Chapter 49
Temporally and Spatially Resolved Air Pollution in Georgia Using Fused Ambient Monitor Data and Chemical Transport Model Results

Sheila A. Sororian, Heather A. Holmes, Mariel Friberg, Cesunica Ivey, Yongtao Hu, James A. Mulholland, Armistead G. Russell, and Matthew J. Strickland

Abstract Health data geo-coded with residential coordinates are being used to investigate the relationship between ambient air quality and pediatric emergency department visits in the State of Georgia over the time period 2000–2010. Two types of ambient air quality data – observed concentrations from ambient monitors and predicted concentrations from a chemical transport model (CMAQ) – are being fused to provide spatially resolved daily metrics of five air pollutant gases (CO, NO_2, NO_x, SO_2 and O_3) and seven airborne particulate matter measures (PM_{10}, $PM_{2.5}$, and $PM_{2.5}$ constituents SO_4^{2-}, NO_3^-, NH_4^+, EC, OC). The observational data provide reliable temporal trends at and near monitors, but limited spatial information due to the sparse monitoring network; CMAQ data, on the other hand, provide rich spatial information but less reliable temporal information. Four data fusion techniques were applied to provide daily spatial fields of ambient air pollutant concentrations, with data withholding used to evaluate model performance. Two of the data fusion methods were combined to provide results that minimized bias and maximized correlation over time and space with withheld data. Results vary widely across pollutants. These results provide health researchers with complete temporal and spatial air pollutant fields, as well as with temporal and spatial error estimate fields that can be incorporated into health risk models. Future work will apply these methods to five cities for use in ongoing air pollution health studies and to examine strategies for incorporating land use regression variables for spatial downscaling of data fusion results.

S.A. Sororian • H.A. Holmes • M. Friberg • C. Ivey • Y. Hu • J.A. Mulholland (✉) • A.G. Russell
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: heather.holmes@ce.gatech.edu; james.mulholland@ce.gatech.edu; ted.russell@gatech.edu

M.J. Strickland
Department of Environmental Health, Emory University, Atlanta, GA 30322, USA

49.1   Introduction

Large population studies of human health effects associated with air pollution can take advantage of both temporal and spatial variation in ambient air quality levels when such estimates are available. As a part of the Southeastern Center for Air Pollution and Epidemiology (SCAPE), we are examining relationships of health outcomes with ambient air pollution resolved temporally to a daily level and resolved spatially to a 4 km and smaller scale. Presented here is an approach for estimating daily spatial fields of air pollution in the State of Georgia for use in an investigation of the relationship with pediatric emergency department visits, geocoded with residential coordinates, that builds on previous work [2].

49.2   Methods

Observed concentrations of five air pollutant gases (CO, NO_2, NO_x, SO_2 and O_3) and seven airborne particulate matter measures (PM_{10}, $PM_{2.5}$, and $PM_{2.5}$ constituents SO_4^{2-}, NO_3^-, NH_4^+, EC, OC) were obtained from EPA's Air Quality System (AQS), the Southeastern Aerosol Research Characterization (SEARCH) [3], and the Assessment of the Spatial Aerosol Composition in Atlanta (ASACA) [1] networks for Georgia for the 2000–2010 time period. The number of monitors and daily metrics calculated are listed in Table 49.1. Locations of monitors are shown in Fig. 49.1. The observational data provide reliable temporal trends at and near monitors [6], but limited spatial information due to the sparse monitoring network. Predicted hourly concentrations from CMAQ were obtained for the Georgia domain at the 12 km scale for 2002–2008 [5] and at the 4 km scale for 2009–2010 [4]; daily metrics were computed from these datasets. The CMAQ data provide rich spatial information but less reliable temporal information. Four data fusion techniques are applied and evaluated for 2010. Two methods use regression models of observational and CMAQ annual mean data for scaling (Eq. 49.1). One of these methods involves daily kriging of the ratio of the observation (OBS) to its annual mean and then rescaling by the predicted annual mean (C_1^*, Eq. 49.2). The second

Table 49.1 Ambient air pollutant metrics and monitors

	8-h max O_3	1-h max NO_2/NO_x	1-h max CO	1-h max SO_2	24-h avg PM_{10}	24-h avg $PM_{2.5}$	24-h speciated $PM_{2.5}$
North GA mt	3	0	0	2	3	2	2
Metro Atlanta	14	7	5	5	8	16	6
Piedmont	4	0	0	2	1	5	2
Upper coastal	4	0	0	1	4	7	2
Lower coastal	2	0	0	3	3	4	2
Total	27	7	5	13	19	34	14

Fig. 49.1 Map of Georgia showing regions and major cities, major roadways, and locations of speciated PM$_{2.5}$ and ozone monitors

of these methods involves rescaling the CMAQ data (C$^{*}_2$, Eq. 49.3). The other two methods involve kriging error on a multiplicative (C$^{*}_3$, Eq. 49.4) and additive (C$^{*}_4$, Eq. 49.5) basis. Data withholding was used to evaluate model performance. Two of the methods were then selected and combined to provide results that minimized bias and maximized correlation over time and space with withheld data.

$$\langle C^{*}\rangle = \alpha_{yr}\langle CMAQ\rangle^{\beta} \tag{49.1}$$

$$C^{*}_1 = (OBS/\langle OBS\rangle)_{krig} \times \langle C^{*}\rangle \tag{49.2}$$

$$C^{*}_2 = CMAQ \times \langle C^{*}\rangle / \langle CMAQ\rangle \tag{49.3}$$

$$C^{*}_3 = CMAQ \div (CMAQ/OBS)_{krig} \tag{49.4}$$

$$C^{*}_4 = CMAQ - (CMAQ - OBS)_{krig} \tag{49.5}$$

49.3   Results

The four data fusion methods yield similar spatial fields (e.g. Fig. 49.2). However, the methods that involved daily kriging of error (Eqs. 49.4 and 49.5) were less stable, performing less well in terms of both bias and correlation for most pollutants. Therefore, the methods that involved using the annual mean model (Eq. 49.1) were selected. The first approach involving kriging daily observation ratios (OBS method) yields results with a spatial structure of CMAQ and perfectly correlated temporally with observations at monitor locations. As the distance from monitors increases, the Pearson correlation coefficient decreases (exponential correlogram model: $R_1 = e^{\gamma D}$ where D is a weighted distance to monitors). The second approach that involves rescaling CMAQ predictions (CMAQ method) yields results that have similar correlations with observations across monitors (average Pearson correlation coefficient of R_2). Thus, the OBS method performs best near monitors and the CMAQ method performs best far from monitors.

The OBS and CMAQ methods were combined to provide the best overall prediction. A correction was added to the CMAQ method prediction to correct for seasonal bias. The daily spatial fields generated by the OBS and CMAQ methods were averaged by weighting based on their temporal correlation coefficients (R_1 and R_2). Thus, at monitor locations the OBS data fusion method prediction was used and far from monitors the CMAQ data fusion method was used. Combined data fusion method results minimize bias and maximize temporal correlation over space. Results shown in Fig. 49.3 show the percent biases and temporal correlations in
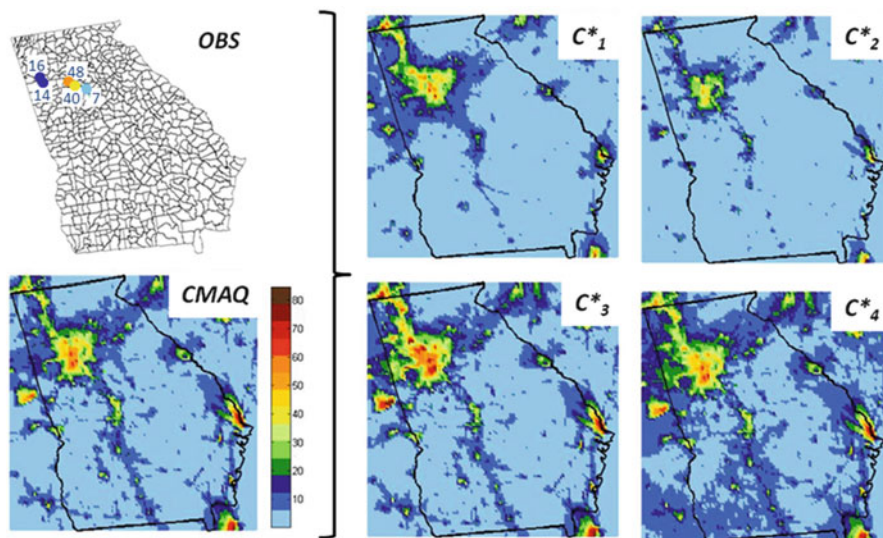

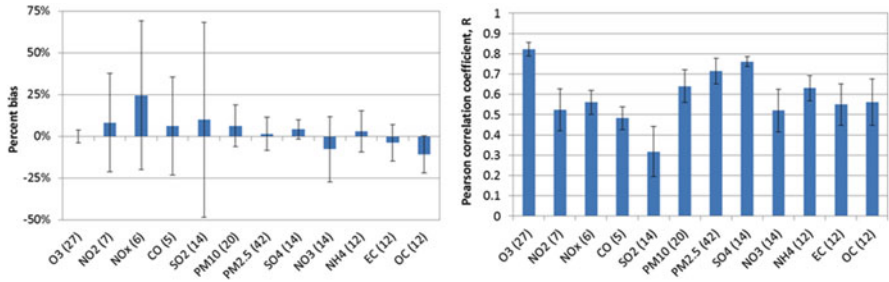**Fig. 49.2**   Predicted 1-h max NO_2 fields by four data fusion methods, 9/21/10

Fig. 49.3 Percent bias (*left*) and Pearson correlation coefficient (*right*) for CMAQ method predictions, 2002–2008. *Bars* represent averages and *error bars* represent standard deviations across all monitors. The number of monitors for each pollutant is given in *parentheses*

the CMAQ method predictions for the 2002–2008 time period using 12 km CMAQ data. When combined with OBS method predictions, temporal correlations increase near monitors.

Results vary widely across pollutants. Estimates of primary pollutants largely from mobile sources (i.e. CO, NO_2, NO_x, and EC) rely largely on CMAQ data because they are spatially heterogeneous and have low spatial autocorrelation, whereas estimates of largely secondary pollutants (i.e. O_3, SO_4^{2-}, NO_3^-, NH_4^+) and pollutants of mixed origin (PM_{10}, $PM_{2.5}$ and OC) rely to a greater extent on observations because they are spatially more homogeneous and have high spatial autocorrelation. SO_2, largely from coal combustion, is the most difficult pollutant to estimate due to the inability of monitors to capture and CMAQ to predict accurately point source plume dispersion.

These results provide health researchers with complete temporal and spatial air pollutant fields, as well as with temporal and spatial error estimate fields that can be incorporated into health risk models. Future work will apply these methods to five cities for use in ongoing air pollution health studies and to examine strategies for incorporating land use regression variables for spatial downscaling of data fusion results.

References

1. Butler AJ, Andrew MS, Russell AG (2003) Daily sampling of PM2.5 in Atlanta: results of the first year of the assessment of spatial aerosol composition in Atlanta study. J Geophys Res 108(D7):8415

2. Darrow LA, Klein M, Flanders WD, Waller LA, Correa A, Marcus M, Mulholland JA, Russell AG, Tolbert PE (2009) Ambient air pollution and preterm birth: a time-series analysis. Epidemiology 20(5):689–698
3. Hansen DA, Edgerton ES, Hartsell BE, Jansen JJ, Kandasamy N, Hidy GM, Blanchard CL (2003) The southeastern aerosol research and characterization study: Part 1 – overview. J Air Waste Manage Assoc 53(12):1460–1471
4. Hu Y, Chang M, Russell A, Odman M (2010) Using synoptic classification to evaluate an operational air quality forecasting system in Atlanta. Atmos Pollut Res 1:280–287
5. PHASE (2013) Air quality data for the CDC national environmental public health tracking network. URL http://www.epa.gov/heasd/research/cdc.html. Accessed 2013
6. Wade KS, Mulholland JA, Marmur A, Russell AG, Hartsell B, Edgerton E, Klein M, Waller L, Peel JL, Tolbert PE (2006) Effects of instrument precision and spatial variability on the assessment of the temporal variation of ambient air pollution in Atlanta, Georgia. J Air Waste Manage Assoc 56(6):876–888