Evaluation of Model Simulated Atmospheric Constituents with Observations in the Factor Projected Space: CMAQ Simulations of SEARCH Measurements

Amit Marmur[2], Wei Liu[1], Yuhang Wang[1], Armistead G. Russell[2] and Eric S. Edgerton[3]

1. Georgia Institute of Technology, School of Earth and Atmospheric Sciences, Atlanta, GA 30332

2. Georgia Institute of Technology, School of Civil and Environmental Engineering, Atlanta, GA 30332

3. Atmospheric Research and Analysis, Inc., Cary, NC 27513

ABSTRACT: Two-year CMAQ simulations of gases and aerosols over the Southeast are evaluated using SEARCH observations for 2000 and 2001, both by direct comparison to observations and by projecting both datasets to the factor space using the Positive Matrix Factorization (PMF) model. Model performance for secondary species (sulfate, ozone) is generally better than for primary species (EC, CO). Nitrate concentrations are overestimated, mainly due to wintertime over-partitioning to the particulate phase. Projecting both observed and simulated constituents to the factor space using PMF, four common factors are resolved for each surface site (two urban sites and two rural sites). The resolved factors include (1) secondary sulfate, (2) secondary nitrate, (3) a fresh motor vehicle factor characterized by EC, OC, CO, NO and NO_y, and (4) a mixed factor characterized by EC, OC, and CO. Performance for the sulfate and nitrate factors follow that of the corresponding driving species, while the motor-vehicle and "mixed" factors exhibit performance corresponding to that of primary species. Comparing observations and

CMAQ simulations in the projected space allow for an evaluation of the co-variability between species, an indicator of source impacts. The fact that similar factors were resolved by PMF from both the observations and the CMAQ simulations suggest that temporal processes related to emissions from specific source-categories, as well as the subsequent dispersion and reactivity, are well captured by the CMAQ model. The ability to identify additional factors can be enhanced by adding tracer species in CMAQ simulations.

1. INTRODUCTION

Air quality models generally fall into one of two classes: receptor-based and emission-based. Here, we combine and compare the two approaches in a novel way: we apply a receptor model to both simulated data from an emission-based model and in situ observations in the southeastern USA over a period of two years.

The Southeastern Aerosol Research and Characterization project (SEARCH) is an ongoing aerosol measurement program beginning in August of 1998. SEARCH data have been used extensively in health effects research (Tolbert et al., 2001), and the Atlanta SEARCH site was an EPA Supersite location (Hansen et al., 2003). We use measurements at four SEARCH sites (Liu et al., 2004, 2006), including North Birmingham [BHM] (urban) and Centreville [CTR] (rural) in Alabama and Atlanta [JST] (urban) and Yorkville [YRK] (rural) in Georgia. These two urban-rural pairs across southeast US are well suited to evaluate the CMAQ model performances in rural and urban areas. For the emission-based model in this project, we use the simulation results for 2000 and 2001 using the EPA's Models3 system (Marmur et al., 2004; Park et al., 2006a), which includes the Fifth-Generation Pennsylvania State University/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) (Grell et al., 1995),

the Sparse Matrix Operator Kernel Emissions Modeling System (SMOKE) (Houyoux et al., 2000) and Community Multiscale Air Quality (CMAQ, v4.3) (Byun and Ching, 1999) model. The long simulation period allows the application of factor analysis-based receptor models that require a relatively long record of data to derive both the source factors as well as daily source apportionment.

A PCA-based receptor model (Cohn and Dennis, 1994 and Li et al., 1995) was used to examine the performance of the regional acid deposition model (RADM) (Chang et al., 1987) and the Acid deposition and oxidant model (ADOM) (Venkatram et al., 1988). Since multiple interacting species are involved in a system, multivariate methods offer a means for characterizing the system by investigating the covariance structures. In this work, the receptor-based model chosen is the positive matrix factorization (PMF) (Paatero and Tapper, 1994). PMF has been widely used in source apportionment studies (Lee et al., 1999; Chueinta et al., 2000, Paterson et al., 1999, Polissar et al., 1999, 2001). Previously it was applied to analyze SEARCH measurements (Kim et al., 2003; Kim 2004 a,b; Liu et al., 2004, 2006; Lee et al., 2008).

In this work, we apply PMF to CMAQ simulation results. To our knowledge this is the first attempt to apply PMF to long-term air quality model simulations. Previously, Shim et al. (2007) applied PMF to analyze aircraft observations of trace gases and corresponding simulations by a global chemical transport model. We took the same approach here to apply the PMF method in the analysis of PM2.5 and selected gas-phase measurements and CMAQ model results in the projected factor space. Compared to direct evaluations, the coherent gas and aerosol structures resulting from the same or co-located sources, as evident in the data covariance, are evaluated simultaneously. In the

analysis, we first briefly describe the evaluation of the model performance on the species level. This "conventional" analysis provides the necessary basis to interpret the PMF analysis results that follow.

2. Measurements and CMAQ simulations

PM$_{2.5}$ composition and gas phase measurements at four sites (BHM, CTR, JST, YRK) from January 1, 2000 to December 31, 2001 are analyzed in this study. Daily integrated PM$_{2.5}$ and gas phase measurements are collected at the JST site. Every third day data are obtained at the others. A total of 600, 224, 230, and 219 samples are available for the JST, BHM, YRK, and CTR sites, respectively.

There are occasional "missing data" (no reported measurements) for one or more species in the observational samples. The analytical uncertainty and detection limit for each chemical species are also provided for the observational data set. More detailed description of these measurements is found elsewhere (Liu et al., 2004). For model evaluation purpose, only those measured species which are simulated by CMAQ are used in this study. Particulate species include PM$_{2.5}$ mass, sulfate (SO$_4^{2-}$), nitrate (NO$_3^-$), ammonium (NH$_4^+$), organic carbon (OC), elemental carbon (EC), and dust elements (calculated as Soil = 2.20*Al + 3.48*Si + 1.63*Ca + 2.42*Fe + 1.94*Ti according to the IMPROVE protocol (Sisler et al., 1996)). The IMPROVE protocol is used because CMAQ only predicts one aggregated soil term, rather than the individual species. The development of trace element modeling within CMAQ is currently underway at Georgia-Tech, but is beyond the scope of this work. Gaseous species include ozone (O$_3$), carbon monoxide (CO), sulfur dioxide (SO$_2$), nitrogen oxides (NO$_{(x)}$), nitric acid gas (HNO$_3$),

and total reactive nitrogen (NO_y). Measurements of NO_x are available only at the JST site; NO measurements are used for the other three sites.

The same chemical constituents are simulated by using CMAQ (Byun and Ching, 1999) during the same period for the eastern US. CMAQ is an Eulerian chemical transport model that simulates the emissions, transport, chemical transformation, and deposition of air pollutants. It employs a ''one-atmosphere'' approach and addresses complex interactions known to occur among multiple pollutants. The model, as applied here, uses a horizontal resolution of 36x36 km^2, with 6 layers. This model setup reduced the computational burden associated with long simulations such as the one used here (2 years). The relatively coarse grid spacing (both horizontally and vertically) are expected to introduce some biases as a result of artificial dilution. However, these are related mainly to spatial variability, while the focus of this work is the temporal variability in source/factor contributions in both ambient data and modeling results. The model domain covers most of North America in order to minimize the effects of boundary conditions on model results (Marmur et al., 2004; Park et al., 2006a). The SAPRC 99 (Carter, 2000) chemical mechanism was used. The EPA 1999 National Emissions Trends (NET99) inventory was used, processed by SMOKE (Houyoux et al., 2000). Meteorology was assimilated using the NCAR/Penn State MM5 model (Grell et al., 1994). Park et al. (2006b) found that MM5 simulated meteorological fields are adequate for air quality simulations.

3. Method

We first compare CMAQ results to the observed gas and aerosol compositions. Among the statistical measures used in the comparison are the mean (M) and standard deviation (SD) of observed (O) and simulated (S) concentrations (MO, SDO, MS, SDS,

respectively), concentration fractional bias (FBC = 2(MS-MO)/(MS+MO), variability fractional bias (FBV = 2(SDS/MS-SDO/MO)(SDS/MS+SDO/MO) (Odman et al., 2002; Bolyan and Russell., 2005), and squared linear correlation coefficient (r^2).

In addition to this rather "conventional" evaluation, PMF was applied to the measurements and model results, respectively. The species chosen in PMF analysis are the same as in the work by Liu et al. (2006) expect for dusts, which do not correlate with other species. The algorithm of PMF was described in detail elsewhere (Paatero and Tapper, 1994; Paatero, 1997). Application of PMF requires that error estimates for the data be chosen judiciously so that the estimates reflect the quality and reliability of each data point because data with high uncertainties are weighted less in the analysis. We follow the approach by Polissar et al. (1998) to estimate the measurement uncertainties. For measurements above the detection limit, the overall uncertainty is the sum of the measurement uncertainty and one third of the detection limit. For measurement data below the detection limit, half of the detection limit is assigned to the concentration and the uncertainty is assigned to be 5/6 of the detection limit. For missing measurement data, the geometric mean is assigned to be the concentration and the uncertainty is assigned to be 4 times the geometric mean. In order to compare PMF projected model results with the observations on a consistent basis, we assign the model values corresponding to the missing measurements to be the geometric mean. For these values, their error estimates are 4 times of geometric mean of the model results. For the model results below the detection limit, we assign an uncertainty of 5/6 of the detection limit to these data. For the other model data, we first calculate the average relative overall uncertainties estimated previously for the measurements above the detection limits. The products of these

relative overall measurement uncertainties and simulated values are taken as model uncertainties.

We follow the normal practice to experiment and find the optimal one with the most physically meaningful results. Analysis of the goodness of model fit, Q, as defined by Paatero (1997), is used to help determine the optimal number of factors. Fkey was used to reduce the rotation uncertainty. The final Q values are: JST (Observation: 6016 with data dimension of 6020, Model: 6113 with data dimension of 6000); BHM (Observation: 2046 with data dimension of 2240, Model: 2156 with data dimension of 2240); YRK (Observation: 2451 with data dimension of 2300, Model: 1974 with data dimension of 2300); CTR (Observation: 2579 with data dimension of 2190, Model: 2634 with data dimension of 2190).

Multiple linear regression (MLR) was performed to regress the mass concentrations against the factor scores (Xie et al., 1999). Because PMF results have a portion of unexplained variation, the mass concentrations excluding the unexplained variation portion from G factors (factor contributions) are used to regress the factor scores to obtain the quantitative factor contributions for each resolved factor. We also exclude the dust components in the regression by using the IMPROVE protocol to estimate the dust contributions. The regression coefficients are used to transform the factor contribution results into the particle source contributions with physically meaningful units. The corresponding factors derived from the observed and simulated datasets are compared.

4. Results and Discussion

4.1 Composition Statistics

Detailed comparison statistics of the model simulations with the measurements are listed in Tables 1a-d. These indicate that the model tends to under-predict concentrations of primary species (such as EC, dust, SO_2, and NO), with the exception of crustal material. The underprediction is likely due to the relatively coarse grid used (both horizontally and vertically). Despite this underprediction in concentrations, variability of primary species (represented here by FBV) is better simulated compared to mass concentrations, the performance of which is represented by FBC (i.e., for most primary species at most sites, the model better simulates variations in concentrations than the actual mass concentrations of these species). The overprediction in crustal material is due to the fact that CMAQ mixes all of the resuspended soil dust within the first layer, while in reality many of them are removed locally by impaction on surfaces such as buildings, vehicles, and vegetation (DRI, 2000). Sulfate concentrations are well simulated compared to most other species (R^2 values from 0.4 to 0.6) at the four sites. However, urban SO_2 concentrations are underpredicted due to artificial dilution of emission sources in the coarse grid. Rural sites exhibit a slight overprediction of SO_2 concentrations. The simulations of NO_y and CO show a similar patterns compared with the measurements: the concentrations at urban sites are significantly underpredicted while the predictions at rural site are reasonably good. The annual average of simulated nitrate is higher than the observations largely because of the overestimates during cold seasons. Nitrate is well correlated at the two Alabama sites ($R^2 \approx 0.5$) and less so at the Georgia sites. The model captures seasonal variations in nitrate levels, but performs rather poorly in capturing intra-season/daily variations in nitrate levels. Comparing these performance findings to a study using a finer grid (Hu et al., 2004; 4 km horizontal resolution with 13 vertical

layers), better agreement with the observations was found for the primary species, but ozone and sulfate performance were similar to those presented here.

Total S [$SO_2 + SO_4^{2-}$, expressed in µg S /m^3] and total oxidized nitrogen [$NO_y + NO_3^-$, expressed in µg N /m^3] are good measures for evaluating the model emission and dispersion estimates, as total S and N are less dependent on the simulated SO_2-SO_4^{2-} and NO_y-NO_3^- partitions. The model is able to simulate the observed variability in general as reflected by the reasonable R^2 values (from 0.35 to 0.68). CMAQ simulated gas to particle phase partitions of sulfur and oxidized nitrogen can be evaluated with the observations by investigating the ratios of [SO_2 (µg S /m^3)]/[total S concentration (µg S /m^3)] and [NO_y (µg N /m^3)] / [total oxidized N concentration (µg N/m^3)], respectively. In order to compare the seasonal variations between the observations and CMAQ, ratios of SO_2 to total sulfur and NO_y to total oxidized nitrogen are processed using sym8 wavelet decomposition with soft heuristic thresholding and scaled noise options (Eskridge et al., 1997, Lou and Loparo, 2004). Linear interpolation is used in places of missing data. Considering the amount of missing data at each site, only year 2000 data for the JST, YRK and CTR sites and year 2001 data for BHM site are processed (Figure 1). Seasonal cycles of the observed SO_2 fractions are well captured by the model, although the model cannot capture all the higher-frequency variations. The observed NO_y fractions show little variations, indicating that gas-phase NO_y accounts for the bulk of atmospheric oxidized nitrogen. In comparison, the model results show a clear winter minimum. The gas-particle partition of the oxidized nitrogen species is particularly sensitive to temperature and the availability of ammonia gas. The underestimated CMAQ NO_y fractions in winter imply that CMAQ apportions too much gas-phase N to nitrate during

cold seasons. It appears that either the nitrate mechanism of the model is overly sensitive to the variations in temperature or ammonia is overestimated in the model.

4.2 PMF RESULTS

The PMF analysis was carried out using nine common species that are both measured and simulated (SO_2 was excluded from the analysis due to a high number of missing observations). Four factors are resolved for the observed and simulated datasets for the four sites, based on the goodness of model fit, Q (Paatero, 1997) and physical representation of the identified factors. The factors identified are: secondary sulfate, secondary nitrate, fresh motor vehicle, and a mixed factor. The linear correlations (R^2 values) between the factors from observed and simulated data (Table 2) typically follow the same patterns as in the "conventional" evaluation (Table 1), with the sulfate factor exhibiting the highest correlation, followed by the nitrate factor, and factors driven by primary species (motor vehicle and "mixed") exhibiting poorer performance. Comparing observations and CMAQ simulations in the projected space allow for an evaluation of the co-variability between species, an indicator of source impacts. The fact that similar factors were resolved by PMF from both the observations and the CMAQ simulations suggest that temporal processes related to emissions from specific source-categories, as well as the subsequent dispersion and reactivity, are well captured by the CMAQ model.

We use explained variations (EV) to define the factor profiles because the gas and particulate-phase concentrations of different species are not directly comparable. The value of EV_{ij} is the fraction of species i that can be explained by factor j (Paterson et al., 1999). The average seasonal factor contributions were also compared at the four sites

between the observed and simulated datasets, though only results for the urban sites are shown (Figures 6,7), as the rural sites showed fairly similar seasonal patterns.

The secondary sulfate factor has high concentrations of sulfate and ammonium (Figure 2) in both the simulated and the observed data among the four sites. This factor shows a strong seasonal variation with high concentrations during summertime (Figures 6,7), reflecting, in part, the more active photochemical production of sulfate from SO_2 in the summer. Not surprisingly, the factor accounts for a major fraction of ammonium. Sulfate in the southeast is neutralized (e.g., Liu et al., 2004). This factor is also mixed with HNO_3 in both the simulated and observed data. The high correlation between the secondary sulfate and HNO_3 is likely due to three reasons. First, sulfate and HNO_3 are secondary photochemical products, and regional recirculation mixes these pollutants. Furthermore, the major primary sources of SO_2 also tend to be NO_x sources. Second, HNO_3 competes with H_2SO_4 gas for available ammonia (Russell et al., 1983). The more abundant H_2SO_4 in summer leaves little ammonia for the conversion of HNO_3 to particulate nitrate. Third, the cool temperatures in winter tend to favor nitrate formation. Hence, both sulfate and HNO_3 tend to be higher in summer. The resolved CMAQ factor contributions are generally consistent with those for the measurements. The R^2 values between the two datasets are relatively high, in the range of 0.5-0.6 (Table 2); the good correlations are reflected in the time series of the corresponding factors. The correlations are better at the rural sites than the urban sites. The meteorological model, MM5, has been shown to capture more of the inter-annual and synoptic-scale variability of important meteorological parameters such as surface temperature and wind speed, compared to fluctuations on the intra-day and diurnal time scales (Hogrefe et al., 2001,

2004). Not fully capturing the shorter time scales in the meteorological fields by MM5 may partly explain the poorer performances in urban areas, where the variations on the intra-day and diurnal time scales are larger than rural sites. However, the CMAQ based factors are more "smeared" compared to the observation based ones. For example, in most cases, the CMAQ based sulfate factors contain more EC, a species not typically associated with sulfate formation, compared to the observation based factors. This is due mainly to the relatively coarse spatial setup of CMAQ. The CMAQ based factors are derived based on modeling results representing an average grid cell of 36 by 36 km, while the observation based factors are derived from point measurements (though even a finer grid, such as 12x12 km or 4x4 km would be much courser compared to a point measurement at a monitoring site). The point measurements typically allow for a more distinct breakdown of factors affecting trends in ambient concentrations at the receptor site, compared to the more mixed spatial representation of CMAQ.

The secondary nitrate factor dominates the contributions to nitrate, as well as ~20% of the ammonium (Figures 3). Nitrate is formed in the atmosphere through oxidation of NO_x. Nitric acid gas tends to condense to particle phase nitrate at low temperatures, high humidity, and in the presence of ammonia gas. Therefore high concentrations of nitrate occur mainly during winter in both of the simulated and observed data (Figure 6,7). However, the wintertime peaks are significantly higher for CMAQ. Time series of this factor for the observed and simulated datasets show similar structures but the day-to-day variations are not always the same. As a result, the R^2 values between observed and simulated datasets are lower than for the sulfate factor, ranging from 0.2 to 0.5 (Table 2). High-frequency peaks likely influenced by local sources are not simulated well by the

model. This is partly because of the meteorological model's poor performance in simulating shorter time-scale variations of the meteorological parameters.

A fresh motor vehicle factor is resolved from both the observed and simulated datasets. It is represented by high concentrations of CO, NO, and NO_y and the inclusion of OC and EC (Figure 4). It is labeled here as "fresh motor vehicle factor" because of its high concentrations of NO which has a short lifetime in the atmosphere. Both the factor profiles and factor contributions are different between the simulated and the observational data. Comparing to the observations, smaller portions of NO and greater portions of CO reside in this factor from the CMAQ results at the urban sites. This, as in the case of the sulfate factor, can be the result of "smearing" in the CMAQ results, limiting the ability to distinctly apportion NO between the nitrate and motor vehicle factors. This mixing or "smearing" of factors, as well as the relatively coarse setup of the vertical layers in CMAQ (limiting the model's ability to simulate the effect of mixing layer heights on ground level concentrations) explain to some extent the discrepancies in seasonal patters between observation and modeling results (Figures 6,7). The OC/EC ratios of this factor are in the range of 1.14 to 1.58 (Table 3). The reported OC/EC ratio is typically 2.05 in fresh gasoline exhaust and 0.72 in diesel emissions (Cadle et al., 1999). Therefore, this factor represents a mixture of diesel emissions and gasoline-powered vehicles emissions. On average, the CMAQ results are lower at all of the four sites (Figure 6,7), likely due to the coarse grid used. Discrepancies between observed and simulated temporal variability are due, in part, by the use of "typical" emissions in CMAQ, while atypical traffic events are unaccounted for.

A mixed factor is also resolved at the four sites from simulated and observed data. This factor is represented by high concentrations of OC, EC and CO (Figure 5). Wood smoke and some local industry factors with high OC/EC ratios were resolved in a previous study (Liu et al., 2004). The mixed factor resolved in the current study is probably a combination of wood smoke, industry, and aged motor vehicle sources. The resolved OC/EC ratios from CMAQ for this factor are higher than those from the observational data (Table 3) at the urban sites, but are lower at the rural sites, which could be related to differences in secondary organic aerosol formation process between rural and urban locations. Without trace elements like those used by Liu et al. (2004, 2006), we cannot resolve the contributions of various sources to this factor, which leads to additional uncertainties in the comparison.. The input emission inventory uncertainties (Mendoza-Dominguez and Russell, 2001) and the meteorological model's ability to capture fine temporal scales may be the other reasons contributing to the disagreement between CMAQ and observation data. Better coherence might result from a finer grid resolution in the model, especially for the urban sites, where there are more intense emissions with significant chemical gradients (Sillman et al., 1990). Higher uncertainties and poorer spatial representation have also been reported for measuring EC and OC, which are the major elements in this factor (Wade et al., 2006).

5. Summary

CMAQ simulations for a period of two years (2000 and 2001) are evaluated using SEARCH observations in the Southeast, both by direct comparison to observations and by projection to the factor space sing PMF. Model performance for secondary species (sulfate, ozone) is generally better than for primary species (EC, CO). Nitrate concentrations are overestimated, mainly due to wintertime over-partitioning to the

14

particulate phase. Projecting both observed and simulated constituents to the factor space using PMF, four common factors, including sulfate, nitrate, a fresh motor vehicle, and a mixed factor, are resolved and compared. Performance for the sulfate and nitrate factors follow that of the corresponding driving species, while the motor-vehicle and "mixed" factors exhibit performance corresponding to that of primary species. The OC/EC ratios of the fresh motor vehicle factor reflect a mixture of gasoline and diesel vehicle exhausts. The mixed factor is likely a combination of aged motor vehicle, wood smoke, and some industry factors. The comparison between observations and CMAQ simulations in the projected space allow for an evaluation of the co-variability between species, an indicator of source impacts. The fact that similar factors were resolved by PMF from both the observations and the CMAQ simulations suggest that temporal processes related to emissions from specific source-categories, as well as the subsequent dispersion and reactivity, are well captured by the CMAQ model. Adding trace elements in CMAQ would allow a more detailed evaluation emissions and transport processes in the model.

References:

Boylan, J. W. and Russell, A. G., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmospheric Environment 40, 563-573.

Byun, D. and . Schere, K. L., 2006, Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, Applied Mechanics Reviews, Vol. 59 (2), 51-77.

Cadle, S.H., Mulawa, P.A., Hunsanger, E.C., Nelson, K.E., Ragazzi, R.A., Barrett, R., Gallagher, G.L., Lawson, D.R., Knapp, K.T., Snow, R. 1999. Composition of Light-Duty Motor Vehicle Exhaust Particulate Matter in the Denver, Colorado Area, Environment Science and Technology. 33, 2328-2339.

Carter, W. P. L. (2000): "Implementation of the SAPRC-99 Chemical Mechanism into the Models-3 Framework" Report to the United States Environmental Protection Agency, January 29.

Chang, J.S., Brost, R.A., Isakson, I.S.A., Madronic, S., Middleton, P., Stockwell, W.R., Walcek, C.J., 1987. A three-dimensional Eulerian acid deposition model: physical concepts and formulation. Journal of Geophysical Research 92, 14681-14700.

Chueinta, W., Hopke, P.K., and Paatero, P., 2000. Investigation of Sources of Atmospheric at Aerosol Urban and Suburban Residential Areas in Thailand by Positive Matrix Factorization. Atmospheric Environment. 34, 3319-3329.

Cohn, R.D., Dennis, R.L., 1994. The evaluation of acid deposition models using principal component spaces. Atmospheric Environment 28, 2531-2543.

DRI, 2000. Reconciling Urban Fugitive Dust Emissions Inventory and Ambient Source Contribution Estimates: Summary of Current Knowledge and Needed Research, Desert Research Institute Document No. 6110.4F, Reno, NV 89512.

Eskridge, R.E., Ku, J.Y., Rao, S.T., Porter, P.S., Zurbenko, I.G., 1997. Separating different scales of motion in time series of meteorological variables. Bulletin of the American Meteorology Society. 78, 1473–1483.

Grell, G.A., Dudhia, J., Stauffer, D.R., 1995. A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note, NCAR/TN-398+STR, 122pp.

Hansen, D.A., Edgerton, E.S., Hartsell, B.E., Jansen, J.J., Kandasamy, N., Hidy, G.M., Blanchard, C.L. 2003. The Southeastern Aerosol Research and Characterization Study: part 1—Overview. Journal of the Air and Waste Management Association, 53, 1460-1471

Hogrefe, C., Rao, S.T., Kasibhatla, P., Kallos, G., Tremback, C., Hao, W., Olerud, D., Xiu, A., McHenry, J., Alapaty, K., 2001. Evaluating the performance of regional-scale photochemical modeling systems: Part I—meteorological predictions. Atmospheric Environment. 35, 4159–4174.

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

Houyoux, M.R., Vukovich, J.M., Coats, Jr. C.J., Wheeler, N.J.M., Kasibhatla, P., 2000. Emission inventory development and processing for the seasonal model for regional air quality. Journal of Geophysical Research. 105, 9079–9090.

Hu, Y., Cohan, D. S., Odman, M.T., and Russell, A. G., 2004. Final Report: Air Quality Modeling of the August 11-20, 2000 Episode for the Fall Line Air Quality Study. Prepared for Georgia Department of Natural Resources, prepared by Georgia Institute of Technology, Atlanta, GA.

Kim, E., Hopke, P.K., Edgerton, E.S., 2003. Source Identification of Atlanta Aerosol by Positive Matrix Factorization. Journal of the Air and Waste Management Association. 53, 731-739.

Kim, E., Hopke, P.K., Edgerton E.S., 2004a. Improving Source Identification of Atlanta Aerosol using Temperature Resolved Carbon Fractions in Positive Matrix Factorization. Atmospheric Environment. 38, 3349-3362.

Kim, E., Hopke, P.K., Paatero, P., Edgerton, E.S., 2004b. Incorporation of Parametric Factors into Multilinear Receptor Model Studies of Atlanta Aerosol. Atmospheric Environment. 37, 5009-5021

Lee, E., Chan C.K., Paatero P., 1999. Application of Positive Matrix Factorization in Source Apportionment of Particulate Pollutants in Hong Kong. Atmospheric Environment. 33, 3201-3212.

Lee S., W. Liu, Y. Wang, A. Russell and E. S. Edgerton, 2008. Source Apportionments of PM2.5: Comparing PMF and CMB Results for 4 Southeast US Sites, in press, Atmos. Environ..

Li, S., 1995. Using a principal component analysis technique to evaluate ratios in the NAPAP area emissions. Pure and Applied Chemistry 67, 1407-1486.

Liu, W., Y. Wang, A. Russell, and E. Edgerton, 2004. Atmospheric Aerosol over Two Urban-Rural Pairs in the Southeastern United States: Chemical Composition and Sources, Atmos. Environ., 39, 4453-4470.

Liu, W., Y. Wang, A. Russell, and E. Edgerton, 2006. Enhanced source identification of Southeast aerosols using temperature resolved carbon fractions and gas phase components, Atmos. Environ., 40, S445-S466.

Lou, X., Loparo, K.A., 2004. Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mechanical Systems and Signal Processing 18 (5) 1077–1095.

Marmur A., Russell, A.G., and Mulholland J. A., 2004. Air-Quality Modeling of $PM_{2.5}$ Mass and Composition in Atlanta: Results from a Two-Year Simulation and Implications for Use in Health Studies. Proceeding of 2004 Models-3 conference, Chapel Hill, NC

Mendoza-Dominguez A. and Russell A.G., 2001. Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling. Atmospheric Environment. 35, 2879-2894.

Odman M.T., Boylan J.W., Wilkinson, J.G., Russell A .G., 2002. Integrated modeling for air quality assessment: The Southern Appalachians mountains initiative project. Journal of Physics IV 12 (PR10): 211-234.

Paatero, P., 1997. Least Squares Formulation of Robust, Non-Negative Factor Analysis. Chemometrics and Intelligent Laboratory System. 37, 23-35.

Paatero, P., Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics. 5, 111–126.

Park, S.K., Marmur, A., Kim, S.B., Tian, D., McMurry, P., Russell, A.G. (2006a), Evaluation of fine particle number concentrations in CMAQ, Aerosol Science and Technology, 40: 985-996.

Park, S. K., Cobb C. E., Wade K., Mulholland J., Hu Y., and Russell, A.G. (2006b), Uncertainty in air quality model evaluation for particulate matter due to spatial variations in pollutant concentrations, Atmospheric Environment. 40, S563–S573.

Paterson, K.G., Sagady, J.L., Hooper, D.L., Bertman, S.B., Carroll, M.A., Shepson, P.B., 1999. Analysis of Air Quality Data Using Positive Matrix Factorization. Environment Science and Technology. 33, 635-641.

Polissar, A.V., Hopke, P.K., Malm, W.C., Sisler, J.F., 1998. Atmospheric Aerosol over Alaska: 2. Elemental Composition and Sources, Journal of Geophysical Research. 103, 19,045-19,057.

Polissar, A.V., Hopke, P.K., Paatero, P., Kaufman, Y.J., Hall, D.K., Bodhaine, B.A., Dutton, E.G., Harris, J.M., 1999. The Aerosol at Barrow, Alaska: Long-Term Trends and Source Locations. Atmospheric Environment. 33, 2441-2458.

Polissar, A.V., Hopke, P.K., Poirot, R.L., 2001. Atmospheric Aerosol over Vermont: Chemical Composition and Sources. Environment Science and Technology. 35, 4604-4621.

Russell, A.G., McRae, G.J. and Cass, G.R., 1983. Aerosol Nitrate Dynamics in an Urban Basin. Aerosol Science and Technology. 2, 179.

Shim C., Y. Wang, and Y. Yoshida, 2007. Evaluation of model-simulated source contributions to tropospheric ozone with aircraft observations in the factor-projected space, Atmos. Chem. Phys. Discuss., 7,, 15495-15532.

Sillman, S., Logan, J. A., and Wofsy, S. C. A regional-scale model for ozone in the United States with sub-grid representation of urban and power plant plumes. J. Geophys. Res., 95, 5731-5748, 1990.

Sisler, J. F., Malm, W. C., Gebhart, K. A., Pitchford, M. L., 1996. Spatial and seasonal patterns and long term variability of the composition of the haze in the United States: An analysis of data from the IMPROVE network. http://vista.cira.colostate.edu/improve/Publications/Reports/1996/1996.htm

Tolbert, P., Klein, M., Metzger, K., Peel, J., Flanders, W.D., Todd, K., Mulholland, J., Ryan, P.B., Frumkin, H., 2001. Particulate pollution and cardiorespiratory emergency department visits in Atlanta, August 1998-August 2000 (ARIES/SOPHIA studies). Epidemiology, 12:265.

Venkatram, A., Karamchandani, P.K., Misra, P.K., 1988. Testing a comprehensive acid deposition model. Atmospheric Environment 22, 737-747.

Xie, Y. L., Hopke, P., Paatero, P., Barrie, L. A., and Li, S. M. 1999. Identification of source nature and seasonal variations of Arctic aerosol by positive matrix factorization. Journal of Atmospheric Science. 56:249- 260.

Wade, K.S., Mulholland, J.A., Marmur, A., Russell, A.G., Hartsell, B., Edgerton, E., Klein, M., Waller, L., Peel, J.A., Tolbert, P.E. (2006), Instrument Error and Spatial Variability of Ambient Air Pollution in Atlanta, Journal of the Air & Waste Management Association, 56: 876-888.

Table 1. Comparison between trends in observed and simulated concentrations during 2000-2001 at four SEARCH sites (MO – mean observed, MS – mean simulated, SDO – standard-deviation of observed, SDS – standard deviation of simulated, FBC – fractional bias in concentrations, FBV – fractional bias of variability, R^2 – Pearson's correlation coefficient)

a. JST	MO (μg/m3)	SDO (μg/m3)	MS (μg/m3)	SDS (μg/m3)	FBC	FBV	R^2
Mass(μg/m3)	17.18	8.14	22.31	11.27	0.26	0.06	0.42
Total S (μgS/m3)	10.02	7.28	6.17	3.02	-0.48	-0.39	0.35
SO42-(μg/m3)	4.71	3.02	4.72	2.99	0.00	-0.01	0.53
Total N (μgN/m3)	32.9	26.1	13.35	5.71	-0.85	-0.60	0.35
NO3- (μg/m3)	1	0.88	2.89	3.68	0.97	0.37	0.35
NH4+(μg/m3)	2.59	1.18	2.52	1.44	-0.03	0.23	0.2
EC (μg/m3)	1.52	1.04	1.02	0.48	-0.39	-0.37	0.36
OC (μg/m3)	3.98	2.31	2.54	1.44	-0.44	-0.02	0.34
Soil (μg/m3)	0.64	0.52	6.97	10.82	1.66	0.63	0.04
O3(ppb)	23.94	12.48	35.54	18.33	0.39	-0.01	0.63
CO (ppb)	472.77	385.04	329.95	127.18	-0.36	-0.72	0.36
SO2(ppb)	6.79	5.98	3.52	2.09	-0.63	-0.39	0.42
NO(ppb)	26.78	44.6	3.06	3.88	-1.59	-0.27	0.08
HNO3(ppb)	1.29	1.28	1	1.03	-0.25	0.04	0.05
NOy(ppb)	57.02	46.75	22.22	9.01	-0.88	-0.68	0.34

b. BHM	MO (ug/m3)	SDO (ug/m3)	MS (ug/m3)	SDS (ug/m3)	FBC	FBV	R^2
Mass(μg/m3)	18.63	9.85	17.76	8.74	-0.05	-0.07	0.46
Total S (μgS/m3)	8.71	5.54	6.8	3.35	-0.25	-0.25	0.36
SO42-(μg/m3)	4.87	3.32	4.76	3.09	-0.02	-0.05	0.49
Total N (μgN/m3)	25.97	20.97	7.03	2.85	-1.15	-0.66	0.56
NO3- (μg/m3)	1.04	0.9	1.68	2.62	0.47	0.57	0.5
NH4+(μg/m3)	2.9	1.91	1.93	1.12	-0.40	-0.13	0.2
EC (μg/m3)	2.08	1.56	0.5	0.24	-1.22	-0.44	0.35
OC (μg/m3)	4.69	3.09	2.08	1.3	-0.77	-0.05	0.37
Soil (μg/m3)	1.35	1.08	4.62	3.42	1.10	-0.08	0.12
O3(ppb)	22.3	11.14	39.59	15.04	0.56	-0.27	0.57
CO (ppb)	557.11	337.46	200.6	61.8	-0.94	-0.65	0.48
SO2(ppb)	5.73	4.2	3.99	2.24	-0.36	-0.27	0.37
NO(ppb)	25.26	31.6	0.74	0.71	-1.89	-0.26	0.18
HNO3(ppb)	0.63	0.69	0.84	0.81	0.29	-0.13	0.06
NOy(ppb)	44.96	38.66	11.64	4.36	-1.18	-0.79	0.51

c. YRK	MO (ug/m3)	SDO (ug/m3)	MS (ug/m3)	SDS (ug/m3)	FBC	FBV	R^2
Mass(μg/m3)	14.16	8.43	16.16	8.39	0.13	-0.14	0.27
Total S (μgS/m3)	4.88	3.53	7.33	3.33	0.40	-0.46	0.41
SO42-(μg/m3)	4.48	3.04	4.63	3.08	0.03	-0.02	0.59
Total oxidized N (μgN/m3)	4.68	3.63	5.4	3.13	0.14	-0.29	0.68
NO3- (μg/m3)	0.94	0.92	2.38	3.28	0.87	0.34	0.42
NH4+(μg/m3)	2.23	1.53	2.19	1.21	-0.02	-0.22	0.15
EC (μg/m3)	0.63	0.35	0.41	0.26	-0.42	0.13	0.33

OC (μg/m3)	3.1	2.06	1.76	1.14	-0.55	-0.03	0.3
Soil (μg/m3)	0.31	0.32	4.11	2.5	1.72	-0.52	0.002
O3(ppb)	40.47	15.27	42.04	15.88	0.04	0.00	0.66
CO (ppb)	172.52	50.04	154.25	70.16	-0.11	0.44	0.31
SO2(ppb)	2.59	2.52	4.43	2.28	0.52	-0.62	0.41
NO(ppb)	0.68	1.69	0.57	1.01	-0.18	-0.34	0.4
HNO3(ppb)	0.98	0.78	0.75	0.75	-0.27	0.23	0.22
NOy(ppb)	7.83	6.08	8.51	4.62	0.08	-0.35	0.65

d. CTR	MO (ug/m3)	SDO (ug/m3)	MS (ug/m3)	SDS (ug/m3)	FBC	FBV	R^2
Mass(μg/m3)	12.84	7	12.41	6.43	-0.03	-0.05	0.39
Total S (μgS/m3)	4.34	2.89	4.67	2.85	0.07	-0.09	0.56
SO42-(μg/m3)	4.25	2.95	4.39	2.92	0.03	-0.04	0.59
Total oxidized N (μgN/m3)	3.07	1.98	2.76	1.57	-0.11	-0.13	0.59
NO3- (μg/m3)	0.36	0.4	1	1.78	0.94	0.46	0.48
NH4+(μg/m3)	1.4	0.75	1.51	0.86	0.08	0.06	0.32
EC (μg/m3)	0.6	0.38	0.3	0.15	-0.67	-0.24	0.42
OC (μg/m3)	2.9	1.7	1.95	1.15	-0.39	0.01	0.52
Soil (μg/m3)	0.43	0.43	2.7	1.79	1.45	-0.41	0.004
O3(ppb)	37.22	12.92	43.41	11.81	0.15	-0.24	0.64
CO (ppb)	187.77	46.5	105.8	28.44	-0.56	0.08	0.3
SO2(ppb)	2.14	1.96	2.4	1.97	0.11	-0.11	0.6
NO(ppb)	0.42	0.53	0.17	0.24	-0.85	0.11	0.72
HNO3(ppb)	0.69	0.49	0.63	0.49	-0.09	0.09	0.3
NOy(ppb)	5.12	3.32	4.34	2.26	-0.16	-0.22	0.57

Table 2. Correlation coefficient values (R^2) for pairwise comparisons of source contributions between observed and CMAQ simulated data resolved by PMF for the four sites.

	R^2 (Obs. vs. CMAQ)			
	JST	BHM	YRK	CTR
Sulfate	0.55	0.48	0.64	0.61
Nitrate	0.28	0.45	0.38	0.47
Fresh Motor vehicle	0.12	0.17	0.34	0.30
Mixed	0.23	0.31	0.27	0.37

Table 3. OC/EC ratios of source factors between observed and CMAQ simulated
data resolved by PMF for the four sites

	Fresh Motor vehicle (OC/EC)	Mixed factor (OC/EC)
JST Observation	1.48	2.91
JST CMAQ	1.58	3.82
BHM Observation	1.27	2.30
BHM CMAQ	1.44	6.46
YRK Observation	1.35	6.08
YRK CMAQ	1.20	4.83
CTR Observation	1.33	5.95
CTR CMAQ	1.14	5.42

Figure captions

Figure 1. De-noised temporal variations of observed and simulated SO_2 and NO_y fractions at the four sites.

Figure 2. Factor profiles of the sulfate factors in the observed and simulated datasets resolved by PMF at the four sites.

Figure 3. Factor profiles of the nitrate factors in the observed and simulated datasets resolved by PMF at the four sites.

Figure 4. Factor profiles of the fresh motor vehicle factors in the observed and simulated datasets resolved by PMF at the four sites.

Figure 5. Factor profiles of the mixed factors in the observed and simulated datasets resolved by PMF at the four sites.

Figure 6. Comparison of the seasonal averages of factor concentrations for the observed and simulated data at JST site.

Figure 7. Comparison of the seasonal averages of factor concentrations for the observed and simulated data at the BHM site.
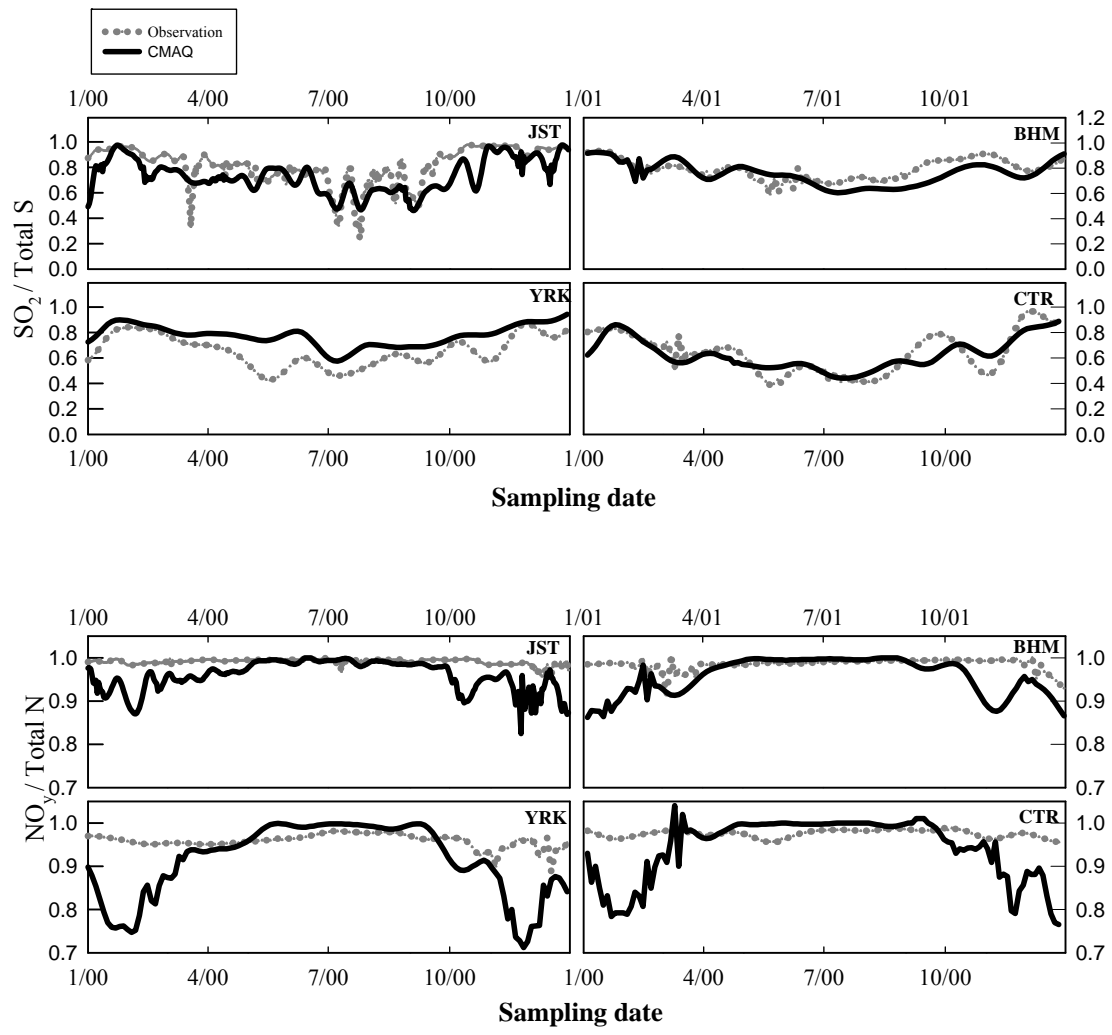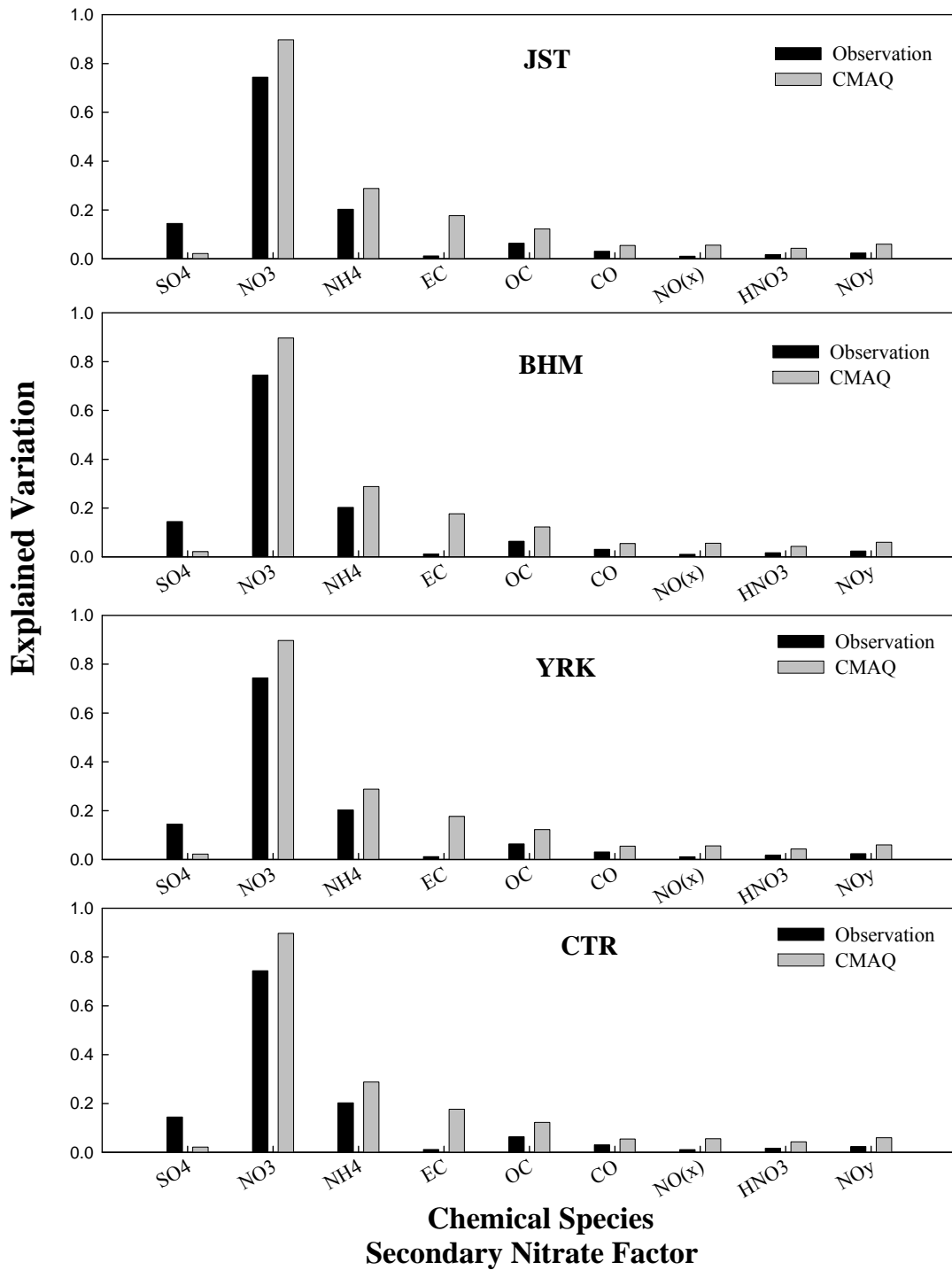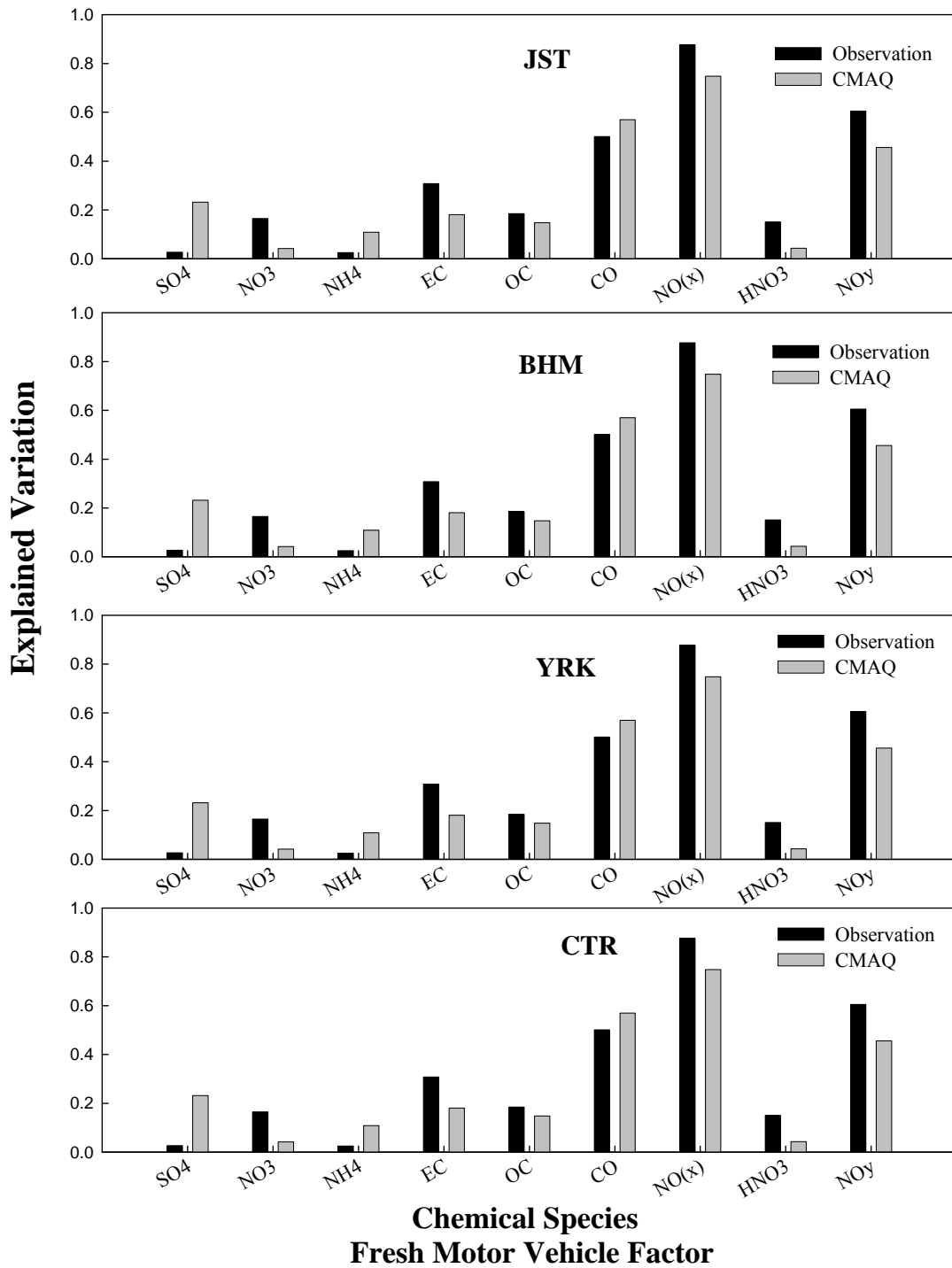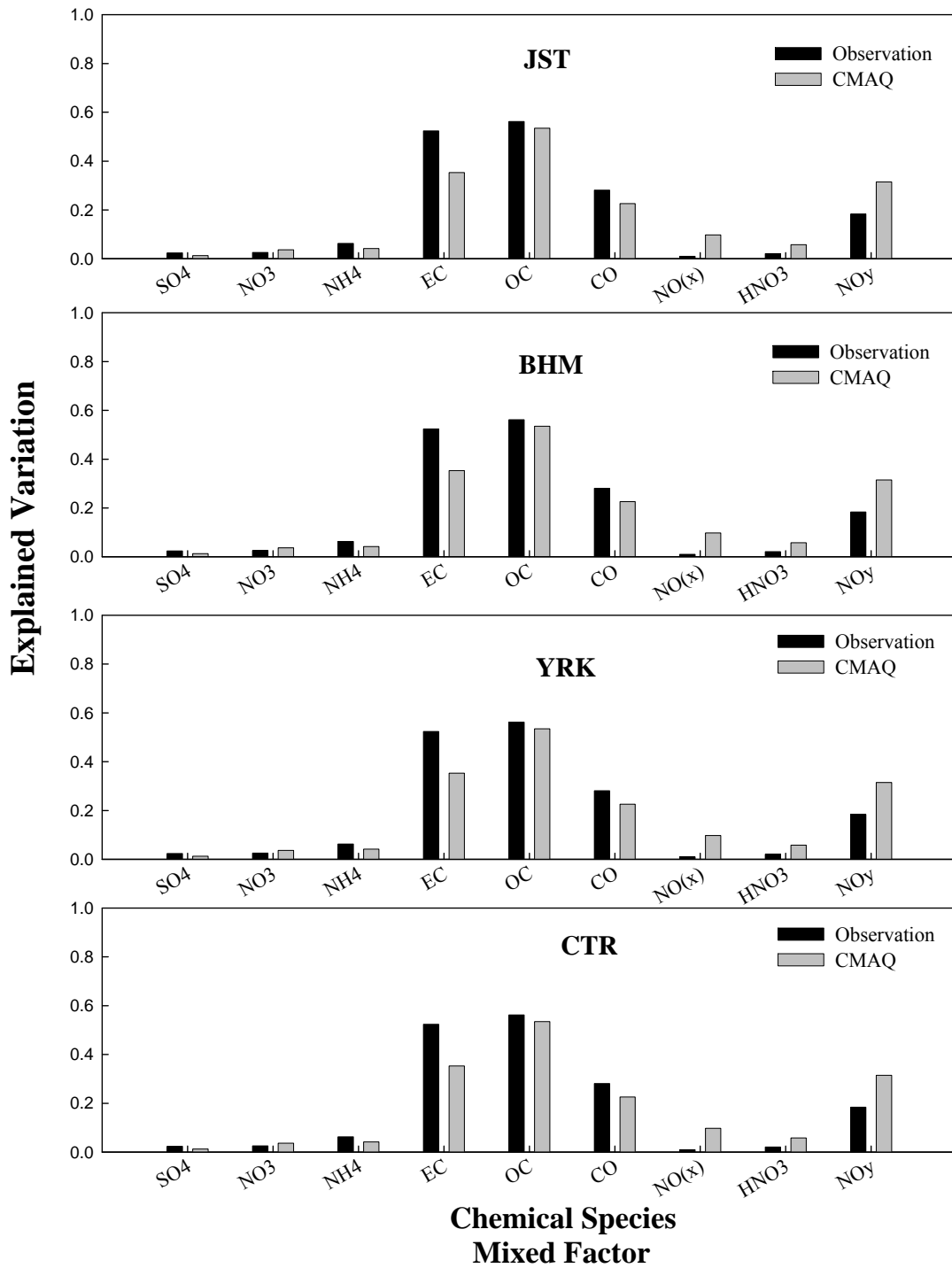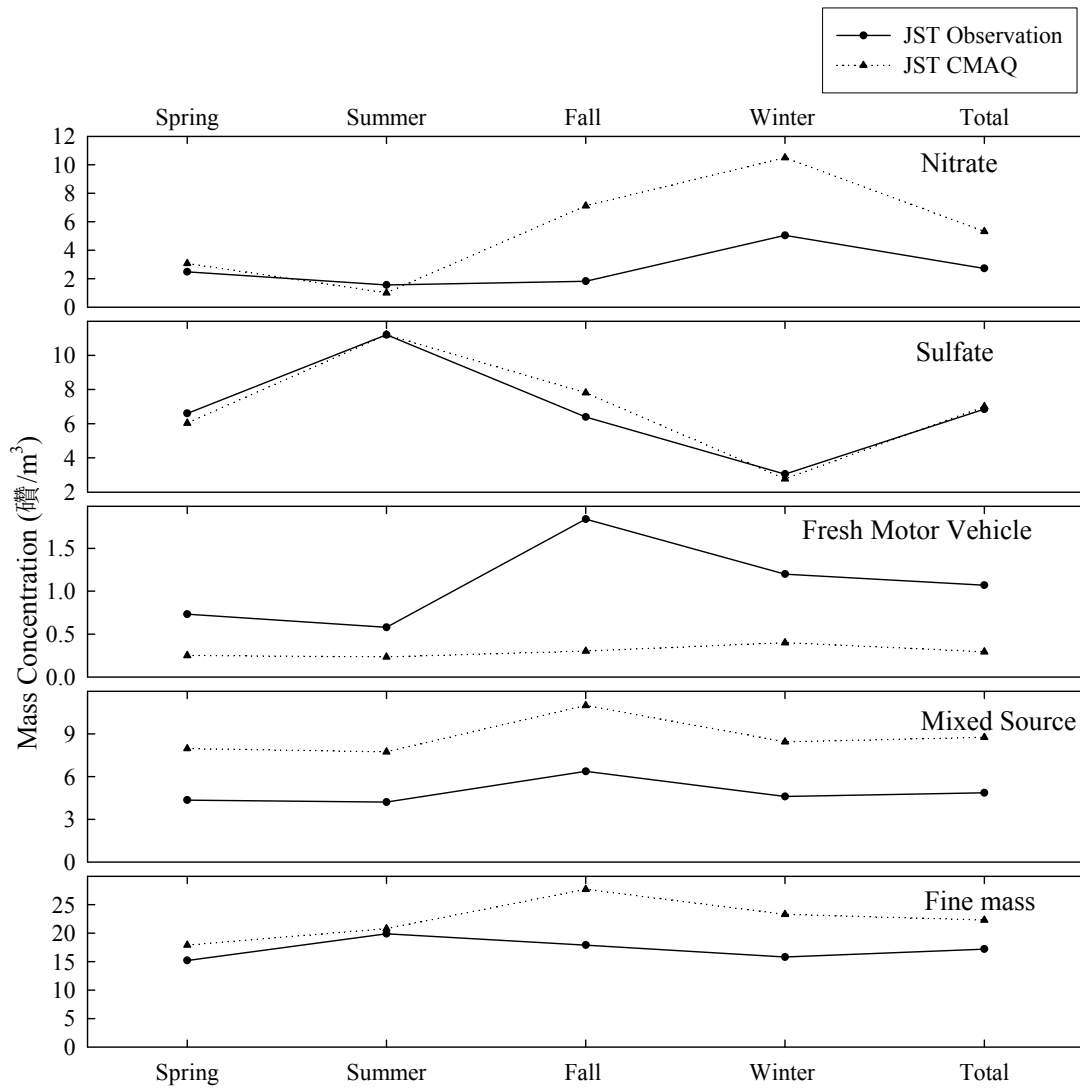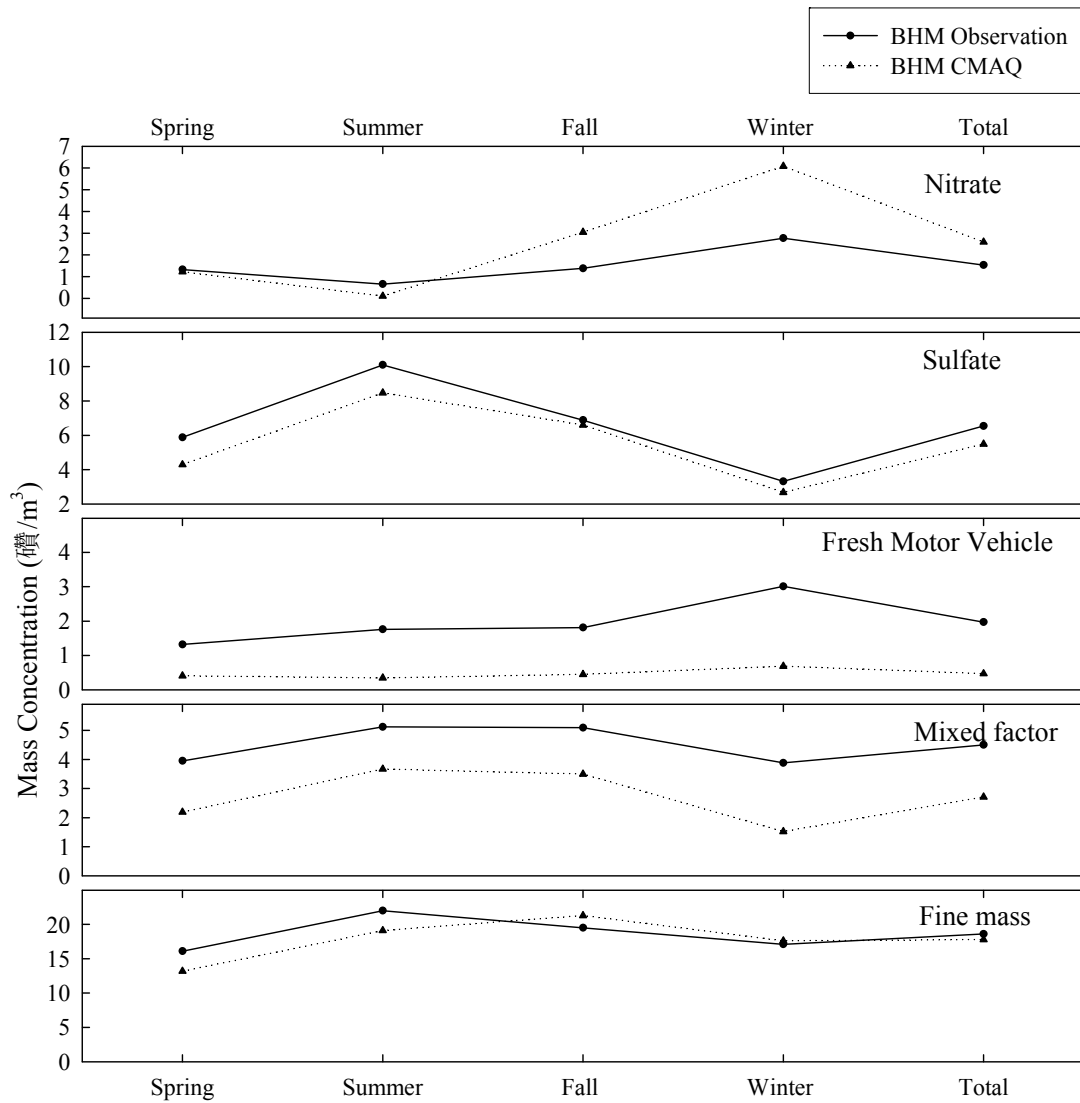
Figure 1
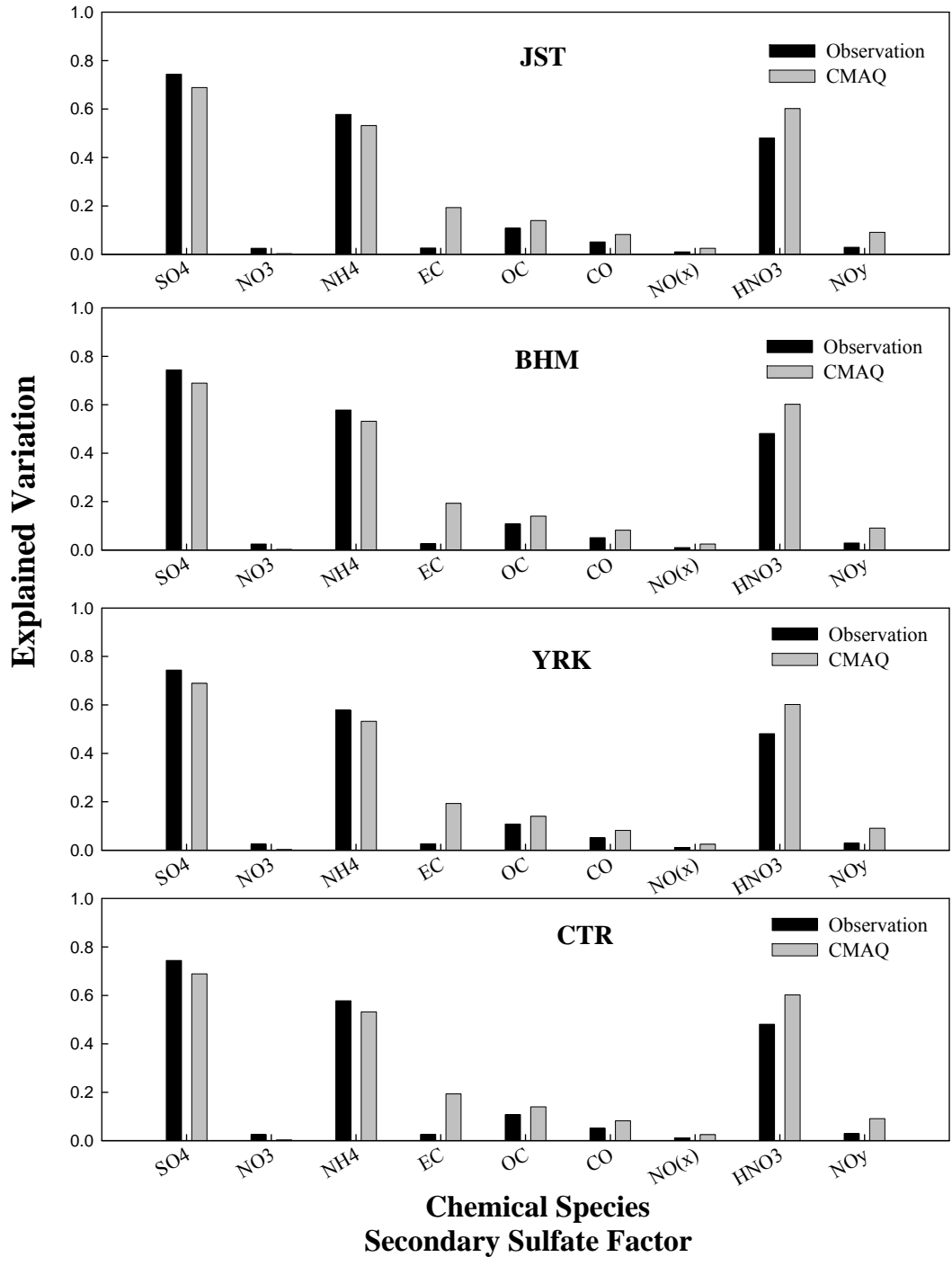
Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7