monoterpenes and sesquiterpenes), because the latter have uncertainties of about one order of magnitude.

Our current understanding of the formation of SOA is limited and large uncertainties are, therefore, associated with any SOA formulation (Pun et al., 2003a). Some processes that have been identified as important for SOA formation (e.g., dissolution of hydrophilic organic compounds in aqueous particles, polymerization of condensed organic compounds, SOA formation via acid-catalyzed oxidation of VOC) are not simulated in the models applied here. The treatment of OM in CAMx, CMAQ and REMSAD should be reviewed and, if warranted, revised based on current knowledge. However, one must keep in mind that uncertainties in model inputs may limit our ability to correctly simulate OM concentrations. Therefore, we recommend that SOA modules be evaluated against smog chamber data prior to their evaluation with 3-D models against ambient data.

#### 5.4.4 Black Carbon

The major issue with BC is the consistency between ambient and source measurements of BC. Also, emissions of BC from non-routine sources such as forest fires contribute to the large uncertainties that are likely associated with BC emissions.

# 5.4.5 Other Fine PM

Concentrations of other fine PM components were not evaluated here because of uncertainties associated with their measurements (by difference between  $PM_{2.5}$  and the measured  $PM_{2.5}$  components) or their definition (i.e., the "soil" component of IMPROVE). Nevertheless, it is worthwhile to point out that current emission inventories do not differentiate between crustal PM emissions (natural and anthropogenic) and industrial, residential and commercial primary PM. Also, emission models typically do not provide sea salt emissions (for example, in the BRAVO study, they were calculated separately). A better resolution of those primary PM emissions in terms of sources and chemical composition will be required to evaluate the performance of models for this PM category.

### 5.4.6 Coarse PM

Emissions of coarse PM appear to be significantly underestimated and there is a dire need for better emission inventories of coarse PM.

# 5.4.7 Wet deposition

Wet deposition is significantly underestimated by all models for all species (i.e., sulfate, nitrate and ammonium). Because wet deposition is a major removal mechanism for PM species, this result indicates that improvements in the modeling of wet deposition are required. Cloud and precipitation fields predicted by the meteorological model MM5 should be evaluated. Consideration should be given to assimilating cloud and

precipitation data in meteorological modeling because the times and locations of precipitation events are typically predicted with significant uncertainties.

## 6. CONCLUSION

Four air quality models (CAMx M4, CAMx 4Sec, CMAQ and REMSAD) were applied by ENVIRON and the University of California at Riverside to simulate PM concentrations over the western United States for the year 1996. We evaluated the performance of these models by comparing the model simulation results with measurements of ozone concentrations, PM concentrations (including several PM species and two PM size fractions) and wet deposition of sulfate, nitrate and ammonium.

The models explain less than 30% of the variance in all PM components. All normalized errors exceed 50% except CMAQ for CASTNet sulfate where the errors are slightly less than 50%. All models show significant overpredictions of nitrate concentrations and significant underpredictions of organic matter (OM). No model shows consistently better performance than the other models. There are, however, differences among the models that suggest possible areas for further investigation.

CAMx M4 (two size sections for PM, fine and coarse) and CAMx 4Sec (four size sections for PM) show similar results except for sulfate, with CAMx 4Sec showing better performance than CAMx M4. The difference is due to the presence of some sulfate in the coarse mode in CAMx 4Sec and a likely faster removal of sulfate by dry deposition in CAMx 4Sec than in CAMx M4. However, it is not clear whether four sections are sufficient to correctly simulate the evolution of the particle size distribution and additional testing of CAMx 4Sec is recommended (e.g., comparison of simulations with various size resolutions – 4 vs. 8 sections, comparison with size distributed field data).

For sulfate, CAMx 4Sec and CMAQ showed comparable performance. REMSAD showed the worst performance. The large differences obtained among the models (factor of two difference for average annual concentrations) suggest that improvements in the model inputs (boundary conditions, cloud and precipitation fields) and model formulation (sulfate formation in clouds, sulfate removal) are warranted.

For both particulate nitrate and total (i.e., gas + particulate) nitrate, CAMx overestimates more than CMAQ and REMSAD but shows similar coefficients of correlation. These results suggest that there is a systematic bias that should be

investigated. However, large uncertainties in the model inputs (ammonia inventory, boundary conditions, ambient temperature and relative humidity) are likely to have large associated uncertainties.

CAMx underestimates OM more than CMAQ and REMSAD. We anticipate that formulation of OM in air quality models will evolve as our understanding of SOA formation continues to improve. Also, efforts should be directed toward improving the emission inventories of primary OM and biogenic SOA precursors.

Model performance for BC reflects uncertainties in the emission inventory, ambient measurements and the formulation of transport and dispersion in the models.

Emissions of coarse PM appear to be significantly underestimated.

All models significantly underestimate wet deposition of sulfate, nitrate and ammonium, which suggests the presence of large uncertainties in the cloud and precipitation fields.

The results of this model performance evaluation point out the need to use the output of a global model for the boundary conditions and to improve various aspects of the emission inventory (ammonia, biogenic SOA precursors, primary PM, including BC and coarse PM). Data assimilation of clouds and precipitation is also recommended for meteorological modeling because clouds affect sulfate and nitrate formation (and possibly SOA formation) and precipitation is a major removal mechanism for PM.

Because of significant uncertainties in the model inputs and the limited amount of data available for diagnostic evaluation, it is not feasible to provide definitive conclusions regarding the adequacy of the model formulations. Nevertheless, areas that deserve further investigation include sulfate formation (e.g., use of multiple droplet sizes), nitrate formation, SOA formation, the size distribution of secondary PM, and PM removal rates (mainly wet deposition rates but also dry deposition rates).

# 7. REFERENCES

- Ashbaugh, L.L. and R.A. Eldred, 2004. Loss of particle nitrate from Teflon sampling filters: effects on measured gravimetric mass in California and in the IMPROVE network, J. Air Waste Manage. Assoc., 54, 93-104.
- Bhave, P., 2004. Postprocessing model output for comparison to ambient data, PM Model Performance Workshop, 10-11 February, 2004, Research Triangle Park, NC.
- Chow, J.C., J.G. Watson, D. Crow, D.H. Lowenthal and T. Merrifield, 2001. Comparison of IMPROVE and NIOSH carbon measurements, *Aerosol Sci. Technol.*, **34**, 23-34.
- EPA, 2001. Guidance for Demonstrating Attainment of Air Quality Goals for PM2.5 and Regional Haze, Draft version 2.1, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC.
- Eatough, D.J., R.W. Long, W.K. Modey and N.L Eatough, 2003. Semi-volatile secondary organic aerosol in urban atmospheres: meeting a measurement challenge, *Atmos. Environ.*, **37**, 1277-1292.
- Fahey, K.M. and S.N. Pandis, 2003. Size-resolved aqueous-phase atmospheric chemistry in a three-dimensional chemical transport model, *J. Geophys. Res.*, **108**, 4690.
- Hering, S. and G. Cass, 1999 The magnitude of bias in the measurement of PM2.5 arising from volatilization of particulate nitrate from teflon filters, *J. Air Waste Manage. Assoc.*, **49**, 725-733.
- Irwin, J., 2004. Model Evaluation; Looking for Spatial and Time Averaged Patterns, EPA PM Model Performance Workshop, February 10-11, 2004, Research Triangle Park, NC.
- Moran, M.D. and R.A. Pielke, 1996. Evaluation of a mesoscale atmospheric dispersion modeling system with observations from the 1980 Great Plains mesoscale tracer experiment- Part II: dispersion simulations, *J. Appl. Meteor.*, **35**, 308-329.
- Pang, Y., N.L. Eatough and D.J. Eatough, 2002. PM2.5, semivolatile organic material at Riverside, California: implications for the PM<sub>2.5</sub> Federal Reference Method sampler, *Aerosol Sci. Technol.*, 36, 277-288.
- Pang, Y., N.L. Eatough, J. Wilson and D.J. Eatough, 2002. Effect of semivolatile material on PM2.5 measurement by the PM<sub>2.5</sub> Federal Reference Method sampler at Bakersfield, California, *Aerosol Sci. Technol.*, 36, 289-299.

- Pun, B., S.-Y. Wu, C. Seigneur, J.H. Seinfeld, R.J. Griffin and S.N. Pandis, 2003a. Uncertainties in modeling secondary organic aerosols: three-dimensional modeling studies in Nashville/western Tennessee, *Environ. Sci. Technol.*, 37, 3647-3661.
- Pun, B., C. Seigneur, S.-Y. Wu and N. Kumar, 2003b. Simulating inert tracers in the BRAVO project using CMAQ, 96th Air and Waste Management Association Annual Meeting, June 22-26, 2003, San Diego, California.
- Pun, B.K., S.-Y. Chen, K. Lohman and C. Seigneur, 2004a. Model performance evaluation data base and software – application to CENRAP, PM Model Performance Workshop, 10-11 February, 2004, Research Triangle Park, NC.
- Pun, B., C. Seigneur, S.-Y. Wu, E. Knipping and N. Kumar, 2004b. Modeling Analysis of the Big Bend Regional Aerosol Visibility Observational (BRAVO) Study, Report 1009283, EPRI, Palo Alto, CA.
- Seigneur, C., A.B. Hudischewskyj, J.H. Seinfeld, K.T. Whitby, E.R. Whitby, J.R. Brock and H.M. Barnes, 1986. Simulation of aerosol dynamics: A comparative review of mathematical models, *Aerosol Sci. Technol.*, **5**, 205-222.
- Seigneur, C., B. Pun, P. Pai, J. F. Louis, P. Solomon, C. Emery, R. Morris, M. Zahniser, D. Worsnop, P. Koutrakis, W. White and I. Tombach, 2000. Guidance for the performance evaluation of three-dimensional air quality modeling systems for particulate matter and visibility, J. Air Waste Manage. Assoc., 50, 588-599.
- Seigneur, C. and M. Moran, 2003. Chapter 8: Chemical transport models, *in* Particulate Matter Science for Policy Makers: A NARSTO Assessment, P.H. McMurry, M. Shepherd and J. Vickery, eds.
- Seigneur, C., 2004. Model evaluation: comparing model output to ambient data, PM Model Performance Workshop, 10-11 February, 2004, Research Triangle Park, NC.
- Seinfeld, J.H. and S.N. Pandis, 1998. Atmospheric Chemistry and Physics, John Wiley & Sons, Inc., New York, NY.
- Solomon, P.A. et al., 2004. PM Model Performance Workshop, 10-11 February, 2004, Research Triangle Park, NC.
- Tonnesen, G., 2004. Private communication, May 2004, University of California, Riverside, CA. May.
- Turpin, B.J., P. Saxena and E. Andrews, 2000. Measuring and simulating particulate organics in the atmosphere: problems and prospects, *Atmos. Environ.*, 34, 2983-3013.

- Turpin, B.J. and H-J. Lim, 2001. Species contributions to PM<sub>2.5</sub> mass concentrations: revisiting common assumptions for estimating organic mass, *Aerosol Sci. Technol.*, **35**, 602-610.
- Zhang, Y., C. Seigneur, J. H. Seinfeld, M. Z. Jacobson and F. Binkowski, 1999. Simulation of aerosol dynamics: A comparative review of algorithms used in air quality models, *Aerosol Sci. Technol.*, **31**, 487-514.
- Zhang, Y., B. Pun, K. Vijayaraghavan, S.-Y. Wu, C. Seigneur, S. Pandis, M. Jacobson, A. Nenes and J.H. Seinfeld, 2004. Development and application of the Model of Aerosol Dynamics, Reaction, Ionization and Dissolution, J. Geophys. Res., 109, D01202, doi:10.1029/2003JD003501.