Chapter 4
Improved Spatiotemporal Source-Based Air Pollutant Mixture Characterization for Health Studies

Heather A. Holmes, Xinxin Zhai, Jeremiah Redman, Kyle Digby, Cesunica Ivey, Sivaraman Balachandran, Sheila A. Sororian, Mariel Friberg, Wenxian Zhang, Marissa L. Maier, Yongtao Hu, Armistead G. Russell, James A. Mulholland, and Howard H. Chang

Abstract The growing availability of spatially resolved health data sets (i.e., resident and county level patient records) requires spatially resolved exposure or air quality metrics to investigate the impact of air pollution on health outcomes. While daily air quality data are essential in time-series epidemiologic analysis, the spatial distribution of the observations is limited. Air pollution modeling (i.e., chemical transport modeling (CTM)) addresses this by producing spatially resolved air quality predictions using terrain, emissions and meteorology inputs. However, predicted concentrations may be biased. This work incorporates unique data fusion approaches to combine air quality observations from regulatory monitoring networks (OBS) with the output from a CTM (CMAQ) to generate spatially and temporally resolved gaseous and PM species concentrations. Species concentrations alone cannot directly identify emission sources or characterize pollutant mixtures, therefore source apportionment (SA) models are required to estimate source impacts. The focus of this work is a comparison of SA results for three U.S. regions with differing air pollution sources, St. Louis, Missouri; Atlanta, Georgia; and Dallas-Fort Worth, Texas.

H.A. Holmes (✉) • X. Zhai • J. Redman • K. Digby • C. Ivey • S. Balachandran • S.A. Sororian • M. Friberg • W. Zhang • M.L. Maier • Y. Hu • A.G. Russell • J.A. Mulholland
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: heather.holmes@ce.gatech.edu

H.H. Chang
Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA

D. Steyn and R. Mathur (eds.), *Air Pollution Modeling and its Application XXIII*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-04379-1_4,

4.1 Introduction

Air pollution concentrations measured from regulatory monitoring networks are commonly used as air quality metrics in time-series epidemiologic analysis to investigate air quality and human health associations. While these data provide useful indicators for air pollution impacts in a region, the data are limited temporally, spatially and chemically. The growing availability of spatially resolved health data sets (i.e., resident and county level patient records) requires spatially resolved exposure or air quality metrics to investigate the impact of air pollution on health outcomes [3]. The species concentration data in combination with chemical transport models (CTM) and source apportionment (SA) techniques can be used to characterize sources and species impacting both individual locations and wider areas. As part of the Southeastern Center for Air Pollution and Epidemiology (SCAPE), a Mixtures Characterization (MC) Toolkit is being developed to effectively analyze air pollution and air quality modeling data to better understand how sources are combining to impact air quality for use in health assessments.

4.2 Methods

Statistical data analysis is done using the daily observations of species concentrations to calculate a time-series of spatial air pollution metrics based on both area and population weighted averages [6]. While daily air quality data are essential in time-series epidemiological analysis, the spatial distribution of the observations is limited. Air pollution modeling (i.e., chemical transport modeling (CTM)) addresses this by producing spatially resolved air quality predictions. The CTM used is CMAQ (U.S. EPA), which uses meteorology outputs from a numerical weather prediction model and emissions modeling to generate a spatial and temporal allocation of the source emissions. The CMAQ model incorporates emissions, meteorological and chemical processes but the model output is not nudged or calibrated to observations therefore predicted concentrations may be biased. Thus, a data fusion approach is utilized to combine air quality observations from regulatory monitoring networks (OBS) with the output from CMAQ to improve the spatially and temporally resolved gaseous and PM species concentrations [13].

Species concentrations alone cannot directly be used to identify emission sources or characterize pollutant mixtures [12, 14]. Hence, the fused OBS-CMAQ data is incorporated into an Emissions Based Integrated Mobile Source Indicator ($IMSI_{EB}$) to obtain spatiotemporal mobile source impact estimates [9]. The method developed by Pachon et al. [9] is for ambient monitor data, this work extends that method and applies it to each grid cell in the OBS-CMAQ modeling domain. Where, concentrations of EC, CO and NO_x are used as tracers to estimate the impact of mobile sources on $PM_{2.5}$. The ratio of the mobile emissions to total emissions for each species (i.e., EC, CO and NO_x) is used with the species concentrations to weight the mobile source contributions in each grid. The estimated mobile source impact is scaled using the relationship between the annual EC concentration and annual mobile source impact estimated by receptor modeling.

Source impacts are also estimated using traditional receptor oriented SA methods, Chemical Mass Balance (CMB) and Positive Matrix Factorization (PMF), and source oriented modeling from CMAQ [4, 7, 10, 15]. As part of the MC Toolkit, two novel SA techniques are being developed to improve the characterization of exposure estimates by source emissions. First, an ensemble method that generates new source profiles for CMB based on an ensemble-trained approach utilizing results from CMB, PMF and CMAQ [1, 2, 8]. The second is a hybrid source-receptor model approach that adjusts the original CMAQ source impact estimates based on scaling factors obtained using simulations and observations in a CMB-fashion optimization [5, 11].

4.3 Results

Results from the methods described above are presented for three regions with differing air pollution sources, St. Louis, Missouri; Atlanta, Georgia; and Dallas-Fort Worth, Texas. There is a significant impact of point source emissions, e.g., chemicals manufacturing and metals processing, on the air pollution in St. Louis. While in Atlanta large biogenic emissions interact with emissions from mobile sources and power plants which lead to a large amount of secondary organic aerosol formation. The primary sources impacting air pollution in Dallas-Fort Worth are mobile sources. Spatially resolved species concentrations and source impact estimates are shown for the state of Georgia because air quality modeling is utilized to provide air quality metrics for a spatially resolved health study in Georgia.

Spatially resolved health studies require more information about the spatial distribution of pollutant concentrations. The OBS-CMAQ method provides the spatial distribution of air quality metrics by combining the monitor data and CMAQ results. Figure 4.1, shows the 8-h ozone concentration from the monitors and the original CMAQ results (left) and the fused OBS-CMAQ predictions are also shown (right). The ratio of the observation and CMAQ results are spatially interpolated to each grid cell in the domain using ordinary kriging which is then multiplied by the original CMAQ field to adjust the concentrations predicted by CMAQ. The result is an unbiased species concentration at a 4-km resolution for the state of Georgia. While 8-h ozone is shown here this is being done for additional pollutant gases (CO, NO_2, NO_x, SO_2) and particulate matter species (PM_{10}, $PM_{2.5}$, SO_4^{2-}, NO_3^-, NH_4^+, EC, OC).

The OBS-CMAQ species concentrations can be used in combination with the emissions inventory to calculate mobile source impacts in Georgia (Fig. 4.2). An emissions model is used to generate spatially and temporally resolved emissions estimates for each grid in the modeling domain. Then, the emissions information and OBS-CMAQ species concentrations for EC, CO and NO_x are used to generate

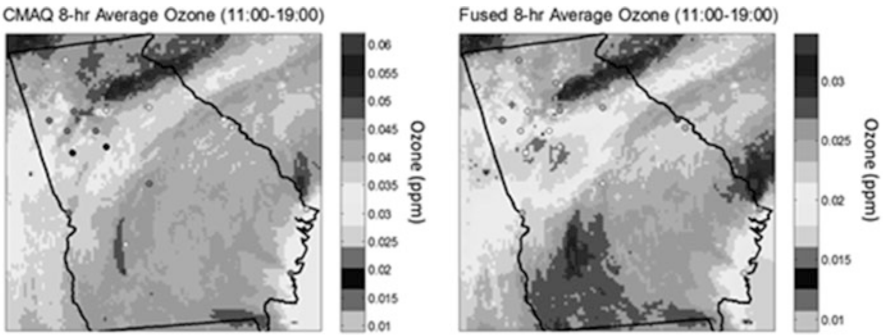
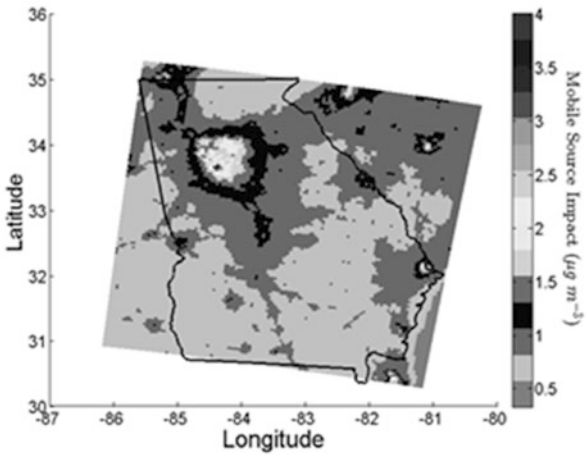


Fig. 4.1 Improved spatial concentration estimates: Georgia 4-km resolution CMAQ simulated 8-h ozone 18 July 2010 (*left*) CMAQ and observations (*right*) Fused OBS-CMAQ, spatially interpolated ratio of observation to CMAQ concentration using ordinary kriging (*Note scale difference on color axis*)

Fig. 4.2 Spatial mobile source impact estimates: 2010 annual average Georgia 4-km resolution using CMAQ-OBS EC, CO and NO_x concentrations and emissions inventory data



daily mobile source impacts at a 4-km resolution for Georgia. The 2010 annual average mobile source impact is shown in Fig. 4.2. The major urban areas can be seen in the figure with the highest mobile source impacts and the structure of the major freeways north and south of Atlanta is also evident.

For time-series health studies, results from multiple SA receptor models are ensemble averaged to generate updated source profiles that are used as inputs to CMB. The ensemble method is an improvement over using a single SA model because there are less zero impact days which is physically unrealistic and the source profiles are more indicative of the sources impacting the location.

The difference in source profiles can be seen by comparing the ensemble based source profiles (EBSP) and measurement based source profiles (MBSP) for multiple receptor locations, shown in Fig. 4.3 for gasoline and diesel vehicles. The ensemble method assumes that there is a seasonality in source impacts and therefore the data is split into summer and winter seasons when generating the EBSPs.

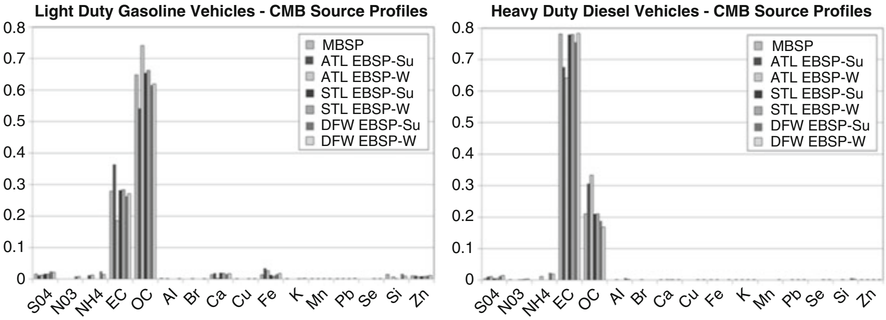


Fig. 4.3 Ensemble source apportionment: Comparison of measurement based (*MBSP*) and ensemble based (*EBSP*) summer and winter source profiles used in CMB for Atlanta (*ATL*), St. Louis (*STL*) and Dallas-Fort Worth (*DFW*)

Future work will incorporate the source impact estimates in each region into epidemiologic studies to investigate the city-to-city variability in pollutant mixtures and associations with health outcomes.

Acknowledgments This research was supported, in part, by USEPA grant R834799. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication. Additional support was made possible by grants from Georgia Power and the Southern Company.

References

1. Balachandran S, Chang HH, Pachon JE, Holmes HA, Mulholland JA, Russell AG (2013) A Bayesian-based ensemble technique for source apportionment of $PM_{2.5}$. Environ Sci Technol 47:13511–13518
2. Balachandran S, Pachon JE, Hu Y, Lee D, Mulholland JA, Russell AG (2012) Ensemble trained source apportionment of fine particulate matter and method uncertainty analysis. Atmos Environ 61:387–394
3. Bell ML (2006) The use of ambient air quality modeling to estimate individual and population exposure for human health research: a case study of ozone in the northern Georgia region of the United States. Environ Int 32(5):586
4. Cohan DS, Hakami A, Hu Y, Russell AG (2005) Nonlinear response of ozone to emissions: source apportionment and sensitivity analysis. Environ Sci Technol 39(17):6739–6748
5. Ivey C, Holmes H, Hu Y, Russell A, Mulholland J (2013) Spatial and temporal extension of a novel hybrid source apportionment model. In: International technical meeting on air pollution modelling and its application, Miami, FL
6. Ivy D, Mulholland JA, Russell AG (2008) Development of ambient air quality population-weighted metrics for use in time-series health studies. J Air Waste Manage Assoc 58(5):711–720
7. Koo B, Dunker AM, Yarwood G (2007) Implementing the decoupled direct method for sensitivity analysis in a particulate matter air quality model. Environ Sci Technol 41(8):2847–2854

8. Maier ML, Balachandran S, Mulholland JA, Russell AG, Ebelt S, Turner JR (2013) Application of an ensemble-trained source apportionment approach at a site impacted by multiple point sources. Environ Sci Technol 47:3743–3751
9. Pachon JE, Balachandran S, Hu Y, Mulholland JA, Darrow LA, Sarnat JA, Tolbert PE, Russell AG (2012) Development of outcome-based, multipollutant mobile source indicators. J Air Waste Manage Assoc 62(4):431–442
10. Reff A, Eberly SI, Bhave PV (2007) Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. J Air Waste Manage Assoc 57(2):146–154
11. Russell A, Holmes H, Friberg M, Ivey C, Hu Y, Balachandran S, Mulholland J, Tolbert P, Sarnat J, Sarnat S, Strickland M, Chang H, Liu Y (2013) Use of air quality modeling results in health effects research. In: International technical meeting on air pollution modelling and its application, Miami, FL
12. Sarnat JA, Marmur A, Klein M, Kim E, Russell AG, Sarnat SE, Mulholland JA, Hopke PK, Tolbert PE (2008) Fine particle sources and cardiorespiratory morbidity: an application of chemical mass balance and factor analytical source-apportionment methods. Environ Health Perspect 116(4):459
13. Sororian S, Holmes H, Friberg M, Ivey C, Hu Y, Mulholland J, Russell A, Strickland M, Chang H (2013) Temporally and spatially resolved air pollution in Georgia using fused ambient monitoring data and chemical transport model results. In: International technical meeting on air pollution modelling and its application, Miami, FL
14. Stanek LW, Sacks JD, Dutton SJ, Dubois JJB (2011) Attributing health effects to apportioned components and sources of particulate matter: an evaluation of collective results. Atmos Environ 45(32):5655–5663
15. Watson JG, Zhu T, Chow JC, Engelbrecht J, Fujita EM, Wilson WE (2002) Receptor modeling application framework for particle source apportionment. Chemosphere 49(9):1093–1136

Questions and Answer

Questioner Name: Clemens Mensink

Q: How far do we have to go in increasing the model resolution and improving temporal aspects (e.g., on traffic emissions) in order to make sure that there is a scientific sound and well understood association with health?

A: The air quality (AQ) data needed to support the epidemiologic analysis depends on the health association being investigated, e.g. if location of residents is being used spatially resolved AQ metrics are necessary and for daily emergency department visits daily AQ metrics are required. The purpose of this work is to provide surrogates of exposure for a spatially resolved health analysis (i.e., 250 m) therefore improving the AQ model resolution to match the health study resolution is desired.