THE DEVELOPMENT, APPLICATION AND EVALUATION OF ADVANCED SOURCE APPORTIONMENT METHODS

A Dissertation
Presented to
The Academic Faculty

by

Sivaraman Balachandran

In Partial Fulfillment
of the Requirements for the Degree
Doctorate of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
December 2013

THE DEVELOPMENT, APPLICATION AND EVALUATION OF ADVANCED SOURCE APPORTIONMENT METHODS

Approved by:

Dr. Armistead G. Russell, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. James A. Mulholland
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael H. Bergin
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Rodeny J. Weber
School of Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. Jeremy A. Sarnat
Rollins School of Public Health
Emory University

Date Approved: August 15, 2013

To my daughters: Maya and Asha. Words cannot express how much I love you.

And

To my parents, for always being there for me.

ACKNOWLEDGEMENTS

I have been very fortunate to have been supported by amazing people without whom it would not have been possible for me to get a PhD. First, I would like to thank my advisors, Dr. Armistead Russell and Dr. James Mulholland, for guiding me with my intellectual growth. I would also like to thank my committee members, Dr. Michael Bergin, Dr. Jeremy Sarnat and Dr. Rodney Weber, for their thoughtful comments, suggestions and input into my research.

I am also very grateful for the contributions of co-colloborators to my research work. I would like to thank Dr. Heather Holmes for being so closely involved with my research, from suggestions to improve methods to help with coding to editing papers. I would also like to thank Dr. Howard Chang of the Rollins School of Public Health at Emory University for helping me set up a Bayesian formulation for the ensemble method. I would like to thank members of the Russell Group past and present who have not only given me research support and but also friendship: Jorge Pachon, Gretchen Goldman, Marcus Trail, Boris Galvis, Fernando Garcia, Sunni Ivey, Aika Yano and Yongtao Hu.

I would also like to thank a number of friends and family for their support during my graduate studies. First, I would like to thank my parents for their physical, emotional and spiritual support – they have been with me 100% of the way. I would like to thank my brother Gopal, his wife Jyoti and my sister Padma for listening to me gripe, encouraging me and carrying the stress of graduate school with me. I would also like to thank my uncles Ram Bishu and V.J. Venkatraman for their encouragement, advice and moral support. I would also like to thank a number of friends who have also provided emotional support for me: Matt Iskra, Devin Theriot-Orr, Lauren O'Toole, Dale and Heather Rose, Andy Cozewith, Mike and Jessy Molina, and Nathan Nobis.

I could not have done this without my daughters Maya and Asha. For most of their lives, I have been in graduate school and they have had to deal with the stress and limitations that comes with that. But they have never complained about it – in fact, they have always inspired me with their boundless energy and playfulness. Their love has provided me with the will and drive to make it through this journey. Finally, I would like to thank their mom, Jenny, for being patient with me, and being a wonderful mom to our kids.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Ambient and indoor air pollution is a major cause of premature mortality, and has been associated with more than three million preventative deaths per year worldwide. Most of these health impacts are from the effects from fine particulate matter. It is suspected that $PM_{2.5}$ health effects vary by composition, which depends on the mixture of pollutants emitted by sources. This has led to efforts to estimate relationships between sources of $PM_{2.5}$ and health effects. The health effects of $PM_{2.5}$ may be preferentially dependent on specific species; however, recent work has suggested that health impacts may actually be caused by the net effect of the mixture of pollutants which make up $PM_{2.5}$. Recently, there have been efforts to use source impacts from source apportionment (SA) studies as a proxy for these multipollutant effects. Source impacts can be quantified using both receptor and chemical transport models (RMs and CTMs), and have both advantages and limitations for their use in health studies.

In this work, a technique is developed that reconciles differences between source apportionment (SA) models by ensemble-averaging source impacts results from several SA models. This method uses a two-step process to calculate the ensemble average. An initial ensemble average is used calculate new estimates of uncertainties for the individual SA methods that are used in the ensemble. Next, an updated ensemble average is calculated using the SA method uncertainties as weights. Finally, uncertainties of the ensemble average are calculated using propagation of errors that includes covariance terms. The ensemble technique is extended to include a Bayesian formulation of weights used in ensemble-averaging source impacts. In a Bayesian approach, probabilistic distributions of the parameters of interest are estimated using prior distributions, along with information from observed data.

Ensemble averaging results in updated estimates of source impacts with lower uncertainties than individual SA methods. Overall uncertainties for ensemble-averaged source impacts were ~45 - 74%. The Bayesian approach also captures the expected seasonal variation of biomass burning and secondary impacts. Sensitivity analysis found that using non-informative prior weighting performed better than using weighting based on method-derived uncertainties. The Bayesian-based source impacts for biomass burning correlate better with observed levoglucosan (R^2=0.66) and water soluble potassium (R^2=0.63) than source impacts estimated using more traditional methods, and more closely agreed with observed total mass. Power spectra of the time series of biomass burning source impacts suggest that profiles/factors associated with this source have the greatest variability across methods and locations.

A secondary focus of this work is to examine the impacts of biomass burning. First a field campaign was undertaken to measure emissions from prescribed fires. An emissions factor of 14±17 g $PM_{2.5}$/kg fuel burned was determined. Water soluble organic carbon (WSOC) was highly correlated with potassium (K) (R^2=.93) and levoglucosan (R^2=0.98). Results using a biomass burning source profile derived from this work further indicate that source apportionment is sensitive to levels of potassium in biomass burning source profiles, underscoring the importance of quantifying local biomass burning source profiles. Second, the sensitivity of ambient $PM_{2.5}$ to various fire and meteorological parameters in was examined using the method of principle components regression (PCR) to estimate sensitivity of $PM_{2.5}$ to fire data and, observed and forecast meteorological parameters. $PM_{2.5}$ showed significant sensitivity to PB, with a unit-based sensitivity of 3.2±1 $\mu g\ m^{-3}$ $PM_{2.5}$ per 1000 acres burned. $PM_{2.5}$ had a negative sensitivity to dispersive parameters such as wind speed

CHAPTER 1: INTRODUCTION

Ambient and indoor air pollution is a major cause of premature mortality, and has been associated with more than three million preventative deaths per year worldwide [*Lim et al.*, 2012]. Most of these health impacts are from the effects from fine particulate matter. Unlike most other air pollutants, fine particulate matter (i.e. particles with an aerodynamic diameter less than 2.5μm, or $PM_{2.5}$) is comprised of a heterogeneous mix of chemical species, some of which are emitted directly from a variety of sources and others that are formed via atmospheric processes which convert gaseous species into condensed-phase compounds. The health concern over $PM_{2.5}$ has grown as associations have been found between $PM_{2.5}$ mass and health outcomes [*Dockery et al.*, 1993; *U.S.EPA*, 2009], and has led EPA to regulate $PM_{2.5}$ as a criteria pollutant as part of the US EPA's National Ambient Air Quality Standards (NAAQS). Controlling fine particulate matter poses unique challenges in developing strategies to improve public health and welfare (e.g., improved visibility), a major goal for states and regional communities.

It is suspected that $PM_{2.5}$ health effects vary by composition and source, and may depend upon the mixture of pollutants, leading to efforts to estimate relationships between sources of $PM_{2.5}$ and health effects [*Hopke et al.*, 2006; *Ito et al.*, 2006; *Mar et al.*, 2006; *Sarnat et al.*, 2008; *Thurston et al.*, 2005]. The health effects of $PM_{2.5}$ may be preferentially dependent on specific species; however, recent work has suggested that health impacts may actually be caused by the net effect of the mixture of pollutants which make up $PM_{2.5}$ [*Solomon et al.*, 2011; *Solomon et al.*, 2012]. Traditional epidemiologic models have generally used $PM_{2.5}$ or individual species in assessing health impacts. Recently, there have been efforts to use source impacts from source apportionment (SA) studies as a proxy for multipollutant effects [*Hopke et al.*, 2006; *Ito et al.*, 2006; *Sarnat et al.*, 2008; *Thurston et al.*, 2005]. There have been several efforts to determine relationships between sources of $PM_{2.5}$ and health outcomes [*Laden et al.*, 2000; *Mar et*

al., 2000; *Marmur et al.*, 2006b; *Sarnat et al.*, 2008; *Stolzel et al.*, 2005; *Thurston et al.*, 2005], though with different results.

Controlling ambient $PM_{2.5}$ concentrations ultimately means controlling sources of $PM_{2.5}$ which requires techniques for estimating source contributions. Source impacts can be quantified using both receptor and chemical transport models (RMs and CTMs), and have both advantages and limitations for their use in health studies. RMs are not computationally intensive, require observational data from a central monitor, and can be used easily in time series health studies. A major limitation of RMs is that their results are only valid for the location of the monitor. Source impacts, as well as central monitor data, are proxies for exposure, an assumption which may not be accurate given, that there is much spatial variability in air pollution within a metro area. Recently, efforts to use CTM SA results have addressed some of these issues because CTMs can provide results over a large spatial domain. In addition, they can provide results at a high temporal frequency (e.g. hourly results). They can also model complex atmospheric chemistry and have a greater number of source categories than RMs. However, CTMs require large computational resources, a major limitation when long time series of source impacts are required.

These different SA approaches often result in source contributions that can differ in magnitude and/or are poorly correlated. Determining which method's set of source contributions is the most accurate is further complicated because source impacts, in general, cannot be directly measured. Without direct measurement of source impacts, methods for estimating uncertainty vary across the SA approaches, making it difficult to directly compare uncertainties across methods. For example, some methods (e.g. CTMs) have not provided source impact estimate uncertainties while others utilize bootstrapping or propagation of errors to estimate uncertainties.

In this work, a technique is developed that reconciles differences between model results by ensemble-averaging source impact results from several SA models. This

method uses a two-step process to calculate the ensemble average. An initial ensemble average is used calculate new estimates of uncertainties for the individual SA methods that are used in the ensemble. Next, an updated ensemble average is calculated using the SA method uncertainties as weights. Finally, uncertainties of the ensemble average are calculated using propagation of errors that includes covariance terms. The ensemble technique is extended to include a Bayesian formulation of weights used in ensemble-averaging source impacts.

Another focus of this dissertation is to examine the effects of biomass burning, specifically prescribed fires, which are a significant contributor to $PM_{2.5}$. Biomass burning, such as wildfires, prescribed burns, and residential wood combustion, are important sources of air pollutants, which can impact health, lead to violations of air quality standards, and impair visibility [*S. Lee et al.*, 2005; *Sandberg et al.*, 2002]. Prescribed burning is widespread, especially in the southeastern US, and is used to manage forest ecosystems and protect endangered species by controlling growth and infestation while minimizing the risk of large-scale forest fires [*Hardy et al.*, 2001]. In addition, the Southeast US has experienced substantial population growth the last few decades [*USCensus*, 2012], causing significant urban sprawl in an otherwise heavily forested region, making the wildland urban interface especially susceptible to air quality impacts from prescribed burning.

The thesis is organized as follows:

Chapter 2: Ensemble-Trained Source Apportionment of Fine Particulate Matter and Method Uncertainty Analysis. This work updates the ensemble-averaging approach by *D Lee et al.* [2009]. Ensemble averaging of SA results is conducted in two steps. In the first step source impact estimates are averaged together. In the second step, the initial ensemble is used to re-estimate SA method uncertainties, which are then used as weights to calculate an updated average. Next, uncertainties for the updated ensemble

source impact are calculated. In part, this can address concerns that the uncertainties provided by the traditional methods are biased.

A compelling reason to quantify uncertainties is that they can be incorporated into epidemiologic studies, which can ultimately lead to improving our understanding of the relationships between $PM_{2.5}$ sources and health outcomes. Further, they can be used to inform policy makers of the effectiveness of control measures.

Chapter 3: Bayesian–Based Ensemble Source Apportionment of PM2.5. In this work, we extend the ensemble technique to include a Bayesian formulation of weights used in ensemble-averaging source impacts. In a Bayesian approach, probabilistic distributions of the parameters of interest are estimated using prior distributions, along with information from observed data. Following this approach, we obtain multiple realizations of ensemble-averaged source impacts, which are subsequently used for deriving multiple realizations of source profiles. We then compare results using this approach to results using our previous ensemble approach as well as to results using individual receptor models.

Chapter 4: Spectral Analysis of $PM_{2.5}$ Source Apportionment Methods. Here we use results from multiple $PM_{2.5}$ source apportionment results at three receptor sites. Two of the three sites are Southeastern Aerosol Research and Characterization (SEARCH) network [*Hansen et al.*, 2003] sites and the third is a Chemical Speciation Trends (CSN) site. We compare results from SDK and JST to assess intra-urban differences in SA. We use results from SEARCH sites, JST and YRK, to compare differences in urban versus rural receptor sites. We apply spectral analysis of source impacts and important tracers at each of these sites to gain insight into how source apportionment methods vary temporally.

Chapter 5: Particulate and Gas Sampling of Prescribed Fires in South Georgia, USA. A major goal of this study was to update emissions factors for gaseous compounds and $PM_{2.5}$ in Georgia with regionally specific biomass burning air emissions data. A second goal was to better understand the role of water soluble organic carbon (WSOC) as a tracer of both biomass burning and secondary organic aerosol. Third, tracers of prescribed burns were studied by characterization of organic chemical compounds. In addition, chemical speciation of $PM_{2.5}$ was used in a source apportionment study to test its applicability as regionally specific biomass burning source profile.

Chapter 6: Verification of Fire Weather Forecasts Using PM2.5 Sensitivity Analysis. In this work, we investigate the sensitivity of ambient $PM_{2.5}$ to various fire and meteorological parameters in a spatial setting that is typical for the wildland urban interface in the southeastern US. We use the method of principle components regression (PCR) to estimate sensitivity of $PM_{2.5}$ to fire data and, observed and forecast meteorological parameters. In PCR, principal components analysis (PCA) is first run on a data set. We ran PCA on 10 data sets that included PB activity data along with meteorological parameters of interest; the meteorological parameters included either observational data only, forecast data only or a combination of observations and forecasts. For each data set, we regressed PCA scores from the first seven principal components against observed $PM_{2.5}$ to quantify sensitivities.

Chapter 7: Summary and Future Work. In this dissertation, a number of inconsistencies and limitations of various source apportionment techniques are addressed by ensemble-averaging results from a short-term application of three receptor-based models and one emissions-based model. The method has a number of benefits over using one model exclusively and provides a way to evaluate different source apportionment

(SA) models, including estimating uncertainties in a consistent manner. A secondary focus of this work is to examine the impacts of biomass burning. Future work includes incorporating results from this work in health assessment models. Also, CTM uncertainties from this work can be compared with other estimates. The Bayesian method developed here can be extended to include non-conjugate priors. The method can also be extended to define selection criteria for sampling source profiles in future source apportionment work.

1.1. References

1. Lim, S. S.; Vos, T.; Flaxman, A. D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Amann, M.; Anderson, H. R.; Andrews, K. G.; Aryee, M.; Atkinson, C.; Bacchus, L. J.; Bahalim, A. N.; Balakrishnan, K.; Balmes, J.; Barker-Collo, S.; Baxter, A.; Bell, M. L.; Blore, J. D.; Blyth, F.; Bonner, C.; Borges, G.; Bourne, R.; Boussinesq, M.; Brauer, M.; Brooks, P.; Bruce, N. G.; Brunekreef, B.; Bryan-Hancock, C.; Bucello, C.; Buchbinder, R.; Bull, F.; Burnett, R. T.; Byers, T. E.; Calabria, B.; Carapetis, J.; Carnahan, E.; Chafe, Z.; Charlson, F.; Chen, H.; Chen, J. S.; Cheng, A. T.-A.; Child, J. C.; Cohen, A.; Colson, K. E.; Cowie, B. C.; Darby, S.; Darling, S.; Davis, A.; Degenhardt, L.; Dentener, F.; Des Jarlais, D. C.; Devries, K.; Dherani, M.; Ding, E. L.; Dorsey, E. R.; Driscoll, T.; Edmond, K.; Ali, S. E.; Engell, R. E.; Erwin, P. J.; Fahimi, S.; Falder, G.; Farzadfar, F.; Ferrari, A.; Finucane, M. M.; Flaxman, S.; Fowkes, F. G. R.; Freedman, G.; Freeman, M. K.; Gakidou, E.; Ghosh, S.; Giovannucci, E.; Gmel, G.; Graham, K.; Grainger, R.; Grant, B.; Gunnell, D.; Gutierrez, H. R.; Hall, W.; Hoek, H. W.; Hogan, A.; Hosgood Iii, H. D.; Hoy, D.; Hu, H.; Hubbell, B. J.; Hutchings, S. J.; Ibeanusi, S. E.; Jacklyn, G. L.; Jasrasaria, R.; Jonas, J. B.; Kan, H.; Kanis, J. A.; Kassebaum, N.; Kawakami, N.; Khang, Y.-H.; Khatibzadeh, S.; Khoo, J.-P.; Kok, C.; Laden, F.; Lalloo, R.; Lan, Q.; Lathlean, T.; Leasher, J. L.; Leigh, J.; Li, Y.; Lin, J. K.; Lipshultz, S. E.; London, S.; Lozano, R.; Lu, Y.; Mak, J.; Malekzadeh, R.; Mallinger, L.; Marcenes, W.; March, L.; Marks, R.; Martin, R.; McGale, P.; McGrath, J.; Mehta, S.; Mensah, G. A.; Merriman, T. R.; Micha, R.; Michaud, C.; Mishra, V.; Hanafiah, K. M.; Mokdad, A. A.; Morawska, L.; Mozaffarian, D.; Murphy, T.; Naghavi, M.; Neal, B.; Nelson, P. K.; Nolla, J. M.; Norman, R.; Olives, C.; Omer, S. B.; Orchard, J.; Osborne, R.; Ostro, B.; Page, A.; Pandey, K. D.; Parry, C. D. H.; Passmore, E.; Patra, J.; Pearce, N.; Pelizzari, P. M.; Petzold, M.; Phillips, M. R.; Pope, D.; Pope Iii, C. A.; Powles, J.; Rao, M.; Razavi, H.; Rehfuess, E. A.; Rehm, J. T.; Ritz, B.; Rivara, F. P.; Roberts, T.; Robinson, C.; Rodriguez-Portales, J. A.; Romieu, I.; Room, R.; Rosenfeld, L. C.; Roy, A.; Rushton, L.; Salomon, J. A.; Sampson, U.; Sanchez-Riera, L.; Sanman, E.; Sapkota, A.; Seedat, S.; Shi, P.; Shield, K.; Shivakoti, R.; Singh, G. M.; Sleet, D. A.; Smith, E.; Smith, K. R.; Stapelberg, N. J. C.; Steenland, K.; Stöckl, H.; Stovner, L. J.; Straif, K.; Straney, L.; Thurston, G. D.; Tran, J. H.; Van Dingenen, R.; van Donkelaar, A.; Veerman, J. L.; Vijayakumar, L.; Weintraub, R.; Weissman, M. M.; White, R. A.; Whiteford, H.; Wiersma, S. T.; Wilkinson, J. D.; Williams, H. C.; Williams, W.; Wilson, N.; Woolf, A. D.; Yip, P.; Zielinski, J. M.; Lopez, A. D.; Murray, C. J. L.; Ezzati, M., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **2012,** *380*, (9859), 2224-2260.
2. Dockery, D. W.; Pope, C. A.; Xu, X. P.; Spengler, J. D.; Ware, J. H.; Fay, M. E.; Ferris, B. G.; Speizer, F. E., An Association between Air-Pollution and Mortality in

6 United-States Cities. *New England Journal of Medicine* **1993,** *329*, (24), 1753-1759.

3. U.S.EPA, Integrated Science Assessment for Particulate Matter (Final Report). . In 2009.
4. Hopke, P. K.; Ito, K.; Mar, T.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Paatero, P.; Thurston, G. D., PM source apportionment and health effects: 1. Intercomparison of source apportionment results. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (3), 275-286.
5. Ito, K.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Neas, L.; Hopke, P. K.; Thurston, G. D., PM source apportionment and health effects: 2. An investigation of intermethod variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (4), 300-310.
6. Mar, T. F.; Ito, K.; Koenig, J. Q.; Larson, T. V.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Neas, L.; Stolzel, M.; Paatero, P.; Hopke, P. K.; Thurston, G. D., PM source apportionment and health effects. 3. Investigation of inter-method variations in associations between estimated source contributions Of PM2.5 and daily mortality in Phoenix, AZ. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (4), 311-320.
7. Sarnat, J. A.; Marmur, A.; Klein, M.; Kim, E.; Russell, A. G.; Sarnat, S. E.; Mulholland, J. A.; Hopke, P. K.; Tolbert, P. E., Fine particle sources and cardiorespiratory morbidity: An application of chemical mass balance and factor analytical source-apportionment methods. *Environmental Health Perspectives* **2008,** *116*, (4), 459-466.
8. Thurston, G. D.; Ito, K.; Mar, T.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Hopke, P. K., Workgroup report: Workshop on source apportionment of particulate matter health effects - Intercomparison of results and implications. *Environmental Health Perspectives* **2005,** *113*, (12), 1768-1774.
9. Solomon, P. A.; Costantini, M.; Grahame, T. J.; Gerlofs-Nijland, M. E.; Cassee, F. R.; Russell, A. G.; Brook, J. R.; Hopke, P. K.; Hidy, G.; Phalen, R. F.; Saldiva, P.; Sarnat, S. E.; Balmes, J. R.; Tager, I. B.; Ozkaynak, H.; Vedal, S.; Wierman, S. S. G.; Costa, D. L., Air pollution and health: bridging the gap from sources to health outcomes: conference summary. *Air Qual. Atmos. Health* **2012,** *5*, (1), 9-62.
10. Solomon, P. A.; Wexler, A. S.; Sioutas, C., Special Issue of Atmospheric Environment for Air Pollution and Health: Bridging the Gap from Sources-to-Health Outcomes Preface. *Atmospheric Environment* **2011,** *45*, (40), 7537-7539.
11. Marmur, A.; Mulholland, J.; Kim, E.; Hopke, P.; Sarnat, J.; Klein, M.; Tolbert, P.; Russell, A., Comparing results from several PM2.5 source-apportionment methods for use in a time-series health study. *Epidemiology* **2006,** *17*, (6), S200-S200.

12. Stolzel, M.; Laden, F.; Dockery, D. W.; Schwartz, J.; Kim, E.; Hopke, P. K.; Neas, L. M., Source apportionment of fine and coarse particulate matter and daily mortality in two US cities - A comparison of different methods. *Epidemiology* **2005,** *16*, (5), S95-S95.
13. Laden, F.; Neas, L. M.; Dockery, D. W.; Schwartz, J., Association of fine particulate matter from different sources with daily mortality in six US cities. *Environmental Health Perspectives* **2000,** *108*, (10), 941-947.
14. Mar, T. F.; Norris, G. A.; Koenig, J. Q.; Larson, T. V., Associations between air pollution and mortality in Phoenix, 1995-1997. *Environmental Health Perspectives* **2000,** *108*, (4), 347-353.
15. Lee, D.; Balachandran, S.; Pachon, J.; Shankaran, R.; Lee, S.; Mulholland, J. A.; Russell, A. G., Ensemble-Trained PM2.5 Source Apportionment Approach for Health Studies. *Environmental Science & Technology* **2009,** *43*, (18), 7023-7031.
16. Hansen, D. A.; Edgerton, E. S.; Hartsell, B. E.; Jansen, J. J.; Kandasamy, N.; Hidy, G. M.; Blanchard, C. L., The southeastern aerosol research and characterization study: Part 1-overview. *Journal of the Air & Waste Management Association* **2003,** *53*, (12), 1460-1471.
17. Lee, S.; Baumann, K.; Schauer, J. J.; Sheesley, R. J.; Naeher, L. P.; Meinardi, S.; Blake, D. R.; Edgerton, E. S.; Russell, A. G.; Clements, M., Gaseous and particulate emissions from prescribed burning in Georgia. *Environmental Science & Technology* **2005,** *39*, (23), 9049-9056.
18. Sandberg, D. V. O., R. D.; Peterson, J. L.; Core, J. , Wildland fire on ecosystems: effects of fire on air; General Technical Report In Service, U. F., Ed. USDA Forest Service: Rocky Mountain Research Station, Fort Collins, CO, 2002.
19. Hardy, C. C. H., S. M.; Mutch, R. E., Smoke Management Guide for Prescribed and Wildland Fire. In Hardy, C. C., et al., Ed. National Wildfire Coordination Group: Boise, ID, 2001.
20. USCensus http://www.census.gov/popest/data/metro/totals/2012/index.html (March 19, 2013),

CHAPTER 2: ENSEMBLE-TRAINED SOURCE APPORTIONMENT OF FINE PARTICULATE MATTER AND METHOD UNCERTAINTY ANALYSIS

Atmospheric Environment 2012, 61, (0), 387-394

Sivaraman Balachandran[*,a], Jorge E. Pachon[a,b], Yongtao Hu [a], Dongho Lee [c], James A. Mulholland [a] and Armistead G. Russell [a]

[a]Georgia Institute of Technology. School of Civil and Environmental Engineering. Atlanta, GA, 30332.

[b]Universidad de La Salle, Programa de Ingenieria Ambiental, Bogota, Colombia.

[c]Gyeongnam Province Institute of Health and Environment, Changwon, Gyeongnam 641-702, Korea

[*]Corresponding author: 311 Ferst Dr., Atlanta, GA, 30332; phone 206.250.6480, fax 404.894.8266; siv@gatech.edu.

2.1. Abstract

An ensemble-based approach is applied to better estimate source impacts on fine particulate matter ($PM_{2.5}$) and quantify uncertainties in various source apportionment (SA) methods. The approach combines source impacts from applications of four individual SA methods: three receptor-based models and one chemical transport model (CTM). Receptor models used are the chemical mass balance methods CMB-LGO (Chemical Mass Balance-Lipschitz global optimizer) and CMB-MM (molecular markers) as well as a factor analytic method, Positive Matrix Factorization (PMF). The CTM used is the Community Multiscale Air Quality (CMAQ) model. New source impact estimates and uncertainties in these estimates are calculated in a two-step process. First, an

ensemble average is calculated for each source category using results from applying the four individual SA methods. The root mean square error (RMSE) between each method with respect to the average is calculated for each source category; the RMSE is then taken to be the updated uncertainty for each individual SA method. Second, these new uncertainties are used to re-estimate ensemble source impacts and uncertainties. The approach is applied to data from daily $PM_{2.5}$ measurements at the Atlanta, GA, Jefferson Street (JST) site in July 2001 and January 2002. The procedure provides updated uncertainties for the individual SA methods that are calculated in a consistent way across methods. Overall, the ensemble has lower relative uncertainties as compared to the individual SA methods. Calculated CMB-LGO uncertainties tend to decrease from initial estimates, while PMF and CMB-MM uncertainties increase. Estimated CMAQ source impact uncertainties are comparable to other SA methods for gasoline vehicles and SOC but are larger than other methods for other sources. In addition to providing improved estimates of source impact uncertainties, the ensemble estimates do not have unrealistic extremes as compared to individual SA methods and avoids zero impact days.

2.2. Introduction

Controlling fine particulate matter poses unique challenges in developing strategies to improve public health and welfare (e.g., improved visibility). Unlike most other air pollutants, fine particulate matter (i.e. particles with an aerodynamic diameter less than 2.5μm, or $PM_{2.5}$) is comprised of a heterogeneous mix of chemical species, some of which are emitted directly from a variety of sources and others that are formed via atmospheric processes which convert gaseous species into condensed-phase compounds. The health concern over $PM_{2.5}$ has grown as associations have been found between $PM_{2.5}$ mass and health outcomes [*Dockery et al.*, 1993; *U.S.EPA*, 2009], and has led EPA to regulate $PM_{2.5}$ as a criteria pollutant as part of the US EPA's National Ambient Air Quality Standards (NAAQS).

Addressing $PM_{2.5}$ levels relies on quantifying source-to-air quality relationships, a process often termed source apportionment (SA). Historically, SA of $PM_{2.5}$ has been conducted using receptor-based modeling approaches such as chemical mass balance (CMB) modeling [*Watson et al.*, 1984] or factor analytic (FA) approaches such as positive matrix factorization (PMF) and UNMIX [*Henry*, 1997; 2003; *Paatero and Tapper*, 1994]. Receptor-based modeling approaches typically solve a mass balance equation that is used to reconstruct the mass of each measured species (Equation 1):

$$C_i = f_{ij} S_j + e_i \qquad \text{(Equation 1)}$$

where C_i, is the measured concentration of species i (μg species i m^{-3}), f_{ij} is the amount of species i emitted per unit amount from source j (μg of species i per μg of $PM_{2.5}$ emitted from j), S_j, is the impact of source j (μg $PM_{2.5}$ m^{-3}), and e_i is the error for the ith species between the measured concentration, C_i, and the calculated concentration, $f_{ij}S_j$. The most commonly used CMB approach, using more routinely available $PM_{2.5}$ observations (elemental and organic carbon: EC/OC, ionic and metal species), and EPA's CMB 8.2 software, is referred to here as CMB-regular, or **CMB-RG**, [*U.S.EPA*, 2004]. CMB has

also been applied using organic molecular markers, referred to here as **CMB-MM**, which allows identification of more primary organic sources than are typically quantified using CMB-RG [*Cass*, 1998; *Zheng et al.*, 2002; *Zheng et al.*, 2007]. Another CMB approach, called **CMB-LGO**, uses CMB and incorporates gaseous species measurements to constrain results [*Marmur et al.*, 2005]. Positive matrix factorization (PMF, version 3.0) [*Paatero and Tapper*, 1994; *Paatero et al.*, 2003; *U.S.EPA*, 2008] is a commonly used factor analytic approach. Receptor models can be readily applied for long time periods for which observational data is available.

Recently, chemical transport models (CTMs), such as the Community Mulitscale Air Quality (**CMAQ**), have been used to quantify source impacts on $PM_{2.5}$ [*Baek et al.*, 2005; *D W Byun et al.*, 1998; *Cohan et al.*, 2005; *Koo et al.*, 2009; *Marmur et al.*, 2006a; *Wang et al.*, 2009; *Yang et al.*, 2000; *Yarwood et al.*, 2007]. CTMs utilize emissions inventories and meteorological information to model transport and atmospheric chemistry in a three dimensional grid, and calculate source impacts over a large spatial domain and over time scales that may not be available from observations. Another advantage of using chemical transport models is that they can directly link and quantify the impacts of gaseous emission sources on particulate matter, a weakness of receptor-based approaches.

There have been several efforts to determine relationships between sources of $PM_{2.5}$ and health outcomes [*Laden et al.*, 2000; *Mar et al.*, 2000; *Marmur et al.*, 2006b; *Sarnat et al.*, 2008; *Stolzel et al.*, 2005; *Thurston et al.*, 2005], though with different results. In [*Thurston et al.*, 2005] traffic sources were not significantly associated with both CVD and non-accidental mortality, and, as the authors note, the factor analytic approaches were limited in their ability to separate gasoline and diesel fractions. Subsequently, Sarnat et al. (2008) compared epidemiologic model results using a factor analytic SA method, PMF, and an optimized CMB method, CMB-LGO (Lipschitz global optimizer) [*Marmur et al.*, 2005] to apportion sources for four years of speciated $PM_{2.5}$ data in Atlanta and using individual compounds that are viewed as reasonable tracers for

various sources. They found good agreement in RRs for CVD and respiratory outcomes between using PMF, CMG-LGO and tracers, implying different SA methods yield similar results when incorporated into epidemiologic models. However, a positive association was shown between biomass burning and CVD outcomes but not respiratory outcomes, whereas a number of previous studies showed positive associations with respiratory but not CVD outcomes [*Ito et al.*, 2006; *Mar et al.*, 2006]. As the authors note, several recent studies corroborate their findings, but there also may be confounding effects across source categories [*Barregard et al.*, 2006; *Barrett et al.*, 2006; *Ostro et al.*, 2007; *Sarnat et al.*, 2008]. Thurston et al. (2005), who incorporated nine factor analytic SA results into epidemiologic models for Phoenix, AZ and Washington D.C., found that variability in SA results across investigators/methods increased 95% confidence intervals (CI) of relative risk ratio (RR) per inter-quartile range by approximately 15%. However, contributions from similar factors sometimes differed by an order of magnitude, making intercomparisons between methods and their associations with health less clear [*Grahame and Hidy*, 2007].

Both receptor and emissions-based SA approaches have limitations that can affect their inclusion in health studies. Receptor-based SA results can vary substantially from method to method, and some approaches lead to bias and increased variability [*Barregard et al.*, 2006; *Barrett et al.*, 2006; *W E Christensen et al.*, 2006; *W F Christensen and Amemiya*, 2003; *Henry*, 1987; *Marmur et al.*, 2006b; *Ostro et al.*, 2007; *Sarnat et al.*, 2008]. With a limited number of factors identified or source profiles available, these methods assign mass from other sources to available factors/sources, leading to bias. Typical receptor model applications use source profiles, or identify factors, associated with only about 80% of the estimated $PM_{2.5}$ emissions [*Baek*, 2009].. The necessary resources required to apply CTMs over long periods inhibit their use, and they are subject to uncertainties in emission and meteorological inputs and model parameterizations.

A number of studies have evaluated SA results [*Brinkman et al.*, 2006; *W F Christensen and Gunst*, 2004; *D Lee et al.*, 2009; *S. Lee et al.*, 2008; *Marmur et al.*, 2006a; *Marmur et al.*, 2006b; *Rizzo and Scheff*, 2007; *Tauler et al.*, 2009]. Marmur et al. (2006a) showed that CMAQ had significantly less variability in fractional source impacts, than CMB-LGO, effectively precluding its use to provide source impact estimates that can be differentiated in terms of health impact associations in acute epidemiologic-based studies [*Marmur et al.*, 2006a; *Marmur et al.*, 2006b]. Christensen and Gunst (2004) evaluated the difference in CMB results for a simulated data set using four approaches to calculating source impacts and found that the weighted least squares approach performed better than the effective variance approach in most cases and was "slightly superior" in cases where the source profile variability is large. Christensen and Schauer [2008] showed that perturbations to species concentration uncertainties can lead to relatively large differences in PMF results. Lee and Russell [2007] found that source impact uncertainties in CMB-RG were more affected by source profile error contributed than measurement error.

Using an ensemble of air quality models has provided a means to evaluate air quality models [*Delle Monache et al.*, 2006; *Dennis et al.*, 2010; *Rao et al.*, 2011; *Wilczak et al.*, 2006]. Ensemble averaging has been limited to CTMs and has often focused on uncertainties in modeling ozone concentrations. However, Lee et al. [2009] showed that using an ensemble average of SA results from four receptor models and one CTM resulted in improved fitting statistics, reduced variability (compared to individual receptor models) and reduced the number of days with no impact from sources that are known to be present. In this work, we build on the work of Lee et al. [2009] by ensemble averaging results from four SA methods and assessing SA uncertainties in the ensemble results. This work updates the approach by Lee et al. [2009] in three ways: this method uses a two step process to calculate the ensemble, uncertainties are calculated using propagation of errors that includes covariance terms, and new estimates of

uncertainties are calculated for the individual SA methods that are used in the ensemble. A compelling reason to quantify uncertainties is that they can be incorporated into epidemiologic studies, which can ultimately lead to improving our understanding of the relationships between $PM_{2.5}$ sources and health outcomes. Further, they can be used to inform policy makers of the effectiveness of control measures.

2.3. Methods

2.3.1. Ensemble Source Apportionment

Ensemble averaging of SA results is conducted in two steps. In the first step source impact estimates and the uncertainties from the SA methods described above (see Appendix A for more on how uncertainties were calculated for each SA method) are averaged together. In the second step, the initial ensemble is used to re-estimate SA method uncertainties, which are then used as weights to calculate an updated average. Next, uncertainties for the updated ensemble source impact are re-calculated. In part, this can address concerns that the uncertainties provided by the traditional methods are biased. This process of re-estimating SA method uncertainties and re-updating the ensemble can be further iterated if desired.

The initial ensemble average, $\overline{S}_j(t_k)$, for source j at time t_k, is calculated using a weighted average:

$$\overline{S}_j(t_k) = \frac{\sum_{l=1}^{L} w_{jl}(t_k) \bullet S_{lj}(t_k)}{\sum_{l=1}^{L} w_{jl}(t_k)} \qquad \text{(Equation 2)}$$

where $w_{jl}(t_k)$ is the weight for source j from method l, and $S_{lj}(t_k)$ is the source impact for source j from method l. The weights (Equation 3) are based on each method's source

impact uncertainties and the value of N determines if and how much source impact uncertainties weight the average:

$$w_{jl} = \frac{1}{\sigma_{S_{lj}}^{N}} \qquad \text{(Equation 3)}$$

While there can be any choice for the weights, here we focus on using the inverse of the individual SA methods' uncertainty squared (i.e. N=2) and equal weighting (N=0). We also evaluate a mixed case, in which we use equal weighting for the initial ensemble and inverse square weighting for the updated ensemble. As discussed below, our focus is on the mixed case since we find that it provides the best results over both seasons. The initial and updated ensemble average uncertainty is calculated using weighted propagation of errors that includes covariance terms (Equation 4, see Appendix A for derivation [*Taylor and Kuyatt*, 1994]):

$$\sigma_{ensemble}^{2} = \left[\frac{\frac{1}{\sigma_{S_1}^{N}}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^{N}}} \quad \cdots \quad \frac{\frac{1}{\sigma_{S_L}^{N}}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^{N}}} \right] \begin{bmatrix} \sigma_{S_1S_1}^{2} & \cdots & \sigma_{S_1S_L}^{2} \\ \vdots & \ddots & \vdots \\ \sigma_{S_1S_L}^{2} & \cdots & \sigma_{S_LS_L}^{2} \end{bmatrix} \left[\frac{\frac{1}{\sigma_{S_1}^{N}}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^{N}}} \quad \cdots \quad \frac{\frac{1}{\sigma_{S_L}^{N}}}{\sum_{l=1}^{L}\frac{1}{\sigma_{S_l}^{N}}} \right]^{T} \qquad \text{(Equation 4)}$$

where S_l is the $PM_{2.5}$ impact of source *j* (source index not shown for clarity) from method *l*. The middle matrix term in the right hand side of equation 4 is a scaled uncertainty covariance matrix which takes into account the source impact uncertainties from the individual SA methods as well as the covariance of source impacts across methods; thus, each element $\sigma_{S_mS_n}^{2}$, where both *m* and *n* index the SA methods that range from 1 to L, is equal to (Equation 5):

$$\sigma_{S_mS_n}^{2} = \frac{\sigma_{S_m}\sigma_{S_n} * Cov(m,n)}{\sqrt{Cov(m,m) * Cov(n,n)}} \qquad \text{(Equation 5)}$$

Where σ_{S_m} and σ_{S_n} are source impact uncertainties from methods *m* and *n*, *Cov (m,n)* is the covariance of source impacts from methods *m* and *n* and *Cov (m,m)* and *Cov (n,n)* are the variances of source impacts from methods *m* and *n*, respectively.

The root mean square error (RMSE) for each method is determined by comparison against the ensemble average (Equations 6-7):

$$RMSE_{jl} = \sqrt{\frac{\sum_{k=1}^{K}\left(S_{jlk} - \overline{S}_{jk}\right)^2}{K}} \qquad \text{(Equation 6)}$$

where S_{jlk} is the source impact for source *j*, from method *l*, on day *k* and $\overline{S}_{jk}$ is the ensemble average for source *j* on day *k*, and *K* is the total number of days used in the ensemble. We then set the RMSE for each method as the updated uncertainty for each day (Equation 7):

$$\sigma'_{jlk} = RMSE_{jl} \qquad \text{(Equation 7)}$$

where σ'_{jlk} is the re-estimate of the source impact uncertainties for source *j*, from method *l*, on day *k*. A major consequence of using Equation 7 is that for a specific source the updated source impact uncertainties are the same for each day. We set new uncertainties in this way because regression analyses between SA method source impacts $\left(S_{jkl}\right)$ and their errors $\left(S_{jkl} - \overline{S}_{jk}\right)$ from the ensemble averages found little correlation. Next, new ensemble averages and uncertainties are calculated based on the weighted propagation of errors using the updated uncertainties for each SA method. The above procedure can be done using both the absolute and fractional source impacts and we focus here on results using absolute source impacts (both approaches were tested with similar results). Finally, we evaluate the individual SA methods and the ensemble by comparing the average source impact (by source category and season). To compare uncertainties between methods, we define the *overall method uncertainty*, $\left(\overline{\sigma}_{S_{jkl}}\right)$, as the root mean

square average of the daily updated source impact uncertainties [*Pachon et al.*, 2010] (Equation 8):

$$\overline{\sigma}_{S_{jl}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\sigma_{S_{jkl}}^2} \qquad \text{(Equation 8)}$$

As discussed previously, the base case was conducted using four SA techniques. SA impacts included previous results for CMB-MM [!!! INVALID CITATION !!!] and CMAQ [*Baek et al.*, 2005] were used as inputs into the ensemble, and we applied CMB-LGO and PMF, for 1999-2004 using speciated $PM_{2.5}$ data from the SEARCH Jefferson St (JST) monitoring site [*Edgerton et al.*, 2005; 2006; *Hansen et al.*, 2003]. The JST data set contains daily speciated concentrations of ions (sulfate, nitrate, and ammonium), organic carbon (OC), elemental carbon (EC), and trace metals. Data also includes speciated organic molecular markers for two one month periods (July 2001, January 2002) used as part of the CMB-MM work [*Zheng et al.*, 2007]. Further details on these methods can be found in Lee et al. [2009]and references therein.

Ensemble-averaging was conducted for July 2001 to represent summer, and January 2002 to represent winter (SA results from CMAQ and CMB-MM were available for these months). Source impacts from individual SA methods used in the ensemble were binned into nine source categories [*D Lee et al.*, 2009], and included five primary sources and four secondary sources. Primary sources include gasoline vehicles (GV), diesel vehicles (DV), dust (DUST), biomass burning (BURN), and coal combustion (COAL). Secondary categories include sulfate, nitrate, ammonium and other organic carbon (Other OC), which was treated as a surrogate for secondary organic carbon (SOC). CMAQ simulations tended to be biased high for sulfate, nitrate and ammonium in winter [*Dennis et al.*, 2010]. To account for this, we did not use CMAQ results for sulfate, nitrate and ammonium to calculate the ensemble impact in the equal weighting case. In addition, we performed a sensitivity analysis of the ensemble by replacing CMB-LGO with CMB-RG since this method is more widely used. We did not use both at

the same time because they are very highly correlated, relying on similar data. A second sensitivity analysis was conducted by not including CMAQ results as such results may not be as readily available or for as long of a period. However, CTM-based source impact files are becoming increasingly available [*Napelenok et al.*, 2006; *Yarwood et al.*, 2007].

2.4. Results

2.4.1. Ensemble Source Impacts and Uncertainties

Comparison of the four methods shows the relative biases of these methods across sources (Figures 2.1, A.1 and Tables A.1, A.2). For example, CMB-LGO has significantly higher SOC impacts, especially in winter. PMF tends to have higher source impacts for DV and BURN with lower impacts for SOC. CMB-MM has higher estimates of SOC in summer and higher estimates of GV in winter. CMAQ has higher DUST impacts in both seasons and higher BURN and COAL impacts in winter. The three receptor models, as expected, have very similar results for ionic species while CMAQ estimates are higher. Ensemble averaging provides daily source apportionment that results in no zero-impact days, reduced variability (Figure 2.1) and updated uncertainties to the daily source impacts in the five individual source apportionment methods. Ensemble averaging overcomes some limitations of the individual SA methods (e.g., when a particular method apportions $PM_{2.5}$ mass poorly for a given source, or does not resolve a set of sources for a given day). The ensemble avoids performing poorly for any particular source, a major limitation of traditional SA methods. The ensemble, for both seasons, has the lowest estimated relative uncertainty for all cases, when averaged across all sources (i.e. the average of the overall relative uncertainties for each source) (Table 2.1).

Table 2.1: Average overall relative uncertainties for equal weighting (EW), inverse square weighting (ISW), and a mixed case (MIX) using both EW and ISW for summer (July 2001) and winter (January 2002). The values shown are averaged over all source categories, excluding sulfate, nitrate and ammonium. Note: For MIX, the base SA methods have uncertainties based on EW.

	Summer				
	CMB-LGO	**PMF**	**CMB-MM**	**CMAQ**	**ENS.**
Initial	97%	38%	143%	222%	-
EW	81%	76%	80%	72%	45%
ISW	76%	69%	72%	93%	52%
MIX	-	-	-	-	45%
	Winter				
Winter	**CMB-LGO**	**PMF**	**CMB-MM**	**CMAQ**	**ENS.**
Initial	172%	53%	143%	388%	-
EW	219%	167%	202%	282%	59%
ISW	152%	124%	88%	409%	74%
MIX	-	-	-	-	62%

In summer, the ensemble, using inverse square weighting, has the lowest overall relative uncertainties (i.e. RMSE divided by average source impact) for BURN (49%), COAL (45%), and SOC (42%) and has the second lowest overall relative uncertainties for GV (77%), DV (36%) and DUST (62%). With equal weighting, the ensemble has the lowest overall relative uncertainties for DV (38%), DUST (48%) and BURN (35%), and has the second lowest uncertainties for GV (65%), COAL (39%) and SOC (40%) . With mixed weighting, the ensemble has the lowest overall relative uncertainties for DV (36%), DUST (55%), BURN (33%), and SOC (29%). CMB-LGO has the lowest overall relative uncertainty for GV and CMAQ for COAL.

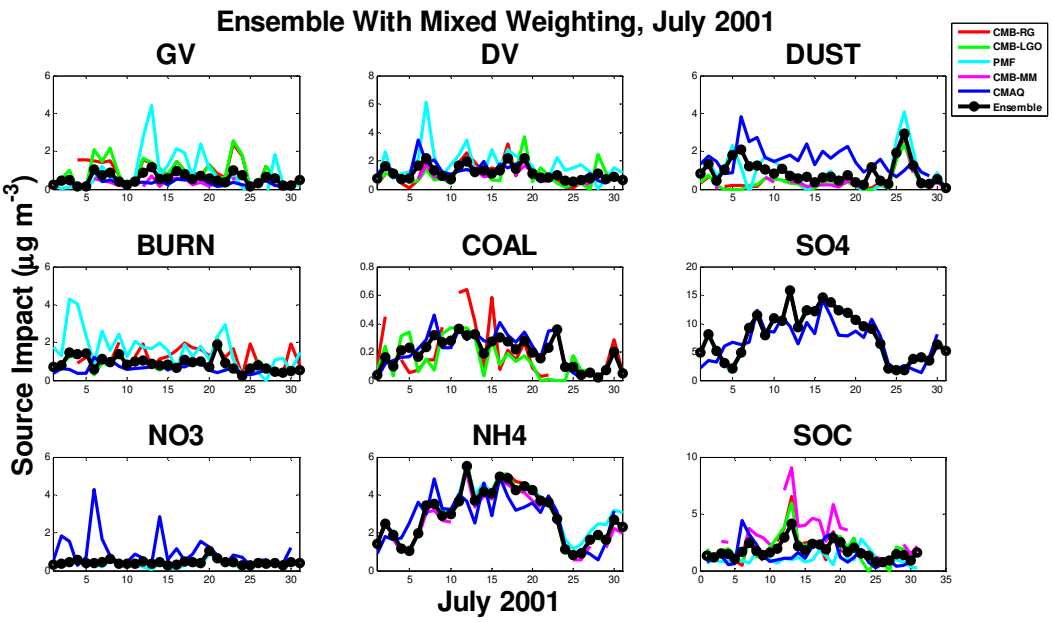


Figure 2.1: Ensemble with mixed weighting for July 2001. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figures A.7 and A.8).

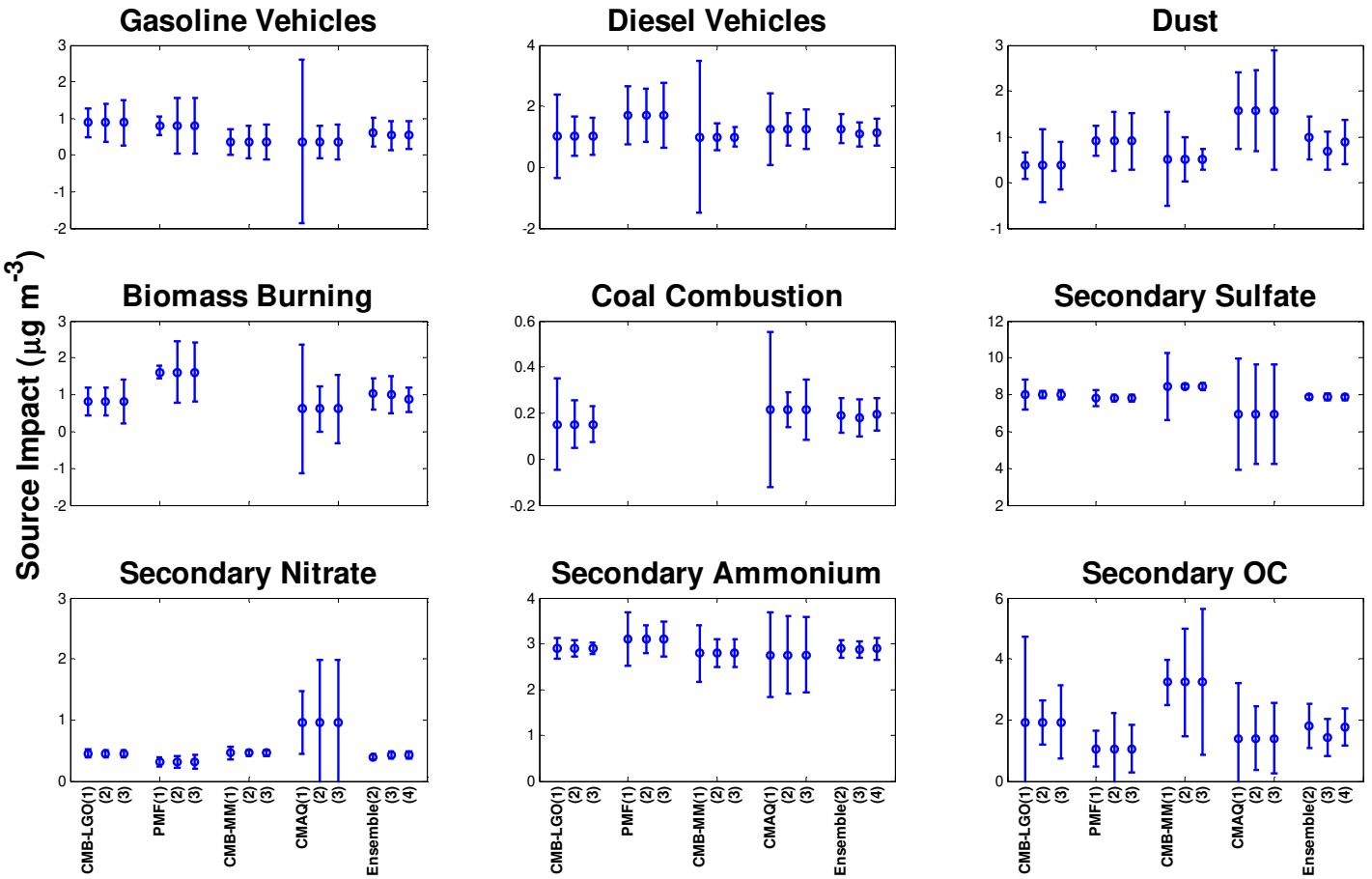


Figure 2.2: Average source impacts and overall uncertainties (Eq. 8) for the four SA methods and the ensemble (error bars represent one sigma) for July 2001. (Note the changes in scales). For each method, the first data point (1) shows source impact and initial uncertainties. The second point (2) shows source impact and updated uncertainties using equal weighting (EW*). The third point (3) shows source impact and updated uncertainties using inverse square uncertainty weighting (ISW). The ensemble has three data point for the EW and ISW and a mixed case (4), respectively. The mixed case uses EW for the initial ensemble and ISW for the updated ensemble. *NOTE: the EW case does not include CMAQ results for secondary sulfate, secondary nitrate and secondary ammonium.

The ensemble overall relative uncertainties in winter are generally higher than in summer (Figure A.1). Also, source impacts in winter are more varied between methods than in summer leading to greater RMSEs between the SA methods and the ensemble.

Choice of weighting does not result in large differences in the overall relative uncertainties in the ensemble averages for primary sources and SOC, though there can be large differences in the magnitude of source impacts (Figures 2.2 and A.1 and Tables A,1 and A.2). For example, the average GV source impact for the summer ensemble with inverse square weighting is 0.53 ±0.21 ($\mu g\ m^{-3}$) and is driven by CMB-MM which has an average impact of 0.36 $\mu g\ m^{-3}$ and an initial overall uncertainty of 0.35 $\mu g\ m^{-3}$. With equal weighting, the ensemble GV has an average of 0.62±0.40 $\mu g\ m^{-3}$. With mixed weighting, the average source impact for the ensemble is 0.55±0.38 $\mu g\ m^{-3}$. However, source impacts across the three cases are in general highly correlated, with low correlations only for SOC in the summer and DUST and SOC in winter (Table A.3).

The ensemble results, as compared to measured $PM_{2.5}$, reconstruct $PM_{2.5}$ mass between 75% and 110% over all cases (Table 2.2). A somewhat low bias may be expected because the typical range of identified sources in receptor models account for only about 80% of the inventoried $PM_{2.5}$ emissions [*Baek*, 2009]. In the work shown here, total mass from receptor models are biased slightly low in summer and slightly high in winter. There were no results for BURN or COAL in summer for CMB-MM, which may be why the predicted to observed $PM_{2.5}$ ratio is low for that method. CMAQ results for total $PM_{2.5}$ are biased low by about 20% in the summer and high by a factor of 2 in the winter. The ensemble, when using inverse square weighting, slightly under estimates $PM_{2.5}$ in both seasons. The ensemble results correlate more strongly with measured $PM_{2.5}$ in both seasons than other methods except PMF, regardless of weighting, having R^2 values between 0.84 and 0.96 (Table 2.2).

Table 2.2: Ratio of calculated to observed $PM_{2.5}$ for July 2001 and January 2002. Calculated $PM_{2.5}$ is defined as the sum of source impacts from the nine source categories. Observed $PM_{2.5}$ is from JST measurements, which use a gravimetric- based method similar to the Federal Reference Method (FRM). (*NOTE: Values are recalculated here because Lee et al, (2009) used a different protocol for calculating measurement uncertainties).

		CMB-LGO	PMF	CMB-MM	CMAQ	Ensemble (EW)	Ensemble (ISW)	Ensemble (MIX)	Le et al. (2009)*
July 2001	Avg. calc./obs. PM2.5	0.78	0.84	0.71	0.77	0.77	0.81	0.79	0.74
	St. Dev. calc./obs. PM2.5	0.10	0.13	0.09	0.29	0.10	0.10	0.10	0.08
	R^2	0.94	0.97	0.93	0.58	0.96	0.96	0.96	0.96
	Slope (Std. Error)	0.68 (0.03)	0.66 (0.02)	0.71 (0.04)	0.53 (0.09)	0.63 (0.02)	0.68 (0.03)	0.65 (0.03)	0.65 (0.02)
	Intercept (Std. Error)	1.77 (0.78)	2.95 (0.53)	-0.08 (1.10)	4.28 (2.05)	2.41 (0.57)	2.30 (0.62)	2.42 (0.61)	1.54 (0.59)
	Reduced Chi-Square	9	60	54	594	158	51	83	20
January 2002	Avg. calc./obs. PM2.5	0.97	1.02	1.01	2.05	0.98	1.13	1.10	0.99
	St. Dev. calc./obs. PM2.5	0.16	0.14	0.15	0.84	0.17	0.19	0.18	0.21
	R^2	0.83	0.88	0.84	0.34	0.84	0.85	0.84	0.76
	Slope (Std. Error)	0.74 (0.06)	0.90 (0.06)	0.76 (0.07)	1.21 (0.32)	0.65 (0.05)	0.78 (0.06)	0.76 (0.06)	0.65 (0.07)
	Intercept (Std. Error)	2.53 (0.83)	1.26 (0.84)	2.84 (0.91)	8.86 (4.28)	3.61 (0.70)	3.78 (0.81)	3.68 (0.83)	3.69 (0.92)
	Reduced Chi-Square	7	107	72	1661	212	71	124	805

Sensitivity analyses were performed by re-running the ensemble in two different ways. First, the ensemble was run using CMB-RG results in lieu of CMB-LGO. In both seasons, using mixed weighting, the ensemble results change little because CMB-RG and CMB-LGO results are highly correlated for all source categories (Figures A.7-A.8). Second, we also ran the ensemble without CMAQ results (i.e. ensemble with CMB-LGO, PMF and CMB-MM). In both seasons, changes are noted for GV, DUST, BURN and SOC since CMAQ is not always strongly correlated with receptor model results. Nevertheless, the changes are within the 67% confidence intervals of the full ensemble (Figures A.9 and A.10).

2.5. Discussion

The ensemble gives insight into how well each SA method works, and provides improved estimates of source impacts and improved estimates of source impact uncertainties by SA method. The ensemble also overcomes poor or unrealistic performance (e.g. high day to day variability or days where source impacts are zero for sources known to present). The ensemble allows for comparison of uncertainties by calculating them in a consistent manner and avoids the need for bootstrapping methods or poorly characterized uncertainties in source profiles. For example, CMB-MM and PMF have very different GV impacts in winter (2.42 and 1.07 $\mu g\ m^{-3}$) with low overall uncertainties when calculated using traditional methods (0.44 and 0.33 $\mu g\ m^{-3}$). Thus, while the average source impacts are very different, the overall relative uncertainties are similar, 26 and 31%, respectively, making it difficult to determine which model provides more accurate estimates. The ensemble reconciles this inconsistency, suggesting uncertainties in both PMF and CMB-MM are larger. In another study using CMB-MM, it was shown that GV source impact uncertainties are sensitive to the percentage of high emitting vehicles for weekend traffic; when smoker vehicles are assumed to be 5% of the GV fleet, GV source impact uncertainties on Saturdays decrease from 51% to 25% while for other days they are below 17% [*Lough and Schauer*, 2007]. Nevertheless, assumptions of fleet composition, vehicle type, driving conditions and driver behavior, all of which are significant sources of uncertainty, affect these types of analyses. Therefore, the uncertainties in Lough and Schauer (2007) should be viewed as tighter than achieved in general applications. In PMF, uncertainties are calculated by bootstrapping, which reflects how similar the bootstrapped data set's correlation structure is to the original data set, and may not reflect the actual factor contribution uncertainty.

Inverse square weighting leads to the ensemble being heavily influenced by a particular method (e.g. CMB-MM for GV), having initial uncertainties that are apparently biased low. This indicates that, given no other information, all methods should be weighted equally, (i.e., using equal weighting). When using mixed weighting, the base case SA methods are also treated equally, but the updated ensemble is weighted by the

new uncertainties to base case SA methods. We recommend mixed weighting because this incorporates the new uncertainties as weights to the updated ensemble average and performed well in the evaluation measures.

Ensemble averaging also allows uncertainties in CTM-based source impacts to be readily estimated. To our knowledge, this is the first work to estimate $PM_{2.5}$ source impact uncertainties in CMAQ. As new techniques are developed to estimate CTM uncertainties, ensemble averaging can provide a means to evaluate these estimates.

Another approach to evaluating the ensemble quantitatively is to compare our results with estimates of secondary organic carbon (SOC) impacts from other work (Table 2.3). Recently, Pachon et al. (2010) found that the regression method for estimating SOC had the lowest overall relative uncertainty, when compared to the EC Tracer Method, CMB-RG and PMF. They showed that both CMB-RG and PMF have high overall uncertainties that ranged from 47% to 56% for CMB-RG and 59% to 120% for PMF in summer and winter, respectively. The regression method estimated SOC to be 1.68 ± 0.14 μg m^{-3} and 0.80 ± 0.11 μg m^{-3} in July 2001 and January 2002, respectively. The ensemble estimates are comparable to the regression method's average impact and overall uncertainty for July 2001, but are higher for January 2002 (Table 2.3). The correlation of the ensemble-based SOC with the regression-based SOC is very encouraging since the regression method includes ozone concentrations, which are not used in any of the receptor models included in the ensemble. In addition, the regression method was more strongly correlated with measured water-soluble organic carbon (WSOC), which is hypothesized to be primarily from secondary reactions. This indicates a better fit with SOC than the other methods. WSOC is, likewise, not used in any of the ensemble methods. Further, it is interesting that the correlation between the ensemble SOC and the PMF SOC is very low ($R^2 = 0.01$ for July 2001).

Table 2.3: Secondary Organic Carbon (SOC). Results for July 2001 and January 2002 (μg m^{-3}). NOTE: Ensemble with MIX uses EW uncertainties in base case SA methods.

	Summer					Winter				
			Uncertainty (±σ)					Uncertainty (±σ)		
	Average SOC		Ens. with EW	Ens. with ISW	Ens. with MIX	Average SOC		Ens. with EW	Ens. with ISW	Ens. with MIX
CMB-LGO	1.93	±	0.72	1.19	-	2.43	±	1.21	2.00	-
PMF	1.06	±	1.17	0.77	-	0.69	±	1.05	0.54	-
CMB-MM	3.23	±	1.73	2.39	-	1.89	±	0.89	1.77	-
CMAQ	1.40	±	1.06	1.15	-	0.97	±	0.71	0.76	-
Ensemble with EW	1.81	±	0.73	-	-	1.45	±	0.68	-	-
Ensemble with ISW	1.42	±	-	0.60	-	0.90	±	-	0.48	-
Ensemble with MIX	1.76	±	-	-	0.60	1.31.	±	-	-	0.63

To evaluate the choice of weighting, we conducted York regression [*Saylor et al.*, 2006; *York et al.*, 2004] between the ensemble and the regression method SOC impacts and found that mixed weighting reproduced regression method results better that equal or inverse square weighting (R^2 = 0.82 and slope = 0.87 for summer 2001) (Figure A.9). A similar analysis was performed for January 2002 (Figure A.10). It has been suggested that CMB based methods overestimate SOC because primary OC from some sources are not considered [!!! INVALID CITATION !!!]. Updated emissions information that include improved estimates of primary OC emissions in the winter, which suggest that gasoline vehicles emit more OC in cold weather than is captured in current inventories, can significantly alter how OC is apportioned [*Donahue et al.*, 2009; *Subramanian et al.*, 2006]. It is expected that improved source profiles for CMB based methods and improved emissions processing in CTMs should lead to improved correlation of SOC estimates between the ensemble and the regression methods.

2.6. Conclusions

Commonly used methods to apportion sources of $PM_{2.5}$ have a number of issues that complicate their appropriate use. Results from the application of different SA methods can disagree substantially. Furthermore, calculation of source impact uncertainties varies from method to method, leading to very different uncertainty estimates and making inter-comparisons of source impacts and their associated

uncertainties difficult. Here we average an ensemble of SA methods, which includes two CMB methods, PMF and CMAQ to estimate updated source impacts and uncertainties. Three weighting cases, equal weighting, inverse square weighting and a mixed case are evaluated.

Ensemble averaging results in source impact estimates that have reduced variability compared to individual SA methods, avoids zero impact days and resolves source impacts for all days. The choice of weighting impacts ensemble-based average source impacts and uncertainties, but in general ensemble source impact uncertainties are lower or very comparable with individual SA method uncertainties. Over both seasons, mixed weighting in the ensemble reproduces $PM_{2.5}$ better than equal or inverse square weighting and agrees better with SOC estimates from a separate approach [*Pachon et al.*, 2010]. In the absence of any prior information which would indicate otherwise, mixed weighting should be used.

The ensemble method provides updated uncertainties for the individual SA methods that are calculated in a consistent way across methods. In general, CMB-LGO and CMB-MM overall uncertainties, averaged over primary sources and SOC, decrease in summer and increase in winter as compared to those found using the traditional approach for these methods. The ensemble method also provides a way to estimate source apportionment uncertainties in CMAQ. CMAQ source impact uncertainties are comparable to other SA methods for GV and SOC and larger than other methods for DV, DUST and BURN.

Acknowledgements

This publication was made possible in part by USEPA STAR grants R833626, R833866, R834799 and RD83479901. 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. We also acknowledge the Southern Company for their support and thank Eric Edgerton of ARA, Inc. for access to the SEARCH data. We also acknowledge the contribution of our colleagues at the Rollins School of Public Health at Emory University for helpful discussions.

2.7. References

Baek, J., Improving aerosol simulations: assessing and improving emissions and secondary organic aerosol formation in air quality modeling. Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA (2009).

Baek, J., Park, S.K., Hu, Y. and Russell, A.G., Source Apportionment of Fine Organic Aerosol Using CMAQ Tracers. The 4th Annual CMAS Models-3 Users' Conference, CMAS, Durham, NC (2005).

Barregard, L., Sallsten, G., Gustafson, P., Andersson, L., Johansson, L., Basu, S. and Stigendal, L., 2006. Experimental exposure to wood-smoke particles in healthy humans: Effects on markers of inflammation, coagulation, and lipid peroxidation, Inhalation Toxicology 18, pp. 845-853.

Barrett, E.G., Henson, R.D., Seilkop, S.K., McDonald, J.D. and Reed, M.D., 2006. Effects of hardwood smoke exposure on allergic airway inflammation in mice, Inhalation Toxicology 18, pp. 33-43.

Brinkman, G., Vance, G., Hannigan, M.P. and Milford, J.B., 2006. Use of synthetic data to evaluate positive matrix factorization as a source apportionment tool for PM2.5 exposure data, Environmental Science & Technology 40, pp. 1892-1901.

Byun, D.W., Young, J., Gipson, G., Godowitch, J., Binkowski, F., Roselle, S., Benjey, B., Pleim, J., Ching, J., Novak, J., Coats, C., Odman, T., Hanna, A., Alapaty, K., Mathur, R., McHenry, J., Shankar, U., Fine, S., Xiu, A.J., Jang, C. and Amer Meteorol Soc, A.M.S., Description of the models-3 community multiscale air quality (CMAQ) modeling system, (1998) 264-268 pp.

Cass, G.R., 1998. Organic molecular tracers for particulate air pollution sources, Trac-Trends in Analytical Chemistry 17, pp. 356-366.

Christensen, W.E., Schauer, J.J. and Lingwall, J.W., 2006. Iterated confirmatory factor analysis for pollution source apportionment, Environmetrics 17, pp. 663-681.

Christensen, W.F. and Amemiya, Y., 2003. Modeling and prediction for multivariate spatial factor analysis, Journal of Statistical Planning and Inference 115, pp. 543-564.

Christensen, W.F. and Gunst, R.F., 2004. Measurement error models in chemical mass balance analysis of air quality data, Atmospheric Environment 38, pp. 733-744.

Christensen, W.F. and Schauer, J.J., 2008. Impact of species uncertainty perturbation on the solution stability of positive matrix factorization of atmospheric particulate matter data, Environmental Science & Technology 42, pp. 6015-6021.

Cohan, D.S., Hakami, A., Hu, Y.T. and Russell, A.G., 2005. Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis, Environmental Science & Technology 39, pp. 6739-6748.

Delle Monache, L., Deng, X.X., Zhou, Y.M. and Stull, R., 2006. Ozone ensemble forecasts: 1. A new ensemble design, Journal of Geophysical Research-Atmospheres 111.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S.T., Scheffe, R., Schere, K., Steyn, D. and Venkatram, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems, Environmental Fluid Mechanics 10, pp. 471-489.

Dockery, D.W., Pope, C.A., Xu, X.P., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G. and Speizer, F.E., 1993. An Association between Air-Pollution and Mortality in 6 United-States Cities, New England Journal of Medicine 329, pp. 1753-1759.

Donahue, N.M., Robinson, A.L., Pandis, S.N., Kroll, J.H. and Worsnop, D.L., 2009. Rethinking organic aerosols: Semivolatile emissions and photochemical aging, Geochimica Et Cosmochimica Acta 73, pp. A299-A299.

Edgerton, E.S., Hartsell, B.E., Saylor, R.D., Jansen, J.J., Hansen, D.A. and Hidy, G.M., 2005. The southeastern aerosol research and characterization study: Part II. Filter-based measurements of fine and coarse particulate matter mass and composition, Journal of the Air & Waste Management Association 55, pp. 1527-1542.

Edgerton, E.S., Hartsell, B.E., Saylor, R.D., Jansen, J.J., Hansen, D.A. and Hidy, G.M., 2006. The Southeastern Aerosol Research and Characterization Study, part 3: Continuous measurements of fine particulate matter mass and composition, Journal of the Air & Waste Management Association 56, pp. 1325-1341.

Grahame, T. and Hidy, G.M., 2007. Pinnacles and pitfalls for source apportionment of potential health effects from airborne particle exposure, Inhalation Toxicology 19, pp. 727-744.

Hansen, D.A., Edgerton, E.S., Hartsell, B.E., Jansen, J.J., Kandasamy, N., Hidy, G.M. and Blanchard, C.L., 2003. The southeastern aerosol research and characterization study: Part 1-overview, Journal of the Air & Waste Management Association 53, pp. 1460-1471.

Henry, R.C., 1987. CURRENT FACTOR-ANALYSIS RECEPTOR MODELS ARE ILL-POSED, Atmospheric Environment 21, pp. 1815-1820.

Henry, R.C., 1997. History and fundamentals of multivariate air duality receptor models, Chemometrics Intell. Lab. Syst. 37, pp. 37-42.

Henry, R.C., 2003. Multivariate receptor modeling by N-dimensional edge detection, Chemometrics Intell. Lab. Syst. 65, pp. 179-189.

Ito, K., Christensen, W.F., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Larson, T.V., Neas, L., Hopke, P.K. and Thurston, G.D., 2006. PM source apportionment and health effects: 2. An investigation of intermethod variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC, Journal of Exposure Science and Environmental Epidemiology 16, pp. 300-310.

Koo, B., Wilson, G.M., Morris, R.E., Dunker, A.M. and Yarwood, G., 2009. Comparison of Source Apportionment and Sensitivity Analysis in a Particulate Matter Air Quality Model, Environmental Science & Technology 43, pp. 6669-6675.

Laden, F., Neas, L.M., Dockery, D.W. and Schwartz, J., 2000. Association of fine particulate matter from different sources with daily mortality in six US cities, Environmental Health Perspectives 108, pp. 941-947.

Lee, D., Balachandran, S., Pachon, J., Shankaran, R., Lee, S., Mulholland, J.A. and Russell, A.G., 2009. Ensemble-Trained PM2.5 Source Apportionment Approach for Health Studies, Environmental Science & Technology 43, pp. 7023-7031.

Lee, S., Liu, W., Wang, Y.H., Russell, A.G. and Edgerton, E.S., 2008. Source apportionment of PM2.5: Comparing PMF and CMB results for four ambient monitoring sites in the southeastern United States, Atmospheric Environment 42, pp. 4126-4137.

Lee, S. and Russell, A.G., 2007. Estimating uncertainties and uncertainty contributors of CMB PM2.5 source apportionment results, Atmospheric Environment 41, pp. 9616-9624.

Lough, G.C. and Schauer, J.J., 2007. Sensitivity of source apportionment of urban particulate matter to uncertainty in motor vehicle emissions profiles, Journal of the Air & Waste Management Association 57, pp. 1200-1213.

Mar, T.F., Ito, K., Koenig, J.Q., Larson, T.V., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Neas, L., Stolzel, M., Paatero, P., Hopke, P.K. and Thurston,

G.D., 2006. PM source apportionment and health effects. 3. Investigation of inter-method variations in associations between estimated source contributions Of PM2.5 and daily mortality in Phoenix, AZ, Journal of Exposure Science and Environmental Epidemiology 16, pp. 311-320.

Mar, T.F., Norris, G.A., Koenig, J.Q. and Larson, T.V., 2000. Associations between air pollution and mortality in Phoenix, 1995-1997, Environmental Health Perspectives 108, pp. 347-353.

Marmur, A., Mulholland, J., Kim, E., Hopke, P., Sarnat, J., Klein, M., Tolbert, P. and Russell, A., 2006a. Comparing results from several PM2.5 source-apportionment methods for use in a time-series health study, Epidemiology 17, pp. S200-S200.

Marmur, A., Park, S.K., Mulholland, J.A., Tolbert, P.E. and Russell, A.G., 2006b. Source apportionment of PM2.5 in the southeastern United States using receptor and emissions-based models: Conceptual differences and implications for time-series health studies., Atmospheric Environment 40, pp. 2533-2551.

Marmur, A., Unal, A., Mulholland, J.A. and Russell, A.G., 2005. Optimization-based source apportionment of PM2.5 incorporating gas-to-particle ratios, Environmental Science & Technology 39, pp. 3245-3254.

Napelenok, S.L., Cohan, D.S., Hu, Y.T. and Russell, A.G., 2006. Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM), Atmospheric Environment 40, pp. 6112-6121.

Ostro, B., Feng, W.Y., Broadwin, R., Green, S. and Lipsett, M., 2007. The effects of components of fine particulate air pollution on mortality in California: Results from CALFINE, Environmental Health Perspectives 115, pp. 13-19.

Paatero, P., Hopke, P.K., Hoppenstock, J. and Eberly, S.I., 2003. Advanced factor analysis of spatial distributions of PM2.5 in the eastern United States, Environmental Science & Technology 37, pp. 2460-2476.

Paatero, P. and Tapper, U., 1994. Positive matrix factorization - a nonnegative factor model with optimal utilization of error-estimates of data values, Environmetrics 5, pp. 111-126.

Pachon, J.E., Balachandran, S., Hu, Y.T., Weber, R.J., Mulholland, J.A. and Russell, A.G., 2010. Comparison of SOC estimates and uncertainties from aerosol chemical composition and gas phase data in Atlanta, Atmospheric Environment 44, pp. 3907-3914.

Rao, S.T., Galmarini, S. and Puckett, K., 2011. Air Quality Model Evaluation International Initiative (AQMEII) Advancing the State of the Science in Regional Photochemical Modeling and Its Applications, Bull. Amer. Meteorol. Soc. 92, pp. 23-30.

Rizzo, M.J. and Scheff, P.A., 2007. Fine particulate source apportionment using data from the USEPA speciation trends network in Chicago, Illinois: Comparison of two source apportionment models, Atmospheric Environment 41, pp. 6276-6288.

Sarnat, J.A., Marmur, A., Klein, M., Kim, E., Russell, A.G., Sarnat, S.E., Mulholland, J.A., Hopke, P.K. and Tolbert, P.E., 2008. Fine particle sources and cardiorespiratory morbidity: An application of chemical mass balance and factor analytical source-apportionment methods, Environmental Health Perspectives 116, pp. 459-466.

Saylor, R.D., Edgerton, E.S. and Hartsell, B.E., 2006. Linear regression techniques for use in the EC tracer method of secondary organic aerosol estimation, Atmospheric Environment 40, pp. 7546-7556.

Stolzel, M., Laden, F., Dockery, D.W., Schwartz, J., Kim, E., Hopke, P.K. and Neas, L.M., 2005. Source apportionment of fine and coarse particulate matter and daily

mortality in two US cities - A comparison of different methods, Epidemiology 16, pp. S95-S95.
Subramanian, R., Donahue, N.M., Bernardo-Bricker, A., Rogge, W.F. and Robinson, A.L., 2006. Contribution of motor vehicle emissions to organic carbon and fine particle mass in Pittsburgh, Pennsylvania: Effects of varying source profiles and seasonal trends in ambient marker concentrations, Atmospheric Environment 40, pp. 8002-8019.
Tauler, R., Viana, M., Querol, X., Alastuey, A., Flight, R.M., Wentzell, P.D. and Hopke, P.K., 2009. Comparison of the results obtained by four receptor modelling methods in aerosol source apportionment studies, Atmospheric Environment 43, pp. 3989-3997.
Taylor, B.N. and Kuyatt, C.E., Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. National Institute for Standards and Technology, Gaithersburg, MD (1994).
Thurston, G.D., Ito, K., Mar, T., Christensen, W.F., Eatough, D.J., Henry, R.C., Kim, E., Laden, F., Lall, R., Larson, T.V., Liu, H., Neas, L., Pinto, J., Stolzel, M., Suh, H. and Hopke, P.K., 2005. Workgroup report: Workshop on source apportionment of particulate matter health effects - Intercomparison of results and implications, Environmental Health Perspectives 113, pp. 1768-1774.
U.S.EPA, EPA-CMB8.2 User's Manual. In: U.S.E.P.A. Office of Air Quality & Standards, Research Triangle Park, NC., Editor, Publication No. EPA-452/R-04-011 (2004).
U.S.EPA, EPA-PMF 3.0 User's Manual. In: U.S.E.P.A. National Exposure Research Laboratory, Research Triangle Park, NC., Editor (2008).
U.S.EPA, Integrated Science Assessment for Particulate Matter (Final Report). . (2009).
Wang, Z.S., Chien, C.J. and Tonnesen, G.S., 2009. Development of a tagged species source apportionment algorithm to characterize three-dimensional transport and transformation of precursors and secondary pollutants, Journal of Geophysical Research-Atmospheres 114.
Watson, J.G., Cooper, J.A. and Huntzicker, J.J., 1984. The effective variance weighting for least-squares calculations applied to the mass balance receptor model, Atmospheric Environment 18, pp. 1347-1355.
Wilczak, J., McKeen, S., Djalalova, I., Grell, G., Peckham, S., Gong, W., Bouchet, V., Moffet, R., McHenry, J., McQueen, J., Lee, P., Tang, Y. and Carmichael, G.R., 2006. Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004, Journal of Geophysical Research-Atmospheres 111, p. 15.
Yang, Y.J., Wilkinson, J.G., Odman, M.T. and Russell, A.G., Ozone sensitivity and uncertainty analysis using DDM-3D in a photochemical air quality model. In: S.E. Gryning and E. Batchvarova, Editors, Air Pollution Modeling and Its Application Xiii (2000), pp. 183-194.
Yarwood, G., Morris, R.E. and Wilson, G.M., Particulate matter source apportionment technology (PSAT) in the CAMx photochemical grid model. In: C. Borrego and A.L. Norman, Editors, Air Pollution Modeling and Its Applications XVII (2007), pp. 478-492.
York, D., Evensen, N.M., Martinez, M.L. and Delgado, J.D., 2004. Unified equations for the slope, intercept, and standard errors of the best straight line, Am. J. Phys. 72, pp. 367-375.
Zheng, M., Cass, G.R., Ke, L., Wang, F., Schauer, J.J., Edgerton, E.S. and Russell, A.G., 2007. Source apportionment of daily fine particulate matter at Jefferson street,

Atlanta, GA, during summer and winter, Journal of the Air & Waste Management Association 57, pp. 228-242.
Zheng, M., Cass, G.R., Schauer, J.J. and Edgerton, E.S., 2002. Source apportionment of PM2.5 in the southeastern United States using solvent-extractable organic compounds as tracers, Environmental Science & Technology 36, pp. 2361-2371.

APPENDIX A: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1. Uncertainty Calculations

Uncertainties in receptor models are typically calculated by techniques specific to each method. In CMB-RG and CMB-MM, uncertainties are traditionally calculated using the effective variance approach, which incorporates both measurement and source profile uncertainties, $\sigma_{f_{ij}}$, [*Watson et al.*, 1984]:

$$\sigma_{CMB-RG,CMB-MM} = \left(\sum_{i=1}^{n} \frac{f_{ij}^2}{\sigma_{C_i}^2 + \sum_{j=1}^{m} \sigma_{f_{ij}}^2 S_j^2} \right)^{-1/2} \quad \text{(Equation A1)}$$

In CMB-LGO, uncertainty is calculated accounting for uncertainties in measurements (CITE CMB-LGO Lee and Russell 2007); this is the same as setting the $\sigma_{f_{ij}}$ term equal to zero in the CM-RG effective variance calculation of uncertainty (Equation 3):

$$\sigma_{CMB-LGO} = \left(\sum_{i=1}^{n} \frac{f_{ij}^2}{\sigma_{C_i}^2} \right)^{-1/2} \quad \text{(Equation A2)}$$

There is no commonly accepted method for calculating uncertainties in factor analytic approaches such as PMF. Here, we utilize a bootstrapping method to estimate source impact uncertainty, in a manner similar to previous work [*D Lee et al.*, 2009; *Pachon et al.*, 2010]. The uncertainty for each daily factor contribution, g_{jk}, was taken to be the factor contribution times the standard deviation of the $PM_{2.5}$ fitting species ($\sigma_{PM_{2.5}j}$) from the bootstrapping results (Equation 4):

$$\sigma_{g_{jk}} = g_{jk}\sigma_{PM_{2.5}j} \quad \text{(Equation A3)}$$

There is no generally accepted method for determining uncertainties in CTM PM source apportionment results. In this study, we use the method employed by Lee et al. (2009) to develop initial estimates of CMAQ uncertainties (Equation 5):

$$\sigma_{sj} = \left[\sigma^2_{S_{j,CMB-RG}} + \frac{(S_{j,CMAQ} - S_{j,CMB-RG})^2}{4} \right]^{1/2} \quad \text{(Equation A4)}$$

Table A.1: Ensemble average source impacts and overall uncertainties (as defined by eq12), for July 2001. For each SA method, overall uncertainties are shown for (1) the initial application of the SA method, (2) the ensemble with equal weighting and (3) with inverse square weighting. For the ensemble, results show average source impacts and overall uncertainties for with (2) equal weighting, (3) with inverse square weighting and (4) mixed weighting. In mixed weighting, the initial ensemble uses equal weighting and the updated ensemble uses inverse square weighting.

	CMB-LGO					PMF					CMB-MM					CMAQ				
Summer	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)
GV	0.88	±	0.41	0.53	0.62	0.80	±	0.26	0.75	0.77	0.36	±	0.35	0.45	0.49	0.36	±	2.24	0.44	0.47
DV	1.03	±	1.37	0.65	0.61	1.71	±	0.95	0.87	1.06	0.99	±	2.50	0.44	0.32	1.25	±	1.18	0.52	0.64
DUST	0.37	±	0.29	0.79	0.52	0.90	±	0.33	0.65	0.62	0.51	±	1.02	0.49	0.22	1.57	±	0.83	0.89	1.30
BURN	0.81	±	0.38	0.38	0.60	1.61	±	0.17	0.84	0.79	-	±	-	-	-	0.61	±	1.74	0.62	0.93
COAL	0.15	±	0.20	0.10	0.08	-	±	-	-	-	-	±	-	-	-	0.22	±	0.34	0.08	0.13
SO4	8.00	±	0.83	0.18	0.27	7.82	±	0.42	0.18	0.16	8.45	±	1.84	0.11	0.17	6.95	±	3.01	2.71	2.71
NO3	0.45	±	0.07	0.06	0.05	0.32	±	0.08	0.10	0.11	0.46	±	0.10	0.05	0.05	0.97	±	0.51	1.01	1.02
NH4	2.90	±	0.23	0.18	0.12	3.10	±	0.58	0.30	0.39	2.79	±	0.62	0.31	0.31	2.76	±	0.92	0.86	0.83
SOC	1.93	±	2.80	0.72	1.19	1.06	±	0.59	1.17	0.77	3.23	±	0.75	1.76	2.39	1.40	±	1.82	1.06	1.15

	Ensemble with Equal Weighting			Ensemble with Inverse Square Weighting			Ensemble with Mixed Weighting		
Summer	Avg. Source Impact		Unc. (2)	Avg. Source Impact		Unc. (3)	Avg. Source Impact		Unc. (4)
GV	0.62	±	0.40	0.53	±	0.41	0.55	±	0.38
DV	1.25	±	0.47	1.08	±	0.39	1.15	±	0.42
DUST	0.98	±	0.48	0.70	±	0.42	0.88	±	0.49
BURN	1.02	±	0.42	1.01	±	0.50	0.87	±	0.33
COAL	0.19	±	0.07	0.18	±	0.08	0.20	±	0.07
SO4	7.89	±	0.12	7.85	±	0.19	7.87	±	0.15
NO3	0.40	±	0.05	0.43	±	0.06	0.43	±	0.06
NH4	2.89	±	0.19	2.88	±	0.17	2.89	±	0.23
SOC	1.81	±	0.73	1.42	±	0.60	1.76	±	0.60

Table A.2: Ensemble average source impacts and overall uncertainties (as defined by eq12), for January 2002. For each SA method, overall uncertainties are shown for (1) the initial application of the SA method, (2) the ensemble with equal weighting and (3) with inverse square weighting. For the ensemble, results show average source impacts and overall uncertainties for with (2) equal weighting, (3) with inverse square weighting and (4) mixed weighting. In mixed weighting, the initial ensemble uses equal weighting and the updated ensemble uses inverse square weighting.

	CMB-LGO					PMF					CMB-MM					CMAQ				
Summer	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)	Avg. Source Impact		Unc. (1)	Unc. (2)	Unc. (3)
GV	1.55	±	0.71	0.56	0.62	1.07	±	0.33	0.78	0.85	2.42	±	0.44	1.33	1.35	0.33	±	2.97	1.21	1.23
DV	1.18	±	0.80	0.40	0.45	1.24	±	0.83	0.89	0.78	0.88	±	2.17	0.61	0.63	1.48	±	1.44	0.60	0.82
DUST	0.10	±	0.40	0.74	0.38	0.31	±	0.11	0.55	0.25	0.06	±	0.35	0.67	0.14	1.84	±	1.00	1.22	1.74
BURN	1.27	±	0.66	2.58	1.82	4.82	±	0.51	1.84	2.99	2.86	±	0.85	1.50	1.78	5.65	±	3.33	2.55	4.00
COAL	0.05	±	0.23	0.15	0.13	-	±	-	-	-	-	±	-	-	-	0.23	±	0.36	0.09	0.16
SO4	2.21	±	0.25	0.22	0.28	1.61	±	0.34	0.59	0.59	2.22	±	0.46	0.24	0.29	4.95	±	2.03	3.73	3.76
NO3	1.68	±	0.24	0.10	0.11	1.47	±	0.34	0.15	0.17	1.65	±	0.38	0.06	0.08	5.22	±	2.37	4.78	4.76
NH4	1.28	±	0.10	0.02	0.03	1.32	±	0.32	0.09	0.09	1.24	±	0.26	0.10	0.11	3.33	±	1.29	2.53	2.52
SOC	2.43	±	0.99	1.21	2.00	0.69	±	0.59	1.05	0.54	1.89	±	0.51	0.89	1.77	0.97	±	2.40	0.71	0.76

	Ensemble with Equal Weighting			Ensemble with Inverse Square Weighting			Ensemble with Mixed Weighting		
Summer	Avg. Source Impact		Unc. (2)	Avg. Source Impact		Unc. (3)	Avg. Source Impact		Unc. (4)
GV	1.31	±	0.71	1.35	±	0.62	1.36	±	0.57
DV	1.26	±	0.48	1.19	±	0.49	1.21	±	0.43
DUST	0.78	±	0.74	0.31	±	0.42	0.47	±	0.65
BURN	3.63	±	1.78	2.70	±	1.81	3.58	±	1.60
COAL	0.18	±	0.10	0.17	±	0.13	0.20	±	0.09
SO4	2.04	±	0.24	2.16	±	0.30	2.17	±	0.24
NO3	1.60	±	0.08	1.63	±	0.10	1.63	±	0.08
NH4	1.28	±	0.05	1.28	±	0.03	1.28	±	0.03
SOC	1.45	±	0.68	0.90	±	0.48	1.31	±	0.63

Table A.3: Correlations (R^2) between ensemble results using equal weighting (EW), inverse square weighting (ISW), and the mixed case using EW in the initial average, followed by ISW in the updated average, for summer (July 2001) and winter (January 2002). Values in bold show R^2< 0.70.

		GV	DV	DUST	BURN	COAL	SO4	NO3	NH4	SOC
R^2 (EW vs. ISW)	Summer	0.96	0.87	0.77	0.97	0.90	1.00	0.98	0.99	0.76
	Winter	0.93	0.96	**0.38**	0.89	0.98	0.97	1.00	1.00	**0.67**
R^2 (EW vs. Mixed)	Summer	0.93	0.96	0.93	0.83	0.96	1.00	0.98	1.00	0.94
	Winter	0.92	0.95	0.72	0.95	0.92	0.96	1.00	1.00	0.97
R^2 (ISW vs. Mixed)	Summer	0.99	0.94	0.93	0.93	0.75	1.00	1.00	1.00	**0.67**
	Winter	1.00	0.99	0.86	0.98	0.82	1.00	1.00	1.00	**0.65**

Table A.4: Average correlation, R^2, (range), between methods including ensemble using equal weighting (EW), inverse square weighting (ISW), and a mixed case (MIX) using both EW and ISW for GV, DV, DUST, BURN, COAL and SOC. Values to the left of the diagonal and right are for July, 2001 and January, 2002, respectively. Values in bold show average R^2> 0.50.

	CMB-LGO	PMF	CMB-MM	CMAQ	ENS (EW)	ENS (ISW)	ENS (MIX)
CMB-LGO		0.30 (0.07 – 0.63)	0.47 (0.24 – 0.79)	0.23 (0.16 – 0.39)	**0.53 (0.12 – 0.89)**	**0.52 (0.00 – 0.94)**	**0.56 (0.06 – 0.94)**
PMF	0.23 (0.02 – 0.70)		0.27 (0.02 – 0.56)	0.12 (0.00 – 0.38)	0.50 (0.06 – 0.84)	**0.60 (0.46 – 0.80)**	**0.51 (0.03 – 0.86)**
CMB-MM	**0.64 (0.21 – 0.95)**	0.35 (0.00 – 0.87)		0.09 (0.04 - 0.20)	**0.54 (0.17 – 0.74)**	**0.59 (0.21 - 0.88)**	**0.61 (0.43 – 0.80)**
CMAQ	0.07 (0.00 - 0.27)	0.04 (0.00 - 0.17)	0.09 (0.01 – 0.18)		0.50 (0.13 - 0.77)	0.21 (0.00 - 0.57)	0.44 (0.09 – 0.95)
ENS (EW)	0.48 (0.32 – 0.64)	**0.56 (0.03 - 0.89)**	**0.70 (0.48 – 0.89)**	0.27 (0.03 - 0.65)		**0.78 (0.42 - 0.98)**	**0.87 (0.50 - 0.97)**
ENS (ISW)	**0.57 (0.29 – 0.71)**	**0.58 (0.36 – 0.81)**	**0.72 (0.48 – 1.00)**	0.20 (0.00 – 0.33)	**0.87 (0.76 – 0.97)**		**0.88 (0.65 - 0.99)**
ENS (MIX)	**0.61 (0.29 - 0.88)**	0.44 (0.01 - 0.61)	**0.77 (0.47 - 0.95)**	0.30 (0.00 - 0.83)	**0.92 (0.83 - 0.96)**	**0.87 (0.67 - 0.99)**	

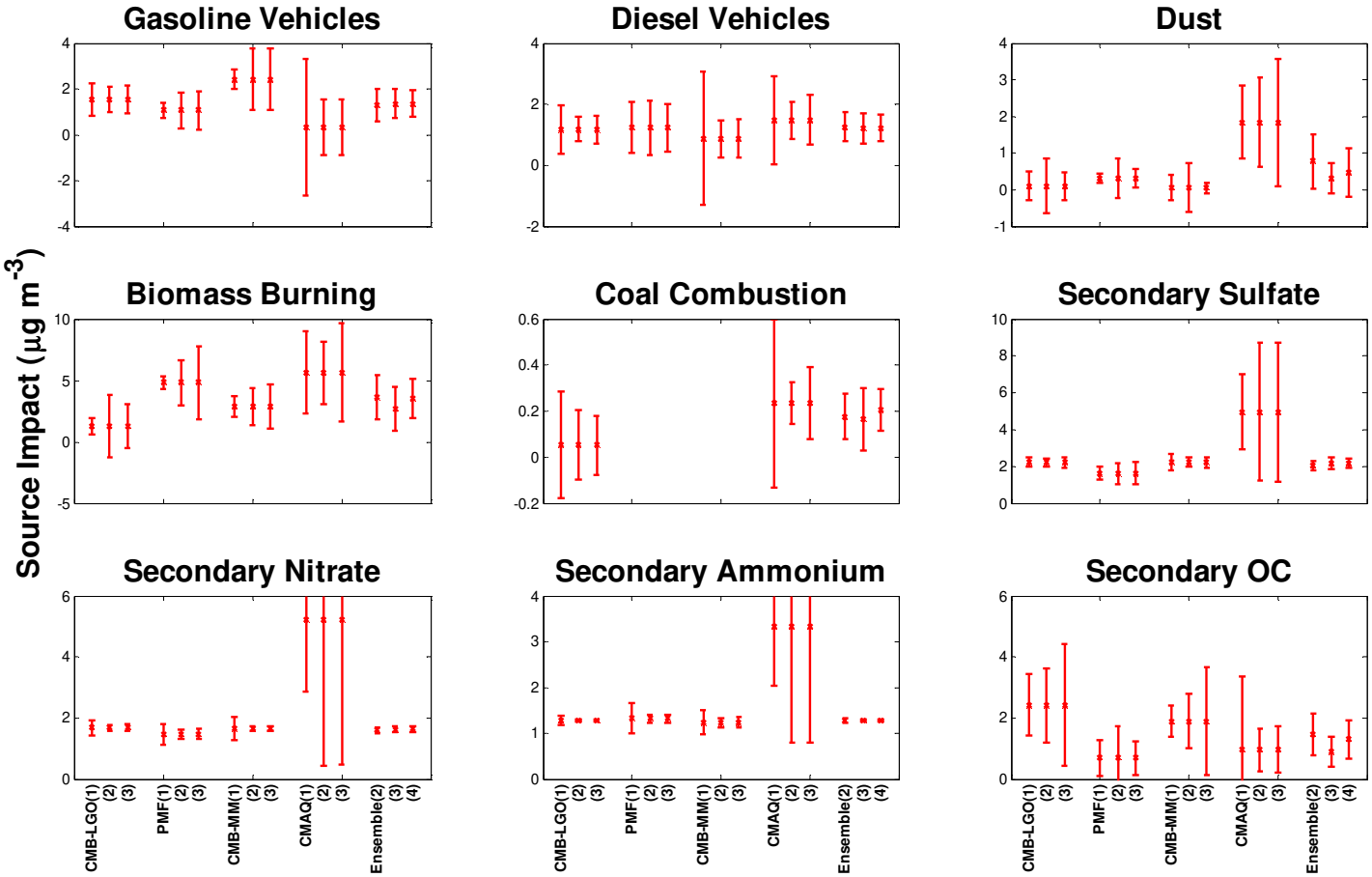


Figure A.1: Average source impacts and overall uncertainties for the five SA methods and the ensemble (error bars represent one sigma) for January 2002. For each method, the first data point (1) shows source impact and initial uncertainties. The second point (2) shows source impact and updated uncertainties using equal weighting (EW*). The third point (3) shows source impact and updated uncertainties using inverse square uncertainty weighting (ISW). The ensemble has three data point for the EW and ISW and a mixed case (4), respectively. The mixed case used EW the initial ensemble and ISW for the updated ensemble. *NOTE: the EW case does not include CMAQ results for Secondary Sulfate, Secondary Nitrate and Secondary Ammonium.

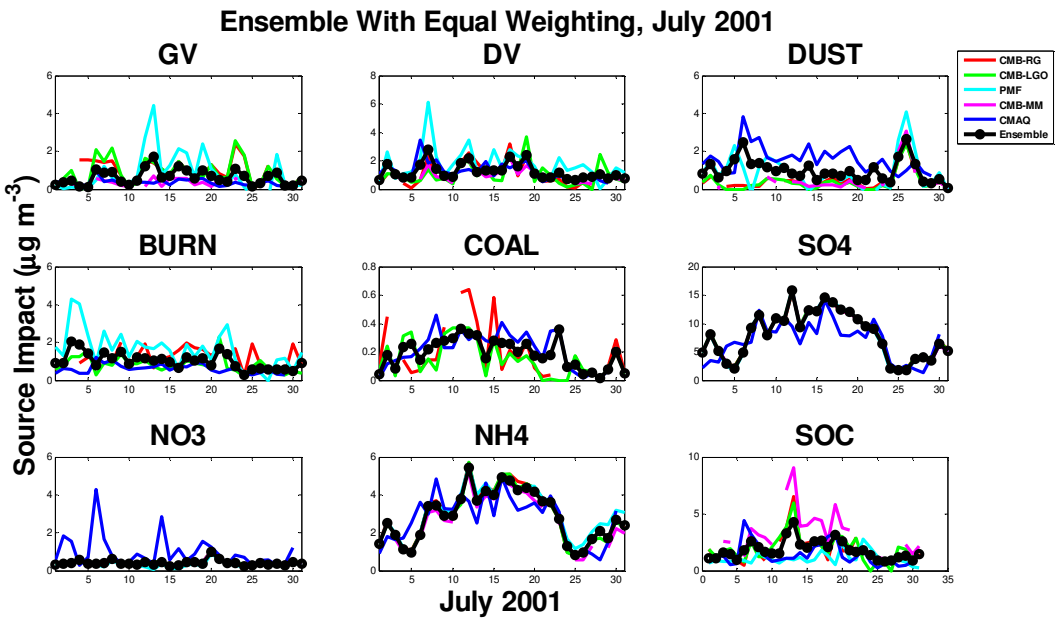


Figure A.2: Ensemble with equal weighting for July 2001. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figures A.7 and A.8).

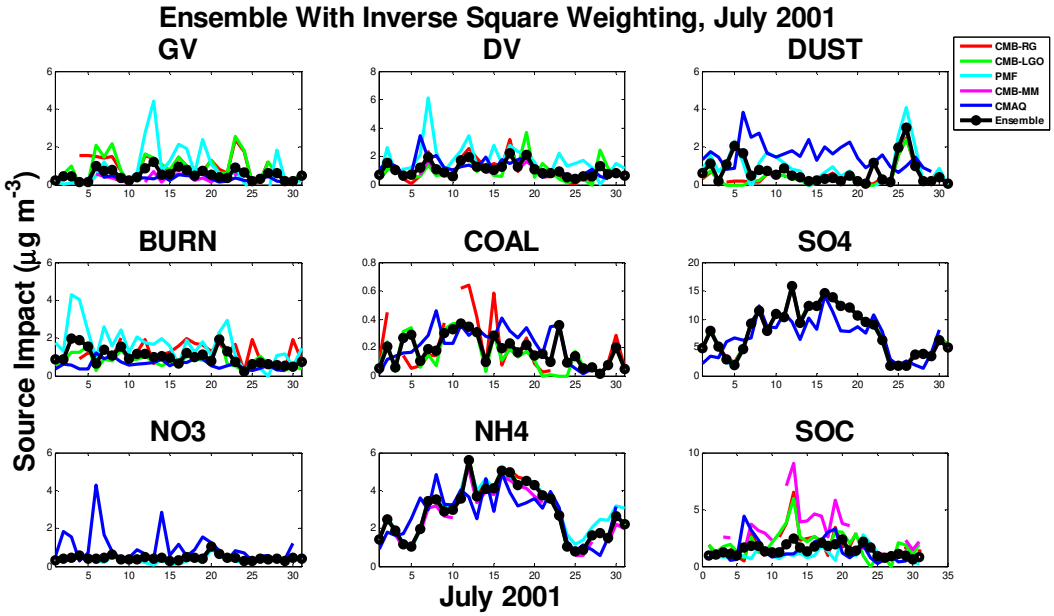


Figure A.3: Ensemble with inverse square weighting for July 2001. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figures A.7 and A.8).

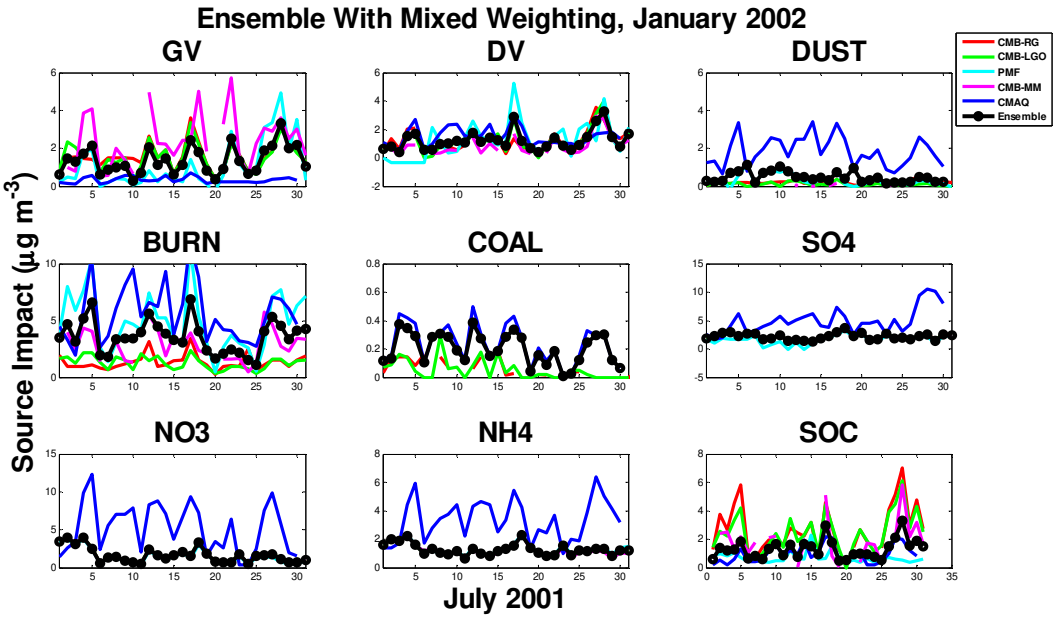


Figure A.4: Ensemble with mixed weighting for January 2002. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figures A.7 and A.8).

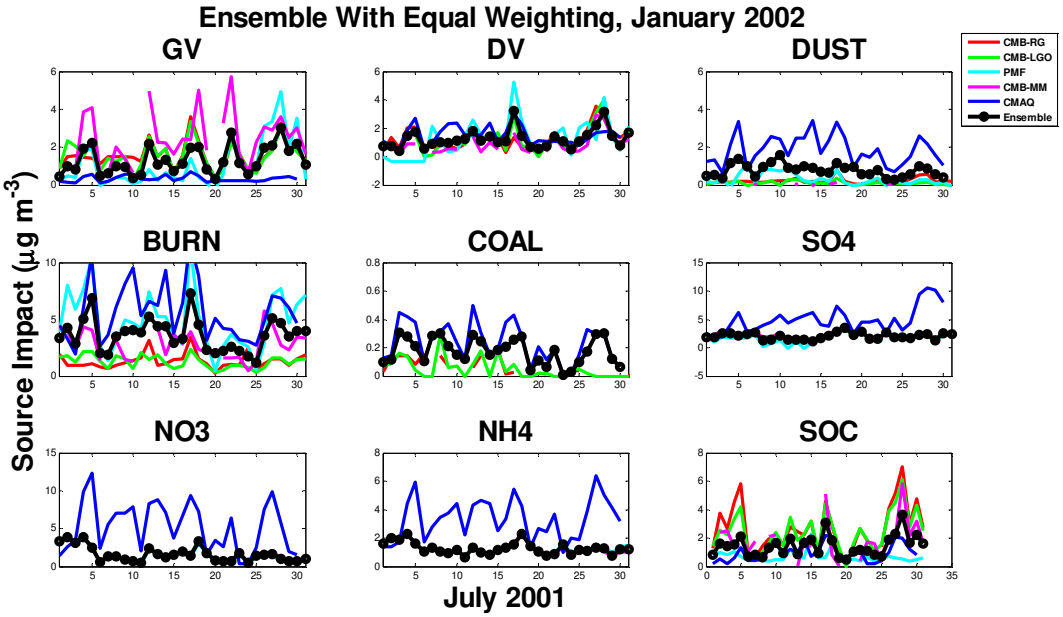


Figure A.5: Ensemble with equal weighting for January 2002. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figures A.7 and A.8).

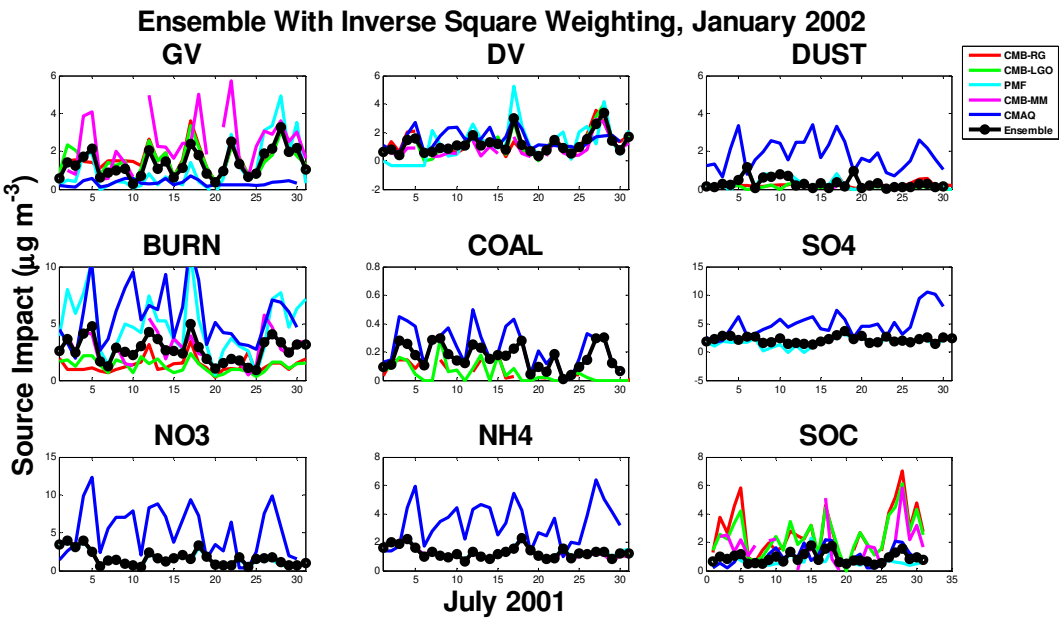


Figure A.6: Ensemble with inverse square weighting for January 2002. NOTE: CMB-RG results shown here are not included in the base ensemble, but are used in the sensitivity analysis (Figures A.7 and A.8).

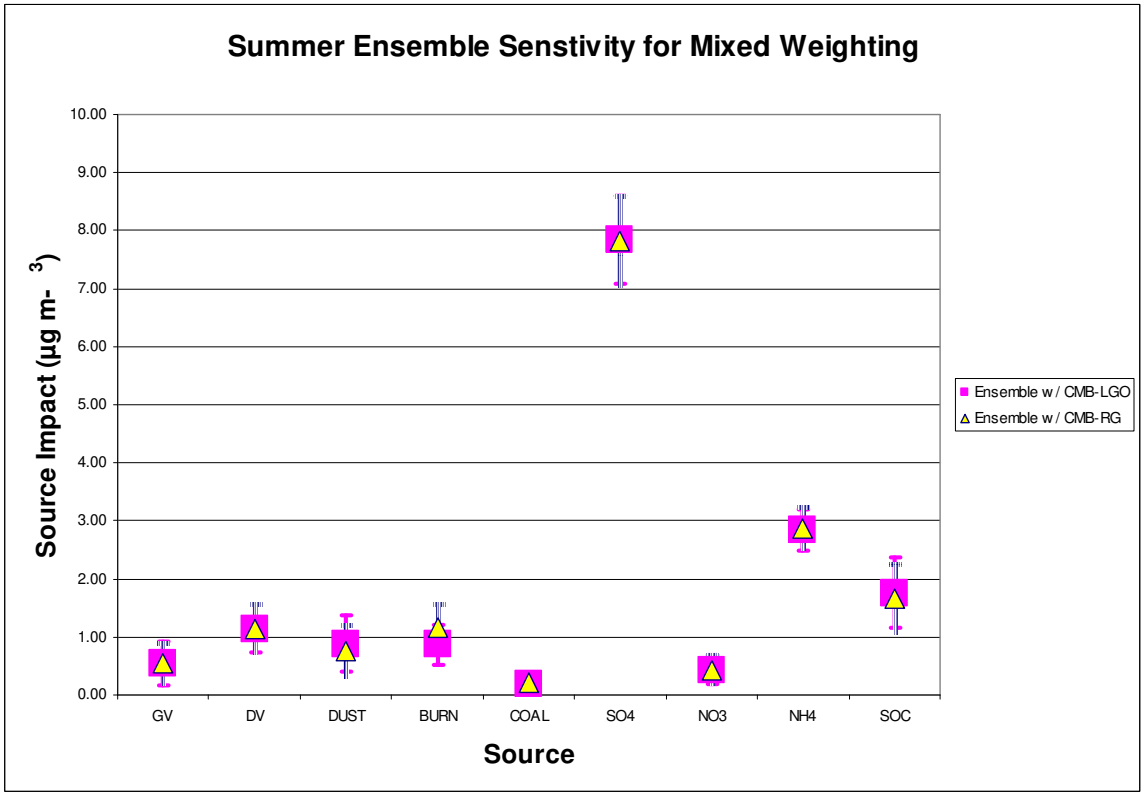


Figure A.7: Ensemble Sensitivity to CMB-RG and CMB-LGO, Summer.

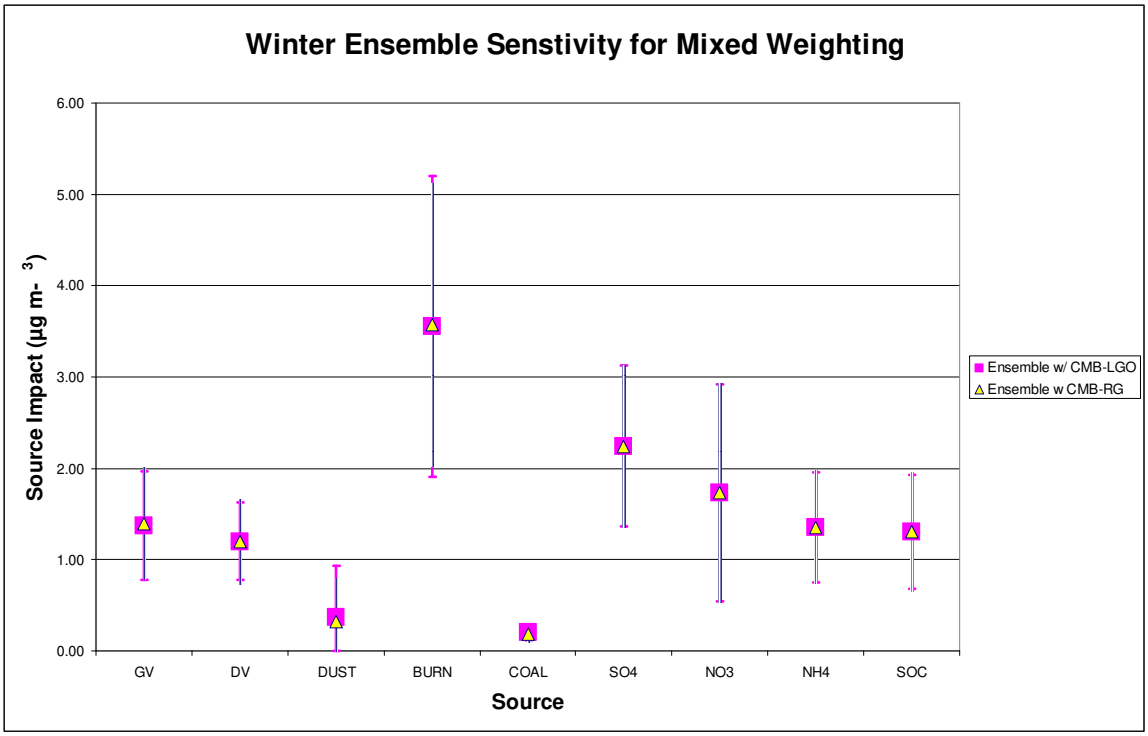


Figure A.8: Ensemble Sensitivity to CMB-RG and CMB-LGO, Winter.

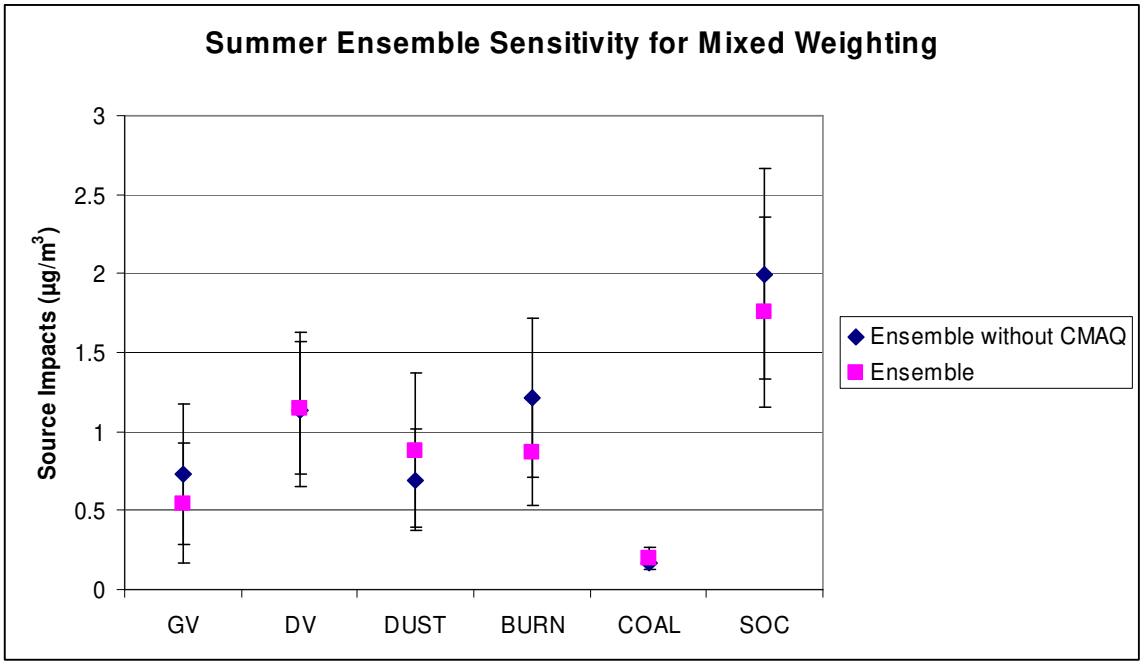


Figure A.9: Ensemble Sensitivity to CMAQ, Summer.

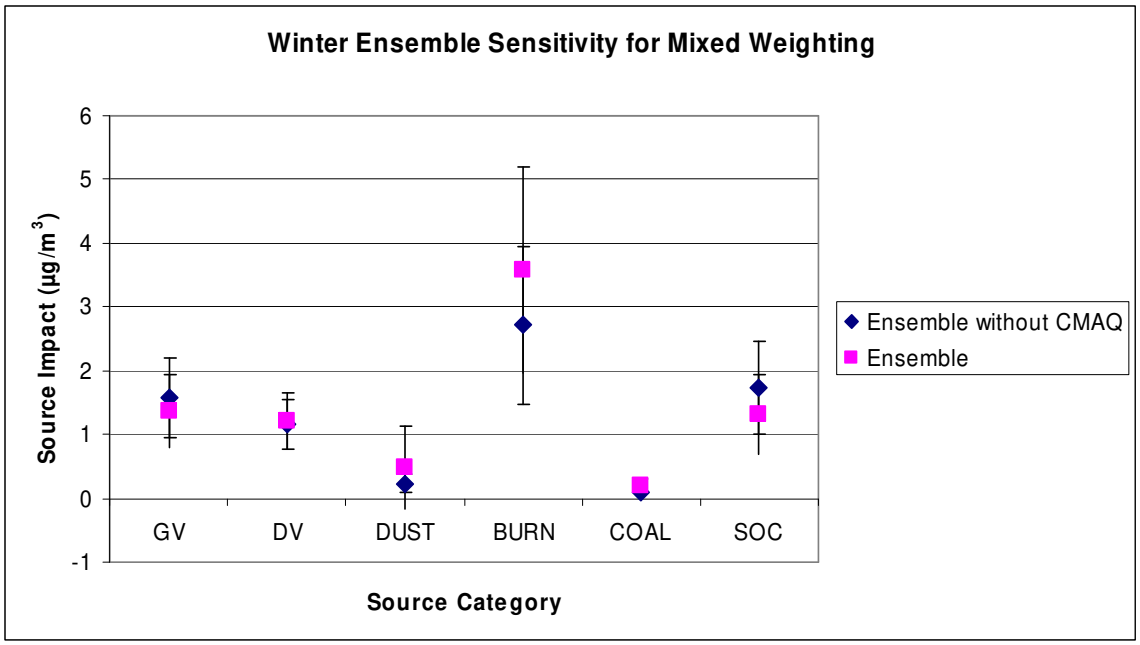


Figure A.10: Ensemble Sensitivity to CMAQ, Winter.

To evaluate the choice of weighting, we conducted York regression between the ensemble and the regression method SOC impacts [*Saylor et al.*, 2006; *York et al.*, 2004] (Figures A.11, A.12 and Table A.5). Equal weighting reproduces the regression method results with high correlation ($R^2 = 0.73$) and a regression slope of 0.88. Mixed weighting had a slightly lower slope of 0.87 but a higher correlation ($R^2 = 0.82$). Inverse square weighting in July 2001 led to a decreased slope and correlation. A similar analysis was performed for January 2002 (Figure A.12), but wintertime SOC is expected to be low, and winter results are impacted by known biases to estimates of both primary and secondary OC impacts in both receptor models and CTMs.

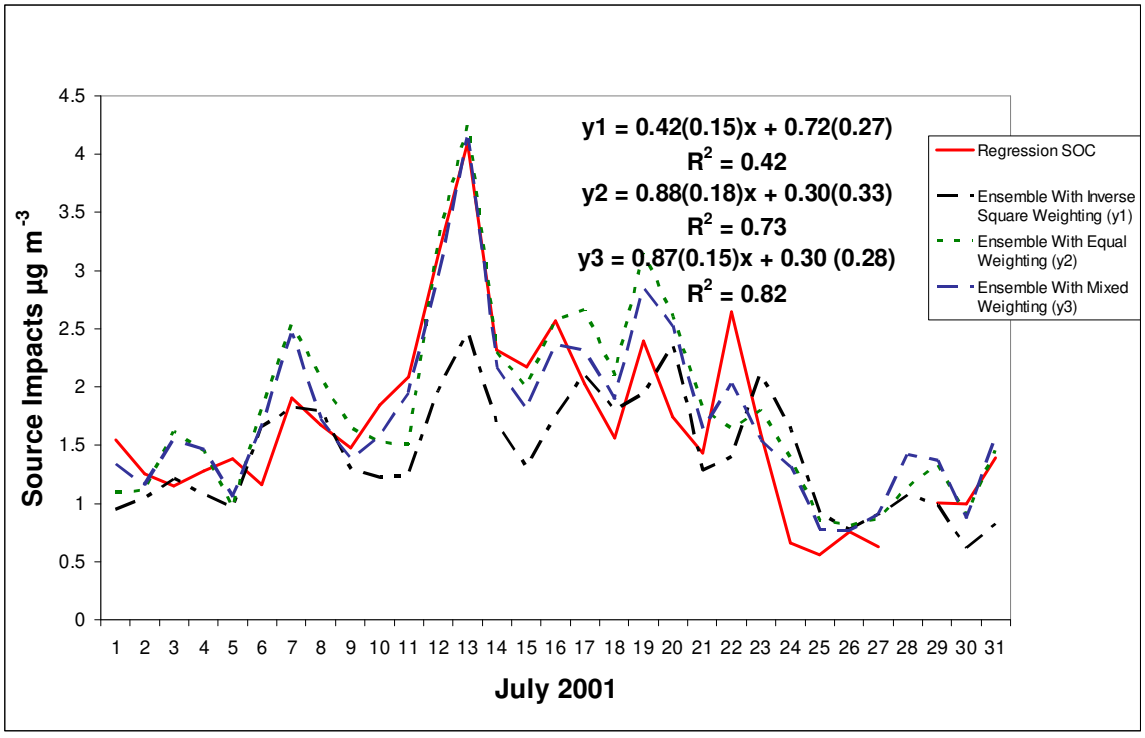


Figure A.11: SOC estimates from the regression method (Pachon et al., 2010), and ensemble results using equal weighting (EW), inverse square weighting (ISW), and a mixed case using both EW and ISW for summer (July 2001) .

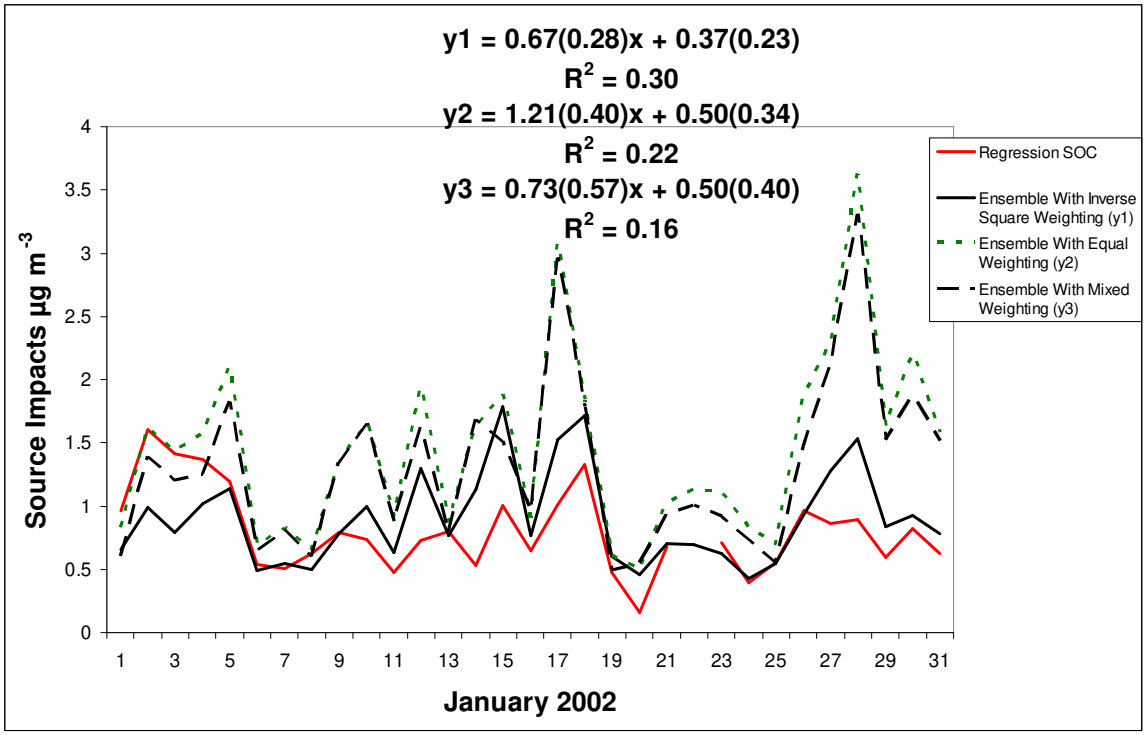


Figure A.12: SOC estimates, January 2002 for the regression method (Pachon et al., 2010), equal weighted ensemble and inverse square weighted ensemble.

Table A.5: Average correlation, R^2, (range), between methods mixed case (MIX) for SOC. Values to the left of the diagonal and right are for July, 2001 and January, 2002, respectively. Values in bold show average $R^2 > 0.50$.

	CMB-LGO	PMF	CMB-MM	CMAQ	ENS (MIX)	Regression Method
CMB-LGO		0.07	0.48	0.39	**0.79**	0.28
PMF	0.02		0.02	0.001	0.027	0.28
CMB-MM	**0.89**	0.002		0.20	**0.74**	0.05
CMAQ	0.002	0.003	0.009		**0.63**	0.03
ENS (MIX)	**0.80**	0.01	**0.88**	0.14		0.16
Regression Method	**0.84**	0.005	**0.90**	0.04	**0.82**	

A.2. Derivation of Ensemble Uncertainty

For each source and for each day, an ensemble-based average, $f_{ensemble}$, is calculated

$$f_{ensemble} = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^N} S_j}{\sum_{j=1}^{J} \frac{1}{\sigma_j^N}}$$

Based on propagation of errors:

$$\sigma_{ensemble}^2 = \left[\frac{\partial f_{ensemble}}{\partial S_l} \right] \begin{bmatrix} \sigma_{S_{11}}^2 & \cdots & \sigma_{S_{1L}} \\ \vdots & \ddots & \vdots \\ \sigma_{L1} & \cdots & \sigma_{LL}^2 \end{bmatrix} \left[\frac{\partial f_{ensemble}}{\partial S_l} \right]^T$$

Where the second matrix term is the variance-covariance matrix of source impacts across the four SA methods (for each day).

Given that:

$$\frac{\partial f_{ensemble}}{\partial S_j} = \frac{\frac{1}{\sigma_j^N}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^N}}$$

Therefore:

$$\sigma_{ensemble}^2 = \begin{bmatrix} \frac{\frac{1}{\sigma_{S_1}^N}}{\sum_{l=1}^{L} \frac{1}{\sigma_{S_l}^N}} & \cdots & \frac{\frac{1}{\sigma_{S_L}^N}}{\sum_{l=1}^{L} \frac{1}{\sigma_{S_l}^N}} \end{bmatrix} \begin{bmatrix} \sigma_{S_{11}}^2 & \cdots & \sigma_{S_{1L}} \\ \vdots & \ddots & \vdots \\ \sigma_{L1} & \cdots & \sigma_{LL}^2 \end{bmatrix} \begin{bmatrix} \frac{\frac{1}{\sigma_{S_1}^N}}{\sum_{l=1}^{L} \frac{1}{\sigma_{S_l}^N}} & \cdots & \frac{\frac{1}{\sigma_{S_L}^N}}{\sum_{l=1}^{L} \frac{1}{\sigma_{S_l}^N}} \end{bmatrix}^T$$

And where the second matrix term on the right hand side is defined by equation 4.

CHAPTER 3: BAYESIAN–BASED ENSEMBLE SOURCE APPORTIONMENT OF $PM_{2.5}$

In Review, *Environmental Science and Technology*, **2013**.

Sivaraman Balachandran[*,a], Howard H. Chang[b], Jorge E. Pachon[c], Heather A. Holmes[a], James A. Mulholland[a] and Armistead G. Russell [a]

[a]Georgia Institute of Technology. School of Civil and Environmental Engineering. Atlanta, GA.

[b]Emory University. Rollins School of Public Health. Atlanta, GA.

[c]Universidad de La Salle, Programa de Ingenieria Ambiental, Bogota, Colombia.

[*]Corresponding author: 311 Ferst Dr., Atlanta, GA, 30332; phone 206.250.6480, fax 404.894.8266; siv@gatech.edu.

3.1. Abstract

A Bayesian source apportionment (SA) method is developed to provide source impact estimates and associated uncertainties. Bayesian-based ensemble averaging of multiple models provides new source profiles for use in a chemical mass balance (CMB) SA of fine particulate matter ($PM_{2.5}$). The approach estimates source impacts and their uncertainties by utilizing a short-term application of four individual SA methods: three receptor-based models and one chemical transport model. For each day of the short term SA application, source impact uncertainties are stochastically sampled from Bayesian-based posterior distributions. The uncertainties for each method are then used as weights in the ensemble-averaged source impacts. A Monte Carlo technique is used to estimate a distribution of Bayesian ensemble–based source impacts for each day in the ensemble. These source impacts are then used to determine two

seasonal distributions of source profiles that are used in SA for a long-term $PM_{2.5}$ dataset. For each day in a long-term $PM_{2.5}$ dataset, 10 source profiles are sampled from these distributions and used in a CMB application resulting in 10 SA results for each day. This formulation results in a distribution of daily source impacts rather than a single value. The average and standard deviation of the distribution are used as the final estimate of source impact and a measure of uncertainty, respectively. The Bayesian-based source impacts for biomass burning correlate better with observed levoglucosan (R^2=0.66) and water soluble potassium (R^2=0.63) than source impacts estimated using more traditional methods, and more closely agreed with observed total mass. The Bayesian approach also captures the expected seasonal variation of biomass burning and secondary impacts. Sensitivity analysis found that using non-informative prior weighting performed better than using weighting based on method-derived uncertainties. This approach can be applied to long-term data sets from EPA's speciation network sites.

3.2. Introduction

Air quality standards are driven, in part, by health impacts of air pollutants and the policies to control sources of air pollutants are often evaluated by improvements to human health. Ambient air pollution has been estimated to contribute to greater than 3,000,000 premature deaths worldwide in 2010; of this burden, the vast majority has been attributed to fine particulate matter ($PM_{2.5}$) [*Lim et al.*, 2012]. $PM_{2.5}$ health impacts include both respiratory and cardiovascular health outcomes [*Dockery et al.*, 1993; *Laden et al.*, 2000]. Given the potential health impacts, the US EPA has set National Ambient Air Quality Standards (NAAQS) for $PM_{2.5}$ and a major goal for states and regional communities is to meet those standards and protect public health. It is suspected that $PM_{2.5}$ health effects vary by composition and source, and may depend upon the mixture of pollutants, leading to efforts to estimate relationships between sources of $PM_{2.5}$ and health effects [*Hopke et al.*, 2006; *Ito et al.*, 2006; *Mar et al.*, 2006; *Sarnat et al.*, 2008; *Thurston et al.*, 2005].

Controlling ambient $PM_{2.5}$ concentrations ultimately means controlling sources of $PM_{2.5}$ which requires techniques for estimating source contributions. However, $PM_{2.5}$ sources typically emit a mixture of pollutants, including gases and particles, which mix in the atmosphere and can undergo chemical transformations prior to impacting a specific receptor location, making it difficult to quantify impacts. Source apportionment (SA) involves one or more techniques that are used to quantify how individual sources contribute to $PM_{2.5}$ concentrations. SA techniques that rely on statistical analysis of observations at monitor sites are referred to as receptor models. These techniques include chemical mass balance (CMB) and positive matrix factorization (PMF). In addition, chemical transport models (CTMs) have utilized sensitivity parameters to estimate source contributions. These different SA approaches often result in source contributions that can differ in magnitude and/or are poorly correlated. Determining which method's set of source contributions is the most accurate is further complicated because source impacts, in general, cannot be directly measured. Without direct measurement of source impacts, methods

for estimating uncertainty vary across the SA approaches, making it difficult to directly compare uncertainties across methods. For example, some methods (e.g. CTMs) have not provided source impact estimate uncertainties while others utilize bootstrapping or propagation of errors to estimate uncertainties.

In this work, we build on an approach to combine multiple SA model results to train a CMB method for long-term application [*Balachandran et al.*, 2012; *D Lee et al.*, 2009; *Maier et al.*, 2013] by extending the ensemble technique to include a Bayesian formulation of weights used in ensemble-averaging source impacts. In a Bayesian approach, probabilistic distributions of the parameters of interest are estimated using prior distributions, along with information from observed data. Bayesian analysis has been used in a variety of applications and can be especially useful for estimating model parameters that are weakly informed by the observed data.

Bayesian techniques have previously been used in SA of $PM_{2.5}$ [*Kashiwagi*, 2004; *Keats et al.*, 2009; *Lingwall and Christensen*, 2007; *Lingwall et al.*, 2008]. These approaches have typically focused on estimating source impacts, which are positive and lognormally distributed. In this work, a method is developed that incorporates Bayesian techniques to estimate SA uncertainties. These uncertainties are then used as weights to estimate an ensemble average of source impacts similar to work by Lee et al. [2009] and Balachandran et al. [2012]. The Bayesian framework for estimating source apportionment uncertainties requires first placing prior distributions about a *subjective (expert-driven)* view of uncertainties associated with each SA method. Next, the root mean square error (RMSE) between an initial ensemble average and each individual method is used as the updated information about source impact uncertainties. Using an inverse gamma prior with a normal data likelihood leads to an inverse gamma posterior distribution of uncertainties for each SA method. These uncertainty distributions are then used as weights to obtain an updated ensemble. One advantage of this method is that it obviates the need to assume lognormally distributed data sets. This assumption can be problematic for receptor models which can result in zero or negative impacts. Also, the approach incorporates

several different models and provides a way to compare methods using a consistent estimation of uncertainties.

The objective of this work is to refine our previously developed ensemble approach for apportioning $PM_{2.5}$ to sources by incorporating a Bayesian technique to obtain multiple realizations of ensemble-averaged source impacts, which are subsequently used for deriving multiple realizations of source profiles. We then compare results using this approach to results using our previous ensemble approach as well as to results using individual receptor models.

3.3 Methods

3.3.1. Ensemble Averaging

The method developed here extends the ensemble method developed by [2009] and [2012] and is comprised of three steps: (1) Bayesian ensemble-averaging source impacts over a short term time period, (2) using these source impacts to develop regionally and seasonal specific source profiles, and (3) using the new source profiles to apportion sources for a long-term data set. We use SA results from three receptor models and one chemical transport model for July 2001 and January 2002. We use two CMB methods: CMB-LGO [*Marmur et al.*, 2005], that incorporates gas based constraints, and CMB-MM [*Zheng et al.*, 2002], which uses molecular marker observations. We use one factor analytic method, PMF [*Paatero and Tapper*, 1994] and one CTM, the community multiscale air quality (CMAQ) model [*D Byun and Schere*, 2006]. We use results from previous work for CMB-MM [*Zheng et al.*, 2007] and CMAQ with tracers [*Baek*, 2009]. We also applied EPA CMB v8.0 (referred to here as CMB-RG, for "regular") [*U.S.EPA*, 2004; *Watson et al.*, 1984], but these results were used for comparison and were not included in the ensemble.

In the work developed by Balachandran et al.[2012], an ensemble average of source impacts is calculated in a two-step process. First, an equally weighted average of source impacts is calculated (Equation 1 with N=0 in Equation 2):

$$\overline{S}_{jk} = \frac{\sum_{l=1}^{L} w_{jlk} \bullet S_{jlk}}{\sum_{l=1}^{L} w_{jlk}} \qquad \text{(Equation 1)}$$

$$w_{jlk} = \frac{1}{\tau_{S_{ijk}}^{N}} \qquad \text{(Equation 2)}$$

where w_{jll} is the weight for source *j* from method *l* on day *k*, and S_{jlk} is the source impact for source *j* from method *l* on day *k*. Next, the root mean square error (RMSE) is calculated between each method and the ensemble average (Equation 3):

$$RMSE_{jl} = \sqrt{\frac{\sum_{k=1}^{K} \left(S_{jlk} - \overline{S}_{jk}\right)^2}{K}} \qquad \text{(Equation 3)}$$

The uncertainty is set to be equal to each method's RMSE and the square is used to weight an updated ensemble average (Equation 2 with N=2 and τ=RMSE). Finally, the uncertainty of the updated ensemble average is calculated using weighted propagation of errors with covariance [2012]. To compare the ensemble with the individual SA methods, we use the root mean square average of the daily source impact uncertainties to reflect the overall method uncertainty, $\left(\overline{\sigma}_{S_{jlk}}\right)$, [*Balachandran et al.*, 2012; *Pachon et al.*, 2010] (Equation 4):

$$\overline{\sigma}_{S_{jl}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} \sigma_{S_{jlk}}^2} \qquad \text{(Equation 4)}$$

3.3.2 Bayesian Ensemble Averaging

One limitation of the method described above is that for any source (for any method), the estimated source impact uncertainty is the same for each day, since the RMSE does not change on a daily basis. A more realistic interpretation is that the RMSE should be viewed as an

"average" uncertainty and that the true uncertainty comes from a distribution whose mean is equal to the RMSE. In Bayesian ensemble averaging, a posterior distribution of uncertainties is calculated using a prior distribution and treating the estimated RMSEs as the data. For each day of the short term application of the four SA methods, source impact uncertainties are sampled from the Bayesian-based posterior distribution using a Monte Carlo technique. These uncertainties are used as weights to calculate ensemble averaged source impacts.

It is assumed that estimates of source impacts vary randomly around "true" source impacts. Therefore, S_{jlk}, the impact from source j and method l on day k, can be viewed as a surrogate measure of the true source impact and that the average of these methods, $\overline{S}_{jk}$, can be treated as the true source impact. A consequence is that that these errors are normally distributed so that for any day k:

$$S_{jlk} - \overline{S}_{jk} \sim Normal\,(0, \tau_{jlk}^2) \qquad \text{(Equation 5)}$$

We wish to obtain posterior samples of τ_{jlk}^2 and use them to calculate an ensemble average using Equations 1 and 2. First we assign an inverse-gamma (scaled-inverse-chi-squared) distribution to each variance component. The inverse-gamma (IG) distribution is specified by a density function with two known parameters α and β, and denoted as IG (α, β):

$$f(\tau_{jlk}^2 \,|\, \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} (\tau_{jlk}^2)^{-\alpha-1} \exp(-\frac{\beta}{\tau_{jlk}^2}) \qquad \text{(Equation 6)}$$

The error of the data (S_{jl} (k=1) . . . S_{jl} (k=K)) with respect to the average $\overline{S}_{jk}$, has a likelihood given by the normal density:

$$f(\text{data} \,|\, \tau_{jlk}^2) = (2\pi\tau_{jlk}^2)^{-\frac{K}{2}} \exp(-\frac{1}{2\tau_{jlk}^2} \sum_{k=1}^{K} \left(S_{jlk} - \overline{S}_{jk}\right)^2) \qquad \text{(Equation 7)}$$

The posterior distribution of τ_{jlk}^2 given the data is found from:

$$f(\text{data}|\tau_{jlk}^2) \propto f(\text{data}|\tau_{jlk}^2) \times f(\tau_{jlk}^2|\alpha,\beta)$$

(Equation 8)

$$= (\tau_{jlk}^2)^{-(\alpha+\frac{K}{2})-1} \exp(-\frac{1}{\tau_{jlk}^2}\left[\beta + \frac{1}{2}\sum_{k=1}^{K}\left(S_{jlk} - \overline{S}_{jk}\right)^2\right])$$

The last expression is proportional to an inverse-gamma distribution:

$$IG\left(\alpha + \frac{K}{2}, \left[\beta + \frac{1}{2}\sum_{k=1}^{K}\left(S_{jlk} - \overline{S}_{jk}\right)^2\right]\right) \quad \text{(Equation 9)}$$

It is important to note that the above distribution has mean:

$$\frac{\left[\beta + \frac{1}{2}\sum_{k=1}^{K}\left(S_{jlk} - \overline{S}_{jk}\right)^2\right]}{\alpha + \frac{K}{2}} \quad \text{(Equation 10)}$$

and for small values of α and β, the mean is approximately the square of the RMSE in Equation 3. Typically, prior information about τ_{jlk}^2 can be incorporated in α and β. We approach this method in two ways. To reflect a lack of knowledge, we can use non-informative priors by setting $\alpha = \beta = 0.0001$. In addition, we can use the distribution of method-specific uncertainties and to have informative priors (Figure B.1). For CMAQ, we use non-informative prior information since uncertainties are not a directly available from the model application. Again, this allows us to sample multiple realizations of weights (i.e. uncertainties) that are used in ensemble-averaging. Ensemble-averaging is conducted for 30 days in summer (July 2001) and 30 days in winter (January 2002). For each day in the ensemble, we used 30 samples from the posterior distributions, resulting in 30 ensemble-averaged source impact estimates for each of 30 days in the short term period.

3.3.3. Development of Seasonal Source Profiles

We develop source profiles in the same manner as Lee et al. [*D Lee et al.*, 2009]. We solve the chemical mass balance equation for the measured chemical species C_i, by treating the

source profile matrix, f_{ij}, as the unknown and treating the source contribution, S_j, as known by using the ensemble based source impacts ($\overline{S}_{jk}$).

$$C_i = f_{ij}S_j + e_i \qquad \text{(Equation 11)}$$

We solve for f_{ij} by minimizing the Chi-squared value:

$$\chi_k^2 = \sum_{i=1}^{I} \frac{\left(C_{ik} - \sum_{j=1}^{J} f_{ij}S_j\right)^2}{\sigma_{C_{ik}}^2} \qquad \text{(Equation 12)}$$

where $\sigma_{C_{ik}}^2$ is the square of the measurement uncertainty of species *i* on day *k*. We use ensemble averaged source impacts from Balachandran et al. [2012] to calculate ensemble-based source profiles (EBSPs) for 30 days in July 2001 (summer) and January 2002 (winter) each. For the Bayesian ensemble, source profiles are derived for both the non-informative prior and informative prior cases. We sample 30 estimates of weights for each of the 30 days in the ensemble; this leads to 900 source profiles for summer and winter each, which represent distributions of two seasonal Bayesian ensemble-based source profiles (BBSPs). For the EBSPs, the average of the 30 source profiles is used in the long-term source apportionment and the standard deviation is treated as the source profile uncertainty. For the Bayesian ensemble, profiles used in the source apportionment are sampled from the distribution of 900 source profiles. Since we have 30 replicates of 30 days in the ensemble, we calculate variability for each species in the source profiles in two ways. We calculate 30 standard deviations across the replicates, *r*, for each day *k* (i.e., within day variation) (Equation 13) and 30 standard deviations across the days, *k*, for each replicate, *r*, (i.e., between day variation) (Equation 14).

$$\sigma_{f_{ij}}(k) = \sqrt{\frac{\sum_{r=1}^{R} (f_{ijr}(k) - \overline{f}_{ij}(k))^2}{R-1}} \qquad \text{(Equation 13)}$$

$$\sigma_{f_{ij}}(r) = \sqrt{\frac{\sum_{k=1}^{K} (f_{ijk}(r) - \overline{f}_{ij}(r))^2}{K-1}} \qquad \text{(Equation 14)}$$

3.3.4. Source Apportionment for a Long-Term Data Set

Long-term SA is conducted for a nine and half year data set (8/1/98 -12/31/07) with 3107 days of measurement data collected from the Jefferson St. SEARCH site (JST) in Atlanta, GA [*Hansen et al.*, 2003]. We use a method that utilizes gas concentrations of SO_2, CO and NO_x, to constrain the solutions and is referred to as CMB-GC and very similar to CMB-LGO, another method that uses gas constraints [*Marmur et al.*, 2005]. We conduct SA using measurement-based source profiles (MBSPs), EBSPs and BBSPs for nine source categories: gasoline vehicles (GV), diesel vehicles (DV), dust (DUST), biomass burning (BURN), coal combustion (COAL), ammonium sulfate, ammonium bisulfate, ammonium nitrate, and other OC, which largely represents secondary organic carbon (SOC). We use winter EBSPs and BBSPs for November through March and summer EBSPs and BBSPs for April through October. When using BBSPs, 10 source profiles are sampled from the 900 distributions and result in 10 source apportionments for each day. This formulation results in a distribution of 10 daily source impacts rather than a single value with an estimated uncertainty. The average and standard deviation of the 10 SA results are treated as the daily source impact and uncertainty, respectively. These are compared with EBSP and MBSP based source impacts and uncertainty, which are calculated using an effective variance approach [*Watson et al.*, 1984]. We also compare results with using the CMB-RG and PMF [*Paatero and Tapper*, 1994; *Watson et al.*, 1984]. The CMB-RG and PMF results were available from 1/1/99-12/31/04 and used in earlier ensemble studies [*Balachandran et al.*, 2012; *D Lee et al.*, 2009].

3.4. Results and Discussion

3.4.1. Ensemble averaging

We evaluate the ensemble method for each of the three steps. First, we evaluate all three cases of the ensemble-averaged source impacts (standard, Bayesian non-informative priors, and Bayesian informative priors). We expect the overall averages and uncertainties to be very similar since the mean of the IG distribution should approximately equal the RMSE; however, this may not always be the case with informative priors. All three cases of ensemble averaging result in average source impacts and overall uncertainties that are very similar, indicating that the ensemble is stable (Table B.1).

3.4.2. Source profile variability

The distribution of species BBSPs, shown as boxplots, of $\sigma_{f_{ij}}(r)$ (between day variation) is greater than $\sigma_{f_{ij}}(k)$ (within day variation), indicating that between day variation is greater than within day variation (Figure 3.1, summer BURN profile using non-informative priors). In addition, the average ratio of the between day variability to within day variation, $\frac{\sigma_{f_{ij}}(r)}{\sigma_{f_{ij}}(k)}$, ranges from 1 (e.g. Pb and Zn in DUST profiles) to more than 16 (Si in summer DUST profile) (Tables B.2 and B.3). BBSPs are expected to be more variable across days than within days because ensemble-averaged source impacts used to derive source profiles have greater variability across days than within days. This indicates that variability in meteorology (e.g. due to changes in source region, atmospheric processing and emission composition) plays a more important role in source profile variability than the uncertainty of ensemble source impacts that were used to derive the source profiles.

The new source profiles derived using Bayesian and standard ensembles are most different from MBSPs for BURN and COAL (Figures B.2a-e). BURN profiles show strong seasonality for Br, Ca, NH4 and K, which are higher in summer profiles (Figure B.2d). This

suggests that seasonal variability may be driven in part by variation in fuel type as summer impacts from biomass burning have contributions from long range transport of western US wildfires whereas winter/early spring impacts are expected to be dominated by the local prescribed fires that occur predominately in the early Spring. In addition, the summer BURN profiles are enriched in Ca, suggesting entrainment of crustal material in summer BURN emissions. Bayesian-derived COAL profiles also have differences from the MBSP profile (Figure B.2e). Most significantly, the Bayesian COAL profiles have lower OC than MBSPs. In addition, there is a distinct seasonality: higher OC in winter vs. summer. This is in contrast with the EBSP COAL profiles derived in Lee et al. [*D Lee et al.*, 2009] which have higher OC in summer than in winter, likely due to the ability of this method to include some secondary OC formation.

New GV source profiles have OC:EC ratios of ~ 2.2, very similar the MBSP ratio of ~2.3. For DV, the EC:OC ratio is approximately ~4.1, slightly higher than MBSP ratios of 3.7. Some species, such as OC, have smaller variation in OC in GV than MBSPs. In addition, the OC:EC ratios in GV profiles do not show a distinct seasonality. DUST profiles are very similar to MBSPs. However, DUST profiles derived in Lee et al. [*D Lee et al.*, 2009] had ~0.2 OC, higher than in this study (~0.07), suggesting that the DUST profiles derived in this work do not reflect a mixed dust source containing traffic dust emissions.

Source profiles are also evaluated by analyzing the distributions of species in the BBSPs (Figure B.4a-e). The limits of species concentrations were set to be between one third and three times the average values in MBSPs. For some species, their values in the BBSPs are distributed between these limits; these are typically major and tracer species for a given source. However, the modes of these distributions are typically the lower limit, and occasionally, the upper allowable limit. For example, for about a third of the days, the Bayesian summer BURN profiles results in EC values of 0.003, the minimum allowable limit (Figure B.4d). This suggests that for those days, BURN profiles may not have converged to a realistic source profile. However, since this occurs only in a minority of days, and 10 out of 900 source profiles are sampled for each day

in the long-term SA, the effect is minimal. A consequence of this is that uncertainties of BBSPs are not necessarily lower than MBSPs. Nevertheless, for some important tracer species such as potassium in BURN, the distributions show distinct seasonality and variation. In winter, the distribution of potassium is tighter and has a lower mean than in summer (Figure B.4d).

3.4.3. Long-term source apportionment

Both ensemble methods affect the amount of mass apportioned to SOC and biomass burning by exhibiting strong seasonal differences. When using CMB-RG and CMB-GC with MBSPs, wintertime SOC levels are comparable to or slightly greater than summertime levels (Figure 3.2, Table B.4). PMF also has little seasonal variation in SOC, but suffers from potentially underestimating SOC in the summer. CMB-GC has a clear summer/winter split for SOC of 2.66/1.41 $\mu g\ m^{-3}$ with BBSPs and 2.55/1.81 $\mu g\ m^{-3}$ with EBSPs. The largest seasonal difference using BBSPs and EBSPs is for biomass burning. The summer/winter split is 1.63/3.95 $\mu g\ m^{-3}$ with BBSPs and 1.21/2.26 $\mu g\ m^{-3}$ with EBSPs. Having more biomass burning impacts in the winter is expected because both prescribed fires and fireplace usage is greater in these months. This seasonal variation is only slightly evident in CMB-GC with MBSPs (1.59/1.73 $\mu g\ m^{-3}$) and PMF (2.70/2.85 $\mu g\ m^{-3}$). Seasonal variation is also seen for GV using BBSPs and EBSPs, which are thought to have greater impacts in winter when cold start emissions contribute significantly to GV emissions, and when meteorological conditions lead to less dispersion.

In CMB, the reduced chi-square value is often used as metric for goodness of fit. Using BBSPs leads to comparable but higher reduced chi-square values than with EBSPs or MPSPs (Table 3.1). Nevertheless, one important limitation of receptor models that is addressed with BBSPs is that zero-impact days are drastically reduced, a consequence of averaging 10 SA results per day. Typical of receptor models, all three predict total mass to approximately 90% of measured $PM_{2.5}$.

Source impact uncertainties using BBSPs are generally smaller than using EBSPs and MBSPs for all source categories, except biomass burning (Figure B.3). Since the uncertainties in BBSPs come for the standard deviation of 10 sets of SA, the higher uncertainties in biomass burning are reflective of a higher variation in BURN source profiles. This indicates that biomass burning impacts are a major source of uncertainty in source apportionment work.

3.4.4. Evaluation of Method

A major assumption in our method is that SA errors between each method's source impact and the ensemble average are normally distributed with a mean of 0. Three SA methods, CMB-GC with MBSPs, CMB-RG, and PMF had results for 1994-2004 of which July 2001 and January 2002 results were used in the ensemble. We compared the 1999-2004 results against the long-term source apportionment from both Bayesian-based ensemble cases. Histograms of errors between Bayesian-based source impacts and CMB-GC with MBSPs, PMF and CMB-RG (Figure 3.3) show that the errors can be reasonably taken to be normally distributed, supporting Equation 5, a major assumption in this work. In addition, the error histograms are not centered at 0 for winter time SOC and BURN impacts from CMB-based methods using MBSPs. This indicates the distinct bias of traditional CMB-based methods: winter time SOC is overestimated and winter time BURN impacts are underestimated. In addition, SOC impact errors from PMF are centered at ~1 in summer, indicating an overall underestimation of summertime SOC from PMF.

To further evaluate the various SA methods, we compare results for BURN and SOC impacts with independent measurements of levoglucosan, water soluble organic carbon (WSOC) and water soluble potassium (K^+). In 2007, a field campaign was undertaken to measure levoglucosan, a tracer for biomass burning, and WSOC, at the South Dekalb (SDK) site located approximately 10 miles southeast of JST. Given this proximity, the measurements of levoglucosan and WSOC at SDK are taken as representative of conditions of JST. There are a total of 55 samples, taken every sixth day, and we compare BURN and SOC impacts from five

SA methods for the corresponding days. CMB-GC with MBSPs, EBSPs, two BBSPs (with non-informative and informative priors) and PMF. It should be noted that PMF was re-run for a data set from 1999-2007 that included fractionated OC data. We make three comparisons: BURN impacts with both levoglucosan and K^+ measurements, and the sum of BURN and SOC impacts with WSOC (Figure 3.4, Table B.5).

All five of the SA methods apportion the sum of BURN and SOC impacts similarly and all methods have similar correlations. The highest correlations for CMB-GC-MBSP and PMF (R^2= ~0.7) and the lowest for CMB-GC-BBSP using informative priors (R^2= ~0.6) (Figure 3.4). However, the methods split the WSOC into BURN and SOC fractions differently. The BBSPs have the highest correlation (R^2=~0.5-0.6), between BURN impacts and levoglucosan, while the other methods have R^2 of approximately ~0.02 - 0.3. The BBSPs also have the highest correlation (R^2=~0.5-0.6) between BURN impacts and water soluble potassium (K^+).

WSOC is viewed as having two major sources: biomass burning and secondary organic aerosol (SOA) formation [*Sullivan and Weber*, 2006; *Weber et al.*, 2007]. The Bayesian approach produces a higher correlation between biomass burning and both levoglucosan and water soluble potassium, than the other methods, suggesting a more accurate split between biomass burning and SOC. Using non-informative priors produces a higher correlation with levoglucosan than using informative priors and may be due to the influence of CMAQ. There is a greater influence from CMAQ when using non-informative priors because all SA methods are essentially treated equally. CMAQ is weighted less when using informative priors. Since there is no accepted method for calculating uncertainties in CMAQ, we still use non-informative priors for CMAQ while the other SA methods use informative priors. This further suggests that uncertainties calculated by the routine-specific approaches are not appropriate in comparing the accuracy of the different SA methods.

One limitation of the ensemble-averaging method is that it is dependent on short term applications of CMAQ (and CMB-MM, but it expected that CTMs will be used more than CMB-MM in ensemble-averaging). As more CTM-based SA is conducted, the Bayesian method should

be applied using short-term applications for different time periods. The use of informative priors led to lower correlations between BURN impacts and measured levoglucosan than with non-informative priors. However, SA results using non-informative priors are, in general, highly correlated with informative priors. In this work, we use inverse gamma priors with a normal likelihood function, in part, because the resulted posterior distributions have closed-form expressions that can be simulated from efficiently. The use of non-conjugate priors may lead to improved results.

3.5. References

1. Lim, S. S.; Vos, T.; Flaxman, A. D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Amann, M.; Anderson, H. R.; Andrews, K. G.; Aryee, M.; Atkinson, C.; Bacchus, L. J.; Bahalim, A. N.; Balakrishnan, K.; Balmes, J.; Barker-Collo, S.; Baxter, A.; Bell, M. L.; Blore, J. D.; Blyth, F.; Bonner, C.; Borges, G.; Bourne, R.; Boussinesq, M.; Brauer, M.; Brooks, P.; Bruce, N. G.; Brunekreef, B.; Bryan-Hancock, C.; Bucello, C.; Buchbinder, R.; Bull, F.; Burnett, R. T.; Byers, T. E.; Calabria, B.; Carapetis, J.; Carnahan, E.; Chafe, Z.; Charlson, F.; Chen, H.; Chen, J. S.; Cheng, A. T.-A.; Child, J. C.; Cohen, A.; Colson, K. E.; Cowie, B. C.; Darby, S.; Darling, S.; Davis, A.; Degenhardt, L.; Dentener, F.; Des Jarlais, D. C.; Devries, K.; Dherani, M.; Ding, E. L.; Dorsey, E. R.; Driscoll, T.; Edmond, K.; Ali, S. E.; Engell, R. E.; Erwin, P. J.; Fahimi, S.; Falder, G.; Farzadfar, F.; Ferrari, A.; Finucane, M. M.; Flaxman, S.; Fowkes, F. G. R.; Freedman, G.; Freeman, M. K.; Gakidou, E.; Ghosh, S.; Giovannucci, E.; Gmel, G.; Graham, K.; Grainger, R.; Grant, B.; Gunnell, D.; Gutierrez, H. R.; Hall, W.; Hoek, H. W.; Hogan, A.; Hosgood Iii, H. D.; Hoy, D.; Hu, H.; Hubbell, B. J.; Hutchings, S. J.; Ibeanusi, S. E.; Jacklyn, G. L.; Jasrasaria, R.; Jonas, J. B.; Kan, H.; Kanis, J. A.; Kassebaum, N.; Kawakami, N.; Khang, Y.-H.; Khatibzadeh, S.; Khoo, J.-P.; Kok, C.; Laden, F.; Lalloo, R.; Lan, Q.; Lathlean, T.; Leasher, J. L.; Leigh, J.; Li, Y.; Lin, J. K.; Lipshultz, S. E.; London, S.; Lozano, R.; Lu, Y.; Mak, J.; Malekzadeh, R.; Mallinger, L.; Marcenes, W.; March, L.; Marks, R.; Martin, R.; McGale, P.; McGrath, J.; Mehta, S.; Mensah, G. A.; Merriman, T. R.; Micha, R.; Michaud, C.; Mishra, V.; Hanafiah, K. M.; Mokdad, A. A.; Morawska, L.; Mozaffarian, D.; Murphy, T.; Naghavi, M.; Neal, B.; Nelson, P. K.; Nolla, J. M.; Norman, R.; Olives, C.; Omer, S. B.; Orchard, J.; Osborne, R.; Ostro, B.; Page, A.; Pandey, K. D.; Parry, C. D. H.; Passmore, E.; Patra, J.; Pearce, N.; Pelizzari, P. M.; Petzold, M.; Phillips, M. R.; Pope, D.; Pope Iii, C. A.; Powles, J.; Rao, M.; Razavi, H.; Rehfuess, E. A.; Rehm, J. T.; Ritz, B.; Rivara, F. P.; Roberts, T.; Robinson, C.; Rodriguez-Portales, J. A.; Romieu, I.; Room, R.; Rosenfeld, L. C.; Roy, A.; Rushton, L.; Salomon, J. A.; Sampson, U.; Sanchez-Riera, L.; Sanman, E.; Sapkota, A.; Seedat, S.; Shi, P.; Shield, K.; Shivakoti, R.; Singh, G. M.; Sleet, D. A.; Smith, E.; Smith, K. R.; Stapelberg, N. J. C.; Steenland, K.; Stöckl, H.; Stovner, L. J.; Straif, K.; Straney, L.; Thurston, G. D.; Tran, J. H.; Van Dingenen, R.; van Donkelaar, A.; Veerman, J. L.; Vijayakumar, L.; Weintraub, R.; Weissman, M. M.; White, R. A.; Whiteford, H.; Wiersma, S. T.; Wilkinson, J. D.; Williams, H. C.; Williams, W.; Wilson, N.; Woolf, A. D.; Yip, P.; Zielinski, J. M.; Lopez, A. D.; Murray, C. J. L.; Ezzati, M., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **2012,** *380*, (9859), 2224-2260.
2. Dockery, D. W.; Pope, C. A.; Xu, X. P.; Spengler, J. D.; Ware, J. H.; Fay, M. E.; Ferris, B. G.; Speizer, F. E., An Association between Air-Pollution and Mortality in 6 United-States Cities. *New England Journal of Medicine* **1993,** *329*, (24), 1753-1759.
3. Laden, F.; Neas, L. M.; Dockery, D. W.; Schwartz, J., Association of fine particulate matter from different sources with daily mortality in six US cities. *Environmental Health Perspectives* **2000,** *108*, (10), 941-947.

4. Hopke, P. K.; Ito, K.; Mar, T.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Paatero, P.; Thurston, G. D., PM source apportionment and health effects: 1. Intercomparison of source apportionment results. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (3), 275-286.
5. Ito, K.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Neas, L.; Hopke, P. K.; Thurston, G. D., PM source apportionment and health effects: 2. An investigation of intermethod variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (4), 300-310.
6. Mar, T. F.; Ito, K.; Koenig, J. Q.; Larson, T. V.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Neas, L.; Stolzel, M.; Paatero, P.; Hopke, P. K.; Thurston, G. D., PM source apportionment and health effects. 3. Investigation of inter-method variations in associations between estimated source contributions Of PM2.5 and daily mortality in Phoenix, AZ. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (4), 311-320.
7. Sarnat, J. A.; Marmur, A.; Klein, M.; Kim, E.; Russell, A. G.; Sarnat, S. E.; Mulholland, J. A.; Hopke, P. K.; Tolbert, P. E., Fine particle sources and cardiorespiratory morbidity: An application of chemical mass balance and factor analytical source-apportionment methods. *Environmental Health Perspectives* **2008,** *116*, (4), 459-466.
8. Thurston, G. D.; Ito, K.; Mar, T.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Hopke, P. K., Workgroup report: Workshop on source apportionment of particulate matter health effects - Intercomparison of results and implications. *Environmental Health Perspectives* **2005,** *113*, (12), 1768-1774.
9. Balachandran, S.; Pachon, J. E.; Hu, Y.; Lee, D.; Mulholland, J. A.; Russell, A. G., Ensemble-trained source apportionment of fine particulate matter and method uncertainty analysis. *Atmospheric Environment* **2012,** *61*, (0), 387-394.
10. Lee, D.; Balachandran, S.; Pachon, J.; Shankaran, R.; Lee, S.; Mulholland, J. A.; Russell, A. G., Ensemble-Trained PM2.5 Source Apportionment Approach for Health Studies. *Environmental Science & Technology* **2009,** *43*, (18), 7023-7031.
11. Maier, M. L.; Balachandran, S.; Sarnat, S. E.; Turner, J. R.; Mulholland, J. A.; Russell, A. G., Application of an ensemble-trained source apportionment approach at a site impacted by multiple point sources. *Environmental Science and Technology (In review).* **2013**.
12. Kashiwagi, N., Chemical mass balance when an unknown source exists. *Environmetrics* **2004,** *15*, (8), 777-796.
13. Keats, A.; Cheng, M. T.; Yee, E.; Lien, F. S., Bayesian treatment of a chemical mass balance receptor model with multiplicative error structure. *Atmospheric Environment* **2009,** *43*, (3), 510-519.
14. Lingwall, J. W.; Christensen, W. F., Pollution source apportionment using a priori information and positive matrix factorization. *Chemometrics Intell. Lab. Syst.* **2007,** *87*, (2), 281-294.
15. Lingwall, J. W.; Christensen, W. F.; Reese, C. S., Dirichlet based Bayesian multivariate receptor modeling. *Environmetrics* **2008,** *19*, (6), 618-629.

16. Marmur, A.; Unal, A.; Mulholland, J. A.; Russell, A. G., Optimization-based source apportionment of PM2.5 incorporating gas-to-particle ratios. *Environmental Science & Technology* **2005,** *39*, (9), 3245-3254.
17. Zheng, M.; Cass, G. R.; Schauer, J. J.; Edgerton, E. S., Source apportionment of PM2.5 in the southeastern United States using solvent-extractable organic compounds as tracers. *Environmental Science & Technology* **2002,** *36*, (11), 2361-2371.
18. Paatero, P.; Tapper, U., Positive matrix factorization - a nonnegative factor model with optimal utilization of error-estimates of data values. *Environmetrics* **1994,** *5*, (2), 111-126.
19. Byun, D.; Schere, K. L., Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* **2006,** *59*, (1-6), 51-77.
20. Zheng, M.; Cass, G. R.; Ke, L.; Wang, F.; Schauer, J. J.; Edgerton, E. S.; Russell, A. G., Source apportionment of daily fine particulate matter at Jefferson street, Atlanta, GA, during summer and winter. *Journal of the Air & Waste Management Association* **2007,** *57*, (2), 228-242.
21. Baek, J. Improving aerosol simulations: assessing and improving emissions and secondary organic aerosol formation in air quality modeling. Dissertation, Georgia Institute of Technology, Atlanta, GA, 2009.
22. U.S.EPA, EPA-CMB8.2 User's Manual. In Office of Air Quality & Standards, U. S. E. P. A., Research Triangle Park, NC., Ed. Publication No. EPA-452/R-04-011: 2004.
23. Watson, J. G.; Cooper, J. A.; Huntzicker, J. J., The effective variance weighting for least-squares calculations applied to the mass balance receptor model. *Atmospheric Environment* **1984,** *18*, (7), 1347-1355.
24. Pachon, J. E.; Balachandran, S.; Hu, Y. T.; Weber, R. J.; Mulholland, J. A.; Russell, A. G., Comparison of SOC estimates and uncertainties from aerosol chemical composition and gas phase data in Atlanta. *Atmospheric Environment* **2010,** *44*, (32), 3907-3914.
25. Hansen, D. A.; Edgerton, E. S.; Hartsell, B. E.; Jansen, J. J.; Kandasamy, N.; Hidy, G. M.; Blanchard, C. L., The southeastern aerosol research and characterization study: Part 1-overview. *Journal of the Air & Waste Management Association* **2003,** *53*, (12), 1460-1471.
26. Sullivan, A. P.; Weber, R. J., Chemical characterization of the ambient organic aerosol soluble in water: 2. Isolation of acid, neutral, and basic fractions by modified size-exclusion chromatography. *Journal of Geophysical Research-Atmospheres* **2006,** *111*, (D5).
27. Weber, R. J.; Sullivan, A. P.; Peltier, R. E.; Russell, A.; Yan, B.; Zheng, M.; de Gouw, J.; Warneke, C.; Brock, C.; Holloway, J. S.; Atlas, E. L.; Edgerton, E., A study of secondary organic aerosol formation in the anthropogenic-influenced southeastern United States. *Journal of Geophysical Research-Atmospheres* **2007,** *112*, (D13), 13.
28. Chow, J. C.; Watson, J. G.; Kuhns, H.; Etyemezian, V.; Lowenthal, D. H.; Crow, D.; Kohl, S. D.; Engelbrecht, J. P.; Green, M. C., Source profiles for industrial, mobile, and area sources in the Big Bend Regional Aerosol Visibility and Observational study. *Chemosphere* **2004,** *54*, (2), 185-208.

29. Zielinska, B.; McDonald, J. D.; Hayes, T.; Chow, J. C.; Fujita, E. M.; Watson, J. G. *Northern Front Range Air Quality Study Final Report*; Report to Colorado State University by Desert Research Institute: Reno, NV, 1998.

3.6. Tables

Table 3.1: Statistical metrics of CMB-GC using four types of source profiles for 8/31/98 -12/31/07 (3107 days of SA results out of 3149 total days): BBSPs with informative priors (BBSP-IP), BBSPs with non-informative priors (BBSP-NIP), EBSPs and MBSPs

	BBSP-IP	BBSP-NIP	EBSP	MBSP
Reduced Chi Square	5.28	5.70	3.45	4.86
Pred./Obs. PM Mass	0.94	0.93	0.90	0.87
	Zero Impact Days			
	Inf. Priors	Non-Inf. Priors	EBSP	MBSP
GV	0	0	0	0
DV	3	6	204	154
DUST	0	0	15	54
BURN	0	0	4	5
COAL	9	9	184	267
SOC	24	25	60	25

3.7. Figures

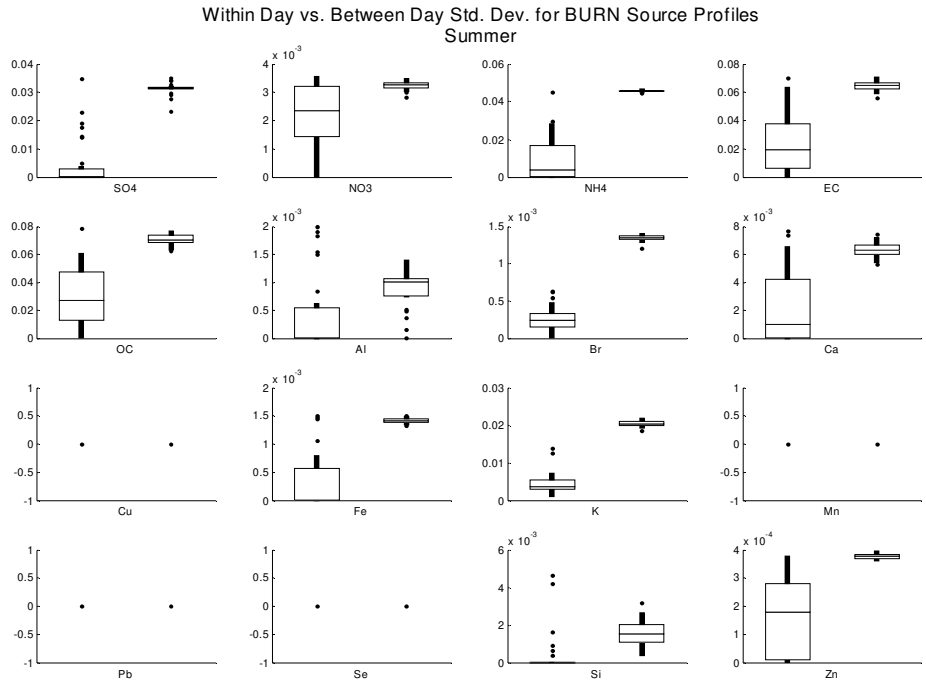


Figure 3.1: Boxplots of within-day ($\sigma_{f_{ij}}(k)$) and between-day variation ($\sigma_{f_{ij}}(r)$) for 16 species in the BURN summer Bayesian profile using non-informative priors (BBSP-NIP).

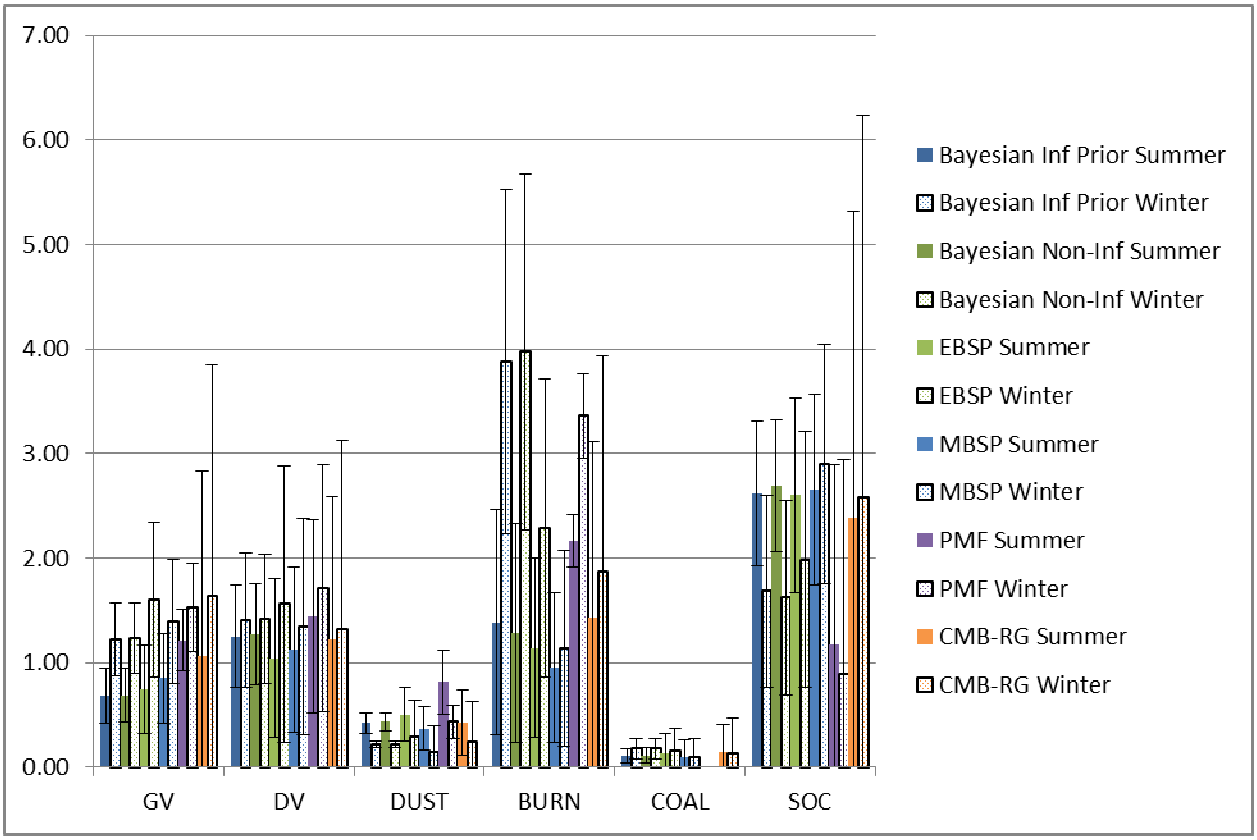


Figure 3.2: Average source impacts and overall uncertainty (as defined in Equation 12) for source apportionment from 1999-2004.

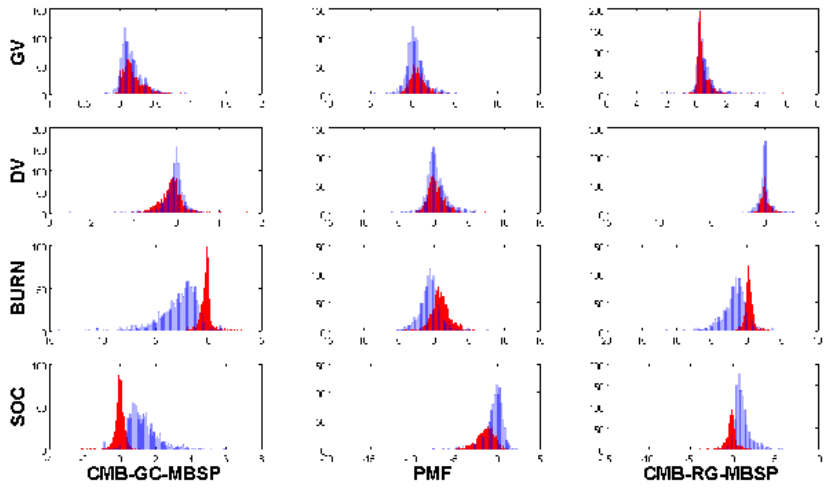


Figure 3.3: Histograms (red = summer, blue = winter) of errors between SA method impacts and Bayesian based SA ($S - \overline{S}$) with profiles derived using non-informative priors (BBSP-NIP) for GV, DV, BURN and SOC for 1999-2004.

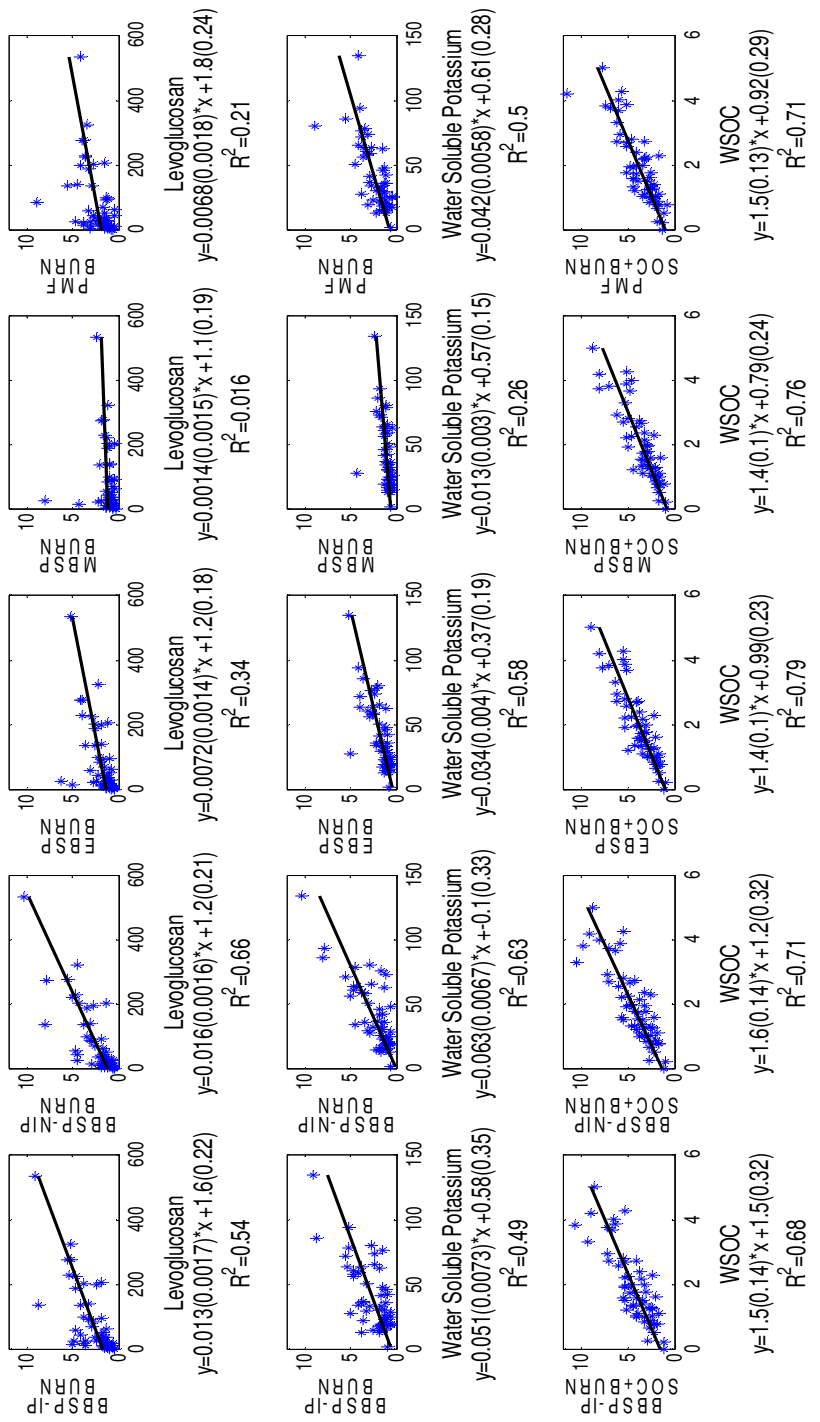


Figure 3.4: Comparison of source impacts for BURN and SOC and water soluble organic carbon (WSOC), levoglucosan and water soluble potassium (K^+). The first row compares BURN and levoglucosan. The second row compares BURN and water-soluble potassium. The last row compares the sum of SOC and BURN impacts and WSOC.

APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Table B.1: Average source impacts and overall relative uncertainties for the standard ensemble, Bayesian ensemble with non-informative priors and Bayesian ensemble with informative priors for summer (July 2001) and winter (January 2002). NOTE: Uncertainty determined by taking the root mean square average of each day's ensemble average uncertainty

	Summer								
	Standard Ensemble			**Bayesian Ens. With Non-Inf. Prior**			**Bayesian Ens. With Informative Prior**		
	Avg. Source Impact		**Unc.**	**Avg. Source Impact**		**Unc.**	**Avg. Source Impact**		**Unc.**
GV	0.55	±	0.38	0.55	±	0.38	0.53	±	0.39
DV	1.15	±	0.42	1.15	±	0.43	1.08	±	0.38
DUST	0.88	±	0.49	0.89	±	0.49	0.69	±	0.41
BURN	0.87	±	0.33	0.87	±	0.34	1.02	±	0.48
COAL	0.20	±	0.07	0.20	±	0.07	0.18	±	0.09
SULFATE	7.87	±	0.15	7.87	±	0.15	7.86	±	0.30
NITRATE	0.43	±	0.06	0.43	±	0.06	0.43	±	0.06
AMMONIUM	2.89	±	0.23	2.89	±	0.24	2.88	±	0.20
SOC	1.76	±	0.60	1.76	±	0.61	1.41	±	0.60
	Winter								
	Standard Ensemble			**Bayesian Ens. With Non-Inf. Prior**			**Bayesian Ens. With Informative Prior**		
	Avg. Source Impact		**Unc.**	**Avg. Source Impact**		**Unc.**	**Avg. Source Impact**		**Unc.**
GV	1.36	±	0.57	1.36	±	0.58	1.37	±	0.57
DV	1.21	±	0.43	1.22	±	0.44	1.20	±	0.45
DUST	0.47	±	0.65	0.48	±	0.66	0.45	±	0.62
BURN	3.58	±	1.60	3.59	±	1.61	3.54	±	1.51
COAL	0.20	±	0.09	0.20	±	0.09	0.20	±	0.09
SULFATE	2.17	±	0.24	2.17	±	0.25	2.16	±	0.27
NITRATE	1.63	±	0.08	1.63	±	0.08	1.62	±	0.14
AMMONIUM	1.28	±	0.03	1.28	±	0.03	1.28	±	0.06
SOC	1.31	±	0.63	1.31	±	0.63	1.33	±	0.61

Table B.2: Ratio of between day variability and within day variability, $\frac{\sigma_{f_{ij}}(r)}{\sigma_{f_{ij}}(k)}$, of Bayesian-based source profiles (BBSP) using non-informative priors.

	Summer					Winter				
	GV	**DV**	**DUST**	**BURN**	**COAL**	**GV**	**DV**	**DUST**	**BURN**	**COAL**
SO4	5.42	4.32	4.99	5.62	6.39	1.76	1.54	1.57	3.07	2.21
NO3	1.75	1.44	1.67	1.42	1.90	2.21	1.69	1.65	2.17	1.93
NH4	2.63	2.11	-	3.51	3.08	1.52	1.44	-	4.71	2.62
EC	1.92	2.73	2.94	2.21	3.46	1.71	1.69	2.16	2.38	2.45
OC	2.22	2.18	3.23	1.82	3.04	1.39	1.64	2.44	2.72	2.53
Al	3.29	3.20	7.75	4.23	9.05	0.96	1.01	1.00	1.38	1.14
Br	1.99	-	-	5.99	2.29	1.31	-	-	4.14	1.42
Ca	1.93	1.91	2.89	2.75	5.51	2.89	1.75	3.20	3.41	7.69
Cu	2.98	1.77	2.06	-	2.59	5.02	1.49	1.89	-	2.12
Fe	3.45	2.40	3.79	3.07	4.00	2.80	1.90	2.97	3.05	5.77
K	1.67	1.57	1.62	3.70	1.81	1.34	1.25	1.48	3.98	1.42
Mn	2.65	1.69	2.36	-	2.49	3.22	1.53	2.28	-	2.96
Pb	5.11	2.01	1.00	-	2.62	9.39	1.71	1.00	-	2.63
Se	2.02	1.64	-	-	7.28	1.47	1.56	-	-	4.95
Si	1.05	15.41	16.16	1.50	1.01	0.87	1.00	1.01	1.83	1.33
Zn	3.70	2.69	1.00	2.34	6.19	3.57	2.31	1.00	2.95	3.03

Table B.3: Ratio of between day variability and within day variability, $\frac{\sigma_{f_{ij}}(r)}{\sigma_{f_{ij}}(k)}$,of Bayesian-based source profiles (BBSP) using informative priors.

	Summer					Winter				
	GV	**DV**	**DUST**	**BURN**	**COAL**	**GV**	**DV**	**DUST**	**BURN**	**COAL**
SO4	5.64	5.09	4.41	7.15	6.24	1.66	1.48	1.60	3.65	2.49
NO3	1.70	1.86	1.52	1.48	2.00	2.07	1.62	1.89	2.00	1.94
NH4	2.53	2.41	-	4.50	3.17	1.65	1.71	-	4.52	2.74
EC	1.85	2.52	1.99	2.59	3.49	1.58	1.70	2.21	2.22	2.61
OC	2.09	1.99	2.68	2.31	3.05	1.38	1.65	2.29	2.62	2.44
Al	1.87	2.69	5.50	2.44	6.49	1.32	1.07	1.73	1.27	1.50
Br	1.56	-	-	5.34	2.04	1.31	-	-	3.99	1.66
Ca	1.97	1.77	2.19	2.77	4.09	2.04	1.93	3.02	3.62	6.69
Cu	2.56	1.71	2.02	-	2.61	5.01	1.38	1.82	-	2.16
Fe	5.53	4.32	6.09	4.25	6.32	2.69	2.04	3.16	3.04	6.30
K	1.85	1.48	1.52	4.55	1.83	1.42	1.29	1.50	3.72	1.51
Mn	3.40	1.69	2.72	-	2.74	3.31	1.46	2.27	-	2.77
Pb	5.32	1.83	1.00	-	3.73	9.35	1.68	1.00	-	2.57
Se	1.98	1.54	-	-	6.89	1.33	1.41	-	-	4.34
Si	2.09	3.00	5.86	3.74	2.90	1.13	1.03	1.01	1.89	1.21
Zn	3.51	2.91	1.00	2.34	5.11	3.45	2.38	1.00	3.05	3.20

Table B.4: Average seasonal source impacts from six SA approaches for 1999-2004.

		GV			DV			DUST			BURN			COAL			SOC		
Bayesian Inf. Prior	Summer	0.68	±	0.26	1.25	±	0.49	0.42	±	0.09	1.39	±	1.07	0.11	±	0.07	2.62	±	0.69
	Winter	1.22	±	0.35	1.41	±	0.65	0.22	±	0.03	3.88	±	1.65	0.18	±	0.10	1.69	±	0.92
Bayesian Non-Inf.	Summer	0.69	±	0.26	1.27	±	0.49	0.44	±	0.09	1.28	±	1.05	0.11	±	0.07	2.69	±	0.63
	Winter	1.23	±	0.34	1.42	±	0.62	0.22	±	0.03	3.98	±	1.70	0.18	±	0.10	1.62	±	0.94
EBSP	Summer	0.74	±	0.42	1.04	±	0.76	0.50	±	0.26	1.14	±	0.86	0.13	±	0.19	2.60	±	0.93
	Winter	1.60	±	0.74	1.56	±	1.32	0.29	±	0.34	2.29	±	1.42	0.15	±	0.22	1.98	±	1.22
MBSP	Summer	0.85	±	0.43	1.13	±	0.79	0.37	±	0.21	0.95	±	0.72	0.10	±	0.16	2.66	±	0.91
	Winter	1.39	±	0.60	1.35	±	1.03	0.14	±	0.25	1.14	±	0.93	0.10	±	0.18	2.90	±	1.15
PMF	Summer	1.21	±	0.30	1.44	±	0.92	0.82	±	0.31	2.16	±	0.25	-	±	-	1.17	±	1.71
	Winter	1.53	±	0.42	1.71	±	1.18	0.44	±	0.16	3.36	±	0.40	-	±	-	0.89	±	2.06
CMB-RG	Summer	1.07	±	1.77	1.23	±	1.36	0.42	±	0.31	1.43	±	1.69	0.15	±	0.26	2.38	±	2.93
	Winter	1.64	±	2.21	1.32	±	1.82	0.25	±	0.38	1.87	±	2.07	0.13	±	0.33	2.58	±	3.67

	Levoglucosan	Potassium	WSOC	Bayes Inf. Pri. BURN	Bayes Non Inf. Prior BURN	EBSP BURN	MBSP BURN	PMF BURN	Bayes Inf. Pri. SOC	Bayes Non Inf. Pri. SOC	EBSP SOC	MBSP SOC	PMF SOC	Bayes Inf. Pri. SOC+ BURN	Bayes Non Inf. Pri. SOC + BURN	EBSP SOC + BURN	MBSP SOC + BURN	PMF SOC+ BURN
Levogulcosan	*1*	*0.62*	**0.10**	***0.54***	***0.66***	0.34	0.02	0.21	0.03	0.07	0.01	0.03	0.03	0.28	0.34	0.12	0.05	0.17
Potassium	0.20 (0.021)	*1*	**0.48**	***0.50***	***0.64***	***0.59***	0.27	***0.51***	0.05	0.02	0.09	0.27	0.25	***0.65***	***0.72***	***0.55***	0.42	***0.56***
WSOC	0.0034 (0.0014)	0.029 (0.0044)	*1*	0.15	0.20	0.26	0.18	0.47	0.41	0.36	0.47	***0.54***	***0.70***	***0.68***	***0.71***	***0.79***	***0.76***	***0.71***
Bayes Inf. Pri. BURN	0.014 (0.0017)	0.051 (0.0073)	0.65 (0.211)	***1***	***0.85***	***0.68***	0.22	0.39	0.07	0.08	0.01	0.01	0.04	***0.50***	0.45	0.24	0.15	0.32
Bayes Non Inf. Prior BURN	0.016 (0.0016)	0.063 (0.0068)	0.81 (0.22)	1.0 (0.057)	***1***	***0.66***	0.17	0.38	0.03	0.08	0.01	0.03	0.06	***0.50***	***0.55***	0.26	0.16	0.32
EBSP BURN	0.0072 (0.0014)	0.034 (0.004)	0.57 (0.13)	0.56 (0.052)	0.51 (0.050)	***1***	***0.65***	0.46	0.01	0.01	0.01	0.00	0.07	0.48	0.46	0.41	0.30	0.39
MBSP BURN	0.0014 (0.0015)	0.013 (0.003)	0.41 (0.12)	0.28 (0.072)	0.23 (0.069)	0.71 (0.071)	***1***	0.23	0.01	0.01	0.00	0.01	0.05	0.23	0.19	0.31	0.31	0.21
PMF BURN	0.0068 (0.0018)	0.042 (0.0058)	0.92 (0.135)	0.51 (0.087)	0.46 (0.081)	0.81 (0.12)	0.66 (0.16)	***1***	0.12	0.10	0.16	0.27	0.20	***0.68***	***0.64***	***0.61***	***0.55***	***0.89***
Bayes Inf. Pri. SOC	-0.0027 (0.002)	0.012 (0.0083)	0.87 (0.142)	-0.22 (0.11)	-0.13 (0.101)	-0.092 (0.164)	0.099 (0.19)	0.34 (0.13)	***1***	***0.93***	***0.90***	***0.72***	***0.52***	0.24	0.25	0.46	***0.57***	0.29
Bayes Non Inf. Pri. SOC	-0.0039 (0.002)	0.0067 (0.0083)	0.80 (0.15)	-0.23 (0.11)	-0.21 (0.098)	-0.15 (0.162)	0.10 (0.19)	0.31 (0.13)	0.95 (0.037)	***1***	***0.89***	***0.66***	***0.50***	0.20	0.19	0.41	***0.52***	0.26

	Levoglucosan	Potassium	WSOC	Bayes Inf. Pri. BURN	Bayes Non Inf. Prior BURN	EBSP BURN	MBSP BURN	PMF BURN	Bayes Inf. Pri. SOC	Bayes Non Inf. Pri. SOC	EBSP SOC	MBSP SOC	PMF SOC	Bayes Inf. Pri. SOC+ BURN	Bayes Non Inf. Pri. SOC + BURN	EBSP SOC + BURN	MBSP SOC + BURN	PMF SOC+ BURN
EBSP SOC	-0.0010 (0.0019)	0.015 (0.0074)	0.847 (0.12)	-0.089 (0.10)	-0.056 (0.094)	-0.088 (0.15)	-0.016 (0.17)	0.36 (0.12)	0.87 (0.040)	0.88 (0.041)	***1***	***0.88***	***0.63***	0.34	0.34	***0.51***	***0.60***	0.36
MBSP SOC	0.0026 (0.0020)	0.029 (0.0069)	0.96 (0.12)	0.087 (0.11)	0.13 (0.099)	0.052 (0.16)	-0.12 (0.18)	0.51 (0.11)	0.824 (0.071)	0.80 (0.079)	1.0 (0.050)	***1***	***0.63***	***0.52***	***0.54***	***0.57***	***0.61***	0.49
PMF SOC	0.0012 (0.00098)	0.014 (0.0035)	0.55 (0.050)	0.080 (0.053)	0.090 (0.049)	0.16 (0.0769)	0.15 (0.088)	0.22 (0.060)	0.35 (0.046)	0.35 (0.048)	0.42 (0.044)	0.40 (0.042)	***1***	***0.51***	***0.53***	***0.64***	***0.66***	***0.52***
Bayes Inf. Pri. SOC+BURN	0.011 (0.0024)	0.063 (0.0067)	1.51 (0.143)	0.78 (0.107)	0.72 (0.099)	1.13 (0.162)	0.88 (0.22)	1.12 (0.11)	0.66 (0.16)	0.62 (0.17)	0.87 (0.17)	1.01 (0.13)	2.0 (0.27)	***1***	***0.96***	***0.88***	***0.81***	***0.82***

Bayes Non Inf. Pri. SOC + BURN	0.012 (0.0024)	0.070 (0.0063)	1.61 (0.14)	0.77 (0.12)	0.79 (0.098)	1.16 (0.17)	0.85 (0.24)	1.13 (0.12)	0.72 (0.17)	0.62 (0.18)	0.91 (0.17)	1.07 (0.14)	2.12 (0.28)	1.02 (0.028)	***1***	***0.86***	***0.79***	***0.80***
EBSP SOC + BURN	0.0062 (0.0023)	0.049 (0.0065)	1.42 (0.10)	0.47 (0.115)	0.45 (0.11)	0.91 (0.15)	0.90 (0.19)	0.93 (0.12)	0.81 (0.12)	0.77 (0.13)	0.93 (0.13)	0.92 (0.11)	1.96 (0.20)	0.82 (0.041)	0.78 (0.043)	***1***	***0.96***	***0.82***
MBSP SOC + BURN	0.0040 (0.0023)	0.041 (0.0070)	1.37 (0.11)	0.36 (0.119)	0.354 (0.11)	0.76 (0.161)	0.88 (0.18)	0.86 (0.11)	0.88 (0.11)	0.85 (0.11)	0.99 (0.11)	0.93 (0.11)	1.95 (0.19)	0.77 (0.051)	0.73 (0.052)	0.96 (0.026)	***1***	***0.77***
PMF SOC+BURN	0.0080 (0.0024)	0.056 (0.0072)	1.47 (0.13)	0.59 (0.12)	0.55 (0.11)	0.97 (0.17)	0.81 (0.21)	1.22 (0.06)	0.70 (0.15)	0.67 (0.15)	0.85 (0.15)	0.94 (0.13)	1.92 (0.25)	0.86 (0.056)	0.81 (0.056)	0.99 (0.063)	0.97 (0.073)	***1***

Table B.5: Correlation (R2) between levoglucosan, water-soluble potassium, WSOC, BURN impacts, SOC impacts and the sum of BURN + SOC impacts (values on the diagonal and to the right of the diagonal). Slope and standard error of regression (values to the left of the diagonal).

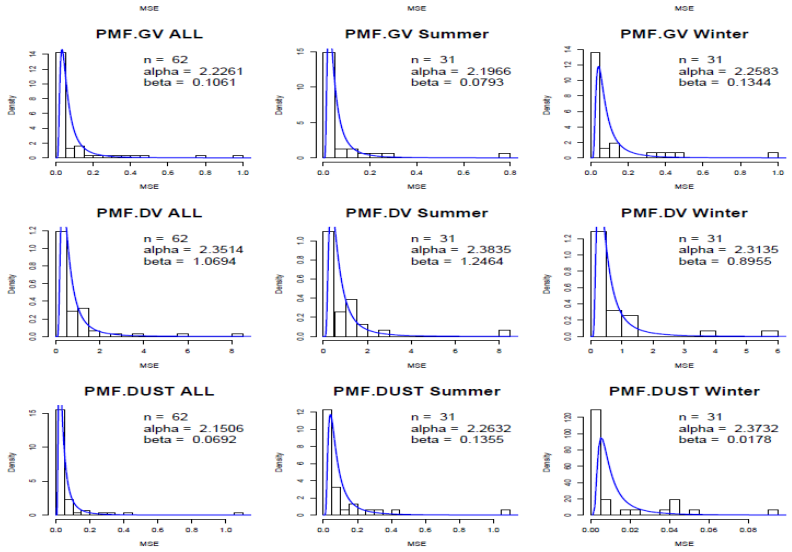
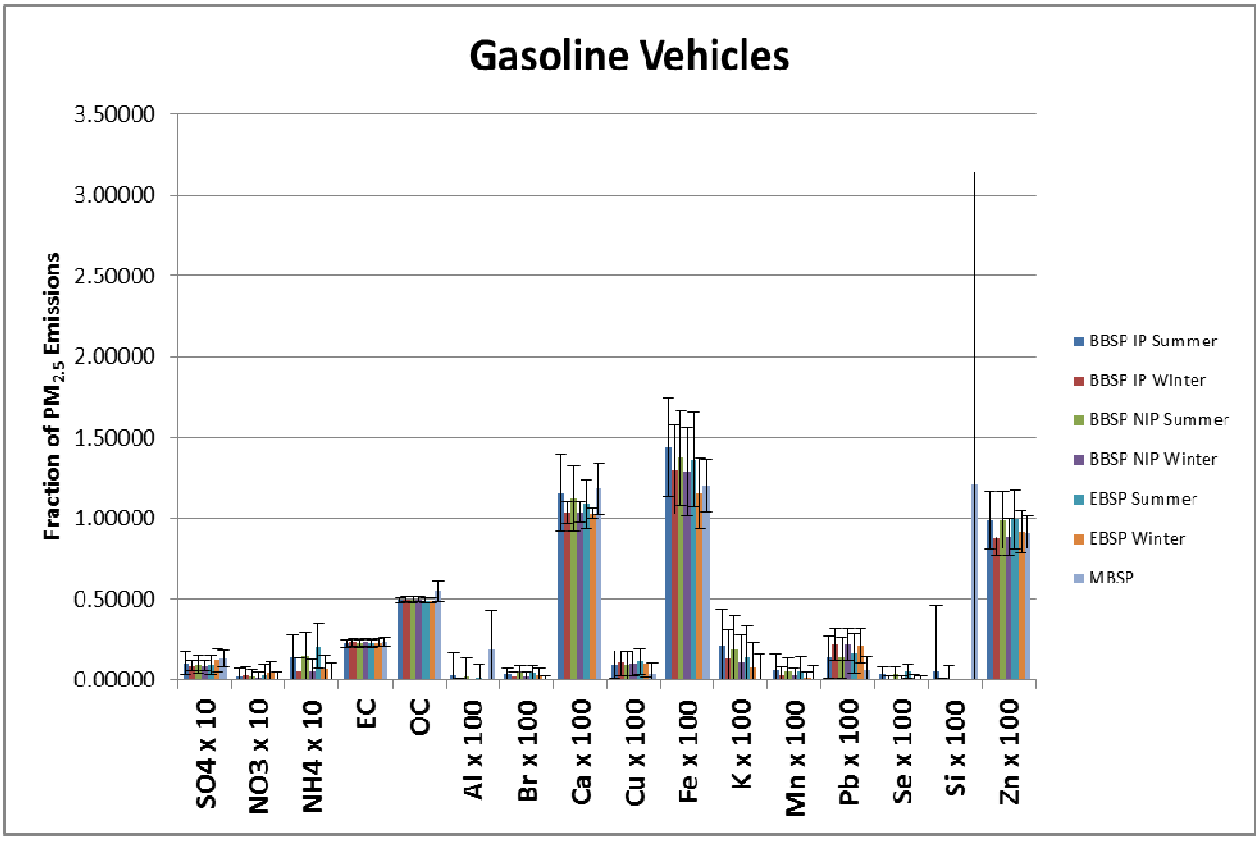
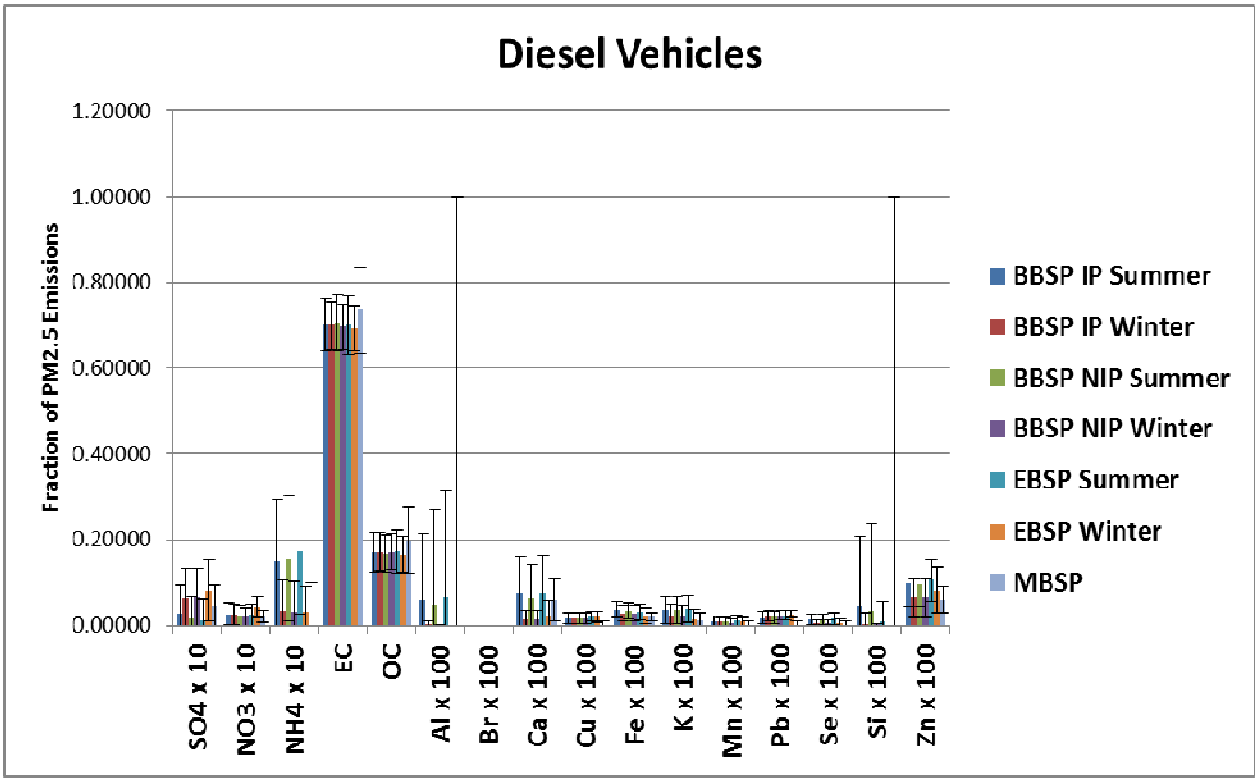
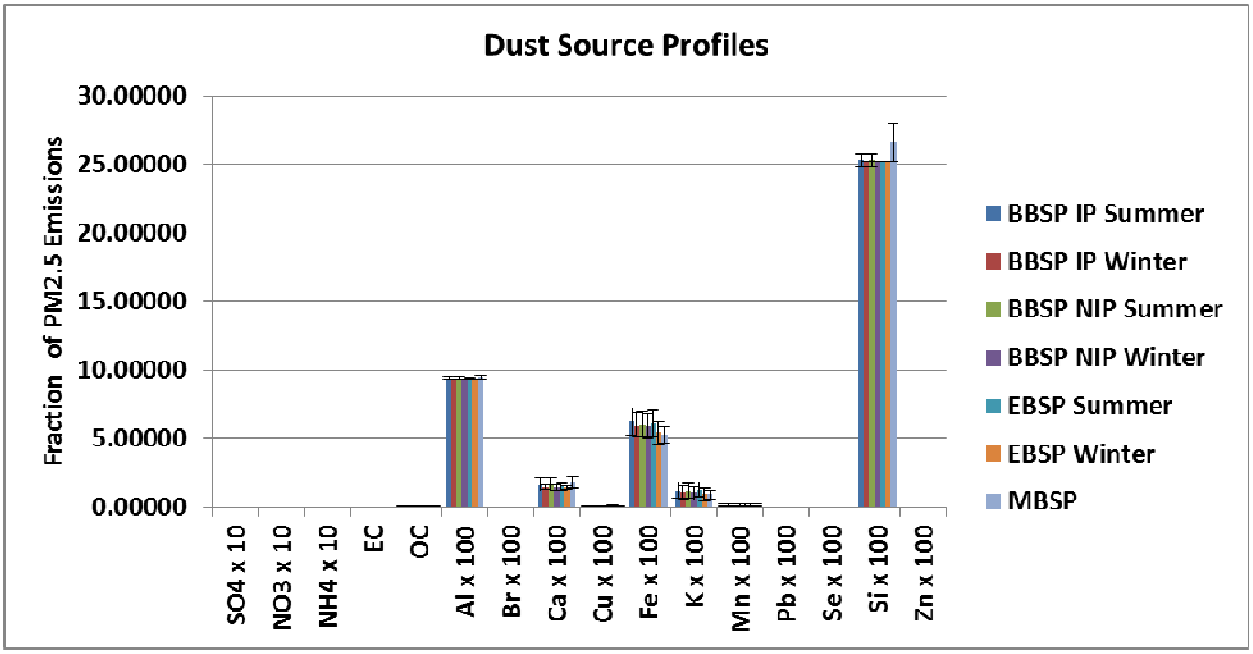
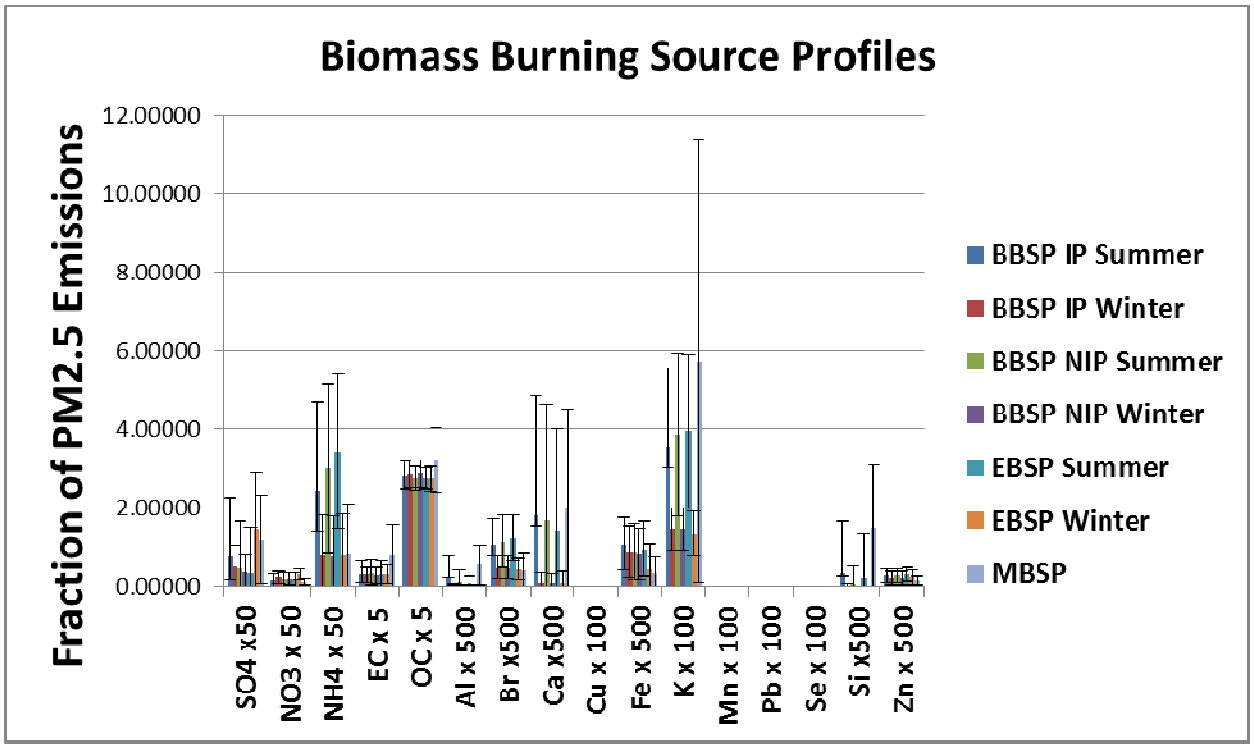


Figure B.1: Example of informative priors.




Diesel Vehicles
Fraction of PM2.5 Emissions
1.20000
1.00000
0.80000
0.60000
0.40000
0.20000
0.00000
SO4 x 10
NO3 x 10
NH4 x 10
EC
OC
Al x 100
Br x 100
Ca x 100
Cu x 100
Fe x 100
K x 100
Mn x 100
Pb x 100
Se x 100
Si x 100
Zn x 100
BBSP IP Summer
BBSP IP Winter
BBSP NIP Summer
BBSP NIP Winter
EBSP Summer
EBSP Winter
MBSP


Dust Source Profiles
Fraction of PM2.5 Emissions
30.00000
25.00000
20.00000
15.00000
10.00000
5.00000
0.00000
SO4 x 10
NO3 x 10
NH4 x 10
EC
OC
Al x 100
Br x 100
Ca x 100
Cu x 100
Fe x 100
K x 100
Mn x 100
Pb x 100
Se x 100
Si x 100
Zn x 100
BBSP IP Summer
BBSP IP Winter
BBSP NIP Summer
BBSP NIP Winter
EBSP Summer
EBSP Winter
MBSP



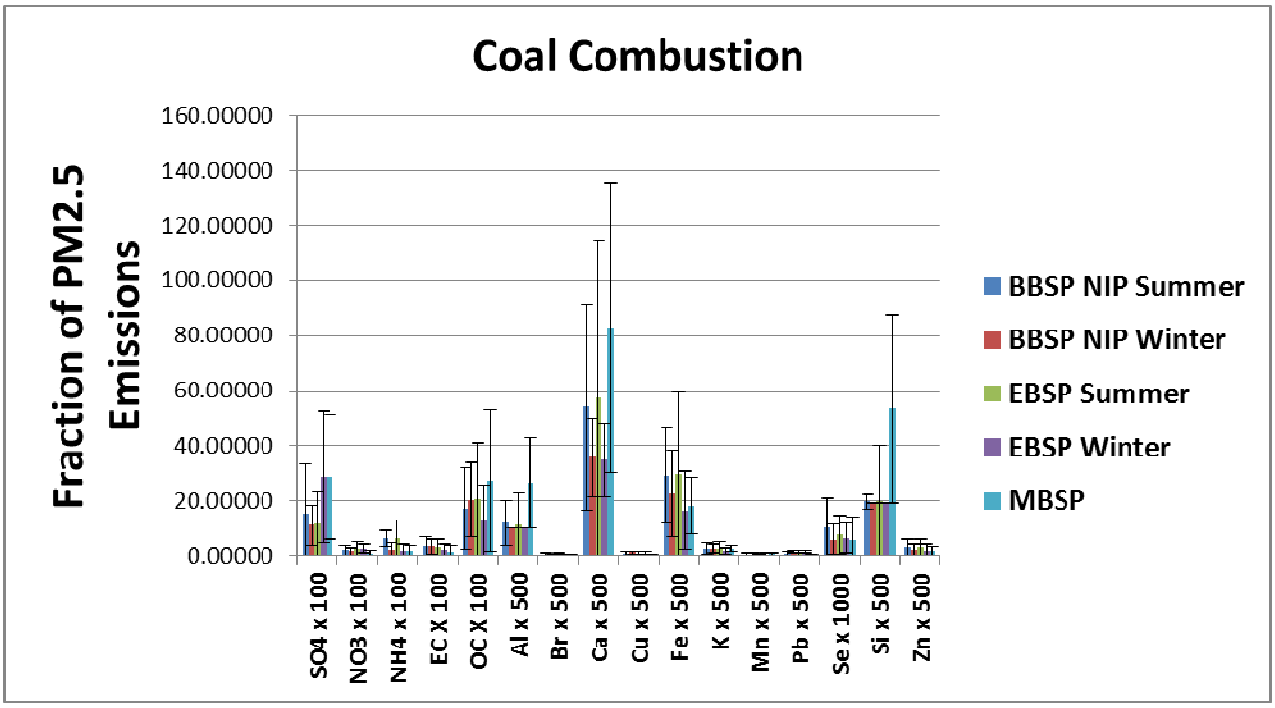


Figure B.2: Source profiles derived from various ensemble methods and compared with MBSPs [*Chow et al.*, 2004b; *Marmur et al.*, 2005; *Zielinska et al.*, 1998b] for (a) GV, (b) DV, (c) DUST, (d) BURN and (e) COAL.

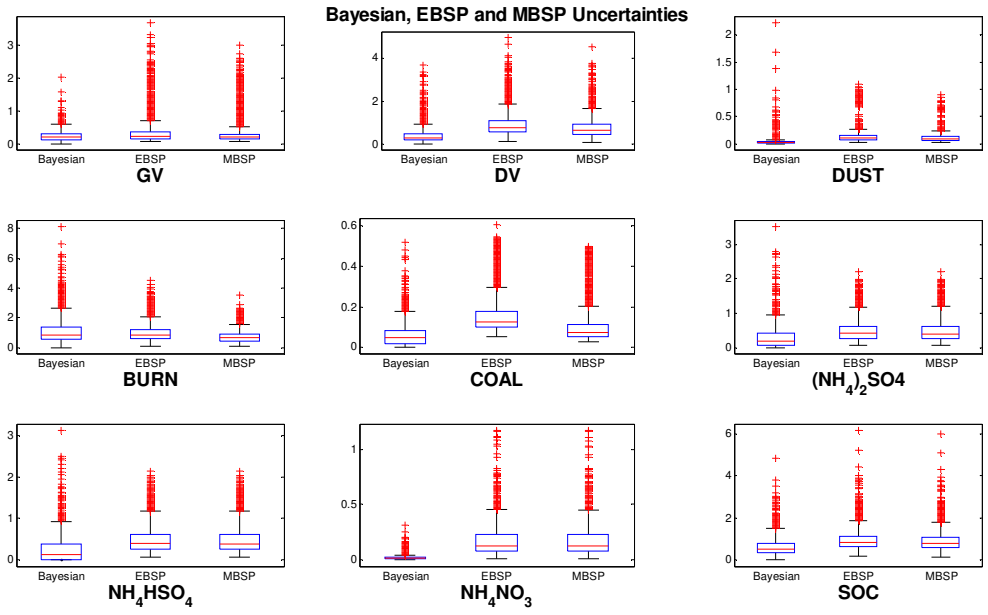
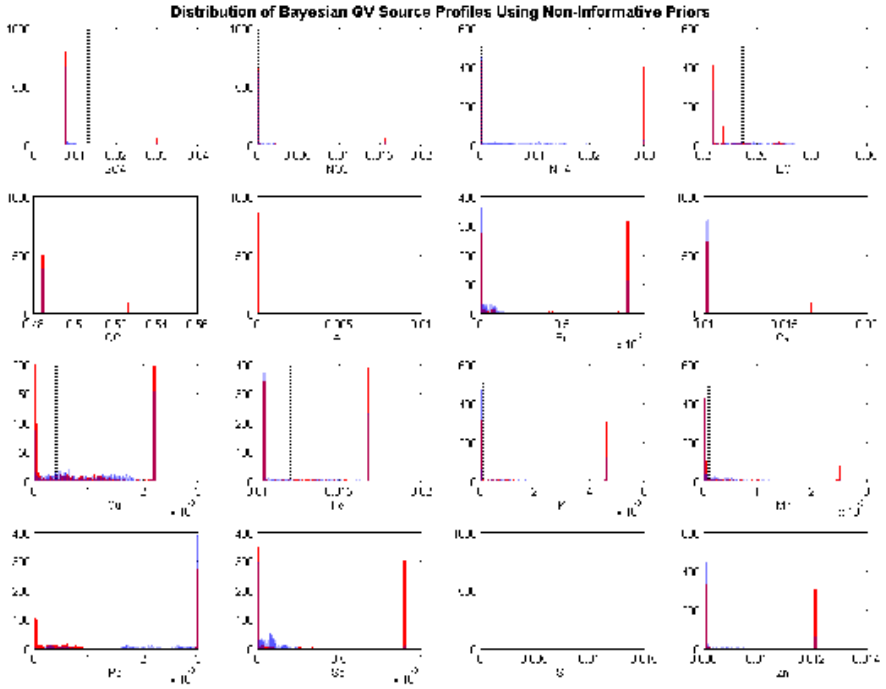
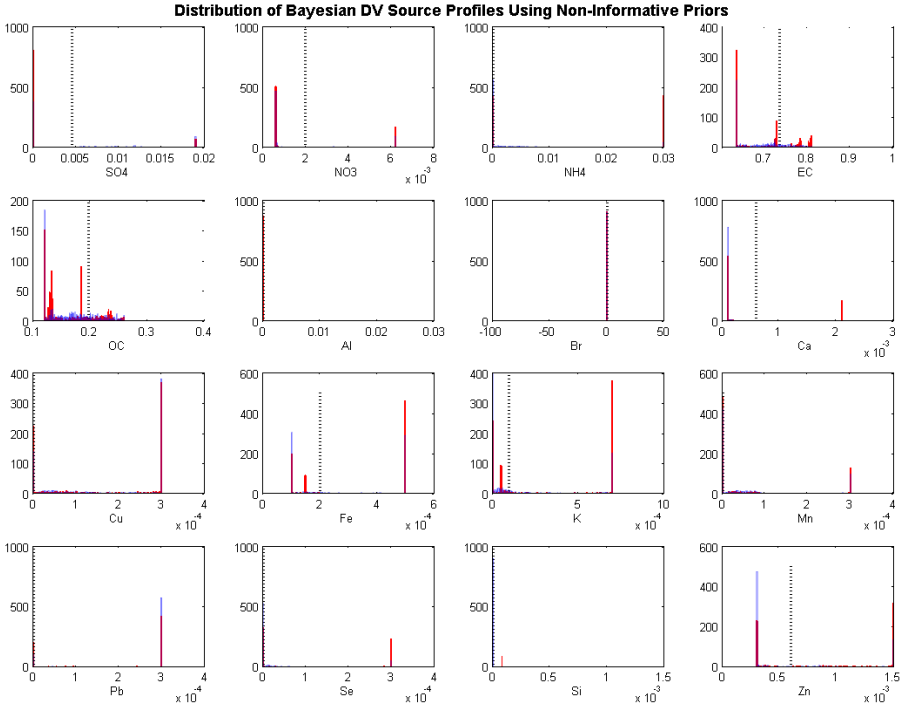
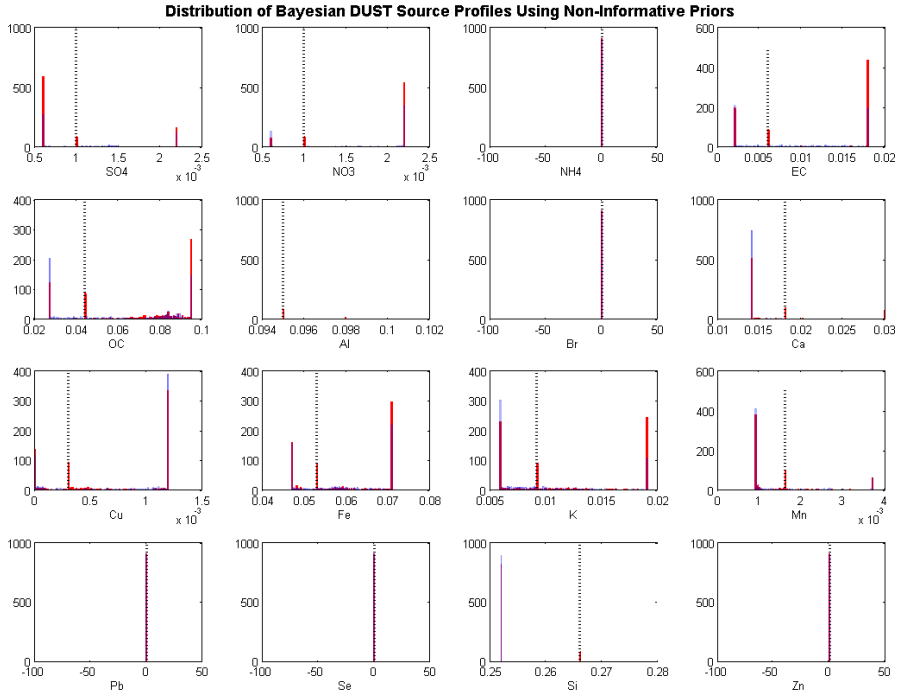
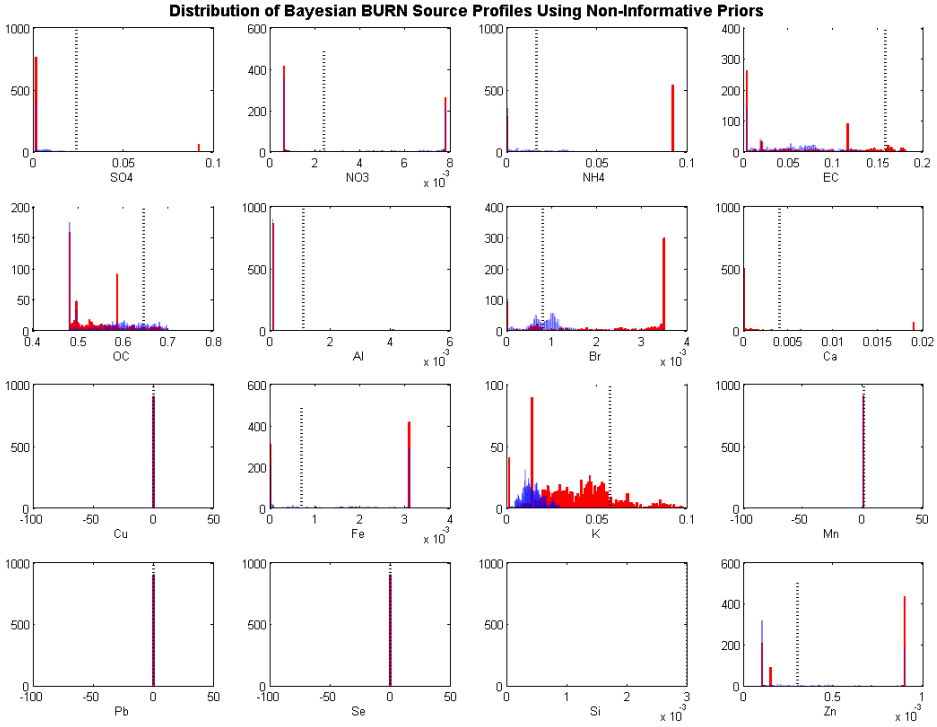


Figure B.3: Comparison of uncertainties for CMB-GC using Bayesian profiles with non-informative priors (BBSP-NIP), standard ensemble (EBSPs) and measurement-based source profiles (MBSPs).


Distribution of Bayesian GV Source Profiles Using Non-Informative Priors


Distribution of Bayesian DV Source Profiles Using Non-Informative Priors
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Distribution of Bayesian DUST Source Profiles Using Non-Informative Priors
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Distribution of Bayesian BURN Source Profiles Using Non-Informative Priors
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn

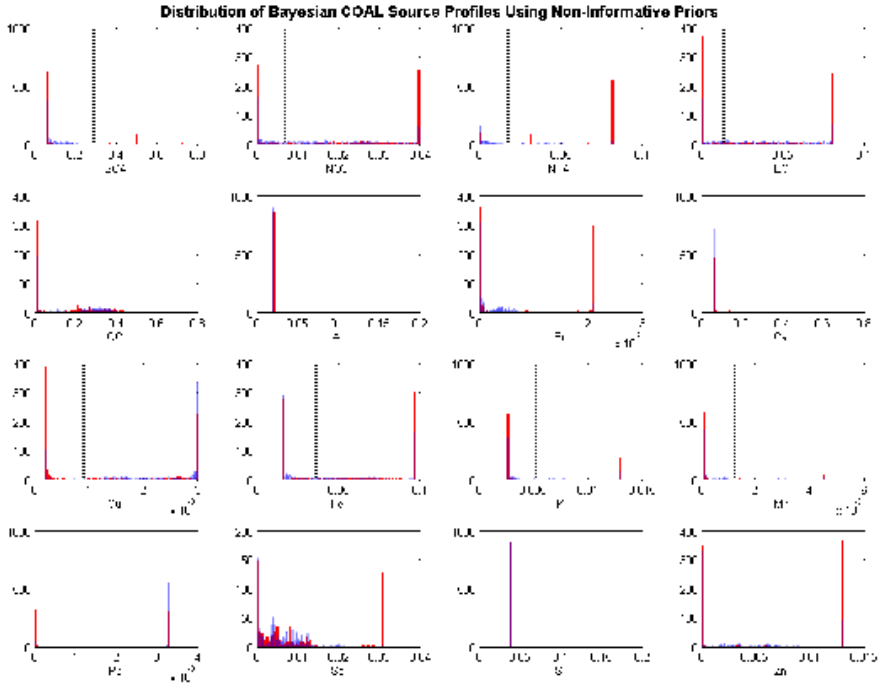
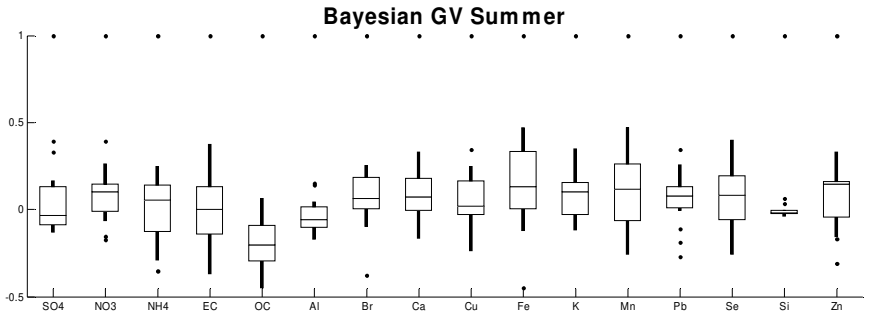
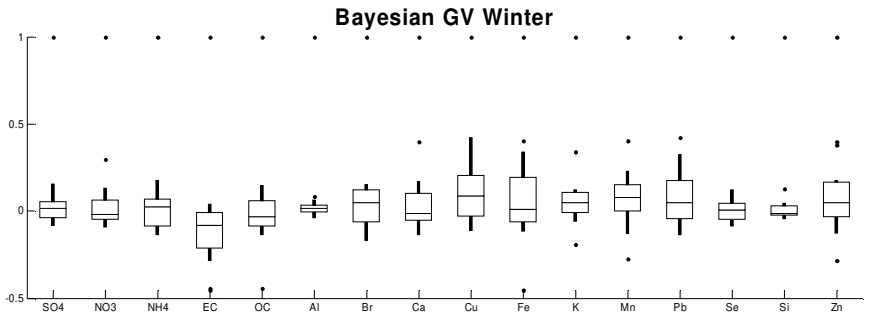
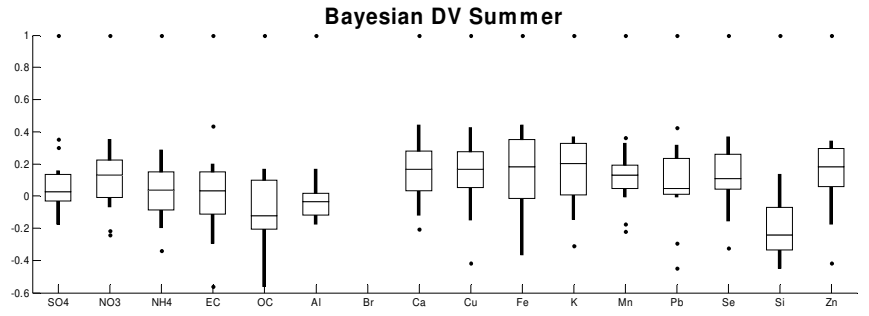
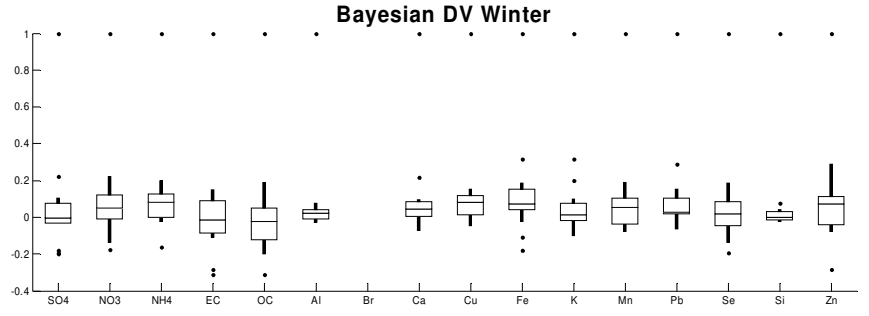


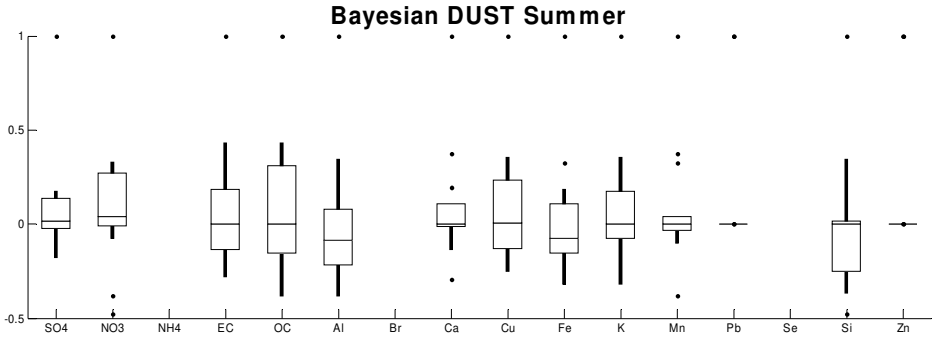
Figure B.4: Histograms of Bayesian COAL source profiles using non-informative priors for (a) GV, (b) DV, (c) DUST, (d) BURN and (e) COAL. Red indicates summer profile, blue indicates winter profile. Dashed line indicates average value of MBSPs.

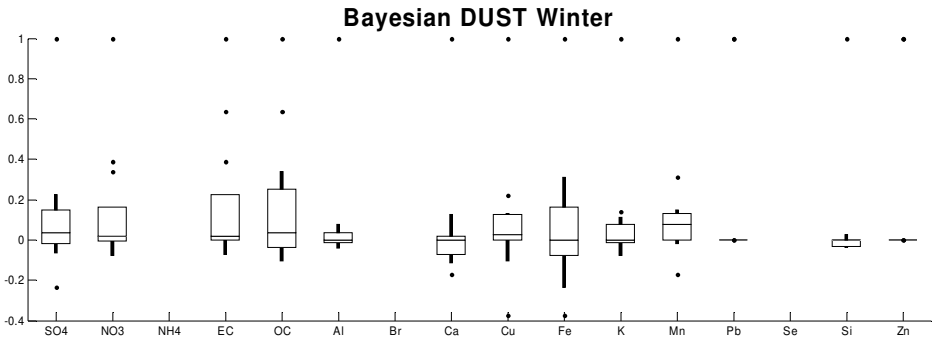

Bayesian GV Summer
1
0.5
0
-0.5
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn

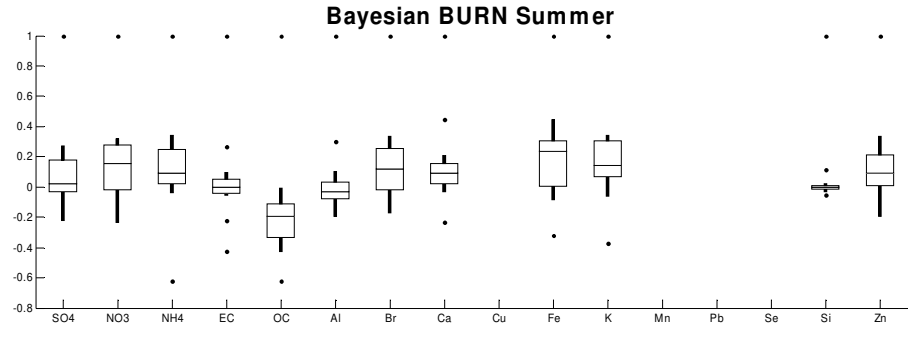

Bayesian GV Winter
1
0.5
0
-0.5
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn

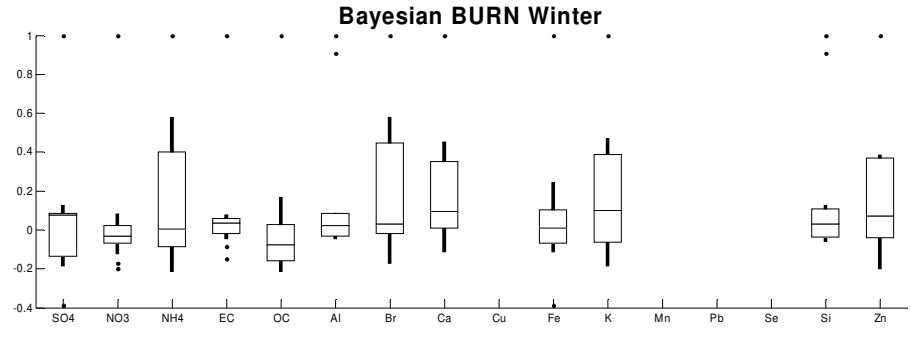

Bayesian DV Summer
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Bayesian DV Winter
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Bayesian DUST Summer
1
0.5
0
-0.5
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Bayesian DUST Winter
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Bayesian BURN Summer
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn


Bayesian BURN Winter
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
SO4
NO3
NH4
EC
OC
Al
Br
Ca
Cu
Fe
K
Mn
Pb
Se
Si
Zn

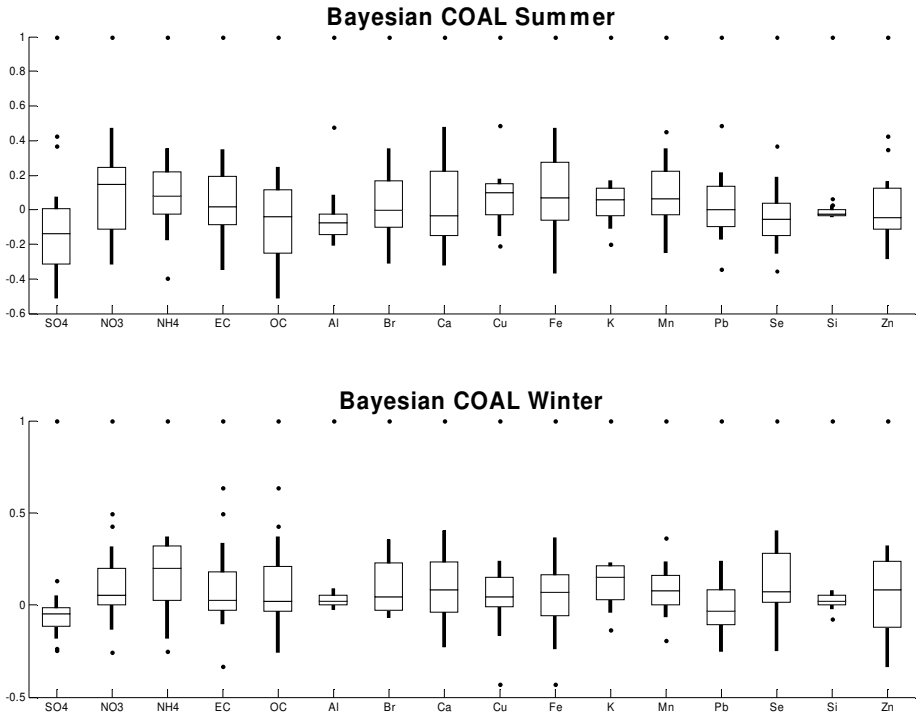


Figure B.5: Boxplots of Bayesian source profiles using non-informative priors for (a) GV, (b) DV, (c) DUST, (d) BURN and (e) COAL.. Boxplots indicate range of correlations of each species in source profile with all other species in profile. NOTE: all values of 1 are self-correlations.

CHAPTER 4: SPECTRAL ANALYSIS OF $PM_{2.5}$ SOURCE APPORTIONMENT METHODS

Sivaraman Balachandran[*,a], Heather A. Holmes[a], James A. Mulholland[a] and Armistead G. Russell [a]

[a]Georgia Institute of Technology. School of Civil and Environmental Engineering. Atlanta, GA.

[*]Corresponding author: 311 Ferst Dr., Atlanta, GA, 30332; phone 206.250.6480, fax 404.894.8266; siv@gatech.edu.

4.1. Abstract

Multiple source apportionment (SA) methods are applied at three receptor sites that measured fine particulate matter ($PM_{2.5}$) composition. These methods include applications of a chemical mass balance with gas constraints (CMB-GC) method using three sets of source profiles and positive matrix factorization (PMF). Source profiles used in CMB-GC include measurement based source profiles (MBSPs), ensemble based source profiles (EBSPs), and Bayesian-based source profile (BBSPs). The EBSPs and BBSPs are derived from ensemble averaging multiple models using a standard and a Bayesian technique, respectively, and then used to derive new source profiles for use in a CMB-GC application. SA is conducted at the Jefferson St. (JST) SEARCH site and the South Dekalb (SDK) CSN site, which are both in Atlanta, GA. We also conduct SA for the rural SEARCH site in Yorkville, GA (YRK). We compare SA method results for JST from 1/3/99-12/31/07, at SDK, from 3/2/01 – 12/10/10, and at YRK, 6/7/98 – 12/29/07. Source impacts from the four SA methods at three sites are compared for temporal trends using spectral analysis. Total reconstructed mass using PMF tends to be biased slightly high whereas that using CMB tends to be biased slightly low. The use of EBSPs and BBSPSs lead to fewer zero impact days as well as stronger seasonal splits for secondary

organic carbon and biomass burning impacts, consistent with expectations. All power spectra derived using the LSPM show a strong peak at one year, independent of SA methods, species and source profiles/factors. Statistically significant peaks (α =0.05) are found for the frequency associated with one week for GV impacts using CMB-GC at JST and both CMB-GC and PMF at SDK. DV impacts estimated using both CMB-GC and PMF at JST and SDK have peaks (α =0.05) for the frequency associated with one week. BURN spectra have the greatest variation intra and inter-method, with low frequency signals at JST and SDK and YRK having both low frequency and weekly signals. Biomass burning profiles/factors have the greatest variability across methods and locations, especially with BBSPs and PMF factors. Across the three sites, OC to EC ratios vary from 3 - 5 in EBSPs, to 3.9 - 17.6 with BBSPs and 3.1 - 10.8 in PMF, suggesting that biomass burning emissions have increased spatial variability as compared to other sources.

4.2. Introduction

Ambient and indoor air pollution is a major suspected cause of premature mortality, and has been associated with more than three million preventative deaths per year worldwide [*Lim et al.*, 2012]. Most of the health impacts estimated are from the effects of fine particulate matter, having aerodynamic diameter less than 2.5 µm ($PM_{2.5}$). $PM_{2.5}$ is comprised of a complex mixture of chemical species, and emitted from a variety of sources. As a result, the health effects of $PM_{2.5}$ may be preferentially dependent on specific species; however, recent work has suggested that health impacts may actually be caused by the net effect of the mixture of pollutants which make up $PM_{2.5}$ [*Solomon et al.*, 2011; *Solomon et al.*, 2012]. Traditional epidemiologic models have generally used $PM_{2.5}$ or individual species in assessing health impacts. Recently, there have been efforts to use source impacts from source apportionment studies as a proxy for multipollutant effects [*Hopke et al.*, 2006; *Ito et al.*, 2006; *Sarnat et al.*, 2008; *Thurston et al.*, 2005].

Source impacts can be quantified using both receptor and chemical transport models (RMs and CTMs), and have both advantages and limitations for their use in health studies. RMs are not computationally intensive, require observational data from a "central" monitor, and can be used easily in time series health studies. A major limitation of RMs is that their results are derived only for the location of the monitor during a period for which sufficient measurements exist. Source impacts, as well as central monitor data, are proxies for exposure, an assumption which may not be accurate given that there is spatial variability in air pollution within a metro area. Another issue with using RMs is that observational data may only be available every third or sixth day, limiting their utility in epidemiologic studies that rely on daily health outcome data. A third issue with RMs is that there are several types of RMs and the magnitude and variability of source impacts differs from model to model. Recently, efforts to use CTM source apportionment (SA) results have addressed some of these issues because CTMs can provide results over a large spatial domain. In addition, they can provide results at a high temporal frequency (e.g. hourly results). They can also model complex atmospheric chemistry and have a greater number of source categories than RMs. However, CTMs require large computational resources, a major limitation when long time series of source impacts are required, and the simulated concentration typically do not agree as closely with observations and receptor models.

Here we use results from multiple $PM_{2.5}$ source apportionment (SA) results at three receptor sites. Two of the three sites are Southeastern Aerosol Research and Characterization (SEARCH) network [*Hansen et al.*, 2003] sites and the third is a Chemical Speciation Trends (CSN) site. The SEARCH sites include Jefferson St. (JST), an urban site located near downtown Atlanta, GA and a rural site at Yorkville, Paulding County, GA (YRK). The CSN site, South Dekalb (SDK), is an urban site located in southeast Dekalb County in the metropolitan Atlanta area. We compare source impacts for the three sites calculated from several receptor models. We compare results from SDK and JST to assess intra-urban differences in SA estimates. We use results from SEARCH sites, JST and YRK, to compare differences in urban versus rural receptor sites. We apply spectral analysis of source impacts and

important tracer species at each of these sites to gain insight into how source apportionment methods vary temporally.

4.3. Methods

CSN network sites span the continental US and measure organic carbon (OC), elemental carbon (EC), ionic species and a suite of elemental species. In addition to $PM_{2.5}$ species, SEARCH sites also measure the gaseous species, CO, SO_2, NO_x and O_3. The analytical methods used in both of these networks are comparable, with the exception of OC and EC. OC and EC are operationally defined parameters and their measured values are consistently different depending on which method is used: thermal optical transmittance (TOT), applying the NIOSH method, or thermal optical reflectance (TOR), applying the IMPROVE method [*Chow et al.*, 2004a]. Prior to May 2007, CSN networks in GA used TOT to measure OC and EC. Subsequently, these CSN network sites began using TOR to measure OC and EC and SDK made this change in April 2009. The entire SEARCH data in this work uses TOR. To account for the changes within SDK dataset and to compare SA results at SDK and SEARCH sites (JST, YRK) with commensurate data, we utilized a regression technique to adjust the TOT-based SDK data to TOR-equivalent values [*Malm et al.*, 2011]. Briefly, the method uses regression coefficients that accounts for sampler type, an additive positive artifact and a multiplicative negative artifact. Using this method, TOT values were converted to TOR equivalent values:

$$EC^{adj} = 1.3 * EC^{TOT} \qquad \text{(Equation 1)}$$

$$OC^{adj} = \frac{\left(OC^{TOT} - 0.3 * EC^{TOT} - a_i\right)}{1 + b_{OC}} \qquad \text{(Equation 2)}$$

where EC^{adj} and OC^{adj} is the TOR equivalent value of EC and OC, respectively and OC^{TOT} and OC^{TOT}, are the TOT-based values, a_i accounts the additive positive artifact, for month *i*, and b_{OC} accounts for the negative artifact [*Malm et al.*, 2011].

In addition to OC and EC, species processed for source apportionment were sulfate, nitrate, ammonium, and the elemental species: Al, As, Ba, Br, Ca, Cl, Cu, Fe, K, Mn, Pb, Sb, Si,

Sn and Zn. All data were processed in a manner similar to other SA work using RMs [*Reff et al.*, 2007] and in companion studies with this work [*Balachandran et al.*, 2012; *Balachandran et al.*, 2013]. As, Ba, Cl, Sb and Sn were not included in the source apportionment because they had a high percentage (>50% of days) with values below detection limit; however, they were included when determining regionally specific source profiles.

Two RM methods and one CTM were applied to provide source impacts at SDK and YRK. A chemical mass balance method (CMB) that utilizes gaseous concentrations to constrain estimated source impacts (referred to as CMB-GC) was used with apriori measurement-based source profiles (MBSPs) [*Marmur et al.*, 2005]. MBSPs used were from previous SA work at JST [*Marmur et al.*, 2005] for nine source categories: gasoline vehicles (GV), diesel vehicles (DV), dust (DUST), biomass burning (BURN), coal combustion (COAL), ammonium sulfate (AMSULF), ammonium bisulfate (AMBSULF), ammonium nitrate (AMNITR) and other OC, which we take to be a surrogate for secondary organic carbon (SOC). Second, positive matrix factorization (PMF), a factor-analytic method, which does not require source profiles was used with solutions ranging from 6 to 10 factors. For the YRK site, fractionated OC and EC data was utilized in the PMF analysis, but not the CMB-GC analysis. With PMF, a GV factor was only derived at JST; at SDK and YRK, only a total motor vehicle factor was derived. We used the result from previous work using the Community Multiscale Air Quality CTM; this work has been used in two previous ensemble studies at JST [*Baek et al.*, 2005] (the SDK site is in the same 36km grid cell as JST). For YRK, we compiled results from *Baek et al.* [2005] in a manner similar to *D Lee et al.* [2009], where source categories from CMAQ were aggregated into the nine source categories used in CMB-GC. Also, at JST, we used results from a CMB method with molecular marker-based observations and source profiles (CMB-MM) [*Zheng et al.*, 2007].

The Bayesian-ensemble averaging method [*Balachandran et al.*, 2013] used in this work has a three step process. First, a weighted average of source impacts from several SA models is calculated using weights sampled from a Bayesian-based posterior distribution. This is done for a short term (i.e. July 2001 to represent summer and January 2002 to represent winter). Second,

these ensemble-averaged source impacts are used to develop source profiles that can be viewed as specific to a location (and season). We develop new source profiles only for primary sources: GV, DV, DUST, BURN and COAL. For secondary pollutant categories (ammonium sulfate, ammonium bisulfate, ammonium nitrate, and secondary organic carbon), we use MBSPs. Here as in the previous work, we develop source profiles from both the standard (non-Bayesian) ensemble [*Balachandran et al.*, 2012] and from the Bayesian ensemble [*Balachandran et al.*, 2013] referred to as ensemble-based source profiles (EBSPs) and Bayesian-Based source profiles (BBSPs), respectively. For the Bayesian ensemble, we use non-informative priors [*Balachandran et al.*, 2013]. Third, the new source profiles are used in an application of CMB-GC to a long-term data set. Details of this method can be found in the companion studies to this work [*Balachandran et al.*, 2012; *Balachandran et al.*, 2013; *D Lee et al.*, 2009].

Ultimately, we wish to gain insight into how different SA techniques might vary spatially (urban, near roadway and rural sites) and temporally. In addition, since source impacts cannot be directly measured, analysis of spatial and temporal trends provides an indication of the relative reasonableness of the different estimates. In this work, we consider two main issues related to variability: the impact of various methods (CMB-GC vs. PMF in this work) and the impact of different source profiles on a particular method (CMB-GC with MBSPs, EBSPs and BBSPs). First, we compare source apportionment results at JST, SDK and YRK. We compare these differences over a range of metrics, including overall mass closure and seasonal averages of source impacts and source impact uncertainties, correlations with tracer species in a manner similar to Balachandran et al. [2013] and the differences in the derived source profiles. CMAQ results are excluded from this analysis because they are only available for the short term periods of July 2001 and January 2002.

We assess temporal variability by conducting spectral analyses using SA results from four methods: MB-GC with non-informative BBSPs, EBSPS, and MBSPs, and PMF. Spectral analysis can be used to determine dominant frequency patterns underlying noisy time series and are typically conducted using fast Fourier transforms (FFTs) (e.g. [*Liu et al.*, 2005]). Since JST

has daily data, the maximum frequency that can be resolved would be 2 days. However, the data at both SDK and YRK are only gathered every three days, so the power spectra generated using FFT can resolve up to a maximum frequency of six days (i.e. the Nyquist frequency). While FFTs can be run for any length of data, the data should be continuous. This condition is typically not met with environmental data and approximately 5-15% of days at JST, SDK and YRK have incomplete results. FFTs can be utilized by interpolating missing values; however, this can add noise to the spectral analysis. For discontinuous, or unevenly spaced data, a commonly employed method is to use a least squares spectral analysis technique (eg. [*Lomb*, 1976; *Vanicek*, 1969]). The method developed by Lomb [*Lomb*, 1976] and further refined by Scargle [*Scargle*, 1982] is referred to in this work as the Lomb-Scargle Periodogram Method (LSPM). We implement this method following previous work [*Press et al.*, 2001; *Pytharouli and Stiros*, 2008] and by modifying a Matlab freeware code [*Shoelson*, 2001].

For a data set $h_j \equiv h(t_j),\ j = 1.\cdots N$, where N is the number of samples , $\bar{h}$ is the mean and σ^2 is the variance, the LSPM defines a normalized periodogram to estimate power, P_N , at different angular frequencies, $\omega \equiv 2\pi f$, [*Press et al.*, 2001]:

$$P_N(\omega) \equiv \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_j^N (h_j - \bar{h})\cos\omega(t_j - \tau)\right]^2}{\sum_j^N \cos^2\omega(t_j - \tau)} + \frac{\left[\sum_j^N (h_j - \bar{h})\sin\omega(t_j - \tau)\right]^2}{\sum_j^N \sin^2\omega(t_j - \tau)}\right\} \quad \text{(Equation 3)}$$

where τ is an offset that is defined as:

$$\tan(2\omega\tau) \equiv \frac{\sum_j \sin(2\omega\tau_j)}{\sum_j \cos t(2\omega\tau_j)} \quad \text{(Equation 4)}$$

The LSPM has several properties that are appealing. First, the method can be applied to unevenly spaced data. Second, if one assumes that data is composed of a sum of period signals and white (i.e., independent or Gaussian) noise, the LSPM can quantify the statistical significance of a particular peak [*Press et al.*, 2001]. Third, since the Nyquist frequency does not

represent an upper limit for unevenly spaced data, the LSPM can give statistically significant peaks at frequencies above the Nyquist frequency [*Pytharouli and Stiros*, 2008].

At JST, CMB-GC was run for data that was available at JST from 8/1/98 – 12/31-07 (from Balchandran et al. [2013]). However, PMF results were only available starting from January 3, 1999 since that is the start date for fractionated OC and EC data, which was needed to separate gasoline and diesel vehicles using PMF. Therefore, we compare SA method results for JST from 1/3/99-12/31/07. At SDK, results are available at a three day interval from 3/2/01 – 12/10/10, with approximately 13% of days having missing data. At YRK, data are available at a three day interval from 6/7/98 – 12/29/07 with about 5% of days with missing data. Also, YRK has daily data from 9/6/98 – 1/2/00. However, similar to JST, fractionated OC and EC data were only available starting on 5/19/99; therefore, our analysis uses a time period from 5/19/99-12/29/07.

It should be noted that we limit our results and discussion to the GV, DV, BURN and SOC source categories because RM results are very similar for sulfate, nitrate and ammonium and the DUST and COAL contributions are relatively small compared to GV, DV, BURN and SOC. We define summer to be April through September and winter to be October through March.

4.4. Results and Discussion

4.4.1. Source impacts

Receptor models typically reconstruct total mass very close to the total measured mass due to model constraints (Table 4.1). At all three sites, the use of BBSPs and EBSPs resulted in higher reconstructed to measured mass ratios than using MBSPs. At JST, the average ratio of reconstructed to measured total mass ranges from 0.87 using CMB-GC with MBSPs to 0.94 with BBSPs. At SDK, this range is from 0.88 to 0.97 using, MBSPs and BBSPs, respectively. At YRK, reconstructed mass is lower, ranging from 0.77 with MBSPs and 0.81 with BBSPs. At all

three sites, PMF total mass is biased high, with an average ratio of 1.1, 1.09, and 1.05 at JST, SDK and YRK, respectively. All CMB-GC SA results have similar reduced chi-square statistics, a goodness-of-fit metric, although values are about an order of magnitude higher at SDK. BBSPs also have fewer zero impact days than EBSPs or MBSPs. As expected, CMB-GC source impact results have higher correlations with each other as compared to PMF (Tables C.1-C.3).

At both JST and SDK, DV impacts are higher than GV impacts (Table 4.2). DV impacts do not have the distinct seasonality shown by GV impacts, which are higher in winter. This seasonality is ostensibly attributable to increased emissions from cold weather starts in gasoline powered engines and reduced dispersion. When using EBSPs and MBSPs, GV and DV impacts are higher in both seasons at JST than with BBSPs. At YRK, mobile source impacts are significantly lower than at the JST or SDK, when using CMB-GC (regardless of source profiles used), indicative of lower mobile source emissions in rural areas. However, with PMF, mobile source estimates are comparable to JST and SDK, and suggest that the PMF mobile source factor at YRK include other sources.

BURN impacts show strong seasonality at both urban sites for all methods except for CMB-GC using MBSPs (Table 4.2). At JST, the use of MBSPs results in slightly lower BURN impacts in summer (0.97 $\mu g\ m^{-3}$) versus winter (1.12 $\mu g\ m^{-3}$) while at SDK they are approximately equal in summer (1.01 $\mu g\ m^{-3}$) and winter (0.95 $\mu g\ m^{-3}$). With BBSP, EBSPs and PMF, average BURN impacts are ~1 $\mu g\ m^{-3}$ in the summer and ~3-4 $\mu g\ m^{-3}$ in the winter. BURN impacts are much lower at YRK than the urban sites at JST and SDK. In addition, there is not as much seasonality at YRK. The lower impacts at YRK are also present in CMAQ results, suggesting that this difference is most likely due to different emission intensities in rural and urban areas. CMB-GC results using BBSPs are less correlated with EBSPs and MBSPs for BURN than other source categories.

In a manner similar to previous work [*Balachandran et al.*, 2013], we compare BURN impacts to levoglucosan measured at SDK in 2007 that was part of a campaign which included measurements of WSOC and water-soluble potassium (K^+). Biomass burning impacts using

BBSPs and EBSPs have the highest correlation (R^2=0.66 and 0.69, respectively) between BURN impacts and levoglucosan, while SA using MBSPs and PMF have an R^2 of approximately of 0.30 and 0.55, respectively (Figure 4.1). All SA methods had high correlations between BURN impacts and water soluble potassium (K^+) ranging from R^2=0.79 with BBSPs and R^2=0.88 with PMF. However, all SA methods sum of BURN +SOC impacts correlated much lower with WSOC, ranging from R^2=0.34 – 0.51. This was surprising since the sum of BURN+SOC from JST correlated higher with WSOC (R^2=0.68 – 0.76) [*Balachandran et al.*, 2013].

SOC impacts are comparable between JST and YRK (2.63 and 2.41 in summer, and, 1.43 and 1.72 in winter, respectively) when using BBSPs. As shown in previous ensemble studies [*Balachandran et al.*, 2012; *Balachandran et al.*, 2013; *D Lee et al.*, 2009], CMB-GC with MBSPs seems to overestimate SOC in the winter, having higher estimates in winter at both JST and SDK. However, SOC should be higher in summer, when there is increased photochemistry and higher biogenic VOC emissions [*Zheng et al.*, 2002]. With BBSPs and EBSPs, SOC impacts are higher in summer. However, SOC impacts are lower, and have greater zero impact days, at SDK than JST, an unexpected result given the regional nature of SOC. This result is likely due to SDK having lower adjusted OC concentrations (~1 μg m^{-3} lower) than at JST. JST and SDK in have 726 overlapping days in this analysis. For these overlapping days, the TOR based OC at JST and adjusted OC SDK have means and standard deviations of 4.05±2.17 and 3.28±1.83 μg m^{-3}, respectively (Figure 4.2), suggesting that the OC artifact correction method at SDK could be overcorrecting and leading to lower values of OC. This may also explain the low correlation of BURN +SOC impacts with WSOC at SDK.

Impacts using CMB-GC have the highest correlations across all three sites (Tables C.1-C.3). Zn has higher correlation at JST and SDK and in general, has highest correlation with GV impacts using MBSP or BBSP. OC has higher correlations with PMF GV than CMB-GC, especially so at YRK. GV impacts using PMF are likely overestimated at YRK since they are on the same scale as at JST and SDK, which are urban sites with much greater motor vehicle traffic

and emissions. EC tends to have highest correlations with both CMB-GC with MBSPs and BBSPs.

4.4.2. Spectral Analysis of Source Impacts

Power spectra derived using the LSPM all show a strong peak at one year, independent of SA methods, species and source profiles/factors. The impact of methods (i.e. CMB vs. PMF) has a greater impact than on spectral results different source profiles in CMB-GC.

4.4.2.1. Gasoline Vehicles (GV)

Statistically significant peaks (α =0.05) are found for the frequency associated with one week for GV and DV at JST and SDK for most methods, but not at the rural YRK; nevertheless, even when statistical significance was not achieved, spectral peaks at one week frequency were noticeable. CMB-GC frequency spectra show strong yearly and weekly cycles for GV at JST (Figure 4.3) due in part to the Zn in GV source profiles, as Zn has a strong weekly peak. The lack of a strong weekly peak in PMF GV impacts suggests that the GV factor in PMF may be comprised of multiple collinear sources.

At SDK, the spectra has a statistically significant frequency associated with one week for CMB-GC with BBSPs and PMF, though, all SA methods have a noticeable peak associated with ~7 days (Figure C.1). A second peak for all SA methods is also noticeable (but not statistically significant at α=0.05) at ~7.6 days. This is due, in part, to a small frequency shift from sampling every three days along with expected weekly traffic patterns. At YRK, all GV peaks are dominated by long term frequencies (~>90 days); a weekly peak is evident, but not statistically significant at (α=0.05) (Figure C.2).

4.4.2.2. Biomass Burning (BURN)

BURN frequency spectra have the greatest variation intra and inter-method. At JST, CMB-GC BURN impacts using both BBSPs and EBSPs result in low frequency signals, (~73-390 days and ~66-390 days with BBSPs and EBSPs, respectively) (Figure C.3). CMB-GC with

MBSPs and PMF both follow the spectra of potassium, and have higher frequency signals ranging from ~13 days. At SDK, all methods and K, show long term signals (~61 days in K to 3570 days in PMF); no spectral signals were statistically significant when using CMB-GC with MBSPs (Figure C.4). At YRK, CMB-GC with BBSPs shows a statistically significant signal at approximately 25 days (peak at 0.04 days $^{-1}$); CMB-GC with EBSPs and MBSPs also show this, but not at α=0.05 significance (Figure 4.4). All methods show a weekly signal, though not significant at α=0.05, likely due to a weekly signal in K, (also not at α=0.05 significance). This weekly signal in K is not apparent at the urban sites.

4.4.2.3. Diesel Vehicles (DV) and Secondary Organic Carbon (SOC)

DV results at JST and SDK are more consistent than GV. At both sites, all SA methods show strong yearly, weekly and intermediary peaks (Figure C.5-C.7). This is largely driven by the spectral signal of EC. Therefore, for major $PM_{2.5}$ constituents that are dominated by a single source, power spectra of source impacts are similar across various SA methods. SOC impacts for all methods have low frequency signals (Figures C.8-C.10). These lower frequency signals, on the order of ~75 days or greater, suggest that temporal variability for SOC, as expected, is associated with time scales indicative of regional and secondary sources. The consistency across methods suggests that the temporal variability of DV and SOC impacts are reasonably captured.

4.4.3. Source Profile Comparison

Since both CMB and PMF solve the same mass balance equation, the main difference between these two methods is that in the former, SA is conducted using a priori source profiles (MBSPs, EBSPs and BBSPs), while in the latter factors intrinsic to each data set are developed. At each of the three sites, two seasonal EBSPs and BBSPs and one PMF factor were developed for each source. Including the MBSPs, which are the same at each site, a total of 16 profiles/factors were used in this study.

Comparison of the GV profiles/factors shows that differences are greater with EBSPs and PMF than BBSPs (Figure 4.5, Table C.5). For example, OC to EC ratios in GV and EC to OC ratios in DV are similar at all three sites with BBSPs (Table 4.3). This should be expected for GV and DV, since fleet characteristics at the three locations should be similar. At both JST, the EBSP OC to EC ratio is about 4 and 1.5 in the winter and summer, respectively. This seasonal split was also found in an earlier ensemble study [*D Lee et al.*, 2009]. However, this seasonal split is not evident using BBSPs or with EBSPs at SDK or YRK. The PMF GV factor at JST had an OC to EC ratio of 3.7. PMF motor vehicle profiles had OC:EC ratios of 0.98 and 2.25 at SDK and YRK, respectively. There was also seasonality in DV OC to EC ratios using EBSPs at YRK with values of 0.2 and 0.5 in winter and summer, respectively.

BURN profiles/factors have the greatest variability across methods and location, especially with BBSPs and PMF factors. Across the three sites, OC to EC ratios vary from 3 to 5 in EBSPs, but vary from 3.9 to 17.6 with BBSPs and 3.1 to 10.8 in PMF (Table 4.3). This suggests that biomass burning emissions have increased spatial and temporal variability as compared to motor vehicles. However, for BBSPs, this may also be an artifact of the source profile derivation, where anywhere from ~30% to 50% of the source profiles had low values of EC in source profiles (e.g. 0.003 µg EC/ µg $PM_{2.5}$). In addition, all BURN profiles/factors had lower levels of potassium than in the MBSP, suggesting that the potassium emitted from biomass burning is overestimated in the MBSPs.

4.5. Conclusions

In this work, we conduct spectral analysis of source impacts and related tracers at three receptor sites to gain insight into how source apportionment methods vary temporally and spatially. PMF total reconstructed mass is biased high, with an average ratio of 1.1, 1.09, and 1.05 at JST, SDK and YRK, respectively, versus CMB-GC, which has ratios ranging from 0.77 - 0.97. The use of EBSPs and BBSPSs lead to few zero impact days as well as stronger seasonal splits for secondary organic carbon and biomass burning impacts. However, SOC impacts are

lower, and have greater zero impact days, at SDK than JST, an unexpected result given the regional nature of SOC. This result is likely due to SDK having lower adjusted OC concentrations (~1 $\mu g\ m^{-3}$ lower) than at JST.

Source impacts from the four SA methods are compared for temporal trends using spectral analysis, using a method developed by Lomb [*Lomb*, 1976] and further refined by Scargle [*Scargle*, 1982] and referred to in this work as the Lomb-Scargle Periodogram Method (LSPM). All power spectra derived using the LSPM show a strong peak at one year, independent of SA methods, species and source profiles/factors. Statistically significant peaks (α =0.05) are found for the frequency associated with one week for GV at JST using CMB-GC, but not with PMF. This suggest that the PMF factor attributed to GV may reflect a mixture sources, highlighting the importance of carefully evaluating PMF factors. Statistically significant peaks (α =0.05) are found for the frequency associated with one week for DV spectra at JST and SDK. At YRK, mobile source spectra do not have statistically significant peaks associated with on week, which contrasts with the urban JST and SDK sites that are likely impacted by weekly commute traffic patterns. BURN spectra have the greatest variation intra and inter-method, with low frequency signals at JST and SDK and YRK having both low frequency and weekly signals. Biomass burning profiles/factors have the greatest variability across methods and locations, especially with BBSPs and PMF factors. OC to EC ratios vary from 3 - 5 in EBSPs, to 3.9 - 17.6 with BBSPs and 3.1 - 10.8 in PMF, suggesting that biomass burning emissions have increased spatial variability as compared to other sources.

Acknowledgements

This publication was made possible in part by USEPA STAR grants R833626, R833866, R834799 and RD83479901. 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. We also acknowledge the Southern Company for their support and thank Eric Edgerton of ARA, Inc. for access to the SEARCH data. We also acknowledge the contribution of our colleagues at the Rollins School of Public Health at Emory University for helpful discussions.

4.6. References

1. Lomb, N. R., LEAST-SQUARES FREQUENCY-ANALYSIS OF UNEQUALLY SPACED DATA. *Astrophysics and Space Science* **1976,** *39*, (2), 447-462.
2. Scargle, J. D., STUDIES IN ASTRONOMICAL TIME-SERIES ANALYSIS .2. STATISTICAL ASPECTS OF SPECTRAL-ANALYSIS OF UNEVENLY SPACED DATA. *Astrophys. J.* **1982,** *263*, (2), 835-853.
3. Lim, S. S.; Vos, T.; Flaxman, A. D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Amann, M.; Anderson, H. R.; Andrews, K. G.; Aryee, M.; Atkinson, C.; Bacchus, L. J.; Bahalim, A. N.; Balakrishnan, K.; Balmes, J.; Barker-Collo, S.; Baxter, A.; Bell, M. L.; Blore, J. D.; Blyth, F.; Bonner, C.; Borges, G.; Bourne, R.; Boussinesq, M.; Brauer, M.; Brooks, P.; Bruce, N. G.; Brunekreef, B.; Bryan-Hancock, C.; Bucello, C.; Buchbinder, R.; Bull, F.; Burnett, R. T.; Byers, T. E.; Calabria, B.; Carapetis, J.; Carnahan, E.; Chafe, Z.; Charlson, F.; Chen, H.; Chen, J. S.; Cheng, A. T.-A.; Child, J. C.; Cohen, A.; Colson, K. E.; Cowie, B. C.; Darby, S.; Darling, S.; Davis, A.; Degenhardt, L.; Dentener, F.; Des Jarlais, D. C.; Devries, K.; Dherani, M.; Ding, E. L.; Dorsey, E. R.; Driscoll, T.; Edmond, K.; Ali, S. E.; Engell, R. E.; Erwin, P. J.; Fahimi, S.; Falder, G.; Farzadfar, F.; Ferrari, A.; Finucane, M. M.; Flaxman, S.; Fowkes, F. G. R.; Freedman, G.; Freeman, M. K.; Gakidou, E.; Ghosh, S.; Giovannucci, E.; Gmel, G.; Graham, K.; Grainger, R.; Grant, B.; Gunnell, D.; Gutierrez, H. R.; Hall, W.; Hoek, H. W.; Hogan, A.; Hosgood Iii, H. D.; Hoy, D.; Hu, H.; Hubbell, B. J.; Hutchings, S. J.; Ibeanusi, S. E.; Jacklyn, G. L.; Jasrasaria, R.; Jonas, J. B.; Kan, H.; Kanis, J. A.; Kassebaum, N.; Kawakami, N.; Khang, Y.-H.; Khatibzadeh, S.; Khoo, J.-P.; Kok, C.; Laden, F.; Lalloo, R.; Lan, Q.; Lathlean, T.; Leasher, J. L.; Leigh, J.; Li, Y.; Lin, J. K.; Lipshultz, S. E.; London, S.; Lozano, R.; Lu, Y.; Mak, J.; Malekzadeh, R.; Mallinger, L.; Marcenes, W.; March, L.; Marks, R.; Martin, R.; McGale, P.; McGrath, J.; Mehta, S.; Mensah, G. A.; Merriman, T. R.; Micha, R.; Michaud, C.; Mishra, V.; Hanafiah, K. M.; Mokdad, A. A.; Morawska, L.; Mozaffarian, D.; Murphy, T.; Naghavi, M.; Neal, B.; Nelson, P. K.; Nolla, J. M.; Norman, R.; Olives, C.; Omer, S. B.; Orchard, J.; Osborne, R.; Ostro, B.; Page, A.; Pandey, K. D.; Parry, C. D. H.; Passmore, E.; Patra, J.; Pearce, N.; Pelizzari, P. M.; Petzold, M.; Phillips, M. R.; Pope, D.; Pope Iii, C. A.; Powles, J.; Rao, M.; Razavi, H.; Rehfuess, E. A.; Rehm, J. T.; Ritz, B.; Rivara, F. P.; Roberts, T.; Robinson, C.; Rodriguez-Portales, J. A.; Romieu, I.; Room, R.; Rosenfeld, L. C.; Roy, A.; Rushton, L.; Salomon, J. A.; Sampson, U.; Sanchez-Riera, L.; Sanman, E.; Sapkota, A.; Seedat, S.; Shi, P.; Shield, K.; Shivakoti, R.; Singh, G. M.; Sleet, D. A.; Smith, E.; Smith, K. R.; Stapelberg, N. J. C.; Steenland, K.; Stöckl, H.; Stovner, L. J.; Straif, K.; Straney, L.; Thurston, G. D.; Tran, J. H.; Van Dingenen, R.; van Donkelaar, A.; Veerman, J. L.; Vijayakumar, L.; Weintraub, R.; Weissman, M. M.; White, R. A.; Whiteford, H.; Wiersma, S. T.; Wilkinson, J. D.; Williams, H. C.; Williams, W.; Wilson, N.; Woolf, A. D.; Yip, P.; Zielinski, J. M.; Lopez, A. D.; Murray, C. J. L.; Ezzati, M., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **2012,** *380*, (9859), 2224-2260.

4. Solomon, P. A.; Costantini, M.; Grahame, T. J.; Gerlofs-Nijland, M. E.; Cassee, F. R.; Russell, A. G.; Brook, J. R.; Hopke, P. K.; Hidy, G.; Phalen, R. F.; Saldiva, P.; Sarnat, S. E.; Balmes, J. R.; Tager, I. B.; Ozkaynak, H.; Vedal, S.; Wierman, S. S. G.; Costa, D. L., Air pollution and health: bridging the gap from sources to health outcomes: conference summary. *Air Qual. Atmos. Health* **2012,** *5*, (1), 9-62.
5. Solomon, P. A.; Wexler, A. S.; Sioutas, C., Special Issue of Atmospheric Environment for Air Pollution and Health: Bridging the Gap from Sources-to-Health Outcomes Preface. *Atmospheric Environment* **2011,** *45*, (40), 7537-7539.
6. Hopke, P. K.; Ito, K.; Mar, T.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Paatero, P.; Thurston, G. D., PM source apportionment and health effects: 1. Intercomparison of source apportionment results. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (3), 275-286.
7. Ito, K.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Neas, L.; Hopke, P. K.; Thurston, G. D., PM source apportionment and health effects: 2. An investigation of intermethod variability in associations between source-apportioned fine particle mass and daily mortality in Washington, DC. *Journal of Exposure Science and Environmental Epidemiology* **2006,** *16*, (4), 300-310.
8. Sarnat, J. A.; Marmur, A.; Klein, M.; Kim, E.; Russell, A. G.; Sarnat, S. E.; Mulholland, J. A.; Hopke, P. K.; Tolbert, P. E., Fine particle sources and cardiorespiratory morbidity: An application of chemical mass balance and factor analytical source-apportionment methods. *Environmental Health Perspectives* **2008,** *116*, (4), 459-466.
9. Thurston, G. D.; Ito, K.; Mar, T.; Christensen, W. F.; Eatough, D. J.; Henry, R. C.; Kim, E.; Laden, F.; Lall, R.; Larson, T. V.; Liu, H.; Neas, L.; Pinto, J.; Stolzel, M.; Suh, H.; Hopke, P. K., Workgroup report: Workshop on source apportionment of particulate matter health effects - Intercomparison of results and implications. *Environmental Health Perspectives* **2005,** *113*, (12), 1768-1774.
10. Hansen, D. A.; Edgerton, E. S.; Hartsell, B. E.; Jansen, J. J.; Kandasamy, N.; Hidy, G. M.; Blanchard, C. L., The southeastern aerosol research and characterization study: Part 1-overview. *Journal of the Air & Waste Management Association* **2003,** *53*, (12), 1460-1471.
11. Malm, W. C.; Schichtel, B. A.; Pitchford, M. L., Uncertainties in PM2. 5 gravimetric and speciation measurements and what we can learn from them. *Journal of the Air & Waste Management Association* **2011,** *61*, (11), 1131-1149.
12. Reff, A.; Eberly, S. I.; Bhave, P. V., Receptor modeling of ambient particulate matter data using positive matrix factorization: Review of existing methods. *Journal of the Air & Waste Management Association* **2007,** *57*, (2), 146-154.
13. Balachandran, S.; Chang, H.; Pachon, J. E.; Holmes, H. A.; Mulholland, J. A.; Russell, A. G., A Bayesian – Based Ensemble Technique for Source Apportionment of PM2.5. *Submitted to Environmental Science and Technology* **2013**.
14. Balachandran, S.; Pachon, J. E.; Hu, Y.; Lee, D.; Mulholland, J. A.; Russell, A. G., Ensemble-trained source apportionment of fine particulate matter and method uncertainty analysis. *Atmospheric Environment* **2012,** *61*, (0), 387-394.

15. Marmur, A.; Unal, A.; Mulholland, J. A.; Russell, A. G., Optimization-based source apportionment of PM2.5 incorporating gas-to-particle ratios. *Environmental Science & Technology* **2005,** *39*, (9), 3245-3254.
16. Baek, J.; Park, S. K.; Hu, Y.; Russell, A. G. In *Source Apportionment of Fine Organic Aerosol Using CMAQ Tracers*, The 4th Annual CMAS Models-3 Users' Conference, Durham, NC, September 27, 2005; CMAS: Durham, NC, 2005.
17. Lee, D.; Balachandran, S.; Pachon, J.; Shankaran, R.; Lee, S.; Mulholland, J. A.; Russell, A. G., Ensemble-Trained PM2.5 Source Apportionment Approach for Health Studies. *Environmental Science & Technology* **2009,** *43*, (18), 7023-7031.
18. Zheng, M.; Cass, G. R.; Ke, L.; Wang, F.; Schauer, J. J.; Edgerton, E. S.; Russell, A. G., Source apportionment of daily fine particulate matter at Jefferson street, Atlanta, GA, during summer and winter. *Journal of the Air & Waste Management Association* **2007,** *57*, (2), 228-242.
19. Liu, W.; Wang, Y. H.; Russell, A.; Edgerton, E. S., Atmospheric aerosol over two urban-rural pairs in the southeastern United States: Chemical composition and possible sources. *Atmospheric Environment* **2005,** *39*, (25), 4453-4470.
20. Vanicek, P., APPROXIMATE SPECTRAL ANALYSIS BY LEAST-SQUARES FIT - SUCCESSIVE SPECTRAL ANALYSIS. *Astrophysics and Space Science* **1969,** *4*, (4), 387-&.
21. Press, W. H.; Teukolksy, S. A.; Vetterling, W. T.; Flannery, B. P., In *Fortran Numerical Recipies*, Cambridge University: Cambridge, UK, 2001; Vol. 2, pp p.569-577.
22. Pytharouli, S. I.; Stiros, S. C., Spectral analysis of unevenly spaced or discontinuous data using the "normperiod" code. *Computers & Structures* **2008,** *86*, (1-2), 190-196.
23. Shoelson, B. *lombsargle.m*, 2001.

4.7. Tables

Table 4.1: Statistical metrics of SA results. NOTE: PMF uses derived factors, and a different fitting statistic, and is therefore not comparable with CMB statistical metrics.

	JST				SDK				YRK			
	BBSP	EBSP	MBSP	PMF	BBSP	EBSP	MBSP	PMF	BBSP	EBSP	MBSP	PMF
Average Mass Ratio (Pred./Obs. $PM_{2.5}$)	0.94	0.91	0.87	1.10	0.97	0.90	0.89	1.09	0.81	0.81	0.77	1.05
Reduced Chi Square	5.40	3.34	4.86	-	26.51	73.41	87.81	-	8.58	4.36	13.97	-
	Zero Impact Days				Zero Impact Days				Zero Impact Days			
GV	0	0	0	-	0	0	0	-	9	11	9	-
DV	6	195	154	-	51	181	195	-	90	400	348	-
DUST	0	15	54	-	0	13	25	-	9	43	60	-
BURN	0	3	5	-	0	1	0	-	0	1	6	-
COAL	9	214	267	-	17	100	216	-	2	24	24	-
SOC	25	78	25	-	138	44	56	-	11	23	21	-

Table 4.2: Average source impacts (μg m^{-3}) for Jefferson St. (JST), South Dekalb (SDK) and Yorkville (YRK). Uncertainties are the root mean square averages of daily source impact uncertainties ($\overline{\sigma}_{S_{jl}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\sigma_{S_{jkl}}^2}$, for source *j* and method *l*).

	SUMMER (April – September)											
JST	**BBSP**			**EBSP**			**MBSP**			**PMF**		
GV	0.67	±	0.25	0.72	±	0.41	0.83	±	0.42	1.22	±	0.26
DV	1.25	±	0.48	1.05	±	0.75	1.11	±	0.78	2.25	±	0.44
BURN	1.31	±	1.06	1.17	±	0.85	0.98	±	0.73	1.70	±	0.15
SOC	2.63	±	0.64	2.53	±	0.93	2.60	±	0.91	1.67	±	1.84
SDK	**BBSP**			**EBSP**			**MBSP**			**PMF**		
GV	0.75	±	0.16	0.67	±	0.21	0.72	±	0.16	1.19	±	0.07
DV	1.09	±	0.53	0.83	±	0.51	0.87	±	0.65	1.29	±	0.07
BURN	1.93	±	1.10	1.26	±	0.49	1.25	±	0.70	3.14	±	0.11
SOC	1.43	±	0.43	1.86	±	0.69	1.71	±	0.78	1.08	±	0.51
YRK	**BBSP**			**EBSP**			**MBSP**			**PMF**		
GV	0.12	±	0.03	0.13	±	0.32	0.14	±	0.33	1.62	±	0.06
DV	0.12	±	0.07	0.16	±	0.51	0.12	±	0.45	1.00	±	0.05
BURN	1.39	±	0.58	1.35	±	0.73	1.01	±	0.65	2.04	±	0.06
SOC	2.41	±	0.31	2.38	±	0.83	2.55	±	0.80	0.94	±	0.74
	WINTER (October – March)											
JST	**BBSP**			**EBSP**			**MBSP**			**PMF**		
GV	1.13	±	0.31	1.45	±	0.73	1.29	±	0.60	1.76	±	0.41
DV	1.36	±	0.57	1.52	±	1.27	1.30	±	0.98	2.38	±	0.54
BURN	3.89	±	1.57	2.22	±	1.36	1.12	±	0.91	2.55	±	0.23
SOC	1.43	±	0.85	1.81	±	1.15	2.67	±	1.07	1.35	±	2.59
SDK	**BBSP**			**EBSP**			**MBSP**			**PMF**		
GV	1.03	±	0.19	1.06	±	0.18	1.05	±	0.18	1.36	±	0.08
DV	1.32	±	0.49	1.30	±	0.55	1.05	±	0.63	1.48	±	0.09
BURN	4.29	±	0.96	1.20	±	0.64	1.13	±	0.63	4.13	±	0.15
SOC	0.64	±	0.32	1.84	±	0.65	1.76	±	0.71	0.85	±	0.51
YRK	**BBSP**			**EBSP**			**MBSP**			**PMF**		
GV	0.25	±	0.05	0.23	±	0.35	0.29	±	0.27	1.10	±	0.02
DV	0.14	±	0.10	0.17	±	0.73	0.15	±	0.44	0.68	±	0.01
BURN	1.54	±	0.64	2.04	±	0.83	0.95	±	0.62	3.45	±	0.10
SOC	1.72	±	0.34	1.42	±	0.80	1.96	±	0.72	0.58	±	0.67

Table 4.3: Average OC:EC ratios in source profiles/factors. For BURN, the ratio of K in the derived EMSPs, BBSPs, and PMF factors to the BURN MBSP is also shown.

GV	MBSP	EBSP SUMMER	EBSP WINTER	BBSP SUMMER	BBSP WINTER	PMF FACTOR
JST OC:EC	2.33	1.49	4.00	2.21	2.14	3.69
SDK OC:EC	2.33	1.92	1.95	2.27	2.19	0.98
YRK OC:EC	2.33	1.49	1.95	1.96	1.98	2.25
DV	**MBSP**	**EBSP SUMMER**	**EBSP WINTER**	**BBSP SUMMER**	**BBSP WINTER**	**PMF FACTOR**
JST EC:OC	3.71	2.21	1.93	4.21	4.08	1.58
SDK EC:OC	3.71	5.25	4.78	4.11	4.81	1.02
YRK EC:OC	3.71	2.21	4.78	4.48	4.56	0.44
BURN	**MBSP**	**EBSP SUMMER**	**EBSP WINTER**	**BBSP SUMMER**	**BBSP WINTER**	**PMF FACTOR**
JST OC:EC	4.09	3.00	3.00	7.59	10.03	5.64
SDK OC:EC	4.09	5.01	4.56	17.64	10.50	10.79
YRK OC:EC	4.09	3.00	4.56	5.65	3.91	3.11
JST K ratio (to MBSP)	-	0.60	0.41	0.68	0.25	0.75
SDK K ratio (to MBSP)	-	0.79	0.87	0.50	0.23	0.39
YRK K ratio (to MBSP)	-	0.60	0.87	0.55	0.59	0.34

4.8. Figures

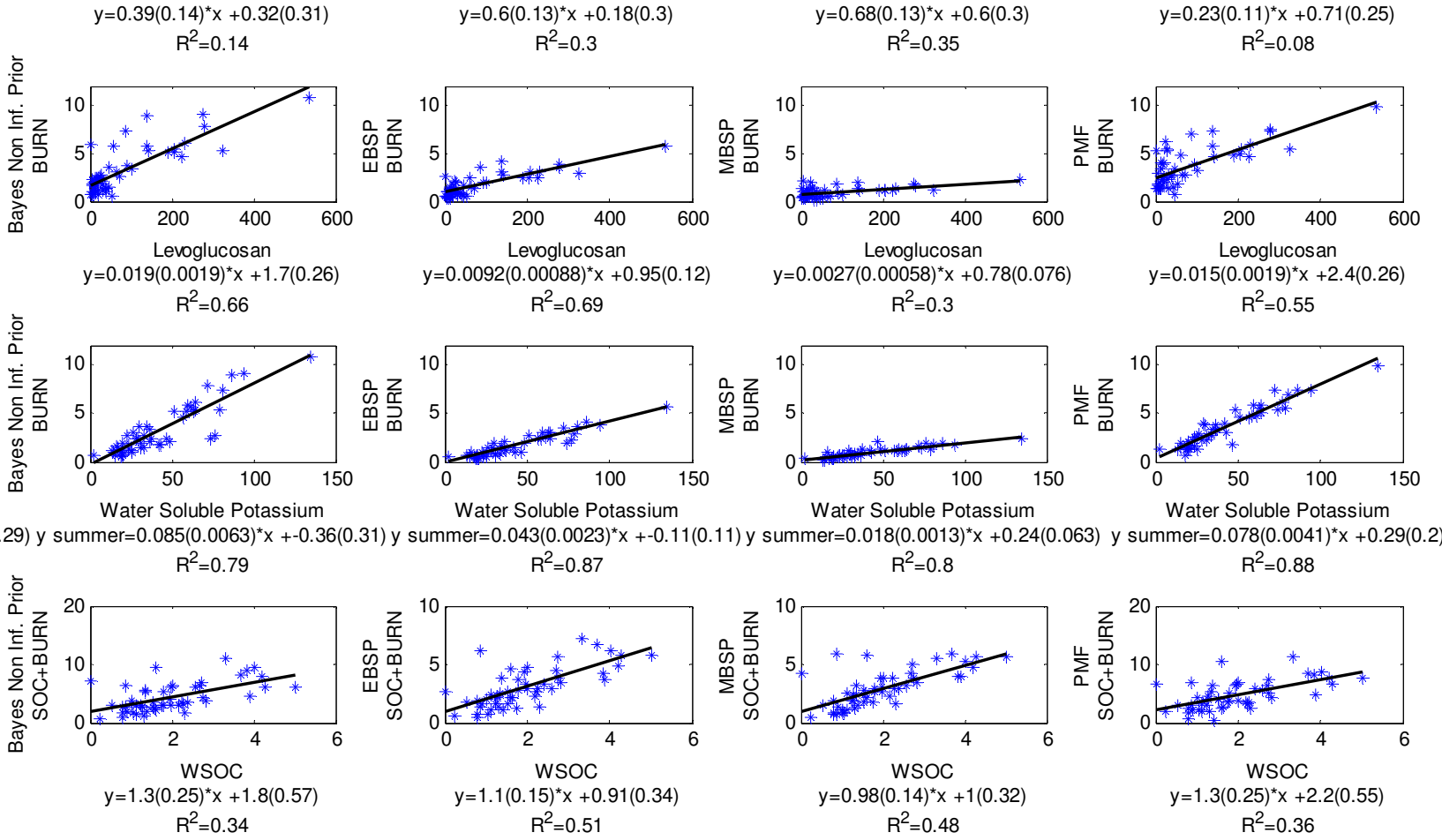


Figure 4.1: Comparison of source impacts for BURN and SOC and water soluble organic carbon (WSOC), levoglucosan and water soluble potassium (K^+). The first row compares BURN and levoglucosan. The second row compares BURN and water-soluble potassium. The last row compares the sum of SOC and BURN impacts and WSOC.

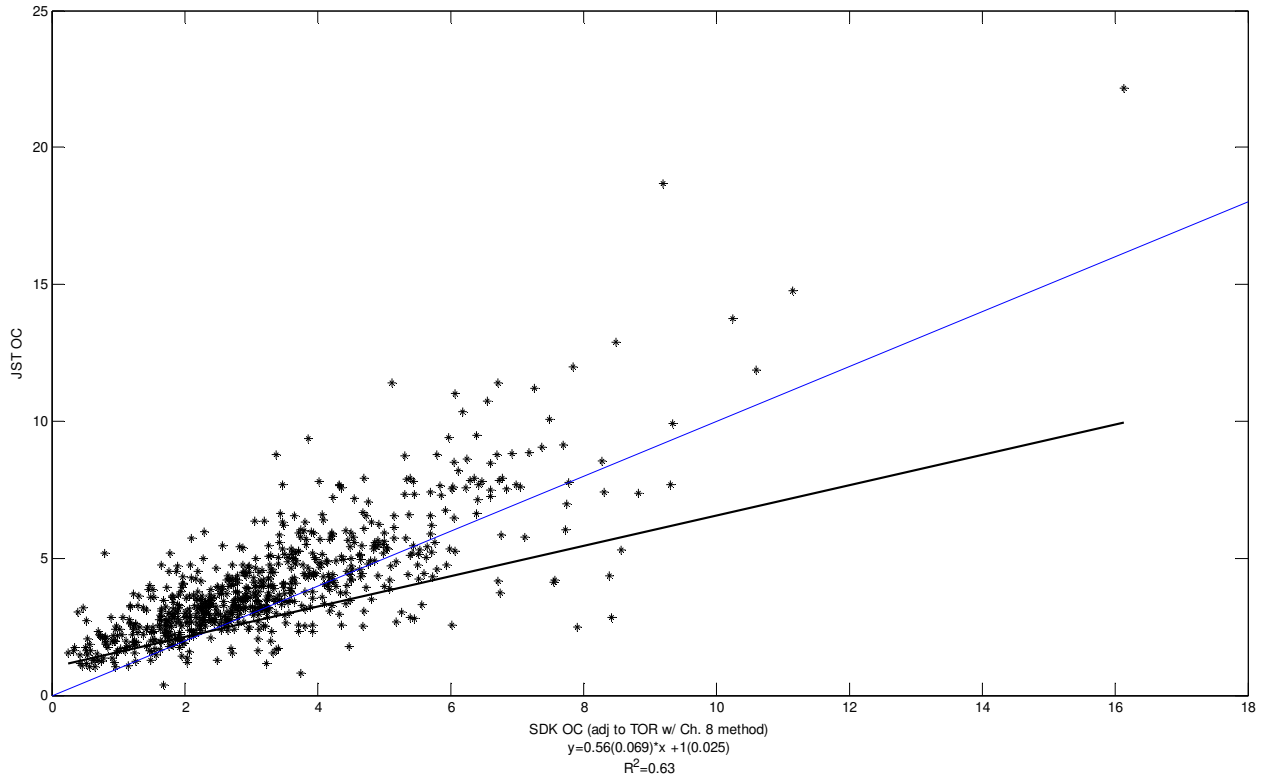


Figure 4.2: JST OC concentrations, based or thermal optical reflectance (TOR), versus SDK OC values adjusted from thermal optical transmittance (TOT) to TOR-equivalent values based on a regression-based adjustment [*Malm et al.*, 2011]. Blue line is a 1:1 line. Black line is York regression line. R^2=0.63.

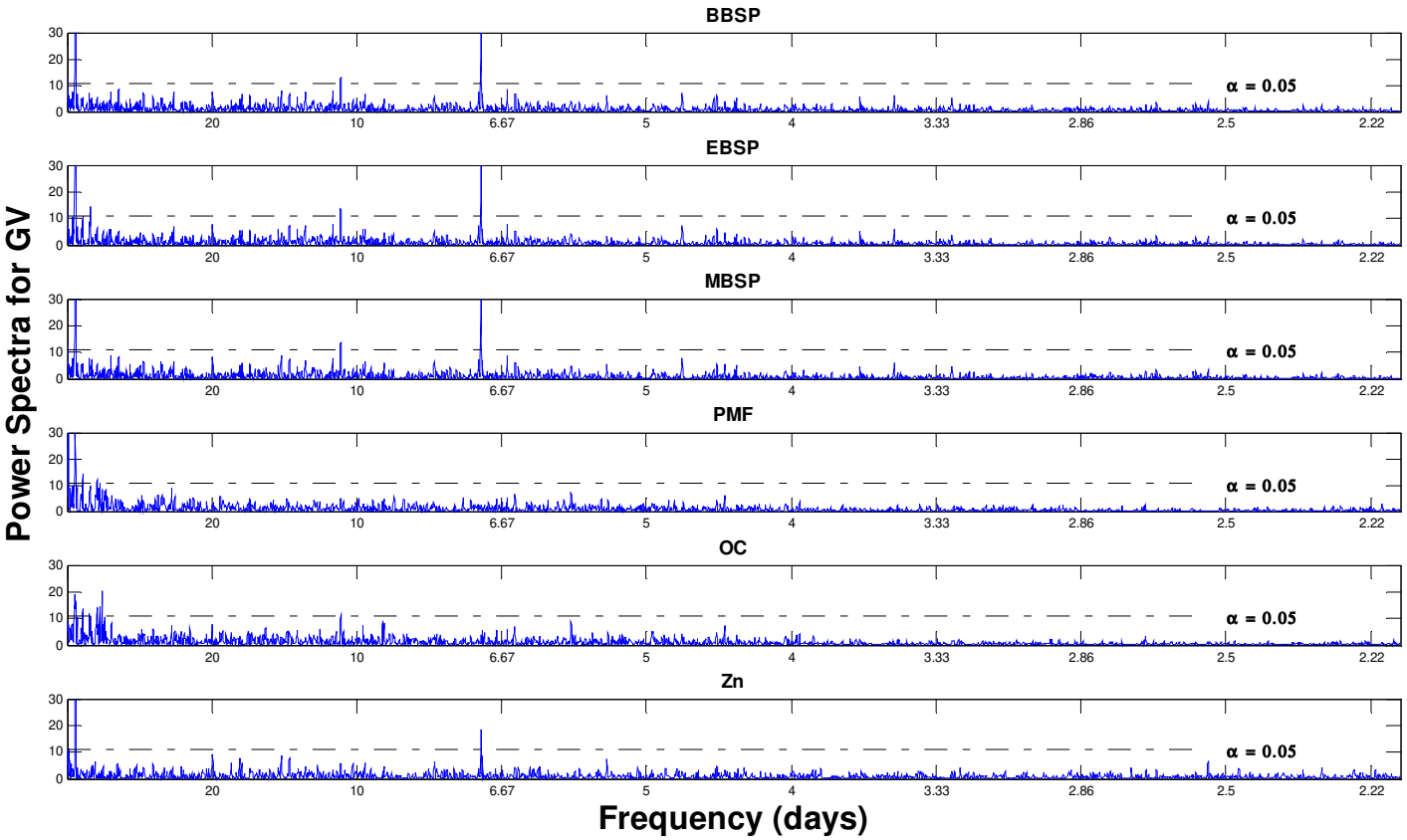


Figure 4.3: JST GV spectral results

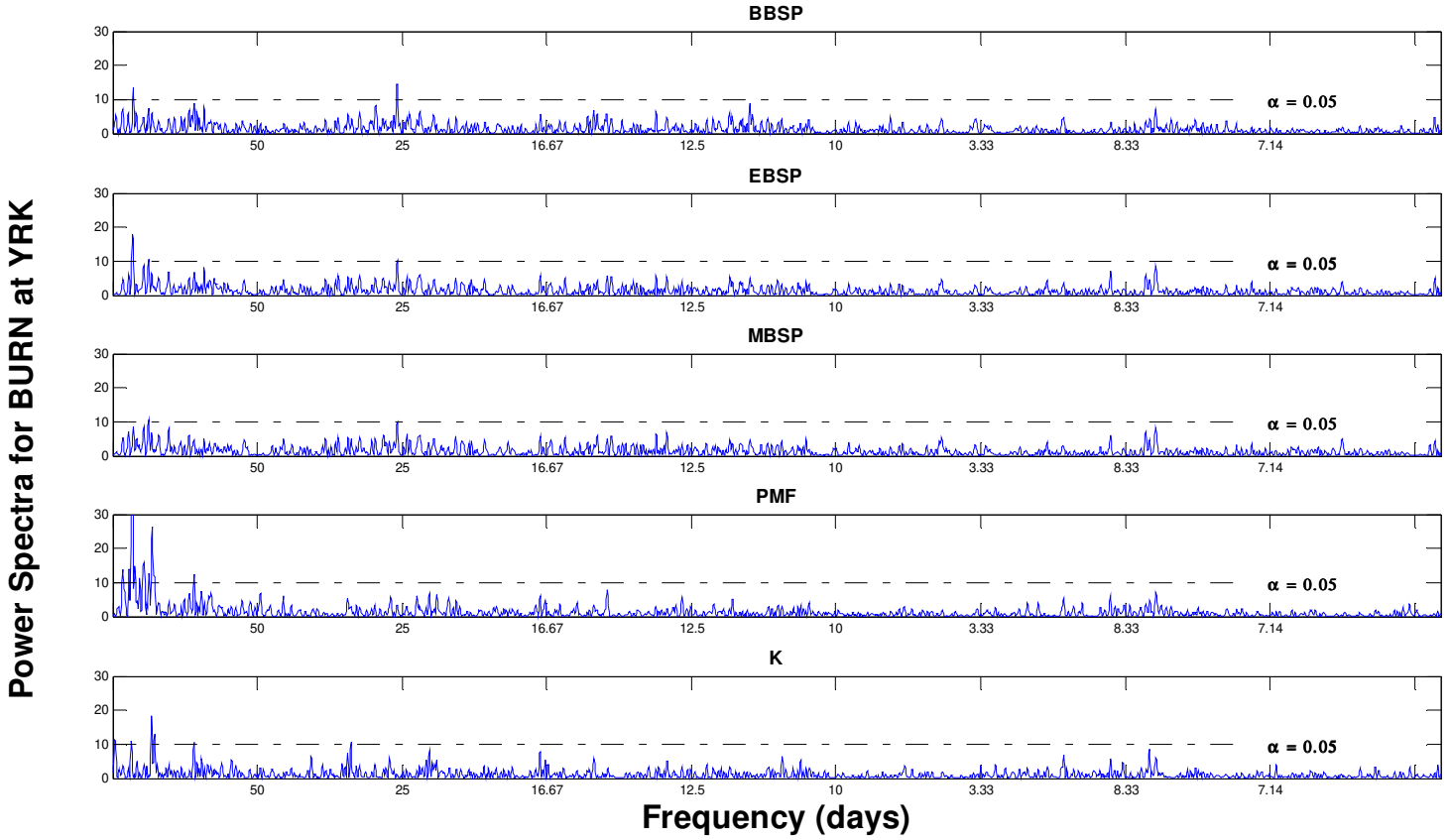


Figure 4.4: YRK BURN spectral results.

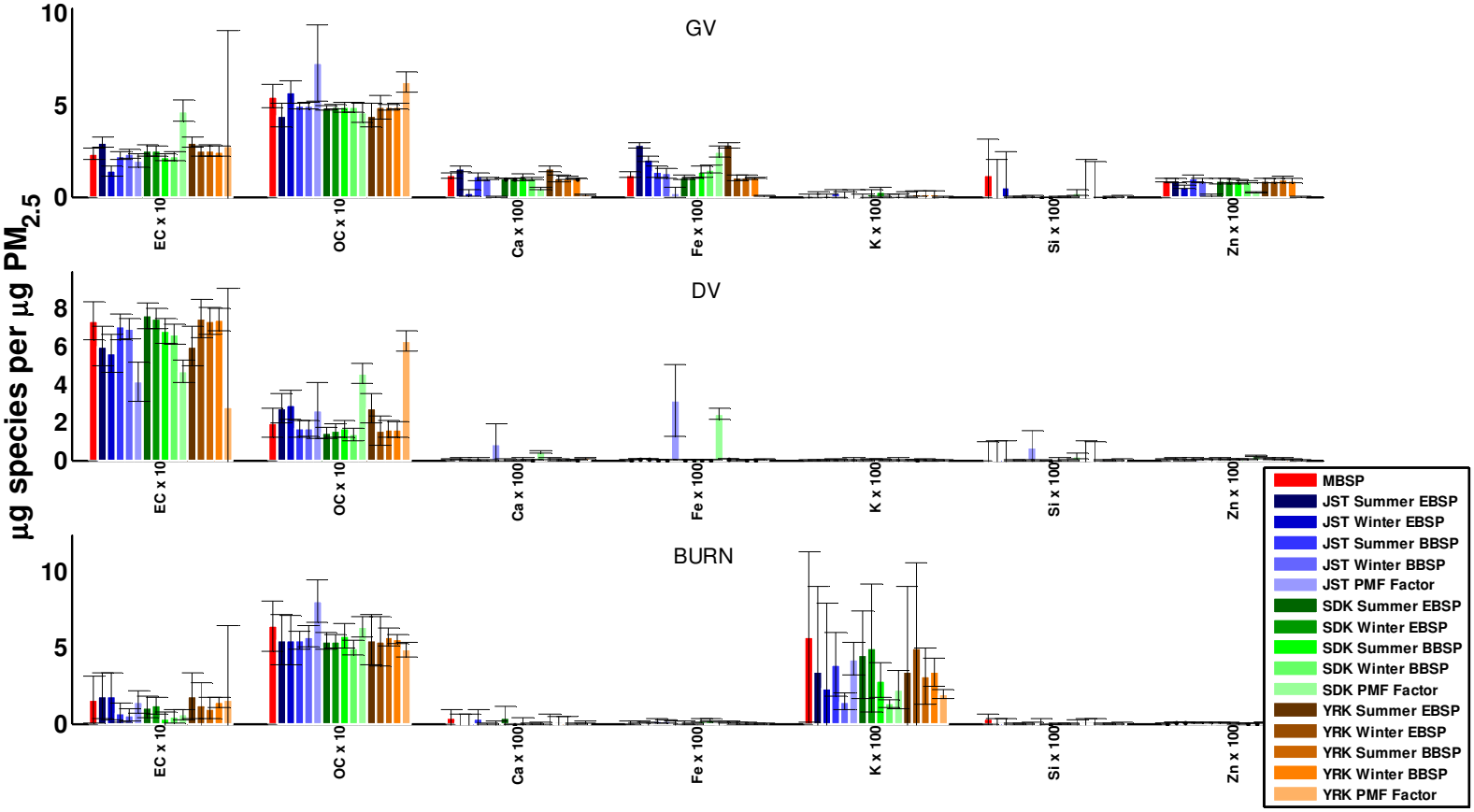


Figure 4.5: Comparison of selected species from 16 source profiles/factors used in this study.

APPENDIX C: SUPPLEMENTAL INFORMATION FOR CHAPTER 4

Table C.1: Correlations (R^2) of source impacts and tracer species across methods at JST. Shaded values indicate $R^2 > 0.75$.

JST	GV	BBSP	EBSP	MBSP	PMF	OC	Zn
	BBSP	1.00	0.91	0.97	0.19	0.29	0.56
	EBSP	0.91	1.00	0.91	0.21	0.31	0.48
	MBSP	0.97	0.91	1.00	0.20	0.32	0.54
	PMF	0.19	0.21	0.20	1.00	0.67	0.08
	OC	0.29	0.31	0.32	0.67	1.00	0.16
	Zn	0.56	0.48	0.54	0.08	0.16	1.00
	DV	BBSP	EBSP	MBSP	PMF	EC	
	BBSP	1.00	0.75	0.92	0.49	0.77	
	EBSP	0.75	1.00	0.77	0.42	0.63	
	MBSP	0.92	0.77	1.00	0.47	0.75	
	PMF	0.49	0.42	0.47	1.00	0.62	
	OC	0.77	0.63	0.75	0.62	1.00	
	BURN	BBSP	EBSP	MBSP	PMF	K	
	BBSP	1.00	0.77	0.38	0.51	0.15	
	EBSP	0.77	1.00	0.69	0.62	0.27	
	MBSP	0.38	0.69	1.00	0.42	0.55	
	PMF	0.51	0.62	0.42	1.00	0.19	
	OC	0.15	0.27	0.55	0.19	1.00	
	SOC	BBSP	EBSP	MBSP	PMF	OC	
	BBSP	1.00	0.92	0.76	0.49	0.60	
	EBSP	0.92	1.00	0.89	0.53	0.71	
	MBSP	0.76	0.89	1.00	0.47	0.89	
	PMF	0.49	0.53	0.47	1.00	0.39	
	OC	0.60	0.71	0.89	0.39	1.00	

Table C.2: Correlations (R^2) of source impacts and tracer species across methods at SDK. Shaded values indicate $R^2 > 0.75$.

SDK	GV	BBSP	EBSP	MBSP	PMF	OC	Zn
	BBSP	1.00	0.91	0.93	0.74	0.34	0.34
	EBSP	0.91	1.00	0.95	0.61	0.32	0.37
	MBSP	0.93	0.95	1.00	0.68	0.33	0.38
	PMF	0.74	0.61	0.68	1.00	0.43	0.26
	OC	0.34	0.32	0.33	0.43	1.00	0.19
	Zn	0.34	0.37	0.38	0.26	0.19	1.00
	DV	BBSP	EBSP	MBSP	PMF	EC	
	BBSP	1.00	0.90	0.93	0.52	0.84	
	EBSP	0.90	1.00	0.91	0.60	0.85	
	MBSP	0.93	0.91	1.00	0.52	0.87	
	PMF	0.52	0.60	0.52	1.00	0.70	
	OC	0.84	0.85	0.87	0.70	1.00	
	BURN	BBSP	EBSP	MBSP	PMF	K	
	BBSP	1.00	0.50	0.45	0.46	0.13	
	EBSP	0.50	1.00	0.97	0.11	0.06	
	MBSP	0.45	0.97	1.00	0.11	0.12	
	PMF	0.46	0.11	0.11	1.00	0.41	
	OC	0.13	0.06	0.12	0.41	1.00	
	SOC	BBSP	EBSP	MBSP	PMF	OC	
	BBSP	1.00	0.76	0.74	0.66	0.51	
	EBSP	0.76	1.00	0.99	0.56	0.86	
	MBSP	0.74	0.99	1.00	0.58	0.83	
	PMF	0.66	0.56	0.58	1.00	0.41	
	OC	0.51	0.86	0.83	0.41	1.00	

Table C.3: Correlations (R^2) of source impacts and tracer species across methods at YRK. Shaded values indicate $R^2 > 0.75$.

YRK	GV	BBSP	EBSP	MBSP	PMF	OC	Zn
	BBSP	1.00	0.91	0.96	0.00	0.00	0.05
	EBSP	0.91	1.00	0.88	0.00	0.00	0.07
	MBSP	0.96	0.88	1.00	0.00	0.01	0.05
	PMF	0.00	0.00	0.00	1.00	0.53	0.04
	OC	0.00	0.00	0.01	0.53	1.00	0.07
	Zn	0.05	0.07	0.05	0.04	0.07	1.00
	DV	BBSP	EBSP	MBSP	PMF	EC	
	BBSP	1.00	0.84	0.95	0.00	0.07	
	EBSP	0.84	1.00	0.76	0.01	0.10	
	MBSP	0.95	0.76	1.00	0.00	0.04	
	PMF	0.00	0.01	0.00	1.00	0.38	
	OC	0.07	0.10	0.04	0.38	1.00	
	BURN	BBSP	EBSP	MBSP	PMF	K	
	BBSP	1.00	0.89	0.87	0.45	0.46	
	EBSP	0.89	1.00	0.72	0.58	0.49	
	MBSP	0.87	0.72	1.00	0.24	0.36	
	PMF	0.45	0.58	0.24	1.00	0.48	
	OC	0.46	0.49	0.36	0.48	1.00	
	SOC	BBSP	EBSP	MBSP	PMF	OC	
	BBSP	1.00	0.98	0.99	0.73	0.91	
	EBSP	0.98	1.00	0.96	0.73	0.86	
	MBSP	0.99	0.96	1.00	0.71	0.92	
	PMF	0.73	0.73	0.71	1.00	0.72	
	OC	0.91	0.86	0.92	0.72	1.00	

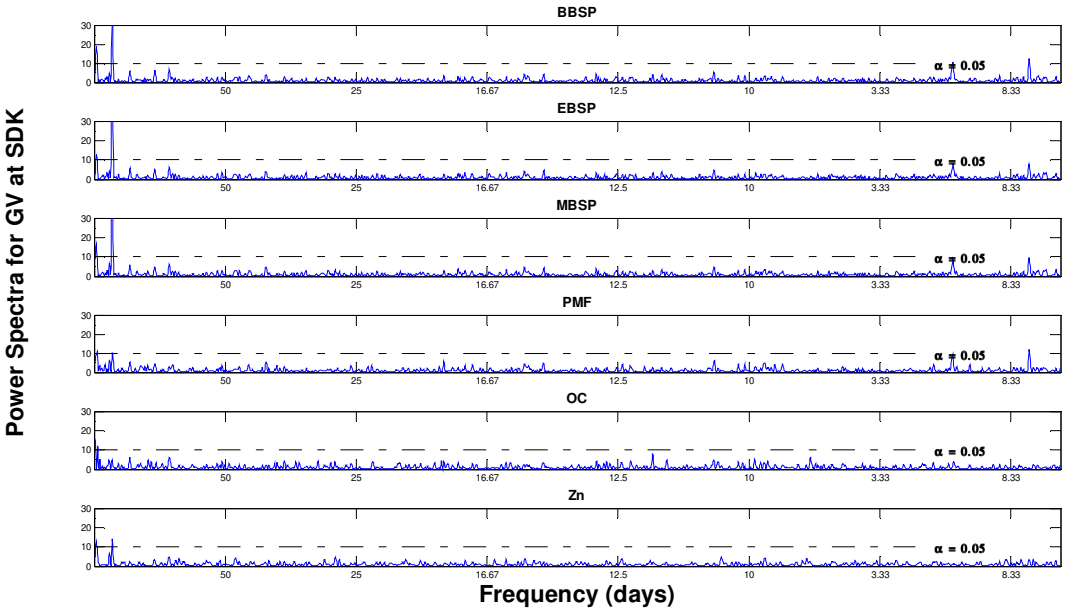


Figure C.1: SDK GV spectral results.

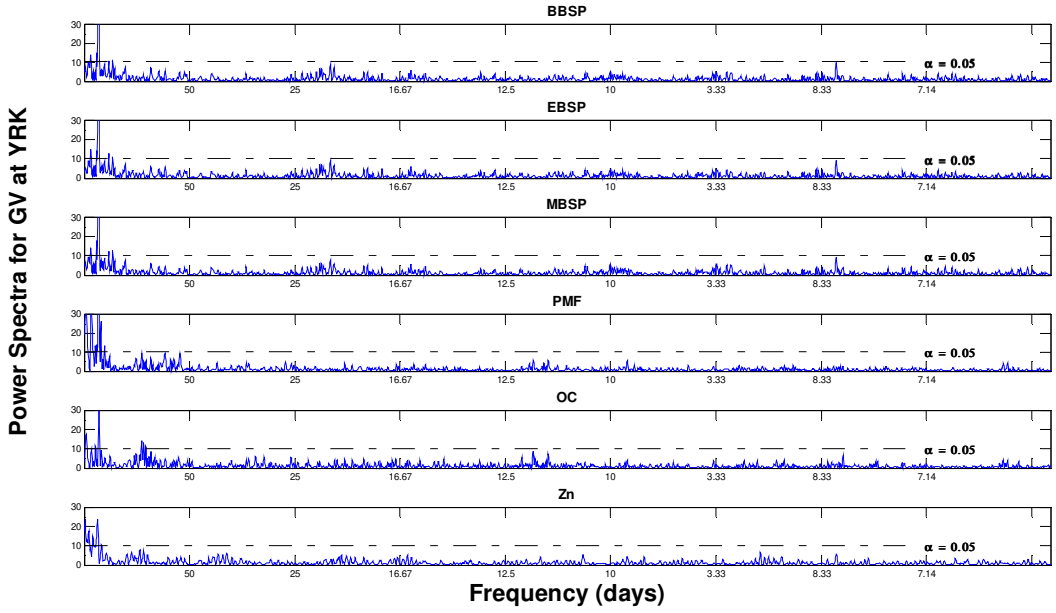


Figure C.2: YRK GV spectral results.

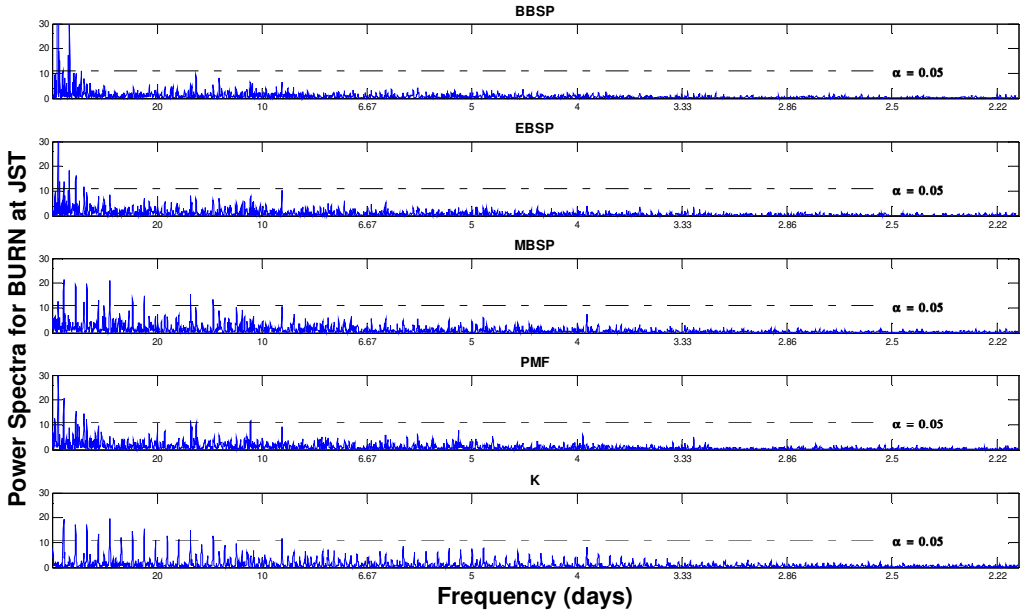


Figure C.3: JST BURN spectral results.

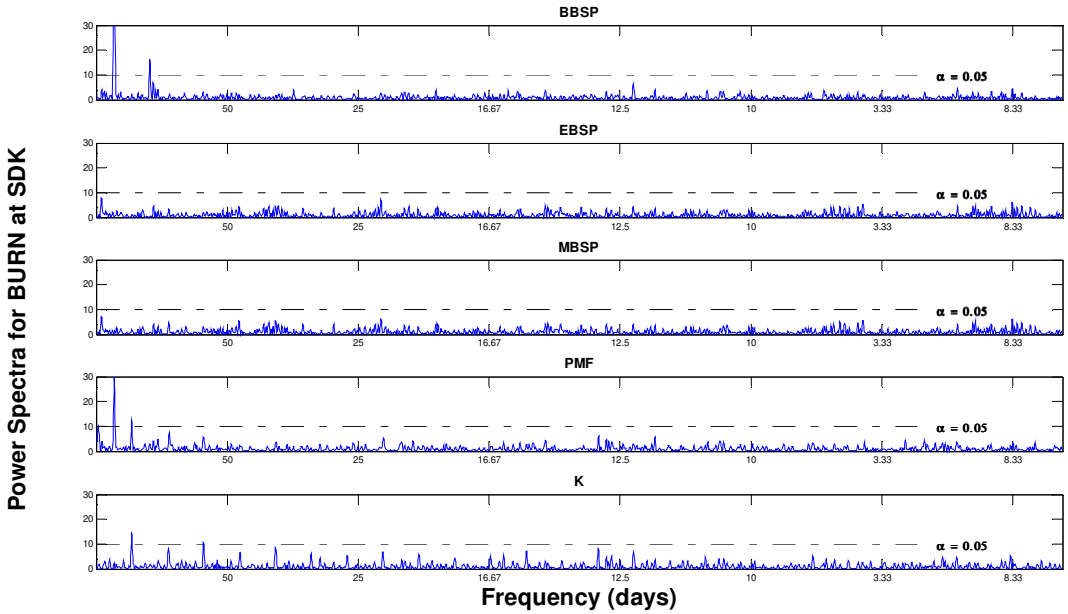


Figure C.4: SDK BURN spectral results.

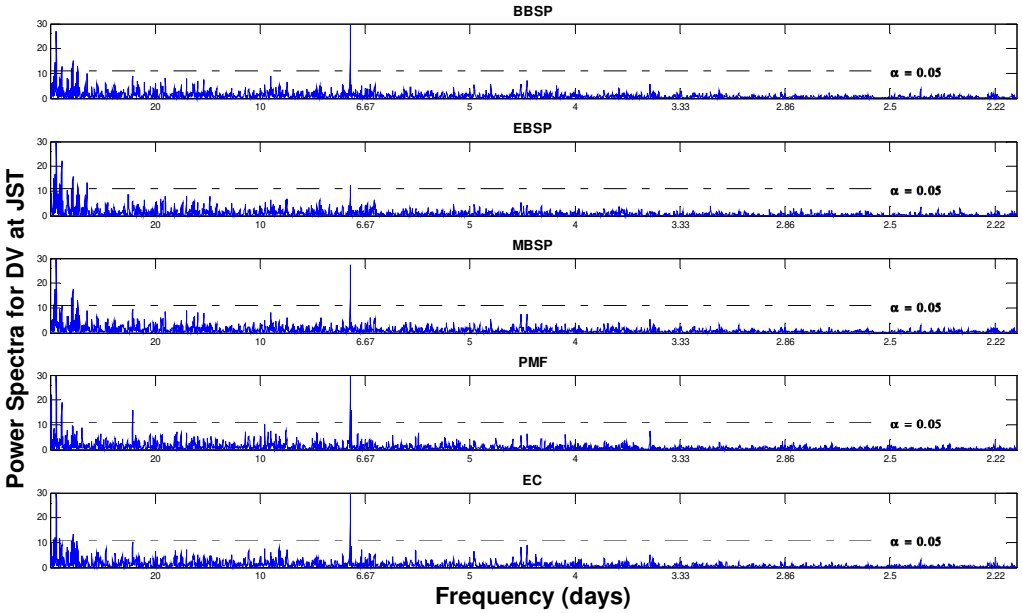


Figure C.5: JST DV spectral results.

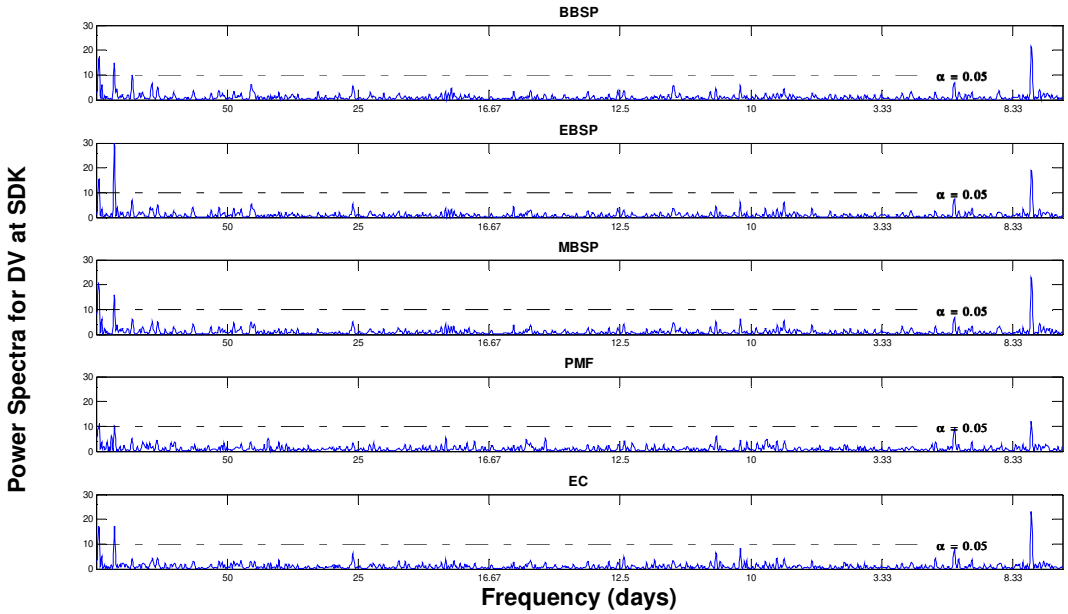


Figure C.6: SDK DV spectral results.

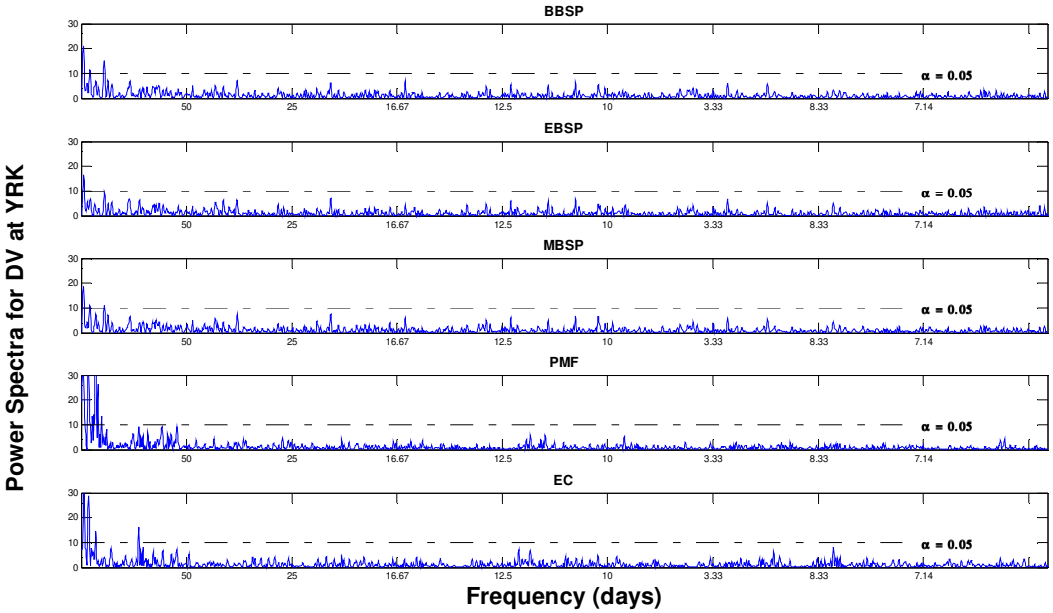


Figure C.7: YRK DV spectral results.

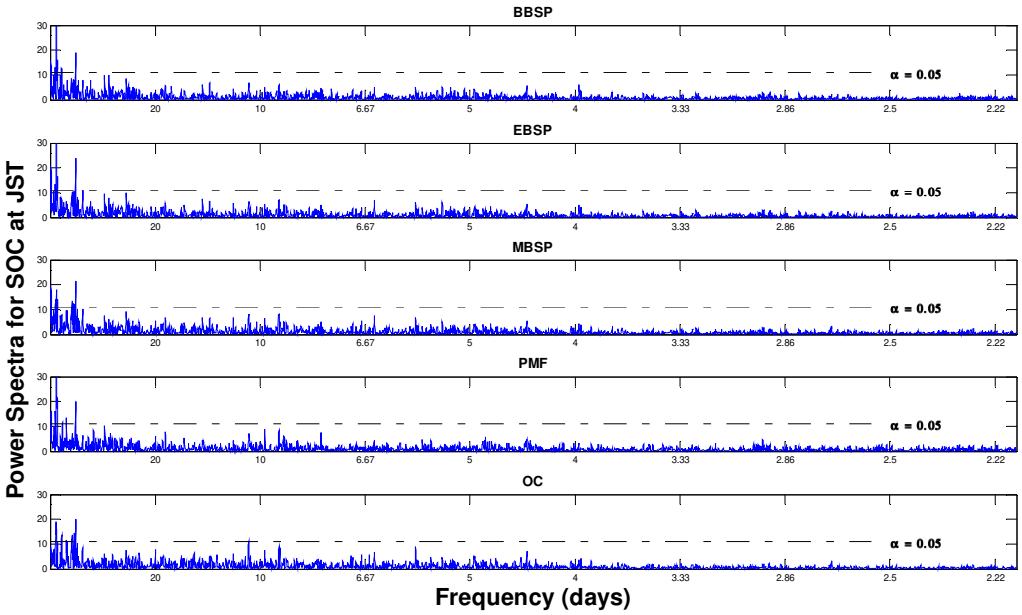


Figure C.8: JST SOC spectral results.

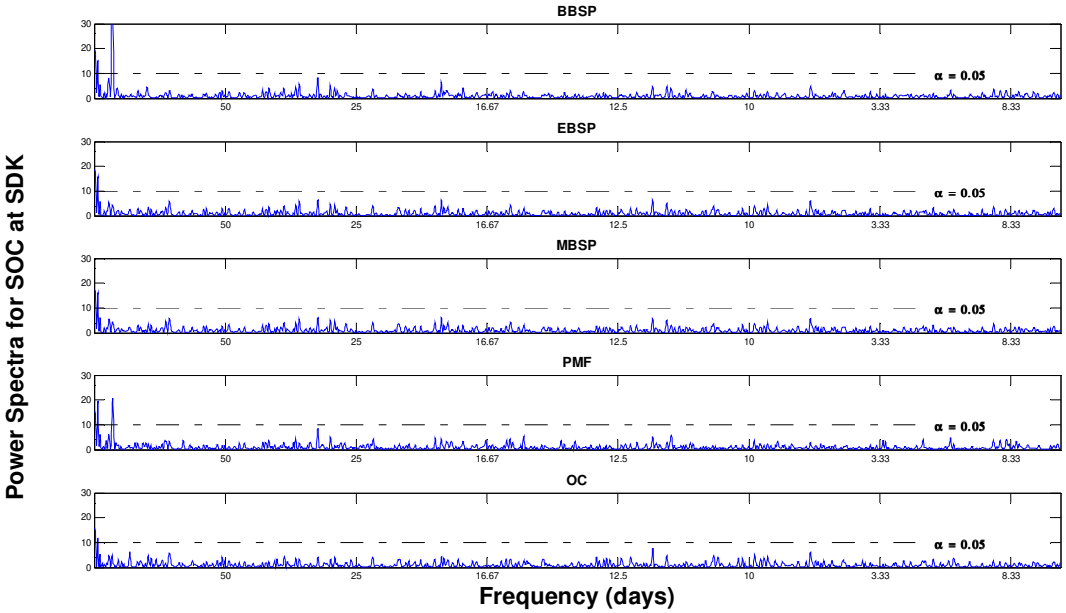


Figure C.9: SDK SOC spectral results.

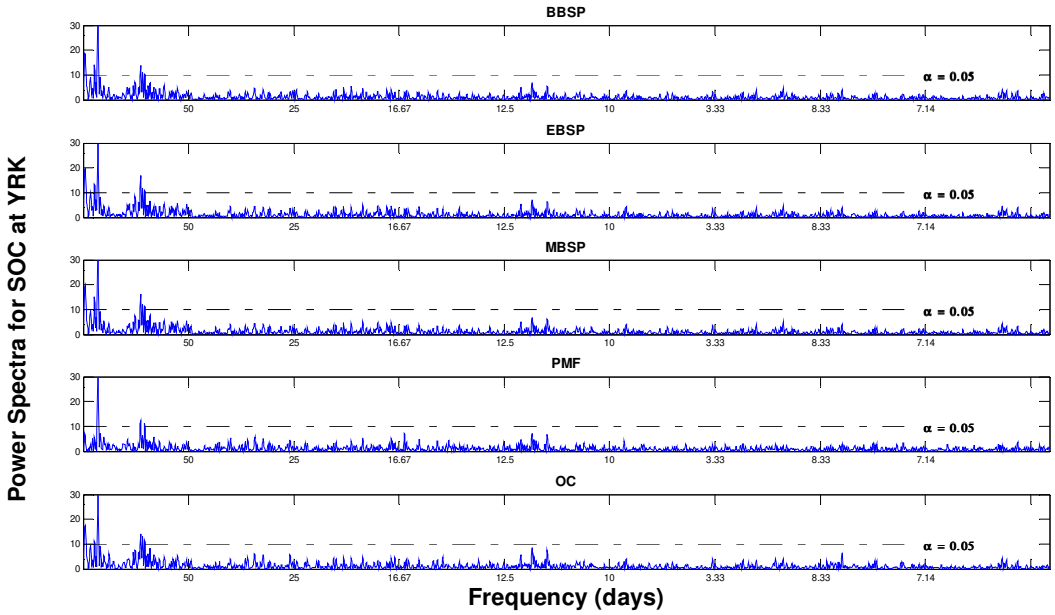


Figure C.10: YRK SOC spectral results.

CHAPTER 5: PARTICULATE AND GAS SAMPLING OF PRESCRIBED FIRES IN SOUTH GEORGIA, USA

Sivaraman Balachandran[a*], Jorge E. Pachon[b], Sangil Lee[c], Michelle Oakes[a+], Neeraj Rastogi[d], Wenyan Shi[a+], Efthimios Tagaris[e], Bo Yan[a+], Aika Davis[a], Xiaolu Zhang[a+], Rodney J. Weber[a+], James A. Mulholland[a], Mike Bergin[a], Mei Zheng[a+, f], and Armistead G. Russell[a]

[a]Georgia Institute of Technology, Department of Civil and Environmental Engineering. Atlanta, GA, 30332

[a+]Georgia Institute of Technology, Department of Earth and Atmospheric Sciences. Atlanta, GA, 30332

[b]Universidad de La Salle, Programa de Ingenieria Ambiental, Bogota, Colombia

[c]Korea Research Institute for Standards and Science, Center for Gas Analysis, Daejon, 305-340, Republic of Korea

[d]Physical Research Laboratory, Ahmedabad, India

[e]Environmental Research Laboratory, Institute of Nuclear Technology-Radiation Protection, NCSR "DEMOKRITOS", 15310, Aghia Paraskevi Attikis, Athens, Greece

[f]Peking University, College of Environmental Science and Engineering, Beijing, China

[*]Corresponding author: 311 Ferst Dr., Atlanta, GA, 30332; phone 206.250.6480, fax 404.894.8266; siv@gatech.edu.

5.1. Abstract

Gaseous and particulate species from two prescribed fires were sampled in situ, to better characterize prescribed burn emissions. Measurements included gaseous and fine particulate matter ($PM_{2.5}$) species, particle number concentration, particulate organic carbon (POC)

speciation, water-soluble organic carbon (WSOC) and water-soluble iron. Major $PM_{2.5}$ components included OC (~57%), EC (~10 %), chloride (~1.6%), potassium (~0.7%) and nitrate (~0.9%). Major gaseous species include carbon dioxide, carbon monoxide, methane, ethane, methanol and ethylene. Particulate organic tracers of biomass burning, such as levoglucosan, dehydroabietic acid and retene, increased significantly during the burns. Water-soluble organic carbon (WSOC) also increased significantly during the fire and levels are highly correlated with total potassium (K) (R^2=0.93) and levoglucosan (R^2=0.98). The average WSOC/OC ratio was 0.51 ± 0.03 and did not change significantly from background levels. Thus, the WSOC/OC ratio may not be a good indicator of secondary organic aerosol (SOA) in regions that are expected to be impacted by biomass burning. Results using a biomass burning source profile derived from this work further indicate that source apportionment is sensitive to levels of potassium in biomass burning source profiles. This underscores the importance of quantifying local biomass burning source profiles.

5.2. Introduction

Biomass burning, such as wildfires, prescribed burns, and residential wood combustion, are important sources of air pollutants, which can impact health, lead to violations of air quality standards, and impair visibility [*S. Lee et al.*, 2005; *Sandberg et al.*, 2002]. Biomass burning emissions can be gaseous or particulates and include species that lead to secondary pollutants. Long-lived primary air pollutants from biomass burning can travel large distances (thousands of km), making populated areas potentially susceptible to impacts from remote fires [*Sapkota et al.*, 2005; *Wotawa and Trainer*, 2000]. In the southeastern United States, emission inventories estimate that biomass burning contributes significantly to air pollutant emissions: ~8-20% of $PM_{2.5}$ (particulate matter that is aerodynamically less than 2.5 μm in diameter), ~8% of carbon monoxide (CO), and ~6% of volatile organic compounds (VOCs) [*Barnard*, 2003; *Kim et al.*, 2004; *Liu et al.*, 2005; *Zhang et al.*, 2010].

Prescribed burning is widespread, especially in the southeastern US, and is used to manage forest ecosystems and protect endangered species by controlling growth and infestation while minimizing the risk of large-scale forest fires [*Hardy et al.*, 2001]. In 2006, a total of 96,385 wild land fires were reported to have burned 9,873,429 acres in the U.S., 125% above the 10-year average [*NIFC*, 2008]. Of that, 2,720,545 acres were treated with prescribed fires, an increase of 410,000 acres from the previous year's total and the second highest since 1998. Most of the prescribed fires occurred in the Southern Geographic Area, which includes the area bounded by Kentucky and Virginia to the north and Texas and Oklahoma to the west [*NIFC*, 2008].

The dynamics of prescribed fires can differ significantly from wild fires and vary by region [*Burling et al.*, 2011; *Burling et al.*, 2010; *S. Lee et al.*, 2005; *Urbanski et al.*, 2011]. Fuels also vary by region. Such differences affect the composition and rate of emissions. Emissions also depend upon fire stage (e.g. flaming vs. smoldering). Since limited data exist on emission characteristics from active prescribed burning events in the U.S., emissions of $PM_{2.5}$

and VOCs from prescribed burning were sampled in situ from two prescribed burns in South Georgia in March 2008. A major goal of this study was to update emissions factors for gaseous compounds and $PM_{2.5}$ in Georgia with regionally specific biomass burning air emissions data. A second goal was to better understand the role of water-soluble organic carbon (WSOC) as a tracer of both biomass burning and secondary organic aerosol. Third, tracers of prescribed burns were studied by characterization of organic chemical compounds. In addition, chemical speciation of $PM_{2.5}$ was used in a source apportionment study to test its applicability as a regionally specific biomass burning source profile. $PM_{2.5}$ constituents were quantified, including organic (OC) and elemental carbon (EC), ionic species, trace elements, water-soluble organic carbon (WSOC). water-soluble iron (FE (II)), and particle number concentration. PM OC speciation identified approximately 100 organic chemical compounds. Gases that were sampled included carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4) and other volatile organic compounds (VOCs).

5.3. Methods and Materials

5.3.1. Site Description

Emissions from two prescribed fires were sampled on March 5 and March 6, 2008 at the Joseph W. Jones Ecological Research Center in Newton, GA (Figure 5.1). On March 5, a 495 acre area with one year of accumulated fuel was burned at Ichauway - North Boundary (N 31° 14' 45.0", W 84° 23' 43.2", Figure D.1). On March 6, a 225 acre area with two years of accumulated fuel was burned at Ichauway - Dub-East (N 31° 12' 4.4", W 84° 26' 35.3", Figure D.1). Fuel characteristics are described in Appendix D (Table D.1).

5.3.2. Measurements and Instrumentation

Two-channel, filter-based, particle composition monitors (PCMs), operating at a flow rate of 16.7 L min^{-1}, were used to collect $PM_{2.5}$ for quantifying metals (Teflon filters, 47 mm diameter, Whatman, Inc., Florham Park, NJ) and ions (nylon filters, 47 mm diameter, Gelman

Sciences, Ann Arbor, MI). Each PCM used two denuders in series (URG, Inc., Chapel Hill, NC) coated with phosphoric acid and sodium carbonate to remove acidic and alkaline gases. In addition, high volume $PM_{2.5}$ samplers (HVSs), operating at a nominal flow rate of 1.13 m^3 min^{-1} and having a pre-baked quartz filter (10 x 8 in), were used for measuring OC/EC, WSOC, and solvent-extractable organic compounds. Carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4), non-methane organic compounds (NMOCs) and VOCs were collected using stainless steel canisters [*Colman et al.*, 2001]. Water-soluble Fe(II) was measured by a particle-into-liquid sampler (PILS) method that utilized the ferrozine technique [*Stookey*, 1970], along with a liquid waveguide capillary flow-through optical cell (LWCC) that allowed for increased sensitivity of the instrument [*Oakes et al.*, 2010a; *Rastogi et al.*, 2009] (see Appendix D for summary of results). Particle number concentrations were measured using an optical particle counter (OPC, Met One, Grants Pass, Oregon) (see Appendix D, Figures D.4-D.5).

Two PCMs and HVSs were placed at each site to allow one PCM/HVS set to operate while filters were replaced in the other set. Electricity was provided by a pair of gasoline generators that were placed far away downwind (~50 m) from the sampling point to minimize any impact on the monitoring. The filter collection schedule was determined onsite, depending on the fire stage and wind direction. On March 5, monitoring started at 8:45, about three hours before the fire, to sample for background concentrations and ended at 16:00. On March 6, monitoring started at 7:50, about three hours before the fire, to sample for background concentrations and ended at 13:15. During the first day, five samples were collected (one for background and four during different fire stages). On the second day, three samples were collected (one for background and two during different fire stages).

5.3.3. Analytical Methods

Total $PM_{2.5}$ mass was measured gravimetrically. Teflon filters (which were also used for measurement of metals) were equilibrated and weighed in a clean room (temperature 21±1ºC, RH 35±3%) prior to and after sampling. OC and EC concentrations were determined by taking a

punch (1.54 x 1 cm) from the HVS quartz filter, followed by analysis using a Sunset Lab Thermal Optical Transmittance (TOT) analyzer employing the NIOSH 5040 method [*Birch*, 1998; *NIOSH*, 1996]. Five punch samples were taken from each HVS filter and results were averaged. Ionic species were analyzed by extraction of the soluble PM components from the nylon filter and using a Dionex ion chromatograph (IC). Metals were measured using x-ray fluorescence (XRF) and VOCs were analyzed by gas chromatography with mass spectrometry (GC/MS). Details of these methods can be found in *S. Lee et al.* [2005].

WSOC was determined with a Sievers Model 800 Turbo TOC analyzer. A punch (1.5 x 1.0 cm) of each filter was extracted in 100 ml of 18-Mohm MilliQ water and sonicated for 30 minutes in a sealed 125 ml Nalgene Amber HDPE bottle. The extract was then filtered using a 0.45µm PTFE syringe filter and transferred to a clean bottle for analysis. Water-soluble organic carbon in the extract was then analyzed using the TOC analyzer. The limit of detection (LOD) of 0.33 µg C m-3 was determined by three standard deviations of blank filter measurements.

Organic compounds in $PM_{2.5}$ were analyzed by using the method detailed in *Yan et al.* [2008] and [*Zheng et al.*, 2006]. Briefly, each filter was spiked with deuterated internal standard (IS) mixtures and then successively extracted using hexane and benzene/isopropanol. After being filtered and concentrated, one half of the extract was silylated with BSTFA (N,O-bis(trimethylsilyl)acetamide) and analyzed using GC/MS to quantify polar organic compounds (levoglucosan, cholesterol and 2-methyltetrols). The other half was methylated with diazomethane and analyzed by GC/MS to speciate the other organic compounds.

5.3.4. Source Apportionment

Source apportionment (SA) of $PM_{2.5}$ is used to quantify impacts from emissions sources at a receptor site. The most common SA approaches are factor analytic (FA) and chemical mass balance (CMB) methods, both based on a mass balance approach [*Friedlander*, 1977; *Paatero and Tapper*, 1994; *Watson et al.*, 1984]. In CMB applications, source profiles are typically taken as known and are usually based on both laboratory and in-situ characterization of emissions,

often termed measurement-based source profiles (MBSPs). One limitation of MBSPs is that they may not be representative of emissions at a particular receptor site, and these source profiles can have much variability, especially for a tracer species such as K (total potassium as measured by XRF), EC and levoglucosan in biomass burning profiles. For example, potassium varied from ~ 2 -12 % [*Chow et al.*, 2004b] and ~0.2 – 13% [*Zielinska et al.*, 1998a], in biomass burning $PM_{2.5}$ emissions. This variability is a major source of uncertainty in source apportionment (SA) [*S. Lee and Russell*, 2007]. Therefore, we tested the applicability of using speciated $PM_{2.5}$ ratios from this study as a regionally specific biomass burning source profile. A CMB method that utilizes gas concentration-based constraints with a Lipschitz Global optimizer (CMB-LGO) [*Marmur et al.*, 2005] was used to apportion $PM_{2.5}$ mass to nine source categories: gasoline vehicles (GV), diesel vehicles (DV), road dust (DUST), biomass burning (BURN), coal combustion (COAL), sulfate, nitrate, ammonium and other organic carbon, which is presumed to be secondary organic carbon (SOC). SA was conducted for daily speciated $PM_{2.5}$ data for 2007 from the Jefferson Street (JST) SEARCH site in Atlanta [*Hansen et al.*, 2003] using both a MBSP-based biomass burning source profile from the literature [*Chow et al.*, 2004b] and a new biomass burning source profile based on emission ratios for $PM_{2.5}$ species from this study.

5.4. Results and Discussion

5.4.1. Emissions of Major $PM_{2.5}$ Species

Total $PM_{2.5}$ mass and major constituents increase significantly during the prescribed fire and consist mainly of OC, EC, "Other" (unidentified material, determined by subtracting OC, EC, and ions from the total $PM_{2.5}$ mass, chloride, nitrate, sulfate and potassium (Figure 5.2, Table 5.1). On March 5, the background concentration of $PM_{2.5}$ (sample B1-B) was 9µg m^{-3}, with concentrations increasing to 524µg m^{-3} during the fire (sample B1-F2). On March 6, the background concentration of $PM_{2.5}$ was higher than on March 5 (samples B2-B), 23 µg m^{-3}, and

levels increased to 377 µg m^{-3} during the fire (sample B2-F1). At the end of both burns there is a large decrease in concentrations (samples B1-F4 and B2-F2) though there was significant smoke visible in the vicinity of the samplers. Over half of the increase in $PM_{2.5}$ mass was OC (57%) while 10% was EC. Major ionic species measured include nitrate (0.94%), ammonium (0.70%) and acetate (0.41%). XRF analysis showed that the primary trace elements in the fire are K (0.69 % by weight of $PM_{2.5}$), Na (0.33%), Cl (0.30%) and Mg (0.14%); trace elements accounted for approximately 2% of $PM_{2.5}$ mass (Table 5.2).

Total $PM_{2.5}$, OC, EC, ionic species and metals results are compared with results from Lee et al. (2005) (Tables 5.1 and 5.2, Figure 5.3) who conducted ambient measurements during April 2004 from two prescribed fires in pine-dominated forest in other parts of Georgia (Fort Benning, SE of Columbus and Fort Gordon, SW of Augusta). The Jones Center is dominated by pine (Pinus spp.) and longleaf pine (Pinus palustri). Results are similar between the studies for OC, Cl and K, but major differences were found for ammonium, Mg, Cu, P, Ca, and Mn, all of which are relatively minor species and typically not used as tracers for biomass burning. While both studies indicate that the biomass burning aerosol is dominated by OC, comprising approximately 60% of $PM_{2.5}$ mass, in this study EC comprised about 10% of the mass whereas *S. Lee et al.* [2005] determined EC to be about 4%.

The fractions of $PM_{2.5}$ components (i.e. mass of chemical species per mass of $PM_{2.5}$) during the flaming and smoldering stages are similar for all major $PM_{2.5}$ components (sulfate, nitrate, ammonium, OC and EC) (Table 5.3). Ratios of trace elements to $PM_{2.5}$ are significantly different in the flaming versus smoldering stages for several species (Table 5.3). The fractions of Na, Mg, Ca and Fe are greater in the smoldering stage. Potassium comprises a greater fraction of $PM_{2.5}$ in flaming versus smoldering, and therefore, the OC/K ratio is not consistent between the two stages. It has been reported that the majority of potassium emissions in a boreal wildfire occurs during the flaming stage [*Cahill et al.*, 2008]. In this study, the OC/K ratio was 44 in the flaming stage and 121 in the smoldering stage, indicating that the majority of potassium released in prescribed fires is released in the flaming stage.

5.4.2. Water Soluble Organic Carbon

Particulate water-soluble organic carbon (WSOC) is generally considered to have two dominant sources: secondary organic aerosol (SOA) formation and biomass burning [*Sullivan and Weber*, 2006; *Weber et al.*, 2007]. In the absence of significant biomass burning impacts, WSOC is often used as a surrogate of secondary organic aerosol (SOA) since the formation of SOA typically includes the addition of oxygen, making it more polar. Although OC levels in rural areas are often dominated by SOA, the WSOC measured during the burn periods in this study was expected to be largely primary organic aerosol (POA) since the samples mainly consisted of smoke from the fire.

High OC loadings, along with high levels of WSOC were measure during the fire. The average WSOC/OC ratio was 0.51 ± 0.033, which was similar to background levels before the first fire (Figure 5.4). However, for the background sample taken before the second fire, the ratio was 0.87. This is potentially due to residual primary WSOC and SOA formation from prescribed burning in the vicinity, including SOA formed from gaseous VOC emissions from the fires measured in this study. The influence of prescribed fires on Day 1 is seen from the background samples Day 2. For example, background levels of OC (3.97 µg m^{-3}) and levoglucosan (11.81 mg/g OC) on Day 1 are elevated to 7.75 µg m^{-3} and 21.27 mg/g OC, respectively, on Day 2 (Table 5.4). These results show that in regions where biomass burning is expected to affect air quality, the WSOC/OC ratio is not a good tracer of SOA.

5.4.3. Particulate Organic Carbon Speciation

During the fire events, large increases were observed for organic tracers of biomass burning such as levoglucosan, resin acids, and PAHs, including retene (Table 5.4 and Figure 5.5). Levoglucosan was the most abundant organic compound in the samples. In the field background samples, the average levoglucosan concentration was 47 ng m^{-3}. During the fire events, the levoglucosan concentrations increased to over 31,000 ng m^{-3}, a factor of approximately 200 above background levels. Levoglucosan contributed, on average, 12% of the

total OC. Concentrations of resin acids also increased, especially dehydroabietic acid and 7-oxodehydroabietic acid, increasing to 3809 and 718 ng m^{-3}, respectively. Resin acids increased by approximately 60 times from background. Retene, a potential softwood burning tracer, increased by a factor of about 38 from the background, and was the dominant PAH (Figure 5.6, Table 5.4). These ratios were higher on the first day of the fire because the background levels for these compounds were lower on that day compared to Day 2.

n-alkanes are associated with both biomass burning (especially plant waxes) and fossil fuel combustion [*Rogge et al.*, 1993; *Yan et al.*, 2008]. To distinguish between the relative impacts of both of these sources, the carbon preference index (CPI) was calculated since biomass burning shows a strong odd carbon number preference whereas fossil fuel combustion shows a strong even carbon number preference in *n*-alkanes [*Simoneit and Mazurek*, 1982]. In the field background samples, *n*-alkanes exhibited a slight odd carbon number predominance with a CPI of 1.1-1.4 However, *n*-alkanes in fire samples had a strong odd carbon number predominance (average CPI=2.6, carbon number maximum, C_{max}=31). For *n*-alkanoic acids, biomass burning has an even carbon number preference (opposite of *n*-alkanes), and both background and fire samples showed strong even carbon number predominance (average CPI=6.1, C_{max}=16), indicative of a biomass source [*Yan et al.*, 2008]. The concentration difference of the sum of even minus odd carbon number series during the prescribed fire was 343 ng m^{-3}, much larger than in the field background samples (on average, 26 ng m^{-3}).

Organic compounds were background corrected, normalized to total OC, and compared with the results from *S. Lee et al.* [2005] (Table 5.4). The OC was determined using a HVS, which has been shown by *S. Lee et al.* [2005] to a have a 40% positive artifact when they sampled two prescribed fires in Georgia. Thus, the values in Table 5.4 can be adjusted accordingly to proximate OC concentrations from a PCM that uses a carbon denuder. In general, the organic compound to OC ratios between the two studies are very comparable. In both studies, the POC is dominated by levoglucosan, resin acids and alkanoic acids. However, in *S. Lee et al.* [2005], the next dominant groups are alkenoic acids and n-alkanes, whereas in this study n-

alkanes are more dominant than alkenoic acids. Overall, the organic compound-to-OC ratios are slightly lower than *S. Lee et al.* [2005] except for levoglucosan/OC and n-alkanes/OC, which are slightly higher.

There is considerable variability between results from studies designed to characterize particulate organic emissions from biomass burning [*Fine et al.*, 2002; *Hays et al.*, 2002; *S. Lee et al.*, 2005; *Schauer et al.*, 2001; *Sinha et al.*, 2004]. While all of these studies show agreement on which compounds are the primary organic species, they differ widely in their organic compound to OC ratios (Figure 5.7). These differences are driven by both fuel characteristics, which can vary between prescribed burning and residential burning, as well as sampling techniques (i.e. laboratory versus field sampling).

5.4.4. VOC Speciation

Emission ratios are used to determine which gaseous species are affected by combustion in the flaming and smoldering stages of the fire. VOCs emitted from fires are important for regional air quality modeling since they can contribute to increases in ozone formation and SOA. Emission ratios of VOCs relative to CO_2 were determined from the slope of the least squares linear regressions between the mixing ratios of the individual VOC compounds and CO_2 measured absolutely (i.e., non-background corrected) in the flaming and smoldering stages (Tables S2-S3, Figures S2-S3). The VOC canister samples were distinguished into flaming (<0.1) and smoldering (>0.1) stages based on $\Delta CO/\Delta CO_2$ ratio s[*S. Lee et al.*, 2005]. The coefficient of determination (R^2 in Table D.2) indicates the extent to which gas emissions are correlated with the intensity of combustion within the fire.

5.4.5. Emission Factors

Results from this sampling provide an opportunity to estimate a regional prescribed fire emission factor for Georgia that can be utilized in air quality models as well as inform air quality management policy. Emission factors, based on the total carbon consumed and emitted to the

atmosphere, were calculated using the carbon mass balance method (Equation 1) [*Sinha et al.*, 2004]. In this method, all carbon consumed in the fire is assumed to be converted to CO_2, CO, CH_4, NMOCs, and particulate carbon (PC).

$$EF(X) = \left(\frac{[\Delta X]}{[\Delta C]_{CO_2} + [\Delta C]_{CO} + [\Delta C]_{CH_4} + [\Delta C]_{NMOC} + [\Delta C]_{PC}} \right) \times \left(\frac{mass_C}{mass_{fuel}} \right) \qquad \text{(Equation 1)}$$

where EF(X) is the emission factor for compound X in mass of X per mass of fuel and the ΔC terms in the denominator are excess (i.e. background corrected) carbon concentrations of CO_2, CO, CH_4, NMOCs and PC respectively (the $[\Delta C]_{PC}$ term was ignored since it contributes very little to total excess carbon). Emission factors were calculated separately for the gaseous and condensed phase components due to different sampling approaches.

Emissions factors for CO_2 (~1380 g /kg fuel burnt), CO (~86 g/kg) and CH_4 (~3 g/kg) are comparable to results found by Lee et al. (2005) (~1343, ~107, and ~4 g/kg, respectively) and Sinha et al. (2004) (~1732, ~58, and ~1 g/kg, respectively) (Table 5.5). VOC emissions are dominated by ethane, methanol and ethyne; methanol and ethanol emissions were found to be significantly greater in the smoldering phase versus the flaming stage. To find speciated $PM_{2.5}$ emission factors, the [ΔC] terms for CO_2, CO and CH_4, in Equation (1) need to be for the same time period as the $PM_{2.5}$ samples. However, CO_2, CO and VOC were measured using canisters that sample for only a few minutes whereas sampling periods for $PM_{2.5}$ ranged from 30 to 105 minutes. Given the lack of continuous CO_2 monitoring at the site, $PM_{2.5}$ samples were differentiated into flaming and smoldering based on the fire stage of the canisters that were closest in time to the $PM_{2.5}$ samples. Four $PM_{2.5}$ samples were identified as flaming and two as smoldering. We used this differentiation to compare fractions of $PM_{2.5}$ components (Table 5.3). For $PM_{2.5}$ emissions factors, we calculated an overall EF (i.e. an average of all samples) because there was approximately a two order of magnitude difference in background corrected (i.e. excess) CO and CO_2 concentrations that skewed results for $PM_{2.5}$ EFs in the smoldering phase,

for which there were only 2 samples. The overall EF for $PM_{2.5}$ was found to be 13.9 (± 17.3) g $PM_{2.5}$/kg fuel (Table 5.6). This is higher than the EFs found by *S. Lee et al.* [2005] of 0.66 and 1.14 g $PM_{2.5}$/kg fuel in the flaming and smoldering stages, respectively. Emission factors for individual $PM_{2.5}$ species were approximately one to two orders of magnitude greater than determined by *S. Lee et al.* [2005]. We also compared the EFs derived from this study with the Fire Emissions Production Simulator (FEPS) [*Anderson et al.*, 2004] and EPA's AP42 emissions inventory [*U.S.EPA*, 1995]. The EFs found here are comparable to both FEPS and AP42 (Table 5.7). However, our CO emissions factor is ~2.7 higher in the smoldering stage than in the flaming stage, while in FEPS the CO EFs ratio is ~1.6. This is ostensibly because FEPS allocates short-term smoldering emissions, which are enriched in CO, to the primary flaming stage. We also compared our VOCs with profile 5560 (Biomass Burning - Extratropical Forest) from EPA's Speciate 4.3 database. We matched Speciate VOCs with VOC measurements from this study and rescaled Speciate weight percentages; our results correlate well (R^2=0.96, slope of 1.01±0.04) (Figure 5.8).

Efforts to estimate source impacts often rely on using tracers specific to a source category, or estimating source impacts using receptor-based approaches, including factor analytic and CMB-based models. CMB models require apriori knowledge of source profiles, which have been shown to be a major source of uncertainty [*S. Lee and Russell*, 2007], and the results are often most sensitive to a few key species in each source. Therefore, several potential tracers of biomass burning were evaluated and emission ratios from this study were used as a new biomass burning source profile.

5.4.6. Comparison of Biomass Burning Tracers

Studies suggest that levoglucosan and retene may be appropriate biomass burning markers [*Simoneit*, 2002] though recent work suggests that levoglucosan may photodegrade [*Hennigan et al.*, 2010]. While potassium is enriched in biomass burning emissions, it may not be an ideal biomass burning marker since local urban sources (e.g., road dust) may be significant

[*Zhang et al.*, 2010] and that potassium may be a better tracer for wood combustion than for underbrush combustion [*Li et al.*, 2009]. However, when biomass burning impacts are expected to be significant, potassium can be a suitable tracer. We compared levoglucosan, potassium and retene, with WSOC (Figure 5.9). WSOC is most strongly correlated with levoglucosan (R^2 = 0.98) and to a lesser extent with potassium (R^2=0.93) and retene (R^2=0.67) (Figures 5.9a, 5.9b and 5.9c). The high correlation with WSOC suggests that levoglucosan may be a better marker for prescribed fires where underbrush is primarily consumed. Levoglucosan is well correlated with both potassium (R^2 = 0.90) and retene (R^2 = 0.80) (Figures 5.9d and 5.9e) while retene and potassium had a lower correlation (R^2 = 0.56). Given that potassium data are generally more available than levoglucosan data, potassium is the most often used tracer of biomass burning in source apportionment work, but its use during times when biomass burning is limited may affect source apportionment results.

5.4.7. Source Apportionment

Applying the CMB-LGO source apportionment method [*Marmur et al.*, 2005] to speciated $PM_{2.5}$ data from the Jefferson St. (JST) SEARCH site [*Hansen et al.*, 2003], we compared results using a biomass burning profile derived from this study with results using the composite profile from *Chow et al.* [2004b]. All other source category source profiles were identical in the two source apportionments. The chi-squared value, an overall goodness of fit metric, is very similar using the composite profile from *Chow et al.* [2004b] (2.80) and using the prescribed fire profile derived from this work (2.85); however, the distribution of mass is very different (Table 5.8). First, use of the prescribed fire profile from this study results in higher estimated to observed $PM_{2.5}$ mass ratios (0.94 versus 0.79). Second, the prescribed fire profile results in substantially more mass to biomass burning (average of 5.42 ± 2.62 µg m^{-3}) than with the *Chow et al.* [2004b] composite biomass burning profile (average of 1.00 ± 0.68 µg m^{-3}). The differences in the two source apportionments are driven by the amount of potassium in the respective biomass burning source profile. The composite profile from *Chow et al.* [2004b] has

potassium levels of 5.73 % ± 5.63%. The prescribed fire biomass burn source profile developed here contains 0.68% ± 0.33% for potassium (Figure 5.10); this has the net effect of leading to an increase in biomass burning impacts to match the measured concentrations of potassium. A range of values for potassium based on fuel type have been reported, where measured levels on a percent basis of $PM_{2.5}$ emissions range from 0.49 % ± 0.06% for Montana grass to 2.9 % ± 2.6% for Dambo grass [*Chen et al.*, 2007] and range from 0.2 – 1.8% for wood combustion [*Fine et al.*, 2002]. *S. Lee et al.* [2005], studying a prescribed fire in Georgia, found a similar amount of potassium as in this study (0.65 % ± 0.37%). These results indicate that regionally specific biomass burning source profiles should be used in source apportionment work. Another consequence is that the chi-squared value is not necessarily a good indicator of goodness of fit since two very different SA results can lead to similar chi-squared values.

5.5. Conclusions

Prescribed burning is increasingly being used as a tool in ecosystem management, underscoring the need to better characterize related emissions. $PM_{2.5}$ emissions during prescribed burn measurements consisted mainly of OC (~57%), EC (~10 %), chloride (~1.6%), potassium (~0.7%) and nitrate (~0.9%). Gaseous emissions were high in CO_2, CO, CH_4, ethane, ethyne, propene, benzene, acetaldehyde, methanol, ethanol and acetone during both flaming and smoldering stages.

WSOC increases significantly during the fire and is more strongly correlated with levoglucosan ($R^2 = 0.98$) and potassium ($R^2 = 0.93$) than retene ($R^2 = 0.67$). The average WSOC/OC ratio was 0.51 ± 0.033 and did not change significantly from background levels. This is likely due to the background containing SOA [*Zhang et al.*, 2012], leading to similar WSOC/OC ratios prior to and during the prescribed fire. Thus, the WSOC/OC ratio may not be a good indicator of SOA in regions that are expected to be impacted by biomass burning.

Organic compound to OC ratios are comparable to Lee et al. (2005) though the alkanoic and alkenoic acids to OC ratio is lower in this study. All of the major organic compound to OC

ratios fall within the broad range of values reported in other studies [*Fine et al.*, 2002; *Hays et al.*, 2002; *S. Lee et al.*, 2005; *Schauer et al.*, 2001; *Sinha et al.*, 2004].

Source profiles derived from the fire were utilized to apportion $PM_{2.5}$ impacts in Atlanta. The new profile has lower levels of K leading to an increase in the calculated amount of $PM_{2.5}$ from biomass burning.

Acknowledgements

This publication was made possible in part by USEPA STAR grants R833626, R833866, R834799 and RD83479901 and the Georgia State Department of Natural Resources (GA DNR). Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA or GA DNR. Further, USEPA and GA DNR do not endorse the purchase of any commercial products or services mentioned in the publication. We are grateful to the Joseph W. Jones Ecological Research Center for allowing access to sample the prescribed fires. We acknowledge the Southern Company for their support and thank Eric Edgerton of ARA, Inc. for access to the SEARCH data. We also acknowledge the Colombian Institute for the Development of Science and Technology for providing support to J.E.P.

5.6. References

Anderson, G., D. Sandberg, and R. Norheim (2004), Fire Emission Production Simulator (FEPS) User's Guide (version 1.0), *Prepared by the USDA Forest Service, Pacific Northwest Research Station, Seattle, WA*.

Barnard, W. S., E. (2003), Review of 1999 NEI and recommendations for developing the 2002 VISTAS inventory for regional haze modeling, edited by N. C. D. o. E. a. N. Resources, Asheville.

Birch, M. E. (1998), Analysis of carbonaceous aerosols: interlaboratory comparison, *Analyst*, *123*(5), 851-857.

Burling, I. R., R. J. Yokelson, S. K. Akagi, S. P. Urbanski, C. E. Wold, D. W. T. Griffith, T. J. Johnson, J. Reardon, and D. R. Weise (2011), Airborne and ground-based measurements of the trace gases and particles emitted by prescribed fires in the United States, *Atmos. Chem. Phys.*, *11*(23), 12197-12216.

Burling, I. R., et al. (2010), Laboratory measurements of trace gas emissions from biomass burning of fuel types from the southeastern and southwestern United States, *Atmos. Chem. Phys.*, *10*(22), 11115-11130.

Cahill, C. F., T. A. Cahill, and K. D. Perry (2008), The size- and time-resolved composition of aerosols from a sub-Arctic boreal forest prescribed burn, *Atmospheric Environment*, *42*(32), 7553-7559.

Chen, L. W. A., H. Moosmüller, W. P. Arnott, J. C. Chow, J. G. Watson, R. A. Susott, R. E. Babbitt, C. E. Wold, E. N. Lincoln, and W. M. Hao (2007), Emissions from laboratory combustion of wildland fuels: Emission factors and source profiles, *Environmental Science & Technology*, *41*(12), 4317-4325.

Chow, J. C., J. G. Watson, H. Kuhns, V. Etyemezian, D. H. Lowenthal, D. Crow, S. D. Kohl, J. P. Engelbrecht, and M. C. Green (2004), Source profiles for industrial, mobile, and area sources in the Big Bend Regional Aerosol Visibility and Observational study, *Chemosphere*, *54*(2), 185-208.

Colman, J. J., A. L. Swanson, S. Meinardi, B. C. Sive, D. R. Blake, and F. S. Rowland (2001), Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Analytical Chemistry*, *73*(15), 3723-3731.

Fine, P. M., G. R. Cass, and B. R. T. Simoneit (2002), Chemical characterization of fine particle emissions from the fireplace combustion of woods grown in the southern United States, *Environmental Science & Technology*, *36*(7), 1442-1451.

Friedlander, S. K. (1977), *Smoke, dust, and haze : fundamentals of aerosol behavior*, New York, Wiley.

Hansen, D. A., E. S. Edgerton, B. E. Hartsell, J. J. Jansen, N. Kandasamy, G. M. Hidy, and C. L. Blanchard (2003), The southeastern aerosol research and characterization study: Part 1-overview, *Journal of the Air & Waste Management Association*, *53*(12), 1460-1471.

Hardy, C. C., S. M. Hermann, and R. E. Mutch (2001), Smoke Management Guide for Prescribed and Wildland Fire, edited by C. C. Hardy, et al., National Wildfire Coordination Group, Boise, ID.

Hays, M. D., C. D. Geron, K. J. Linna, N. D. Smith, and J. J. Schauer (2002), Speciation of gas-phase and fine particle emissions from burning of foliar fuels, *Environmental Science & Technology*, *36*(11), 2281-2295.

Hennigan, C. J., A. P. Sullivan, J. L. Collett, and A. L. Robinson (2010), Levoglucosan stability in biomass burning particles exposed to hydroxyl radicals, *Geophys. Res. Lett.*, *37*, L9806.

Kim, E., P. K. Hopke, and E. S. Edgerton (2004), Improving source identification of Atlanta aerosol using temperature resolved carbon fractions in positive matrix factorization, *Atmospheric Environment*, *38*(20), 3349-3362.

Lee, S., and A. G. Russell (2007), Estimating uncertainties and uncertainty contributors of CMB PM2.5 source apportionment results, *Atmospheric Environment*, *41*(40), 9616-9624.

Lee, S., K. Baumann, J. J. Schauer, R. J. Sheesley, L. P. Naeher, S. Meinardi, D. R. Blake, E. S. Edgerton, A. G. Russell, and M. Clements (2005), Gaseous and particulate emissions from prescribed burning in Georgia, *Environmental Science & Technology*, *39*(23), 9049-9056.

Li, Z., E. N. Porter, A. Sjodin, L. L. Needham, S. Lee, A. G. Russell, and J. A. Mulholland (2009), Characterization of PM2.5-bound polycyclic aromatic hydrocarbons in Atlanta-Seasonal variations at urban, suburban, and rural ambient air monitoring sites, *Atmospheric Environment*, *43*(27), 4187-4193.

Liu, W., Y. H. Wang, A. Russell, and E. S. Edgerton (2005), Atmospheric aerosol over two urban-rural pairs in the southeastern United States: Chemical composition and possible sources, *Atmospheric Environment*, *39*(25), 4453-4470.

Marmur, A., A. Unal, J. A. Mulholland, and A. G. Russell (2005), Optimization-based source apportionment of PM2.5 incorporating gas-to-particle ratios, *Environmental Science & Technology*, *39*(9), 3245-3254.

NIFC (2008), Wildland Fire Summary and Statistics, edited by N. I. C. Center, National Interagency Fire Center.

NIOSH (1996), Elemental carbon (diesel particulate): method 5040, in *NIOSH Manual of Analytical Methods*, edited by P. M. Eller and M. E. Cassinelli, National Institute for Occupational Safety and Health, Cincinatti.

Oakes, M., N. Rastogi, B. J. Majestic, M. Shafer, J. J. Schauer, E. S. Edgerton, and R. J. Weber (2010), Characterization of soluble iron in urban aerosols using near-real time data, *Journal of Geophysical Research-Atmospheres*, *115*, D15302.

Paatero, P., and U. Tapper (1994), Positive matrix factorization - a nonnegative factor model with optimal utilization of error-estimates of data values, *Environmetrics*, *5*(2), 111-126.

Rastogi, N., M. M. Oakes, J. J. Schauer, M. M. Shafer, B. J. Majestic, and R. J. Weber (2009), New Technique for Online Measurement of Water-Soluble Fe(II) in Atmospheric Aerosols, *Environmental Science & Technology*, *43*(7), 2425-2430.

Rogge, W. F., L. M. Hildemann, M. A. Mazurek, G. R. Cass, and B. R. T. Simoneit (1993), SOURCES OF FINE ORGANIC AEROSOL .4. PARTICULATE ABRASION PRODUCTS FROM LEAF SURFACES OF URBAN PLANTS, *Environmental Science & Technology*, *27*(13), 2700-2711.

Sandberg, D. V., R. D. Ottmar, J. L. Peterson, and J. Core (2002), Wildland fire on ecosystems: effects of fire on air; General Technical Report edited by U. F. Service, USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO.

Sapkota, A., J. M. Symons, J. Kleissl, L. Wang, M. B. Parlange, J. Ondov, P. N. Breysse, G. B. Diette, P. A. Eggleston, and T. J. Buckley (2005), Impact of the 2002 Canadian forest fires on particulate matter air quality in Baltimore City, *Environmental Science & Technology*, *39*(1), 24-32.

Schauer, J. J., M. J. Kleeman, G. R. Cass, and B. R. T. Simoneit (2001), Measurement of emissions from air pollution sources. 3. C-1-C-29 organic compounds from fireplace combustion of wood, *Environmental Science & Technology*, *35*(9), 1716-1728.

Simoneit, B. R. T. (2002), Biomass burning — a review of organic tracers for smoke from incomplete combustion, *Applied Geochemistry*, *17*(3), 129-162.

Simoneit, B. R. T., and M. A. Mazurek (1982), Organic matter of the troposphere—II. Natural background of biogenic lipid matter in aerosols over the rural western united states, *Atmospheric Environment (1967)*, *16*(9), 2139-2159.

Sinha, P., P. V. Hobbs, R. J. Yokelson, D. R. Blake, S. Gao, and T. W. Kirchstetter (2004), Emissions from miombo woodland and dambo grassland savanna fires, *Journal of Geophysical Research-Atmospheres*, *109*(D11), D11305.

Stookey, L. L. (1970), FERROZINE - A NEW SPECTROPHOTOMETRIC REAGENT FOR IRON, *Analytical Chemistry*, *42*(7).

Sullivan, A. P., and R. J. Weber (2006), Chemical characterization of the ambient organic aerosol soluble in water: 2. Isolation of acid, neutral, and basic fractions by modified size-exclusion chromatography, *Journal of Geophysical Research-Atmospheres*, *111*(D5), D05314.

U.S.EPA (1995), Compilation of Air Pollution Emission Factors, edited by U. S. E. P. A. Office of Air Quality & Standards, Research Triangle Park, NC.

Urbanski, S. P., W. M. Hao, and B. Nordgren (2011), The wildland fire emission inventory: western United States emission estimates and an evaluation of uncertainty, *Atmos. Chem. Phys.*, *11*(24), 12973-13000.

Watson, J. G., J. A. Cooper, and J. J. Huntzicker (1984), The effective variance weighting for least-squares calculations applied to the mass balance receptor model, *Atmospheric Environment*, *18*(7), 1347-1355.

Weber, R. J., et al. (2007), A study of secondary organic aerosol formation in the anthropogenic-influenced southeastern United States, *Journal of Geophysical Research-Atmospheres*, *112*(D13), D13302.

Wotawa, G., and M. Trainer (2000), The influence of Canadian forest fires on pollutant concentrations in the United States, *Science*, *288*(5464), 324-328.

Yan, B., M. Zheng, Y. T. Hu, S. Lee, H. K. Kim, and A. G. Russell (2008), Organic composition of carbonaceous aerosols in an aged prescribed fire plume, *Atmos. Chem. Phys.*, *8*(21), 6381-6394.

Zhang, X., A. Hecobian, M. Zheng, N. H. Frank, and R. J. Weber (2010), Biomass burning impact on PM2.5 over the southeastern US during 2007: integrating chemically speciated FRM filter measurements, MODIS fire counts and PMF analysis, *Atmos. Chem. Phys.*, *10*(14), 6839-6853.

Zhang, X., Z. Liu, A. Hecobian, M. Zheng, N. H. Frank, S. Edgerton, and R. J. Weber (2012), Spatial and seasonal variations of fine particle water-soluble organic carbon (WSOC) over the southeastern United States: implications for secondary organic aerosol formation, *Atmos. Chem. Phys.*, *12*(14), 6593-6607.

Zheng, M., L. Ke, E. S. Edgerton, J. J. Schauer, M. Y. Dong, and A. G. Russell (2006), Spatial distribution of carbonaceous aerosol in the southeastern United States using molecular markers and carbon isotope data, *Journal of Geophysical Research-Atmospheres*, *111*(D10), D10S06

Zielinska, B., J. D. McDonald, T. Hayes, J. C. Chow, E. M. Fujita, and J. G. Watson (1998), Northern Front Range Air Quality Study Final Report, Volume B: Source Measurements*Rep.*, Report to Colorado State University by Desert Research Institute, Reno, NV.

5.7. Tables

Table 5.1: Particulate Matter Chemical Composition of Emissions from Prescribed Burning: This study versus Lee et al. (2005)

	This study		Lee et al. (2005)	
	Average	**St. Dev.**	**Average**	**St. Dev.**
$PM_{2.5}$ ($\mu g\ m^{-3}$)	161	94	1810	680
OC and EC (weight % of $PM_{2.5}$ mass)				
Organic carbon	57	3	60	18
Elemental carbon	10	0.64	3.9	1.1
WSOC	28	2.3		
Ionic species (weight % of $PM_{2.5}$ mass)				
Acetate	0.41	0.09	0.55	0.16
Formate	*	*	0.45	0.11
Nitrate	0.94	1.00	0.44	0.30
Sulfate	0.23	0.36	0.25	0.11
Oxalic Acid	0.19	0.19	0.069	0.014
Ammonium	0.70	0.37	0.11	0.11

*Below Quantification Limit

Table 5.2: Metals Composition from Prescribed Burning (weight % of PM2.5 mass): This study versus Lee et al. (2005).

	This Study		Lee et al. (2005)	
	Average	**St. Dev**	**Average**	**St. Dev**
Na	0.3286	0.1610	0.0431	0.0175
Mg	0.1405	0.0028	0.0001	0.0003
Al	0.0719	0.0926	0.0229	0.0426
Si	0.0188	0.0707	0.0186	0.0258
P	0.0794	0.0729	0.0010	0.0015
S	0.0888	0.0827	0.1074	0.0403
Cl	0.3009	0.1549	0.4217	0.2295
K	0.6846	0.3336	0.5707	0.3711
Ca	0.0881	0.0138	0.0006	0.0011
Sc	0.0289	0.0002		
Ti	0.0029	0.0224	0.0004	0.0006
Cr	0.0029	0.0224	BL[a]	BL
Mn	0.0319	0.0183	0.0011	0.0010
Fe	0.0586	0.0245	0.0082	0.0137
Cu	0.0137	0.0084	0.0010	0.0010
Zn	0.0099	0.0092	0.0160	0.0089
Se	0.0097	0.0427	0.0001	0.0002
Br	0.0118	0.0003	0.0141	0.0091
Rb	0.0021	0.0093	0.0042	0.0028
Pb	0.0087	0.0220	0.0001	0.0003
Y	0.0100	0.0127	-[b]	-
Zr	0.0019	0.0296	-	-
Nb	0.0024	0.0228	-	-
Mo	BL	0.0203	-	-
Ag	0.0060	0.0355	-	-
Sb	0.0182	0.0622	-	-
La	0.0012	0.0076	-	-
Ce	0.0102	0.0284	-	-
Tb	0.0037	0.0471	-	-
Hf	0.0200	0.1363	-	-
Ta	0.0177	0.1017	-	-
Ir	0.0000	0.0758	-	-
Au	0.0106	0.0673	-	-
Tl	0.0071	0.0508	-	-
V	-	-	BL	BL
Co	-	-	BL	BL
Ni	-	-	BL	BL
Ga	-	-	BL	BL
Ge	-	-	BL	BL
As	-	-	0.0002	0.0003
Sr	-	-	0.0002	0.0003

[a]BL: Below blank levels. [b] -: Not detected or all measured values were 0.

Table 5.3: $PM_{2.5}$ major components (as wt. %) and metals ratios to total $PM_{2.5}$ (mg/g of total $PM_{2.5}$) in flaming and smoldering stages, averaged over two prescribed fires.

	Flaming		Smoldering	
Major Components Units (% of Total $PM_{2.5}$)	**Avg.**	**St. Dev.**	**Avg.**	**St. Dev.**
Other	27%	0.9%	29%	0.29%
OC	55%	1.8%	58%	0.59%
EC	11 %	1.0%	10%	0.74%
Acetate	0.40%	0.16%	0.79%	0.62%
Nitrate	1.45%	0.76%	1.55%	1.36%
Sulfate	0.48%	0.08%	0.61%	0.49%
Oxalic Acid	0.35%	0.34%	0.27%	0.21%
Ammonium	0.86%	0.22%	0.81%	0.07%
$PM_{2.5}$ ($\mu g\ m^{-3}$)	**327.15**	**148.32**	**48.52**	**9.45**
Metals by XRF Units (mg/g of Total $PM_{2.5}$)	**Avg.**	**St. Dev.**	**Avg.**	**St. Dev.**
Na	1.32	N/A	17.54	9.25
Mg	-	N/A	3.87	N/A
Al	1.38	N/A	1.15	0.079
Si	0.37	0.12	2.13	N/A
P	0.34	N/A	3.88	1.38
S	1.32	0.90	-	N/A
Cl	3.22	2.59	0.099	N/A
K	8.84	1.49	0.66	0.89
Ca	1.55	1.52	3.22	0.99
Mn	0.77	0.17	-	N/A
Fe	0.26	0.095	2.61	4.36
Cu	0.39	0.23	0.34	N/A
Zn	0.16	0.11	0.98	N/A
Br	0.10	0.11	0.33	N/A

Table 5.4: Organic compounds summary and comparison with Lee et al. (2005).

	Background Day 1 (µg m^{-3})	**Day 1 Avg.* (µg m^{-3})**	**Background Day 2 (µg m^{-3})**	**Day 2 Avg.* (µg m^{-3})**	**Study Avg. (µg m^{-3})**	**Lee et al., (2005) (µg m^{-3})**
Total OC	3.97	123.89	7.75	55.88	89.88	1090.53
	Background Day 1 (mg/g OC)	**Day 1 Avg.* (mg/g OC)**	**Background Day 2 (mg/g OC)**	**Day 2 Avg.* (mg/g OC)**	**Study Avg. (mg/g OC)**	**Lee et al., (2005) (mg/g OC)**
n-alkanes	8.41	3.26	3.39	3.62	3.44	2.36
Branched alkanes	0.00	0.05	0.04	0.09	0.07	-
Alkanoic acids	8.10	5.58	4.99	5.76	5.67	27.38
Alkenoic Acids	9.42	1.40	4.11	1.34	1.37	5.20
Hopanes	0.05	0.00	0.06	0.00	0.00	-
Retene	0.10	0.21	0.26	0.31	0.26	0.35
PAHs	0.18	0.93	0.49	1.01	0.97	1.47
Resin Acids	5.43	19.71	18.20	25.49	22.60	38.74
Others	0.00	2.63	0.00	4.80	3.71	-
Levoglucosan	11.81	116.53	21.27	132.57	124.55	94.75
Cholesterol	0.28	0.27	0.54	0.22	0.24	0.81

*background corrected

Table 5.5: Emission factors for selected gaseous components (g per kg fuel burned), averaged over two prescribed fires. (NOTE: See Table D.4 for complete list).

Gaseous Species	Flaming Avg.	Flaming St. Dev.	Smoldering Avg.	Smoldering St. Dev.
CH_4 (g/kg)	2.32	1.17	3.48	1.48
CO (g/kg)	48.08	35.02	133.30	54.16
CO_2 (g/kg)	1425.14	63.78	1324.67	88.28
OCS (g/kg)	0.0061	0.0030	0.0088	0.0003
CS_2 (g/kg)	0.0014	0.0002	0.0016	0.0011
CH_3Cl (g/kg)	0.0167	0.0120	0.0181	0.0048
CH_2Cl_2 (g/kg)	0.0008	0.0007	0.0002	0.0003
Ethane (g/kg)	0.1459	0.0817	0.3494	0.1144
Ethene (g/kg)	1.0372	0.3747	1.1285	0.2070
Ethyne (g/kg)	0.4285	0.1609	0.4576	0.0742
Propane (g/kg)	0.0359	0.0365	0.0856	0.0551
Propene (g/kg)	0.2742	0.1108	0.3975	0.0094
Benzene (g/kg)	0.2197	0.0859	0.2625	0.0278
Toluene (g/kg)	0.1018	0.0613	0.1474	0.0178
p-Xylene* (g/kg)	0.0021	-	-	-
o-Xylene* (g/kg)	0.0007	-	-	-
Acetaldehyde (g/kg)	0.2875	0.1767	0.6771	0.2669
Methanol(g/kg)	0.6301	0.5132	2.0997	1.2559
Ethanol (g/kg)	0.1239	0.1676	0.4195	0.0514
Acetone (g/kg)	0.3111	0.2417	0.4180	0.1693
MAC g/kg)	0.0351	0.0335	0.0668	0.0301
MVK (g/kg)	0.0284	0.0350	0.0797	0.0569
Total NMOCs (g/kg)	3.08	2.43	7.14	1.94

*No standard deviation was calculated in the flaming stage because there was one sample with level above detection. Also, no samples above detection were measured in the smoldering stage.

Table 5.6: Emission Factors for major $PM_{2.5}$ species and selected trace elements (g/kg fuel burned) (NOTE: See Table D.5 for complete list).

	Avg.	St. Dev.
PM2.5 (g/kg)	13.87	17.27
OC (g/kg)	7.15	9.32
EC (g/kg)	1.40	1.80
Acetate (g/kg)	0.05	0.07
Nitrate (g/kg)	0.19	0.27
Sulfate (g/kg)	0.07	0.09
Oxalic Acid (g/kg)	0.04	0.05
Ammonium (g/kg)	0.12	0.18
Na (mg/kg)	31.33	26.37
Mg (mg/kg)	12.94	-[a]
Al (mg/kg)	2.12	2.37
Si (mg/kg)	2.76	5.52
P (mg/kg)	5.36	6.20
S (mg/kg)	35.80	43.74
Cl (mg/kg)	72.97	120.10
K (mg/kg)	116.45	180.49
Ca (mg/kg)	12.21	16.85
Ti (mg/kg)	0.01	-[a]
V (mg/kg)	BL[b]	n/a
Cr (mg/kg)	0.04	0.02
Mn (mg/kg)	16.85	4.91
Fe (mg/kg)	3.18	5.80
Ni (mg/kg)	BL[b]	n/a
Cu (mg/kg)	6.06	9.81
Zn (mg/kg)	1.82	2.00
As (mg/kg)	BL[b]	n/a
Se (mg/kg)	1.29	1.63
Br (mg/kg)	2.50	4.01

-[a]only one sample had measured concentrations greater than detection limit.

[b]BL: Below blank levels

Table 5.7: Comparison of Emission Factors (EFs) (g/kg fuel burned) with Fire Emissions Production Simulator FEPS and AP42

	Carbon Balance Method (This Study)			FEPS[a]			AP42[b]		
EF	**Flaming**	**Smoldering**	**Overall**	**Flaming**	**Smoldering**	**Overall**	**Flaming**	**Smoldering**	**Overall**
CO2	1420 ± 64	1320 ± 88	1380 ± 80	-	-	-	-	-	-
CO	48 ± 35	130 ± 54	86 ± 61	141 ± 5	214 ± 1	166 ± 42	45	166	126
CH_4	2.3 ± 1.2	3.5 ± 1.5	2.8 ± 1.4	6.7 ± 0.2	11± 2	7.8 ± 1.8	1.5	7.7	5.7
VOCs	3.1 ± 2.4	7.1 ± 1.9	4.9 ± 3.0	-	-	-	1.7	5.4	4.2
$PM_{2.5}$	-	-	14 ± 17	12± 0.3	17 ± 1.3	13± 2.7	6	16	13

[a]**See Appendix D (Table D.6) for additional model run information.**
[b]**For prescribed burning in long leaf conifer forests.**

Table 5.8: Source Apportionment Results for Jefferson St. (JST), Atlanta, GA for 1/1/07 – 12/31/07 using two different biomass burning profiles.

	Composite Source Profile* [Chow et al., 2004]	Biomass Burning Profile from this Study
χ^2	2.80 (2.23)	2.85 (2.20)
Predicted PM2.5	12.05 (5.92)	14.06 (6.45)
PM2.5 ratio	0.79 (0.16)	0.94 (0.20)
Gasoline Vehicles	0.77 (0.45)	0.75 (0.44)
Diesel Vehicles	0.96 (0.81)	0.53 (0.68)
Dust	0.34 (0.32)	0.32 (0.32)
Biomass Burning	1.00 (0.68)	5.42 (2.62)
Coal Combustion	0.07 (0.07)	0.07 (0.07)
Ammonium Sulfate	3.61 (3.55)	3.55 (3.54)
Ammonium Bisulfate	1.86 (1.35)	1.92 (1.31)
Ammonium Nitrate	1.01 (0.97)	0.95 (0.97)
Other OC (SOC)	2.42 (2.24)	0.53 (1.69)

*This profile is the BURN profile from *Chow et al.* [2004b] and is an average of 19 vegetative burning profiles.

5.8. Figures

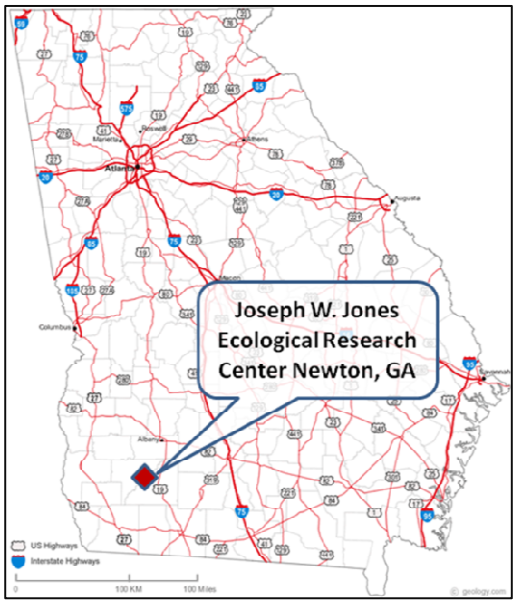


Figure 5.1: Location of Jones Ecological Research Center in Southwestern Georgia

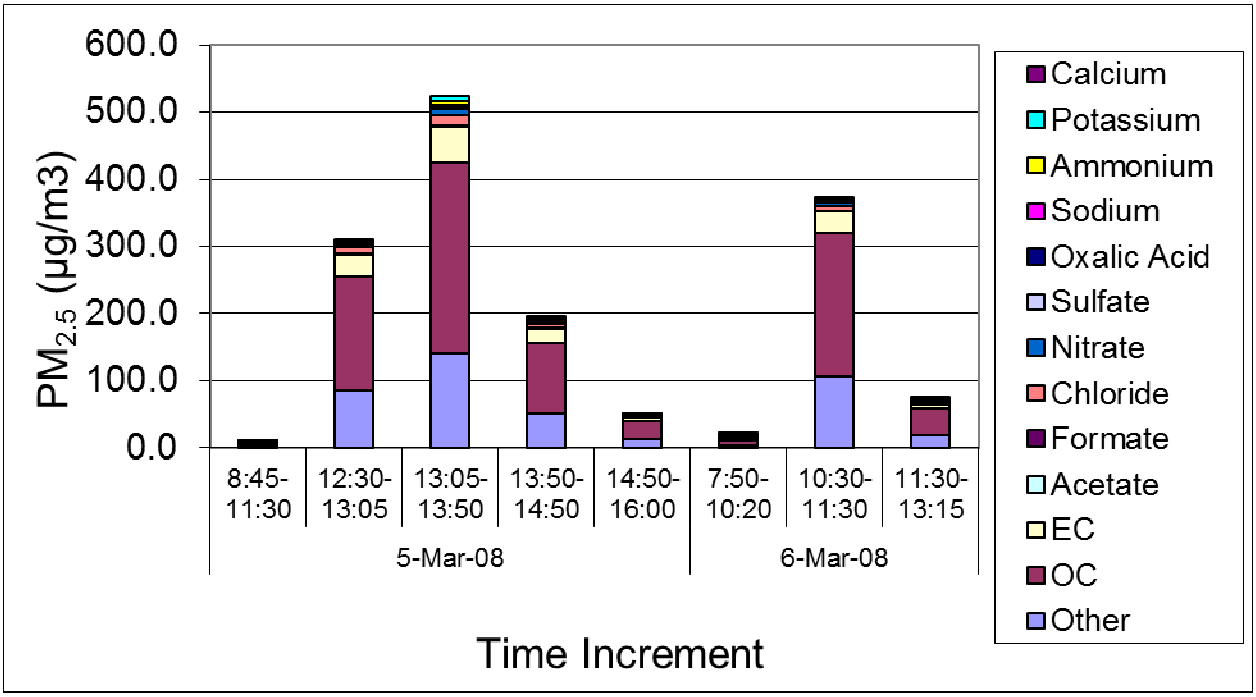


Figure 5.2: PM2.5 composition of the prescribed fires sampled in this study.

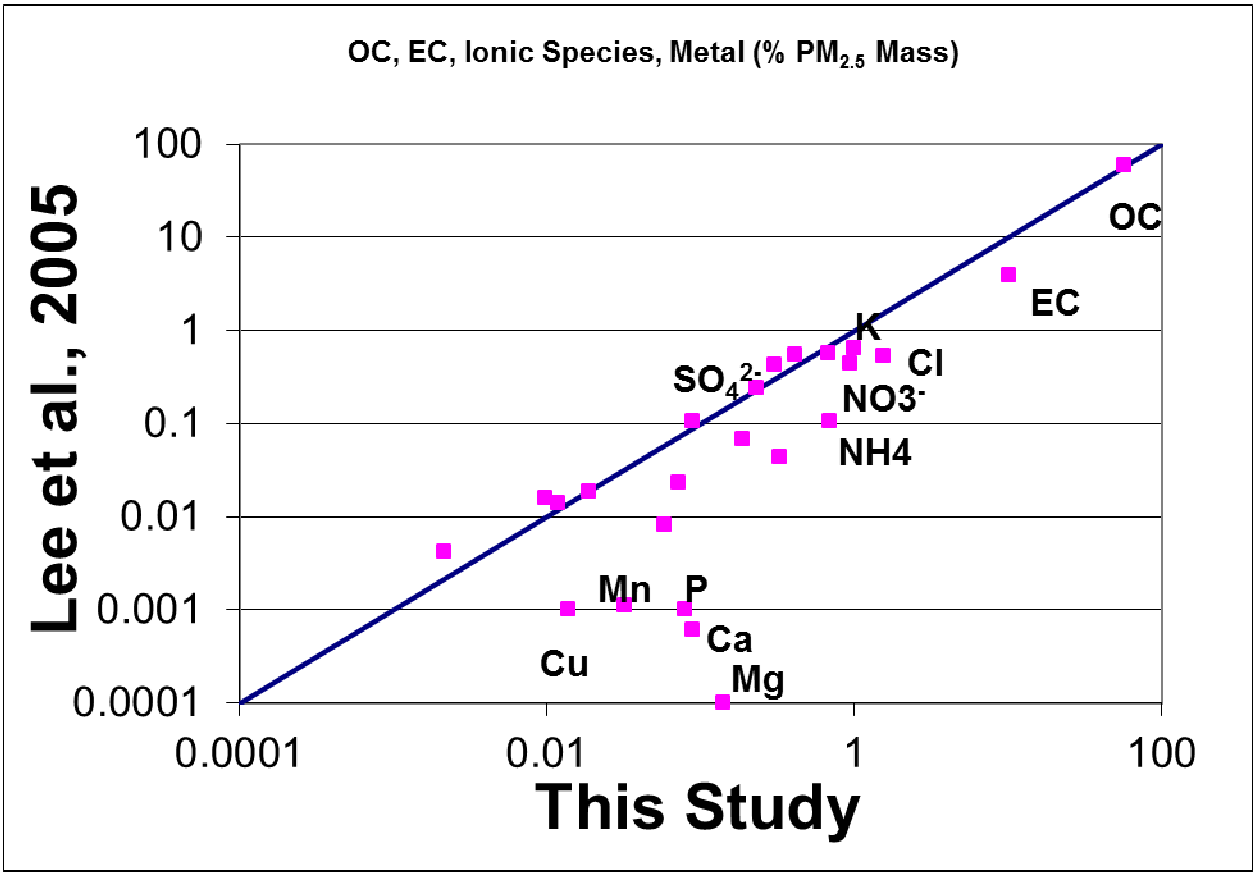


Figure 5.3: Comparison of the Chemical Composition of Particle-Phase Emissions from Prescribed Burning between this study and Lee et al. (2005)

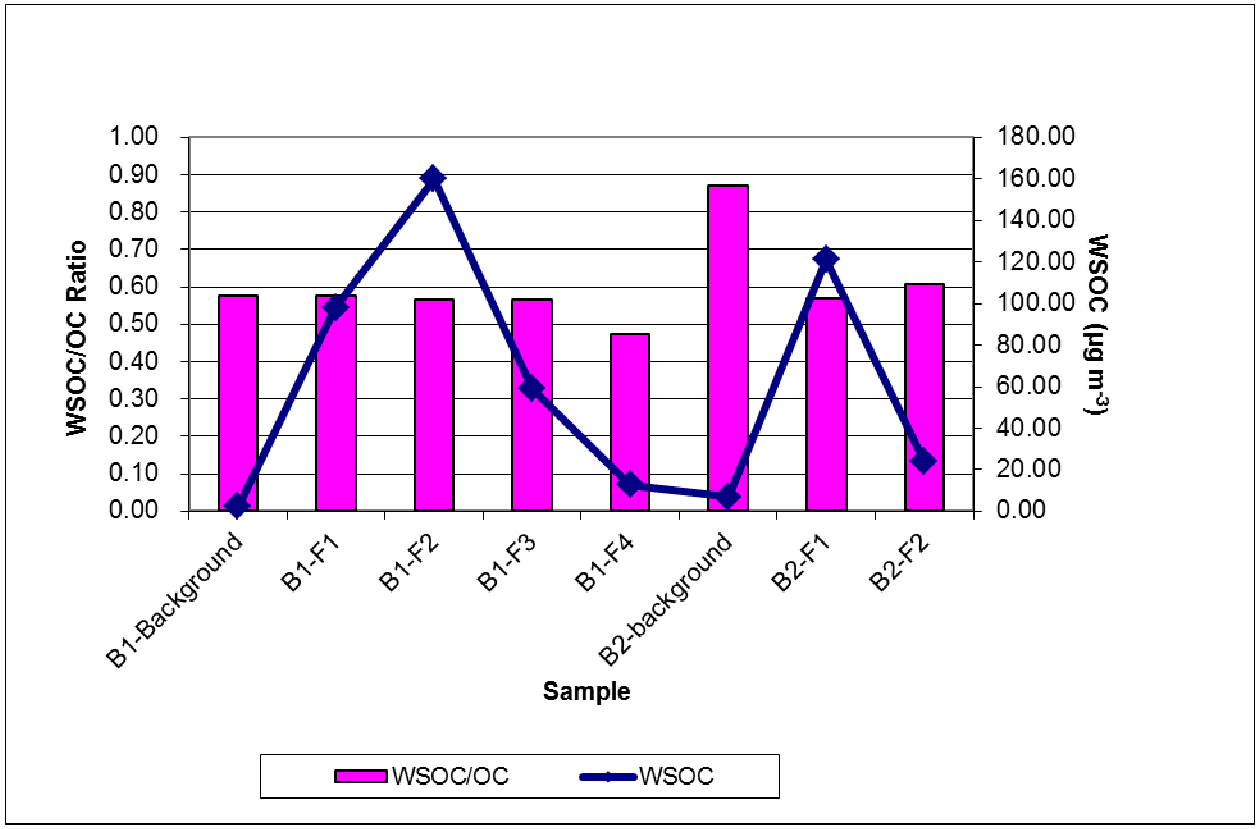


Figure 5.4: WSOC and WSOC/OC ratios of the prescribed fires sampled in this study.

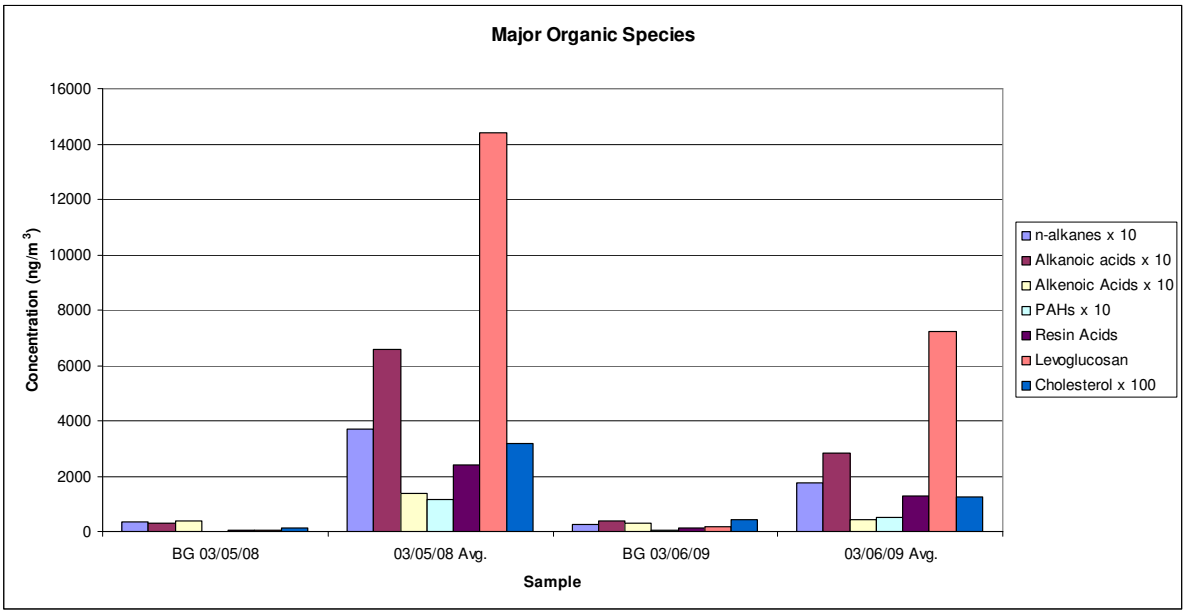


Figure 5.5: Major organic compound groups in background (BG) and averaged during burning events for both days.

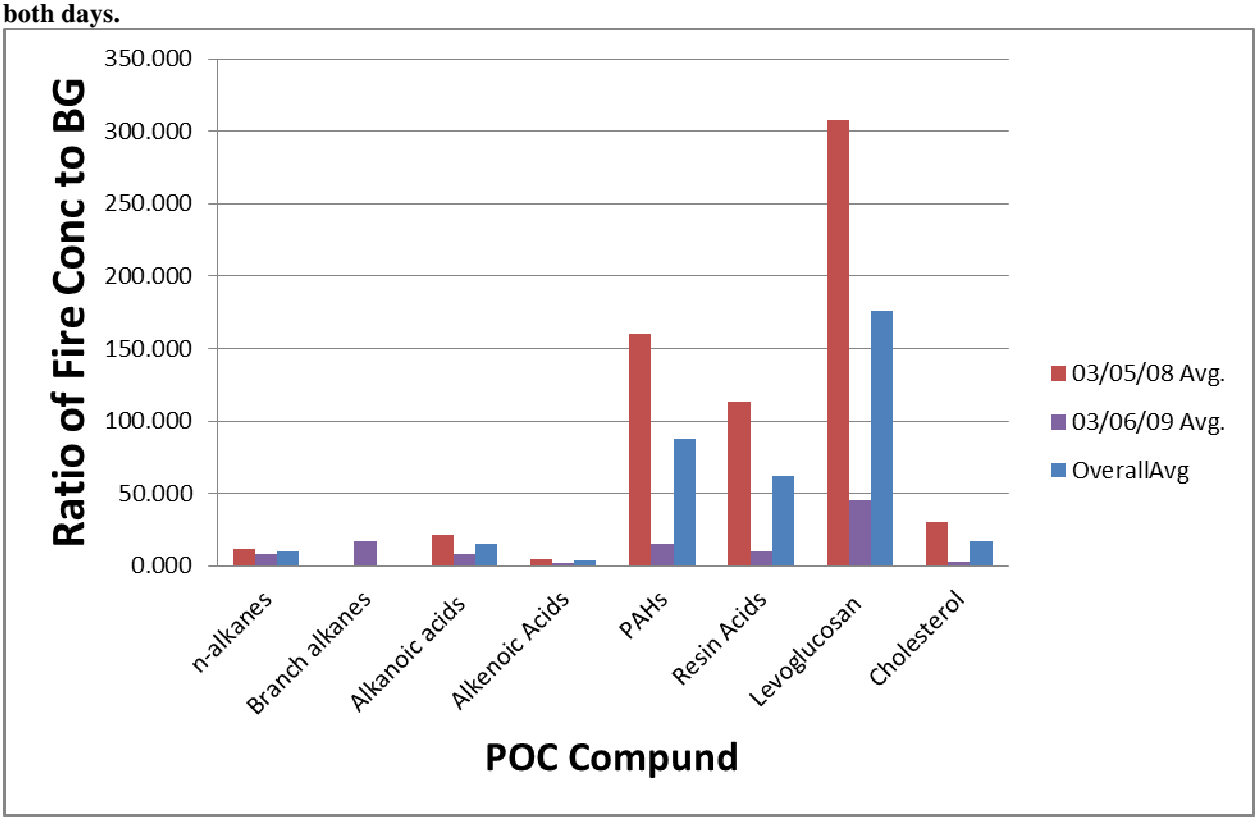


Figure 5.6: Ratio of major organic compound (averaged during burn events) groups to background (BG).

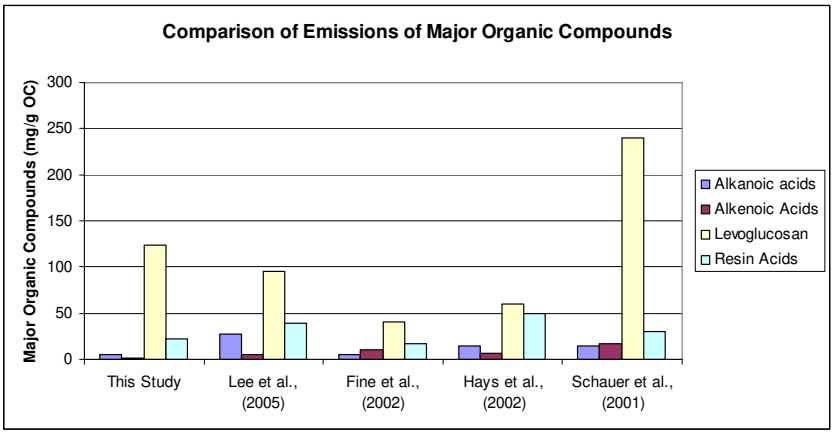


Figure 5.7: Comparison of emissions of major organic compounds between this study and previous studies.

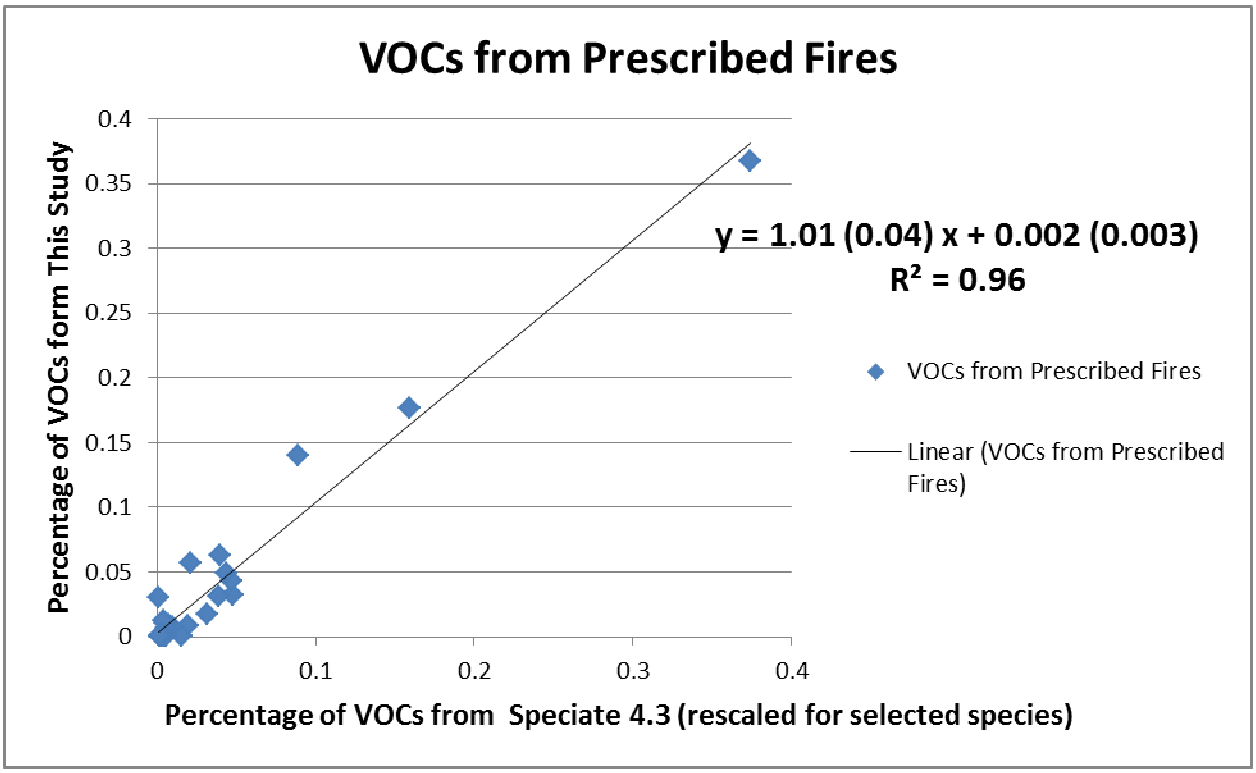


Figure 5.8: Comparison of VOCs between this study and EPA's Speciate 4.3 database for profile 5560: Biomass Burning - Extratropical Forest

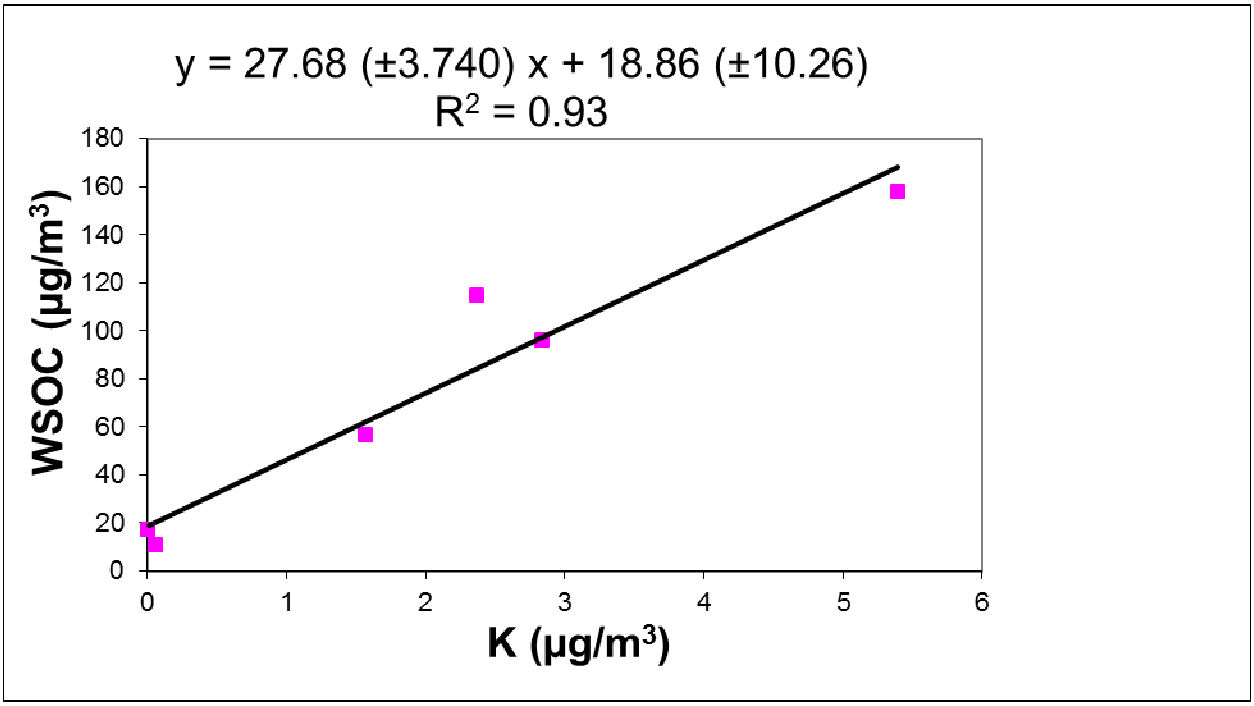


Figure 5.9a

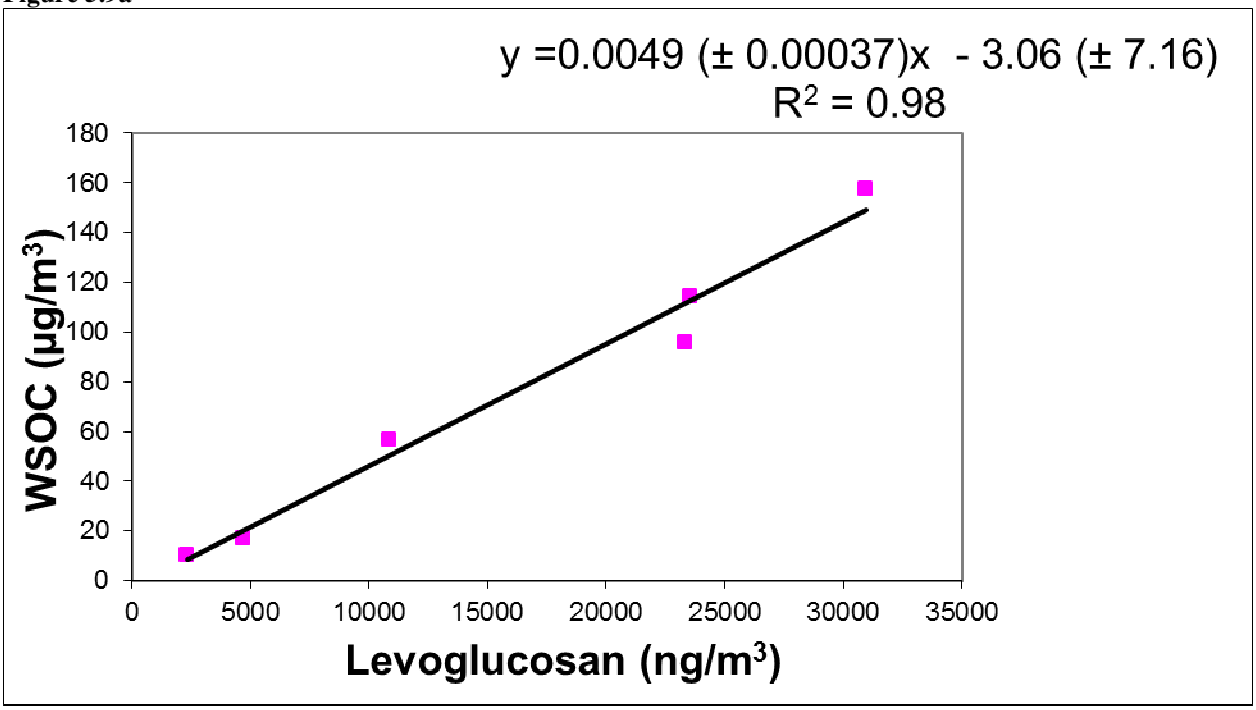


Figure 5.9b

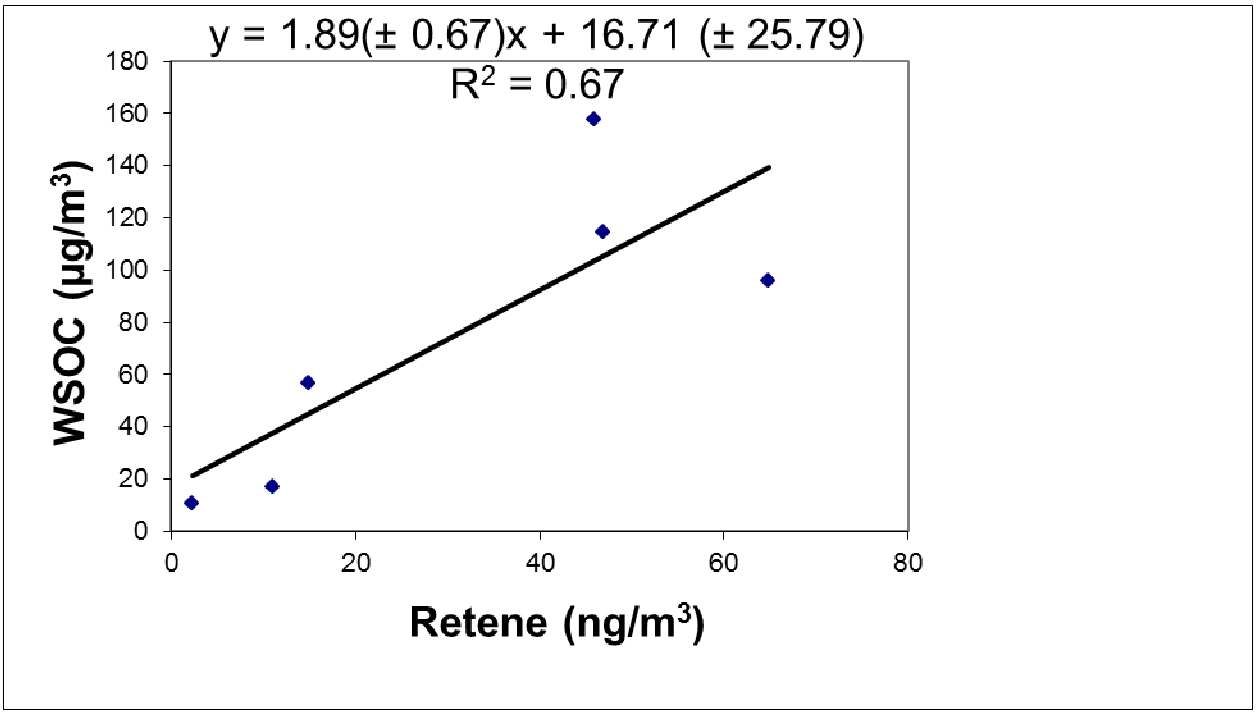


Figure 5.9c

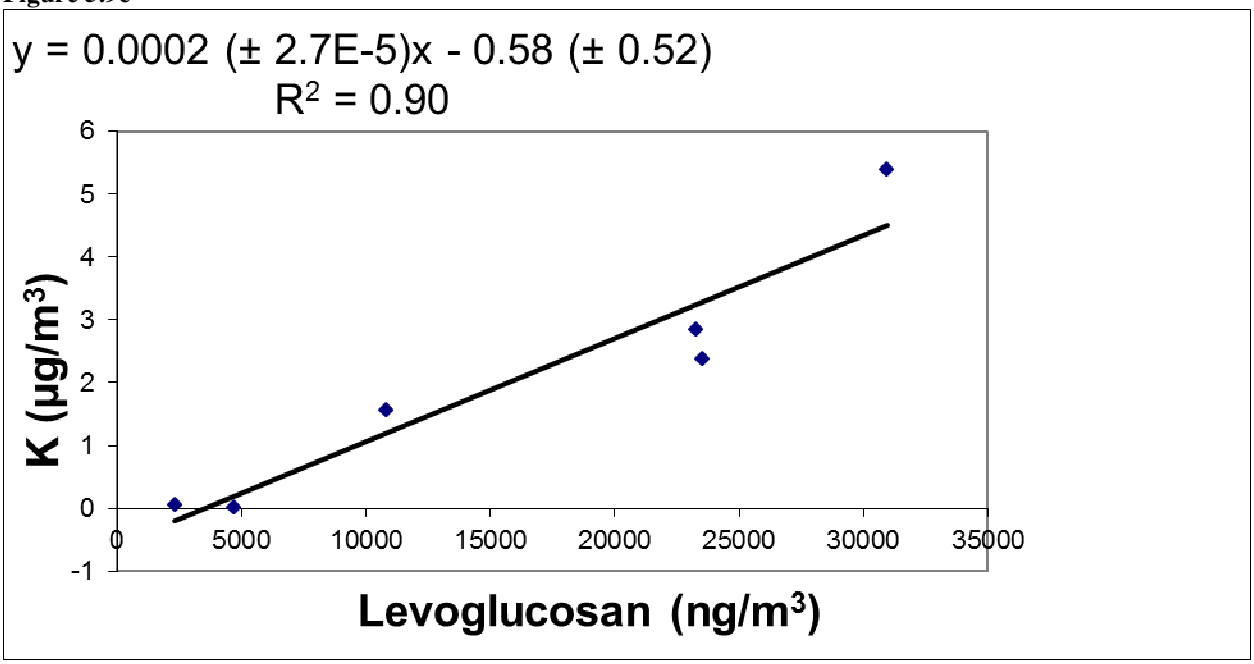


Figure 5.9d

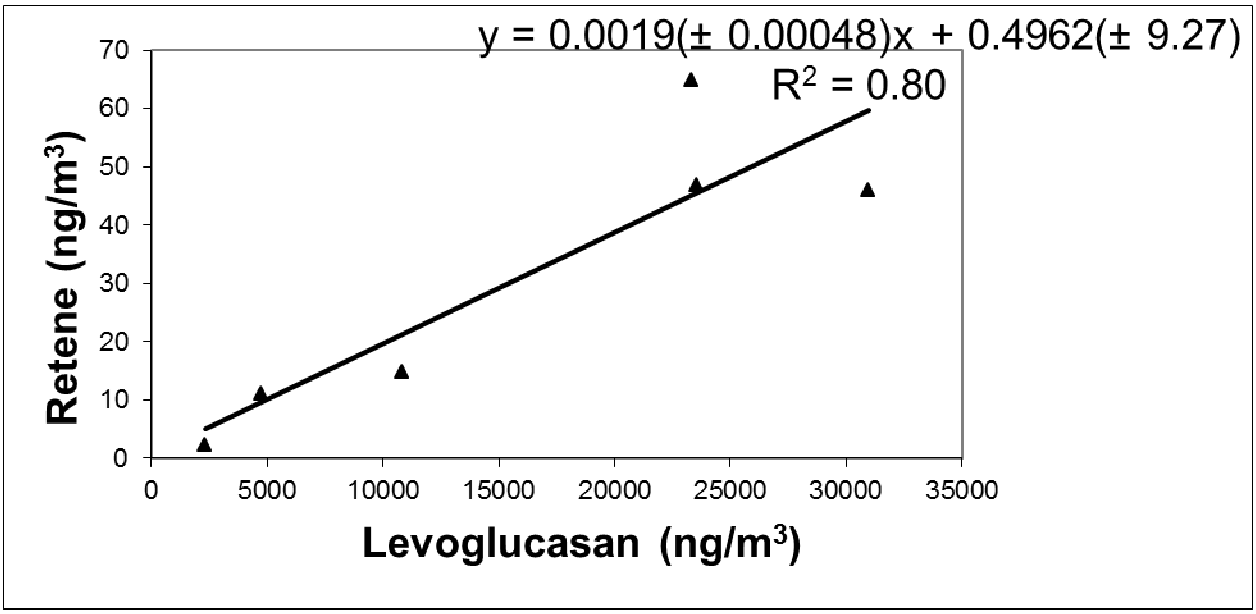


Figure 5.9e

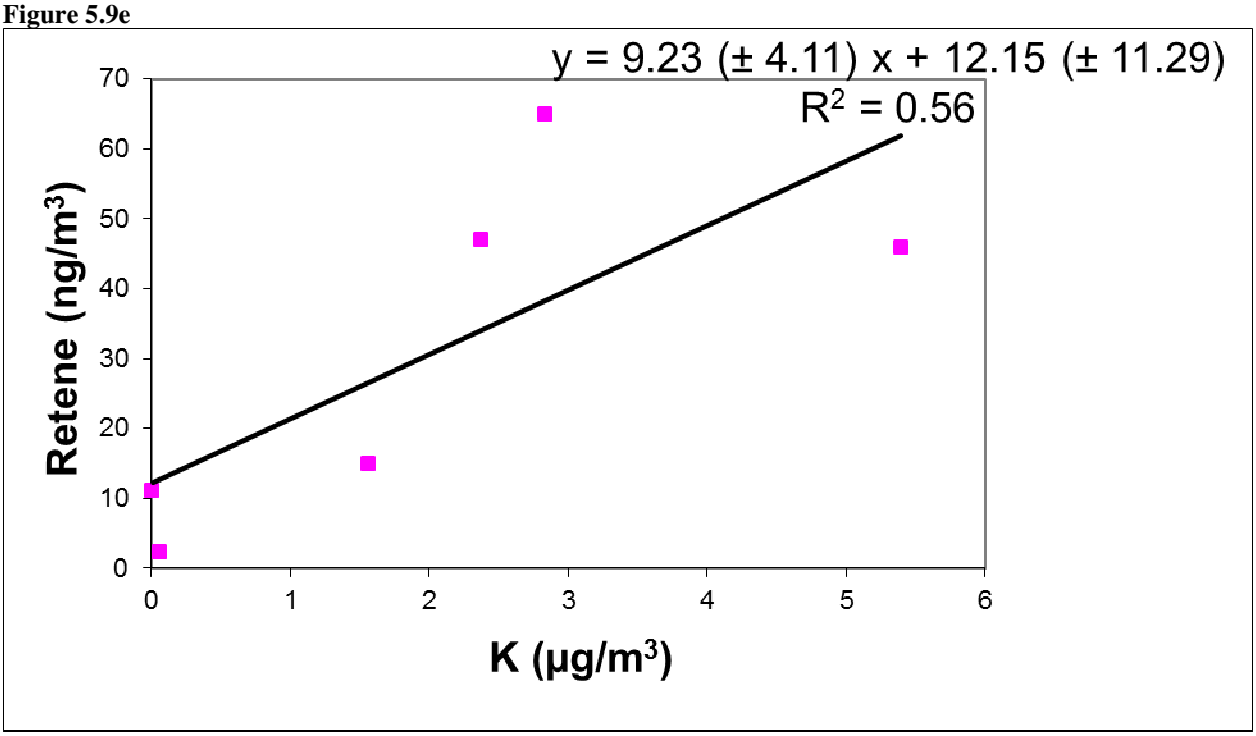


Figure 5.9f

Figure 5.9: (a) Water-Soluble Organic Carbon (WSOC) vs. Total Potassium , (b) WSOC vs. **Levoglucosan (c) WSOC vs. Retene, (d) Total Potassium** vs. **Levoglucosan, (e) Retene** vs. **Levoglucosan and (f) Retene** vs. **Total Potassium**

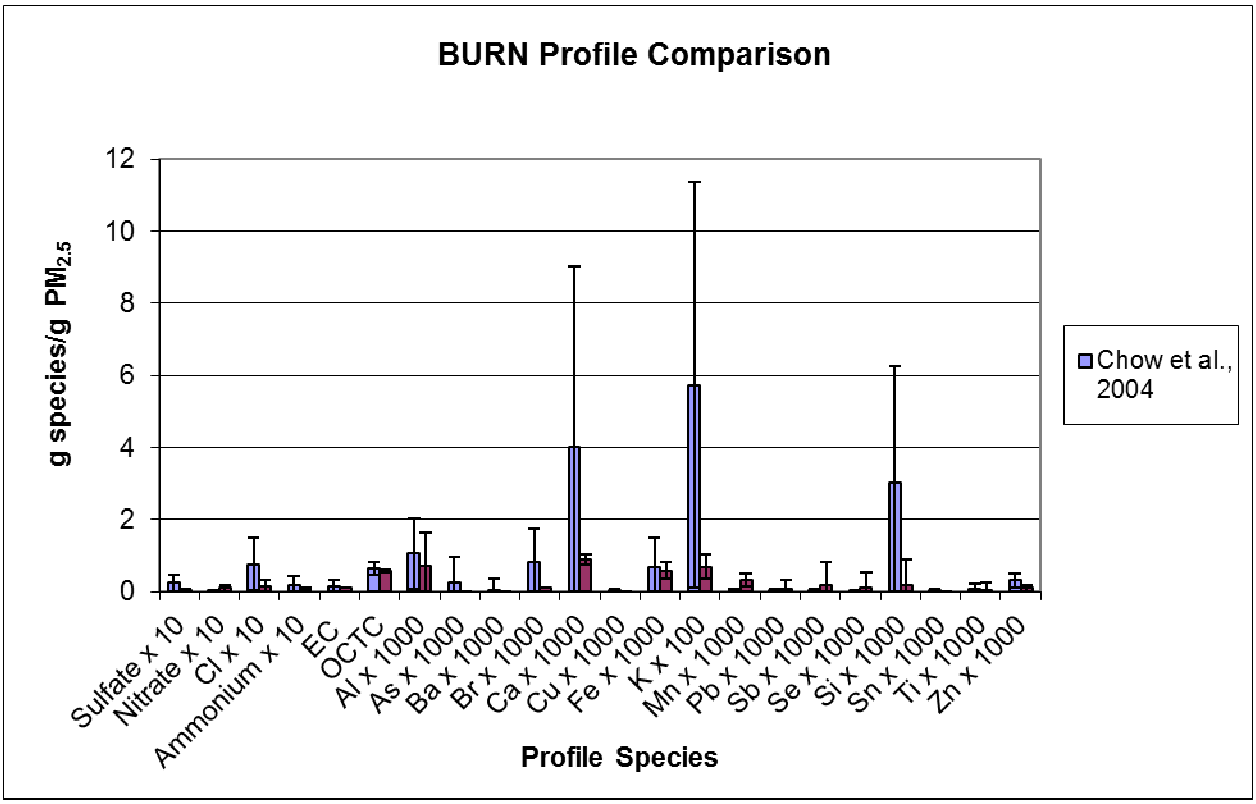


Figure 5.10: Biomass burning source profiles derived from this study compared with composite profile from Chow et al. (2004b).

APPENDIX D: SUPPLEMENTAL INFORMATION FOR CHAPTER 5

D.1. Fuel characteristics

Fuel characteristics from the two burning sites were compiled by Jones Ecological Center staff and are summarized in Table D.1. The dominant overstory species are pine (Pinus spp.) and longleaf pine (Pinus palustri) at Ichauway - North Boundary and Ichauway - Dub-East, respectively. The pre- and post-burn characteristics indicate that grasses and forbs are nearly entirely consumed whereas shrubs and pine are partially consumed. At both sites, pine cones were only partially (Dub East) or minimally (N. Boundary) consumed.

Table D.1: Fuel characteristics from the two burning areas

Table D.1.a: Ichauway – N. Boundary Fuel Characteristics

Loading	Pre-burn Loading				Post-burn Loading				Consumption	
	Tons/Acre	±	Estimated error		tons/acre	±	Estimated error		tons/acre	Percent
Wire Grass[2]	0.32	±	0.06		0.00	±	0.00		0.32	100%
Misc. Grass[2]	0.11	±	0.02		0.00	±	0.00		0.11	100%
Forbs[2]	0.07	±	0.01		0.00	±	0.00		0.07	100%
Litter[2]	1.08	±	0.11		0.89	±	0.08		0.19	18%
Pine Cone[2]	0.30	±	0.12		0.31	±	0.13		-0.01	-3%
Woody shrubs[1]	0.10	±	0.02		0.05	±	0.02		0.05	50%
Down Woody[2]										
1 hr	0.00	±	0.00		0.00	±	0.00		0.00	0%
10 hr	0.33	±	0.06		0.39	±	0.12		-0.06	-18%
100 hr	0.20	±	0.08		0.12	±	0.09		0.08	40%
TOTAL	**2.51**	**±**	**0.24**		**1.76**	**±**	**0.35**		**0.75**	**30%**
Day-of-burn Fuel Moisture	Percent	±	Estimated error		Sample Size (n)					
Litter-	22.5%	±	1.08			7				

Early										
Litter-Middle	19.8%	±	0.39			3				
Litter-Late	13.4%	±	0.15			3				
Grass	54.7%	±	3.04			10				
1 hr	20.5%	±	2.62			2				
10 hr	71.9%	±	3.42			5				

[1]Loading and consumption based on 20 pre-burn and 20 post-burn clip plots (4 m^2 each).

[2] Loading and consumption based on 20 pre-burn and 20 post-burn clip plots (1 m^2 each).

Table D.1.b: Ichauway – Dub-East Fuel Characteristics

Loading[1]	**Pre-burn Loading**				**Post-burn Loading**				**Consumption**	
	tons/acre	**±**	**Estimated error**		**tons/acre**	**±**	**Estimated error**		**tons/acre**	**Percent**
Wire Grass	0.15	±	0.06		0.00	±	0.00		0.15	100.0%
Misc. Grass	0.22	±	0.08		0.00	±	0.00		0.22	100.0%
Forbs	0.10	±	0.07		0.01	±	0.00		0.09	90.0%
Litter	1.96	±	0.22		0.94	±	0.13		1.02	52.0%
Pine Cone	0.47	±	0.12		0.34	±	0.14		0.13	27.7%
Woody shrubs	1.92	±	0.64		**0.95**	±	**0.25**		**0.97**	**50.5%**
Down Woody										
1 hr	0.03	±	0.02		0.03	±	0.01		0.00	0.0%
10 hr	0.18	±	0.04		0.30	±	0.06		-0.12	-66.7%
100 hr	0.13	±	0.10		0.04	±	0.04		0.09	69.2%
Down Woody Subtotal	**0.34**	±			**0.37**	±			**-0.03**	**-8.8%**
TOTAL	**5.16**	**±**	**0.72**		**2.61**	**±**	**0.30**		**2.55**	**49.4%**
Day-of-burn Fuel Moisture	**Percent**	**±**	**Estimated error**		**Sample Size (n)**					
Litter	18.6%	±	0.46			10				
Perched Litter	20.1%	±	2.42			2				
Oak Foliage	12.2%	±	0.68			5				
Grass	39.4%	±	2.68			10				
1 hr	21.3%	±	0.11			2				
10 hr	53.3%	±	4.42			5				

[1] Loading and consumption based on 20 pre-burn and 20 post-burn clip plots (1 m^2 each).

D.2. VOC Speciation

Table D.2: Emission ratios relative to CO_2 of gaseous emissions (± standard error, coefficient of determinations R^2 and number of samples N) from least squares linear regressions between mixing ratios of individual VOCs and CO_2 measured in 5 flaming and 4 smoldering emission samples, averaged over two prescribed fires.

	Flaming				Smoldering			
	$\Delta X/\Delta CO_2$	Std. Error	R^2	N	$\Delta X/\Delta CO_2$	Std. Error	R^2	N
CH_4 (ppmv/ppmv)	0.0007	0.0004	0.49	5	0.0074	0.0040	0.63	4
CO (ppmv/ppmv)	0.0241	0.0054	0.87	5	0.1705	0.0929	0.63	4
OCS (ppbv/ppmv)	-0.0010	0.0007	0.44	5	0.0046	0.0006	0.97	4
CS_2 (ppbv/ppmv)	0.0006	0.0001	0.91	5	0.0002	0.0001	0.84	4
methyl chloride (pptv/ppmv)	4.6247	1.2726	0.81	5	11.0146	5.0080	0.71	4
dichloromethane (pptv/ppmv)	0.1283	0.0515	0.67	5	0.0103	0.0035	0.81	4
chloroform (pptv/ppmv)	0.0000	0.0138	0.00	5	0.0012	0.0034	0.06	4
tetrachloroethylene (pptv/ppmv)	0.0525	0.0303	0.50	5	0.0001	0.0041	0.00	4
methyl nitrate (pptv/ppmv)	0.0495	0.0179	0.72	5	0.1016	0.0759	0.47	4
ethyl nitrate (pptv/ppmv)	0.0106	0.0056	0.55	5	0.0220	0.0066	0.85	4
i-propyl nitrate (pptv/ppmv)	0.0335	0.0200	0.48	5	0.0405	0.0183	0.71	4
n-propyl nitrate (pptv/ppmv)	0.0040	0.0016	0.68	5	0.0022	0.0005	0.92	4
2 butyl nitrate (pptv/ppmv)	0.0523	0.0328	0.46	5	0.0213	0.0170	0.44	4
ethane (ppbv/ppmv)	0.0738	0.0144	0.90	5	0.4661	0.2021	0.73	4
ethene (ppbv/ppmv)	0.4910	0.1149	0.86	5	1.2166	0.3275	0.87	4
ethyne (ppbv/ppmv)	0.1797	0.0524	0.80	5	0.5708	0.0810	0.96	4
propane (ppbv/ppmv)	0.0179	0.0042	0.86	5	0.0847	0.0321	0.78	4
propene (ppbv/ppmv)	0.1020	0.0199	0.90	5	0.3032	0.0104	1.00	4
i-butane (ppbv/ppmv)	0.0027	0.0003	0.95	5	0.0046	0.0011	0.89	4
n-butane (ppbv/ppmv)	0.0058	0.0022	0.70	5	0.0138	0.0050	0.79	4
1-butene (ppbv/ppmv)	0.0150	0.0028	0.91	5	0.0409	0.0016	1.00	4
i-butene (ppbv/ppmv)	0.0095	0.0016	0.92	5	0.0310	0.0058	0.93	4
trans-2-butene (ppbv/ppmv)	0.0029	0.0003	0.98	4	0.0134	0.0039	0.86	4
cic-2-butene (ppbv/ppmv)	0.0025	0.0005	0.91	4	0.0089	0.0024	0.87	4
i-pentane (ppbv/ppmv)	0.0032	0.0012	0.69	5	0.0061	0.0028	0.71	4
n-pentane (ppbv/ppmv)	0.0014	0.0005	0.73	5	0.0042	0.0018	0.72	4
1,3-butadiene (ppbv/ppmv)	0.0227	0.0049	0.88	5	0.0512	0.0049	0.98	4
1-Pentene (ppbv/ppmv)	0.0103	0.0015	0.96	4	0.0939	0.0450	0.68	4
trans-2-Pentene (ppbv/ppmv)	0.0010	0.0004	0.73	5	0.0037	0.0000	1.00	4
cis-2-Pentene (ppbv/ppmv)	0.0007	0.0001	0.96	4	0.0026	0.0001	1.00	4

2-Methyl-1-butene (ppbv/ppmv)	0.0014	0.0001	0.98	4	0.0056	0.0014	0.89	4
2-Methyl-2-butene (ppbv/ppmv)	0.0020	0.0004	0.90	5	0.0148	0.0071	0.68	4
2-Methyl-1-Pentene (ppbv/ppmv)	0.0061	0.0010	0.93	5	0.0104	0.0043	0.75	4
isoprene (ppbv/ppmv)	0.0202	0.0025	0.95	5	0.0372	0.0247	0.53	4
2,3-Dimethylbutane (ppbv/ppmv)	0.0003	0.0000	0.93	5	0.0002	0.0001	0.74	4
2-methylpentane (ppbv/ppmv)	0.0007	0.0002	0.85	5	0.0005	0.0002	0.68	4
3-methylpentane (ppbv/ppmv)	0.0003	0.0002	0.41	5	0.0001	0.0000	0.72	4
n-hexane (ppbv/ppmv)	0.0014	0.0002	0.96	5	0.0021	0.0017	0.44	4
n-heptane (ppbv/ppmv)	0.0006	0.0001	0.89	5	0.0018	0.0007	0.77	4
n-octane (ppbv/ppmv)	0.0000	0.0000	0.00	1	0.0012	0.0005	0.71	4
2,2,4-Trimethylpentane (ppbv/ppmv)	0.0007	0.0001	0.98	5	0.0002	0.0001	0.68	4
benzene (ppbv/ppmv)	0.0307	0.0095	0.78	5	0.1009	0.0101	0.98	4
toluene (ppbv/ppmv)	0.0169	0.0033	0.90	5	0.0462	0.0053	0.97	4
ethylbenzene (ppbv/ppmv)	0.0016	Inf	1.00	2	0.0060	0.0012	0.92	4
m-xylene (ppbv/ppmv)	0.0017	0.0001	1.00	3	0.0167	0.0073	0.72	4
p-xylene (ppbv/ppmv)	0.0014	0.0001	1.00	3	0.0061	0.0024	0.77	4
o-xylene (ppbv/ppmv)	0.0013	0.0000	1.00	3	0.0047	0.0016	0.81	4
1,3,5-Trimethylbenzene (ppbv/ppmv)	0.0000	0.0000	0.00	0	0.0000	0.0000	0.51	4
1,2,4-Trimethylbenzene (ppbv/ppmv)	0.0000	0.0000	0.00	1	0.0003	0.0002	0.49	4
1,2,3-Trimethylbenzene (ppbv/ppmv)	0.0000	0.0000	0.00	0	0.0001	0.0001	0.42	4
a-pinene (ppbv/ppmv)	0.0000	0.0000	0.00	1	0.0015	0.0013	0.41	4
b-pinene (ppbv/ppmv)	0.0000	0.0000	0.00	0	0.0002	0.0002	0.51	4
Acetaldehyde (ppbv/ppmv)	0.1365	0.0231	0.92	5	0.6815	0.3102	0.71	4
Methanol (ppbv/ppmv)	0.3010	0.0402	0.95	5	3.2297	1.8728	0.60	4
Ethanol (ppbv/ppmv)	0.0302	0.0048	0.93	5	0.3208	0.0605	0.93	4
Acetone (ppbv/ppmv)	0.0719	0.0110	0.93	5	0.3134	0.0708	0.91	4
MAC (ppbv/ppmv)	0.0118	0.0008	0.99	5	0.0485	0.0131	0.87	4
MVK (ppbv/ppmv)	0.0116	0.0007	0.99	5	0.0626	0.0244	0.77	4

Table D.3: Comparison of Gaseous and VOC Emissions from prescribed burning between this study and Lee et al. (2005).

	This Study	Lee et al. (2005)		This Study	Lee et al. (2005)
CO (ppmv/ppmv)	0.0241	0.0709	n-heptane (ppbv/ppmv)	0.0006	0.0018
CH_4 (ppmv/ppmv)	0.0007	0.003	n-octane (ppbv/ppmv)		0.0012
chloroform (pptv/ppmv)		0.0016	ethene (ppbv/ppmv)	0.4910	1.2414
Dichloromethane (pptv/ppmv)	0.1283	-0.1606	ethyne (ppbv/ppmv)	0.1797	0.3888
tetrachloroethylene	0.0525	0.0074	propene (ppbv/ppmv)	0.1020	0.2447
methyl chloride (pptv/ppmv)	4.6247	8.6976	1-butene (ppbv/ppmv)	0.0150	0.0374
methyl nitrate (pptv/ppmv)	0.0495	0.8219	i-butene (ppbv/ppmv)	0.0095	0.024
ethyl nitrate (pptv/ppmv)	0.0106	0.0579	trans-2-butene	0.0029	0.0083
i-propyl nitrate (pptv/ppmv)	0.0335	0.1025	cic-2-butene (ppbv/ppmv)	0.0025	0.0063
n-propyl nitrate (pptv/ppmv)	0.0040	0.0075	1,3-butadiene	0.0227	0.0232
2 butyl nitrate (pptv/ppmv)	0.0523	0.0531	benzene (ppbv/ppmv)	0.0307	0.0952
ethane (ppbv/ppmv)	0.0738	0.2621	toluene (ppbv/ppmv)	0.0169	0.0431
propane (ppbv/ppmv)	0.0179	0.0525	ethylbenzene		0.0053
i-butane (ppbv/ppmv)	0.0027	0.0029	m-xylene (ppbv/ppmv)	0.0017	0.009
n-butane (ppbv/ppmv)	0.0058	0.0091	p-xylene (ppbv/ppmv)	0.0014	0.0042
i-pentane (ppbv/ppmv)	0.0032	0.0007	o-xylene (ppbv/ppmv)	0.0013	0.0035
n-pentane (ppbv/ppmv)	0.0014	0.0034	isoprene (ppbv/ppmv)		0.001
2-methylpentane (ppbv/ppmv)	0.0007	0.0007	a-pinene (ppbv/ppmv)		0.0012
3-methylpentane (ppbv/ppmv)	0.0003	0.0002	b-pinene (ppbv/ppmv)		0.0017
n-hexane (ppbv/ppmv)	0.0014	0.0023			
	Smoldering ($\Delta X/\Delta CO_2$)				
CO (ppmv/ppmv)	0.1705	0.2337	n-heptane (ppbv/ppmv)	0.0018	0.0118
CH_4 (ppmv/ppmv)	0.0074	0.0107	n-octane (ppbv/ppmv)	0.0012	0.0091
chloroform (pptv/ppmv)	0.0012	0	ethene (ppbv/ppmv)	1.2166	0.8568
dichloromethane (pptv/ppmv)	0.0103	-0.0669	ethyne (ppbv/ppmv)	0.5708	0.0969
tetrachloroethylene	0.0001	0.0039	propene (ppbv/ppmv)	0.3032	0.3982
methyl chloride (pptv/ppmv)	11.0146	32.67	1-butene (ppbv/ppmv)	0.0409	0.0621
methyl nitrate (pptv/ppmv)	0.1016	0.0113	i-butene (ppbv/ppmv)	0.031	0.089
ethyl nitrate (pptv/ppmv)	0.022	0.0044	trans-2-butene	0.0134	0.0299
i-propyl nitrate (pptv/ppmv)	0.0405	0.0352	cic-2-butene (ppbv/ppmv)	0.0089	0.022
n-propyl nitrate (pptv/ppmv)	0.0022	0.0004	1,3-butadiene	0.0512	0.028
2 butyl nitrate (pptv/ppmv)	0.0213	0.0095	benzene (ppbv/ppmv)	0.1009	0.1885
ethane (ppbv/ppmv)	0.4661	0.9095	toluene (ppbv/ppmv)	0.0462	0.1044
propane (ppbv/ppmv)	0.0847	0.2445	ethylbenzene	0.006	0.0133
i-butane (ppbv/ppmv)	0.0046	0.0177	m-xylene (ppbv/ppmv)	0.0167	0.0362
n-butane (ppbv/ppmv)	0.0138	0.0651	p-xylene (ppbv/ppmv)	0.0061	0.008
i-pentane (ppbv/ppmv)	0.0061	0.0022	o-xylene (ppbv/ppmv)	0.0047	0.0127
n-pentane (ppbv/ppmv)	0.0042	0.0255	isoprene (ppbv/ppmv)	0.0372	0.025
2-methylpentane (ppbv/ppmv)	0.0005	0.0051	a-pinene (ppbv/ppmv)	0.0015	0.0202
3-methylpentane (ppbv/ppmv)	0.0001	0.0011	b-pinene (ppbv/ppmv)	0.0002	0.0123
n-hexane (ppbv/ppmv)	0.0021	0.0162			

Table D.4: Emission factors for gaseous components (g per kg fuel burned).

Gaseous Species	Flaming Avg.	Flaming St. Dev.	Smoldering Avg.	Smoldering St. Dev.
CH_4 (g/kg)	2.32	1.17	3.48	1.48
CO (g/kg)	48.08	35.02	133.30	54.16
CO_2 (g/kg)	1425.14	63.78	1324.67	88.28
OCS (g/kg)	0.0061	0.0030	0.0088	0.0003
CS2 (g/kg)	0.0014	0.0002	0.0016	0.0011
CH3Cl (g/kg)	0.0167	0.0120	0.0181	0.0048
CH2Cl2 (g/kg)	0.0008	0.0007	0.0002	0.0003
CHCl3 (C/D)(g/kg)	0.000007	N/A*	-	N/A
C2Cl4 (C/D)(g/kg)	0.0000	N/A	-	N/A
MeONO2 (C/D)(g/kg)	0.0003	0.0001	0.0002	0.0001
EtONO2 (C/D)(g/kg)	0.0000	0.0000	0.0000	0.0000
i-PrONO2 (C/D)(g/kg)	0.0002	N/A	0.0001	0.0000
n-PrONO2 (C/D)(g/kg)	0.0000	N/A	0.0000	0.0000
2-BuONO2 (C/D)(g/kg)	0.0001	0.0000	0.0000	0.0000
Ethane (g/kg)	0.1459	0.0817	0.3494	0.1144
Ethene (g/kg)	1.0372	0.3747	1.1285	0.2070
Ethyne (g/kg)	0.4285	0.1609	0.4576	0.0742
Propane (g/kg)	0.0359	0.0365	0.0856	0.0551
Propene (g/kg)	0.2742	0.1108	0.3975	0.0094
i-Butane (g/kg)	0.0061	N/A	0.0074	0.0003
n-Butane (g/kg)	0.0122	0.0012	0.0196	0.0027
1-Butene (g/kg)	0.0475	0.0244	0.0741	0.0055
i-Butene (g/kg)	0.0353	0.0225	0.0544	0.0101
trans-2-Butene (g/kg)	0.0110	0.0055	0.0209	0.0061
cis-2-Butene (g/kg)	0.0091	0.0039	0.0131	0.0084
i-Pentane (g/kg)	0.0037	0.0045	0.0081	N/A
n-Pentane (g/kg)	0.0045	0.0030	0.0104	0.0010
1,3-Butadiene (g/kg)	0.0689	0.0269	0.0982	0.0172
1-Pentene (g/kg)	0.0473	0.0366	0.1448	0.0597
trans-2-Pentene (g/kg)	0.0064	0.0027	0.0077	0.0020
cis-2-Pentene (g/kg)	0.0039	0.0017	0.0054	0.0003
2-Methyl-1-butene (g/kg)	0.0079	0.0033	0.0112	0.0047
2-Methyl-2-butene (g/kg)	0.0078	0.0034	0.0211	0.0085
2-Methyl-1-Pentene (g/kg)	0.0262	0.0113	0.0353	0.0077
Isoprene (g/kg)	0.0576	0.0320	0.0973	0.0215
2,3-Dimethylbutane (g/kg)	0.0022	0.0013	0.0002	0.0003
2-Methylpentane (g/kg)	0.0054	0.0049	0.0004	0.0005
3-Methylpentane (g/kg)	0.0012	0.0001	0.0001	N/A
n-Hexane (g/kg)	0.0035	0.0026	0.0106	0.0020
n-Heptane (g/kg)	0.0028	0.0018	0.0059	0.0007
n-Octane (g/kg)	-	-	-	-
2,2,4-Trimethylpentane (g/kg)	0.0007	0.0009	0.0003	0.0002
Benzene (g/kg)	0.2197	0.0859	0.2625	0.0278
Toluene (g/kg)	0.1018	0.0613	0.1474	0.0178

Ethylbenzene (g/kg)	-	N/A	-	N/A
m-Xylene (g/kg)	-	N/A	-	N/A
p-Xylene (g/kg)	0.0021	N/A	-	N/A
o-Xylene (g/kg)	0.0007	N/A	-	N/A
1,3,5-Trimethylbenzene (g/kg)	-	N/A	-	N/A
1,2,4-Trimethylbenzene (g/kg)	-	N/A	-	N/A
1,2,3-Trimethylbenzene (g/kg)	-	N/A	-	N/A
alpha Pinene (g/kg)	-	N/A	-	N/A
beta Pinene (g/kg)	-	N/A	-	N/A
Acetaldehyde (g/kg)	0.2875	0.1767	0.6771	0.2669
Methanol (g/kg)	0.6301	0.5132	2.0997	1.2559
Ethanol (g/kg)	0.1239	0.1676	0.4195	0.0514
Acetone (g/kg)	0.3111	0.2417	0.4180	0.1693
MAC (g/kg)	0.0351	0.0335	0.0668	0.0301
MVK (g/kg)	0.0284	0.0350	0.0797	0.0569

* N/A - No standard deviation was calculated because there was only one sample with level above detection or no samples above detection (denoted by -).

D.3. $PM_{2.5}$ Emission Factors

Table D.5: Emission Factors for $PM_{2.5}$ (g/kg fuel burned).

	Avg.	St Dev
PM2.5 (g/kg)	13.87	17.27
OC (g/kg)	7.15	9.32
EC (g/kg)	1.40	1.80
acetate (CH3COO-) (g/kg)	0.05	0.07
Nitrate (g/kg)	0.19	0.27
Sulfate (g/kg)	0.07	0.09
Oxalic Acid (g/kg)	0.04	0.05
Ammonium (g/kg)	0.12	0.18
Na (mg/kg)	31.33	26.37
Mg (mg/kg)	12.94	-[a]
Al (mg/kg)	2.12	2.37
Si (mg/kg)	2.76	5.52
P (mg/kg)	5.36	6.20
S (mg/kg)	35.80	43.74
Cl (mg/kg)	72.97	120.10
K (mg/kg)	116.45	180.49
Ca (mg/kg)	12.21	16.85
Sc (mg/kg)	2.66	-[a]
Ti (mg/kg)	0.01	-[a]
V (mg/kg)	BL[b]	n/a
Cr (mg/kg)	0.04	0.02
Mn (mg/kg)	16.85	4.91
Fe (mg/kg)	3.18	5.80
Co (mg/kg)	BL[b]	n/a
Ni (mg/kg)	BL[b]	n/a
Cu (mg/kg)	6.06	9.81
Zn (mg/kg)	1.82	2.00
Ga (mg/kg)	BL[b]	n/a
As (mg/kg)	BL[b]	n/a
Se (mg/kg)	1.29	1.63
Br (mg/kg)	2.50	4.01
Rb (mg/kg)	1.88	-[a]
Sr (mg/kg)	BL[b]	n/a
Y (mg/kg)	3.02	5.97
Zr (mg/kg)	3.25	3.21
Nb (mg/kg)	1.71	-[a]
Mo (mg/kg)	0.00	-[a]
Pd (mg/kg)	BL[b]	n/a
Ag (mg/kg)	4.20	-[a]
Cd (mg/kg)	BL[b]	n/a
In (mg/kg)	BL[b]	n/a

Sn (mg/kg)	BL[b]	n/a
Sb (mg/kg)	11.30	8.21
Cs (mg/kg)	BL[b]	n/a
Ba (mg/kg)	BL[b]	n/a
La (mg/kg)	1.84	-[a]
Ce (mg/kg)	0.68	0.51
Sm (mg/kg)	BL[b]	n/a
Eu (mg/kg)	BL[b]	n/a
Tb (mg/kg)	0.34	-[a]
Hf (mg/kg)	0.14	-[a]
Ta (mg/kg)	15.60	-[a]
W (mg/kg)	BL[b]	n/a
Ir (mg/kg)	0.00	-[a]
Au (mg/kg)	9.38	-[a]
Hg (mg/kg)	BL[b]	n/a
Tl (mg/kg)	0.52	0.30
Pb (mg/kg)	6.95	8.02
U (mg/kg)	BL[b]	n/a

-[a]only one sample had measured concentrations greater than detection limit.
[b]BL: Below blank levels

Gas phase emission ratios are similar to Lee et al. (2005), though they tend to be somewhat lower for both the flaming and smoldering stages (Figures D.2, D.3; Table D.3). VOC emission ratios are dominated by methyl chloride, ethane, ethyne, propene, benzene, acetaldehyde, methanol, ethanol and acetone during both flaming and smoldering stages. In the smoldering stage, two VOC species, tetrachloroethylene and 3-methylpentane, are present in Lee et al. (2005) at levels an order of magnitude higher than in this study. Tetrachloroethylene, used primarily for dry cleaning operations but also as an industrial solvent, and 3-methylpentane, indicative of liquid fossil fuel emission, reflect activities in the vicinity of the fires sampled by Lee et al. (2005), which were closer to populated areas. These activities are not expected to be significant in the vicinity of the prescribed fires in this study.

D.3.1. FEPS Simulation

Time varying CO, CH_4, and $PM_{2.5}$ emissions were calculated using FEPS v1.1.0 using a typical fuel loading profile for a 2 year-old fuel dominated by shrubs and litter (Table D.6). Meteorological data input into FEPS was obtained from MesoWest Data (http://mesowest.utah.edu/). Fuel moisture profile was set as moderate. The flaming and short term smoldering involvement percentages were 95% and long term smoldering involvement percentage was set as 10%, which is a typical ratio for prescribed burning. The area of the burn increased linearly for the first 2 hours of the burn, and the last 15 acres were burned in the third hour which was assumed to be in the smoldering phase. FEPS output emission rates (g/s) as well as fuel consumption rate (kg/hr). EFs derived from FEPS are the averages of emission rate/fuel consumption rate for each appropriate phases of the burn.

Table D.6: 2-year-old Fuel Loading Profile used in FEPS simulation.

FEPS Parameter	Canopy	Shrub	Grass	Woody	Litter	Broadcast	Piles	Duff
Loading (tons per acre)	0.00	1.05	0.49	2.93	2.60	0.00	0.00	1.06

D.4. Water-Soluble Iron

Real time soluble Fe (II) and potassium concentrations were measured using a PILS sampler and total particles were measured using an optical particle counter (OPC) [*Rastogi et al.*, 2009]. Very good agreement is found between all three and the R^2 between soluble Fe (II) and K is 0.88 with an emission factor of 0.015 ± 0.022 g soluble Fe (II) per g K; details of this work can be found elsewhere [*Oakes et al.*, 2010b].

D.5. Figures

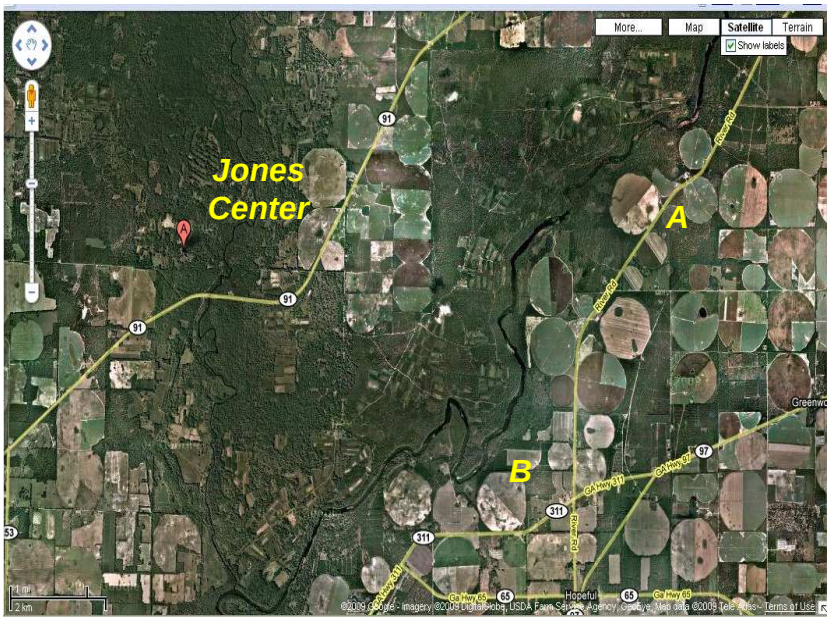


Figure D.1: Sampling locations. North Boundary (A) and Dub-East (B)

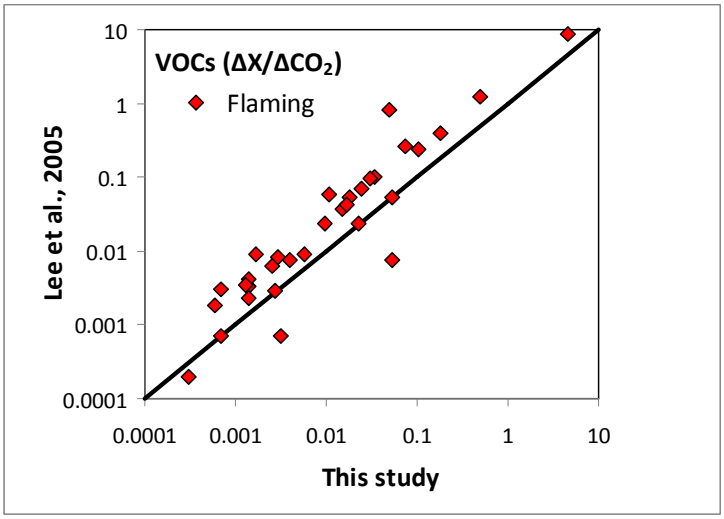


Figure D.2: Comparison of emission ratios relative to CO_2 of gaseous emissions during the flaming stage (this study vs. Lee et al. (2005))

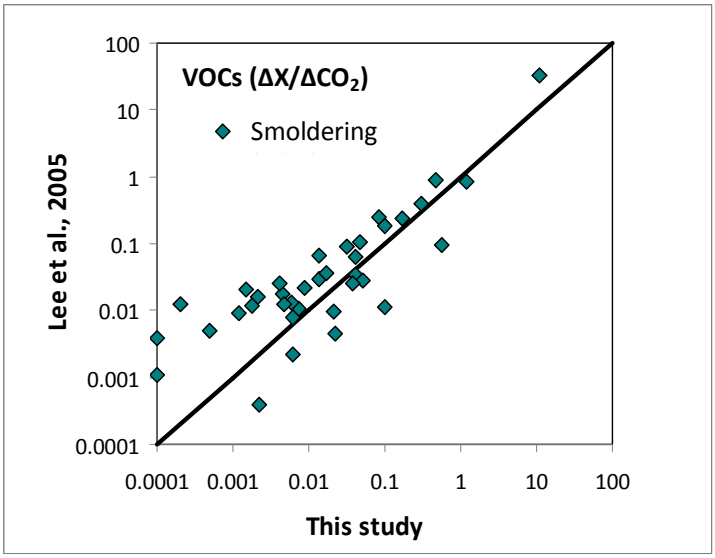


Figure D.3: Comparison of emission ratios relative to CO_2 of gaseous emissions during the smoldering (this study vs. Lee et al. (2005))

D.3. Particle Size Distribution

Particles were classified into six bins from 0.30 µm to >10 µm using an OPC. Both days show similar distributions with predominance in the number concentration as fine particles (Dp < 2.5um (Figure D.4). The volume or mass equivalent distribution, for both days is bi-modal, with one mode for fine particles (X<Dp < 2.5 µm) and one for coarse particles (X<Dp < 10 µm) (Figure D.5).

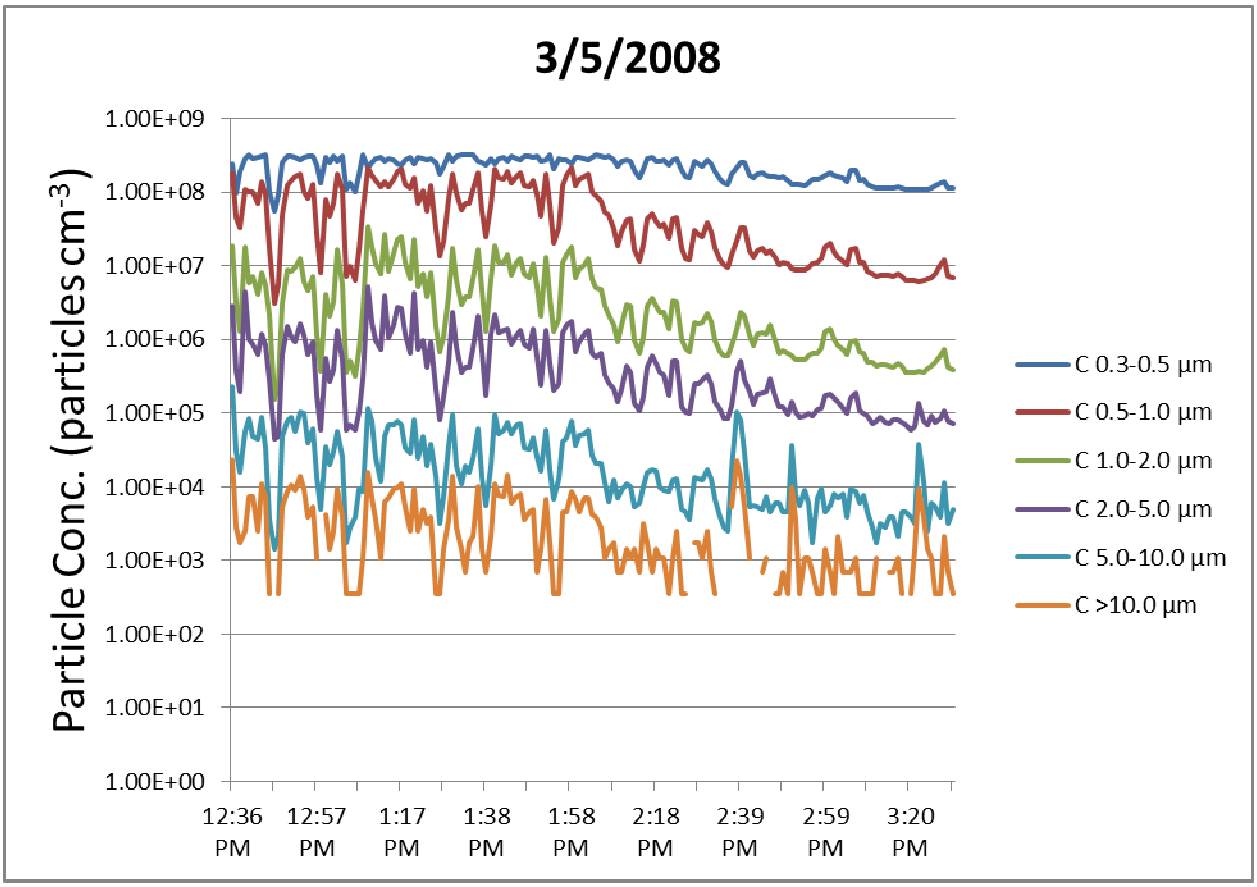


Figure D.4 (a)

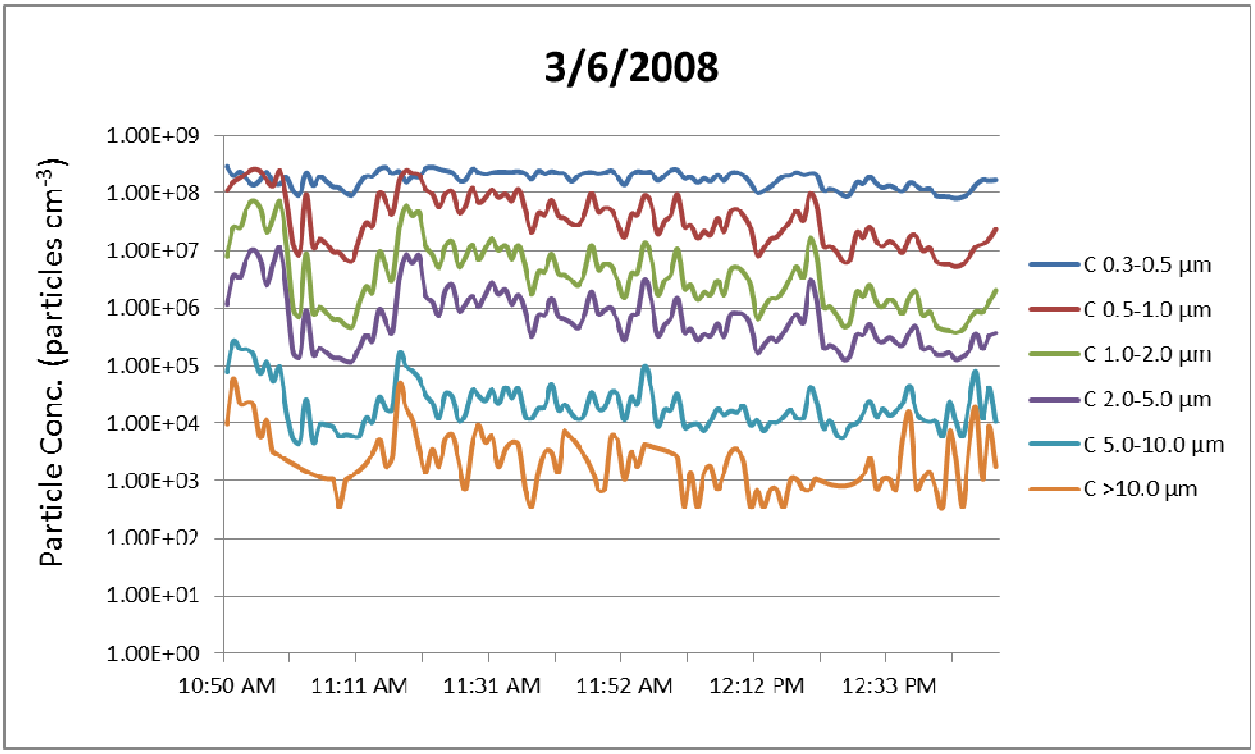


Figure D.4 (b)

Figure D.4: Particle number concentration as measured by OPC for the prescribed fire on (a) 03/05/2008 and (b) **03/06/2008.**

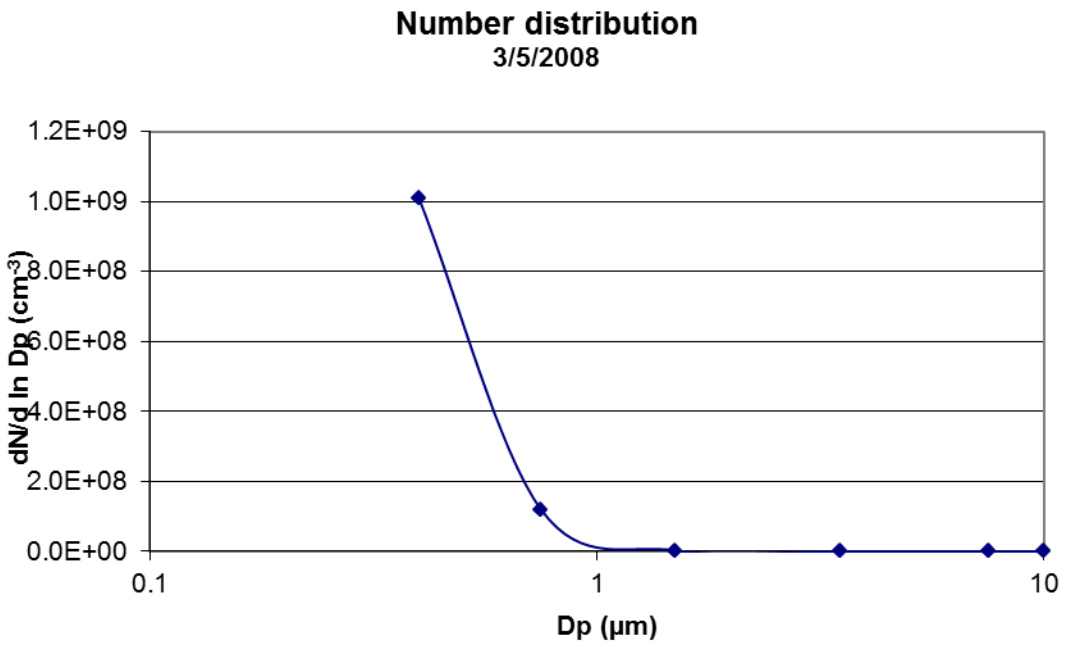


Figure D.5(a)

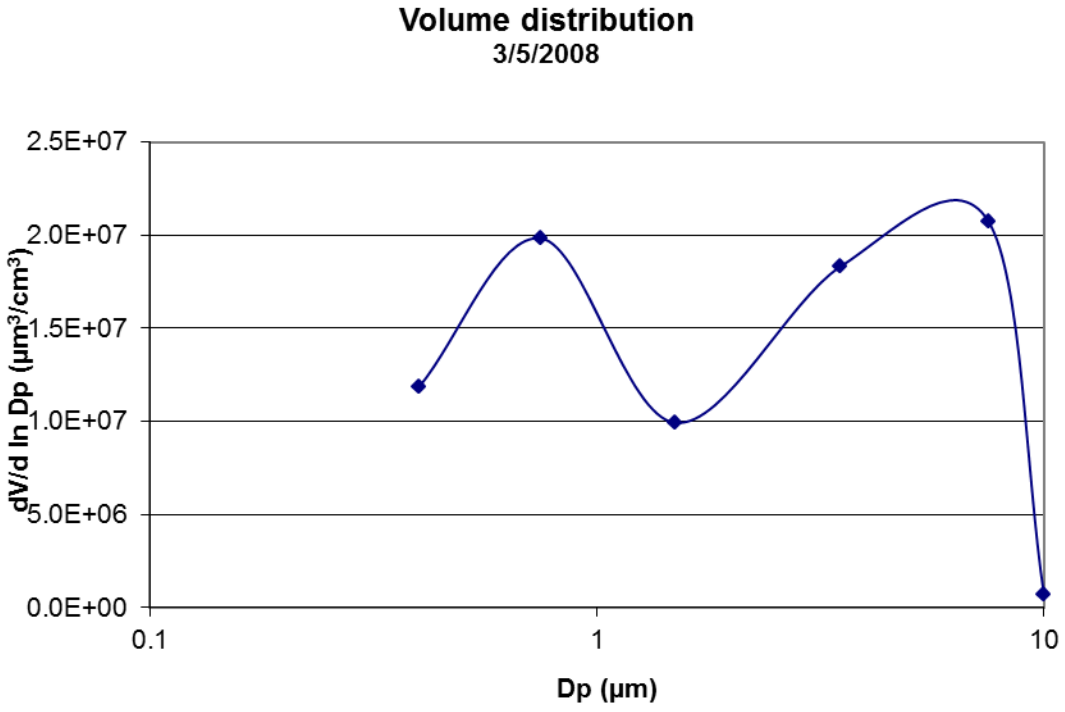


Figure D.5(b)

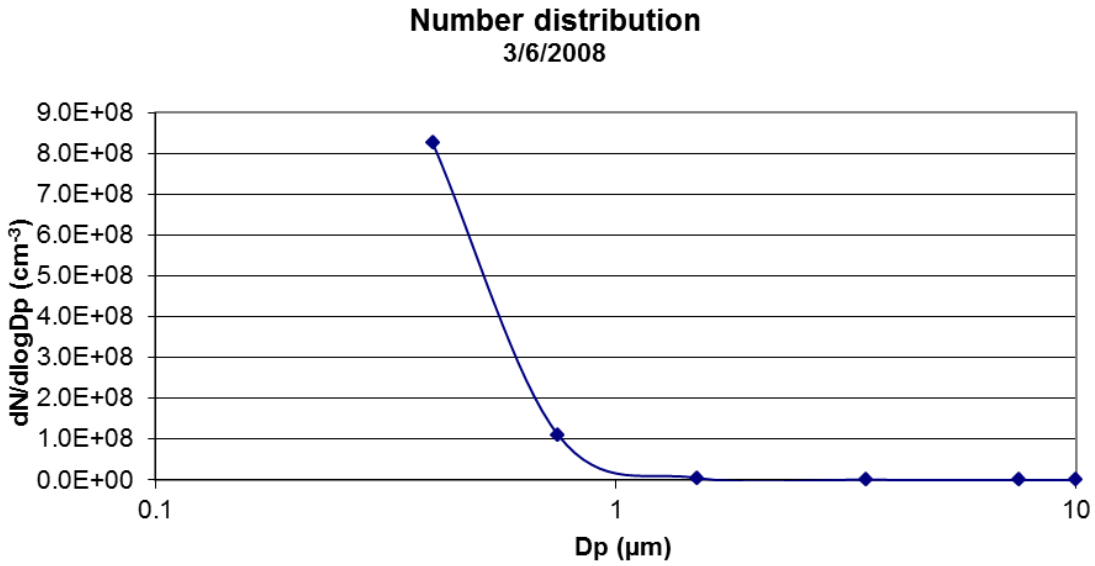


Figure D.5(c) \

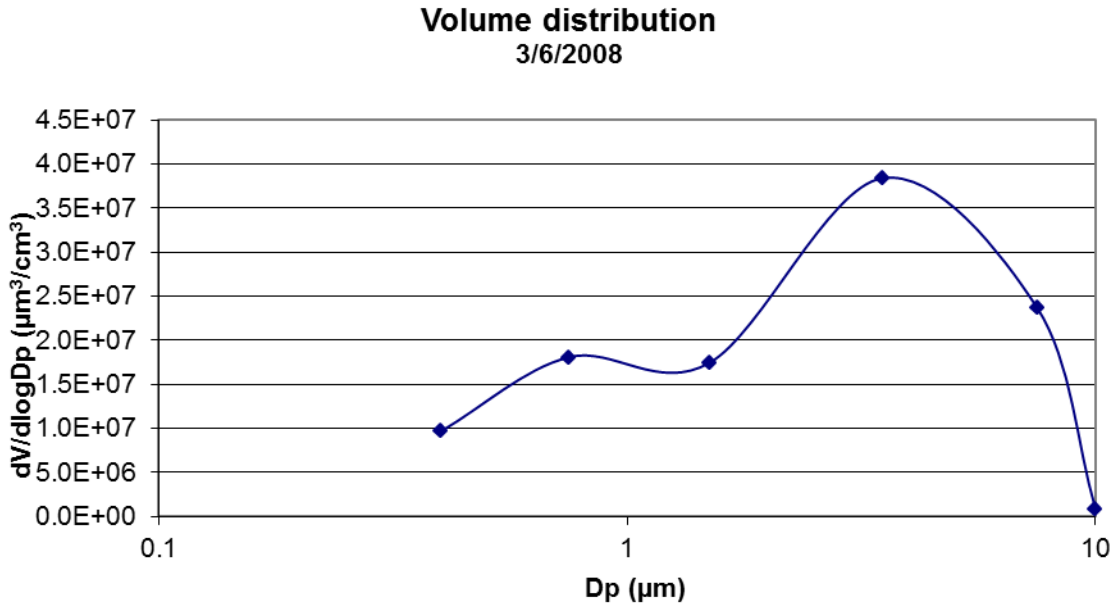


Figure D.5(d)

Figure D.5: ***Average particle number (Figues D.5(a) and (c)) and volume distributions (Figures D.5(b) and (d)) for the prescribed fire on 03/05/2008 and 03/06/2008.***

CHAPTER 6: EVALUATION OF FIRE WEATHER FORECASTS USING PM2.5 SENSITIVITY ANALYSIS

Sivaraman Balachandran[*,a], Karsten Baumann[b], Jorge E. Pachon[c], James A. Mulholland[a] and Armistead G. Russell[a]

[a]Georgia Institute of Technology. School of Civil and Environmental Engineering. Atlanta, GA, 30332.

[b]Atmospheric Research and Analysis, Inc., Cary, NC.

[c]Universidad de La Salle, Programa de Ingenieria Ambiental, Bogota, Colombia.

[*]Corresponding author: 311 Ferst Dr., Atlanta, GA, 30332; phone 206.250.6480, fax 404.894.8266; siv@gatech.edu.

6.1. ABSTRACT

Fire weather forecasts are used by wildlife managers in determining when PB activities are to occur. In this work here, we investigate the sensitivity of ambient $PM_{2.5}$ to various fire and meteorological parameters in a spatial setting that is typical for the wildland urban interface in the southeastern US. We use the method of principle components regression (PCR) to estimate sensitivity of $PM_{2.5}$ to fire data and, observed and forecast meteorological parameters. In PCR, principal components analysis (PCA) is first run on a data set. We ran PCA on 10 data sets that included PB activity data along with meteorological parameters of interest; the meteorological parameters included either observational data only, forecast data only or a combination of observations and forecasts. For each data set, we regressed PCA scores from the first seven principal components against observed $PM_{2.5}$. $PM_{2.5}$ showed significant sensitivity to PB, with a

unit-based sensitivity of 3.2±1 μg m^{-3} $PM_{2.5}$ per 1000 acres burned. $PM_{2.5}$ had a negative sensitivity to dispersive parameters such as wind speed and had positive senstivity to winds coming from the west and the north, the origin of both can be considered continental.

6.2. INTRODUCTION

Fire plays an important role in the management of forest ecosystems of the Southeastern United States (Southeast), where prescribed burning (PB) is employed to manage more than 8 million acres of land every year [*Wade et al.*, 2000]. Over the last few decades, the Southeast has experienced substantial population growth [U.S. Census, 2012], causing significant urban sprawl in an otherwise heavily forested region, making the wildland urban interface (WUI) especially susceptible to air quality impacts from PB. It has been suggested that PB is the third largest source of primary anthropogenic fine $PM_{2.5}$ in the U.S., emitting 12% of the total $PM_{2.5}$ mass [*Davidson et al.*, 2005]. In addition, source apportionment modeling of $PM_{2.5}$ mass concentrations from 23 Speciation Trend Network sites suggests PB may contribute more than 30% of the annual $PM_{2.5}$ mass in the Southeast during winter [*Sangil Lee et al.*, 2007]. Further, individual PB plume events can significantly impact air quality (AQ) in neighboring communities, which can lead to short-term increases of ambient $PM_{2.5}$ and contribute to increases in secondary air pollutants, such as ozone [*Hu et al.*, 2008; *S. Lee et al.*, 2005].

However, meeting Clean Air Act (CAA) rules mandated by the U.S. Environmental Protection Agency (EPA) can be in conflict with the Endangered Species Act, which recommends the use of PB to re-create the natural fire regimes needed to protect the habitat of threatened and endangered species by maintaining the health of its native forest ecosystems. Due to the suspected health impacts, EPA lowered the annual $PM_{2.5}$ standard from 15 μg m^{-3} to 12 μg m^{-3} and retained the 24-hour standard at

35 μg m^{-3} [*U.S.EPA*, 2011], making $PM_{2.5}$ contributions from PB emissions even more important. Before conducting PB on a particular day, land managers across the Southeast consult the Fire Weather Forecast, which is released twice daily by the National Weather Service (NWS). Understanding the NWS fire weather forecast and the association of individual parameters with ambient $PM_{2.5}$ can aid fire managers in making decisions regarding when and if prescribed burning events take place.

In this work, we investigate the association between ambient $PM_{2.5}$ and various fire weather forecast parameters in a spatial setting that is typical for the WUI in the Southeast. Military installations are ideal locations to study such sensitivities because Department of Defense (DoD) lands are intensely managed and neighbored by relatively large civilian communities that are mandated through the CAA to monitor AQ. In addition, the PB activity on military installations is well tracked and recorded, thereby providing adequate data for the analyses in this work. The importance of fire weather forecasts for land management in the WUI is evident when comparing the largest installations in the southeastern US in terms of managed forested land area with the size of the adjacent metropolitan statistical area (MSA) mandated to monitor $PM_{2.5}$ via the CAA (Table 6.1).

Table 6.1: Managed areas employing PB on major military installations in the southeastern US with adjacent MSA population [*USCensus*, 2012] and active $PM_{2.5}$ monitoring site reporting to the AQS repository. (*Marine Corps Base Camp Lejeune. **Site discontinued in January 2008 with Castle Hayne site 371290002 serving as backup)

Mil. Base	Managed Area (acres)	Nearest MSA		Population *2010*	PM2.5 *Site ID*
Eglin AFB	362,000	FWB-Destin	FL	236,058	120730012
Stewart	270,000	Savannah	GA	348,830	130511002
Bragg	162,000	Fayetteville	NC	367,444	370510009
Campbell	140,000	Clarksville	TN	261,868	471251009
Benning	96,000	Columbus	GA	295,741	132150008
MCBCL*	95,000	Jacksonville	NC	179,487	**371330005
Rucker	63,000	Dothan	AL	145,892	010690003
Gordon	56,000	Augusta	GA	566,781	132450091
Jackson	52,000	Columbia	SC	769,819	450790007

Land managers consult fire weather forecasts, released twice daily by the NWS, to determine if conditions are favorable for conducting PBs. The morning (AM) version of the fire weather forecast provides a 24 hour forecast, while the afternoon (PM) version provides a 48 hour forecast. Each posting provides important forecast parameters that are considered in the final preparations and decision process of imminent PB conduct. Among the issued parameters, an area-specific ventilation rate (VR) (i.e. the product of wind speed and mixing height), the probability of precipitation (POP), min/max temperature and relative humidity (RH), inversion burn-off temperature (IBT), boundary layer mixing height (BLH), transport wind speed and direction (TWS, TWD), and Haines Index (HAI) (a lower atmospheric stability index used to forecast the potential for large fire growth and/or erratic fire behavior) are considered more important in this process (Table 6.2).

This work builds on an earlier study in a similar setting investigating sensitivities of ambient $PM_{2.5}$ measured in Columbus, GA to burn activities at Fort Benning (expressed in acres burned) relative to fire forecast data [*Baumann*, 2005]. The method bypasses individual atmospheric processes and looks at statistical links between the source (prescribed fires) and receptor ($PM_{2.5}$ monitoring station), by ranking the relevance and importance of forecast parameters on the receptor's $PM_{2.5}$ concentration relative to that of PB. Such a ranking can inform fire managers of the most important forecast parameters that may influence the PB impact on $PM_{2.5}$ in their district. In other words, the ranking allows a quantitative assessment of each forecast parameter's sensitivity on local $PM_{2.5}$ under a given PB source strength (expressed in acres burned). Note that all available PB records used here are in units of acres that were subject to PB, whereby the amount of fuel actually consumed remains unknown, because post-PB inventories do not exist.

6.3. METHODS

This study utilizes prescribed burning, observed and forecast meteorological, and $PM_{2.5}$ data from December 2002 to March 2007 (Table 6.2). PB activity data were gathered at the Marine Corps Base Camp Lejeune (MCBCL) near Jacksonville, NC. Figure 6.1 shows a map outlining the MCBCL area south from the city of Jacksonville. Of its total land area of 125,000 acres, MCBCL manages 95,000 acres of forested land, employing PB in a 3-year rotation with an annual treatment target of ca. 30,000 acres. MCBCL is surrounded by managed forests outside its borders; i.e. the Hofmann Forest to the north, Croatan National Forest to the north-east and the Holly Shelter Game Land to the south-west. Foresters managing these lands apply similar tools and rely on the same fire weather forecasts as the MCBCL foresters.

The Marine Corps Air Station (MCA in Figure 6.1) operates a suite of meteorological sensors that are typical for most airports in the U.S.; i.e. cloud ceiling and visibility in addition to barometric pressure, temperature, humidity, precipitation, wind speed and direction. Observational data are reported to the NWS and are available through the MesoWest Web site (http://mesowest.utah.edu/index.html) run by the University of Utah's Department of Meteorology. Since the posted data represent different averaging intervals and different reporting frequencies, we processed the data to provide consistent hourly averages, allowing the determination of daily minimum humidity and visibility, and daily max-min temperature difference (see Table 6.2).

Historical fire weather forecast data were extracted from the National Climate Data Center (NCDC) archive in Ashville, NC. The archived data were accessed via NCDC's Service Records Retention System in the Hierarchical Data Storage System Access System [*NOAA*, 2012]. In addition, values of the Keetch-Byram Drought Index (KBDI) [*Keetch and Byram*, 1968], a continuous reference scale for estimating dryness of

soil and duff layers presenting wildfire risks, were gathered. The KBDI is not part of the routine NWS fire weather forecast, but was available at MCBCL.

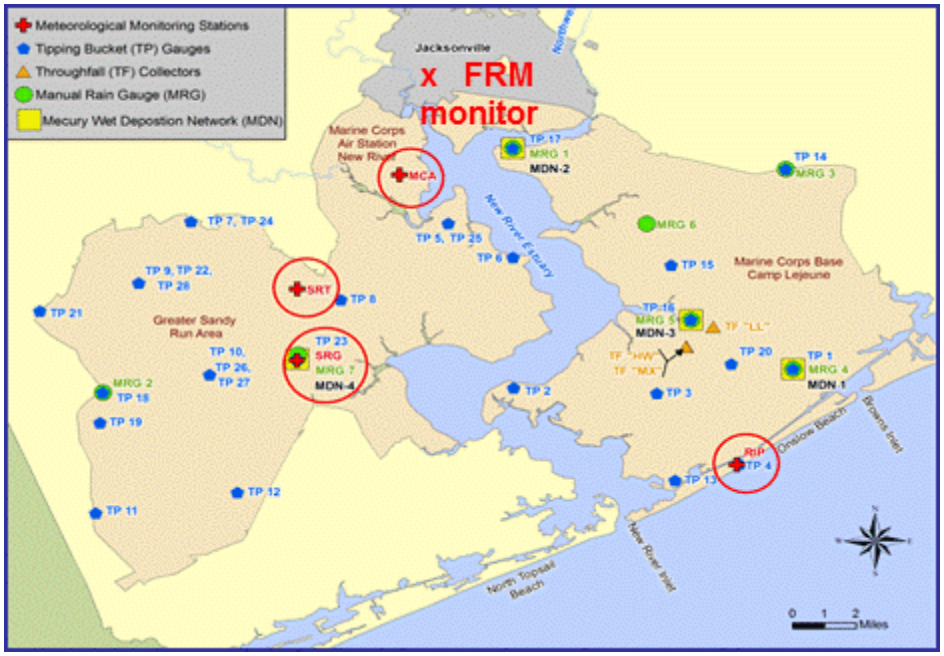


Figure 6.1: Outline of Marine Corps Base Camp Lejune (MCBCL) near the NC coast with the $PM_{2.5}$ (FRM) monitor location at Jacksonville and the nearby meteorological observation site (MCA).

In compliance with the CAA, Jacksonville (population 179,487 per 2010 Census) required regulatory monitoring of ambient $PM_{2.5}$ beginning in 1999. The state's environmental agency (NCDENR-DAQ) measured 24 h integrated $PM_{2.5}$ filter samples to report average 24 h (daily) PM2.5 mass concentration every third day. Most PB activities occur in winter and spring, addressing both dormant and growing season management objectives. We obtained $PM_{2.5}$ data for the several days prior to and when PB activity was documented. However, days with PB activities may occur in between the $PM_{2.5}$ sampling days and might be missed in the analysis. In order to maximize the number of coincidental data points subject to this exercise, the 2 days between each PM measurement have been interpolated in two different ways; one was using linear temporal interpolation, and the other employed an air mass flow dependent spatial correlation with daily averaged $PM_{2.5}$ monitoring data from another State regulatory site (Castle Hayne operating a continuous TEOM) ca. 100 km away to the SW.

Table 6.2: Measured and modeled (from NWS forecast) input variables. *NOTE: Forecast parameters are described with "_am" or "_pm" to denote AM and PM forecasts, respectively. The "_pm" variables are not listed for brevity.

Parameter Type	Variable Name	Units	Description
PM	**jvlPM**	$\mu g/m^3$	24 hour average PM2.5
	jvlPM_si	$\mu g/m^3$	24 hour avg. spatially interpolated PM2.5
	jvlPM_ti	$\mu g/m^3$	24 hour avg. temporally interpolated PM2.5
Fire Data	**PB**	acr	Acres of prescribed burning
	KBDI	0-800	Keech Brynham Drought Index. Continuous reference scale for estimating wildfire risks.
Observations	**mcaT_avg**	deg. C	Daily average temperature
	mcaT_diff	deg. C	Day-Night temperature difference
	mcaRH_avg	%	Relative humidity
	mcaRH_min	%	Daily minimum RH
	mcaPCP	mm daily	Precipitation
	mcaVIS_min	km	Daily minimum visibility
	mcaWS	m/s	Wind speed
	mcaSC	-1 to 1	N-S wind component (1 is from the south)
	mcaEC	_1 to 1	E-W wind component (1 is from the west)
	mcaWD	degN	Direction (0 degrees is from the north)
Forecast*	**POP_am**	pct	Probability/chance of precipitation in %
	dayT_am	F	Daily maximum air temperature in F
	nightT_am	F	Daily minimum air temperature in F
	diffT_am	F	Difference in daily max-min air temperature
	IBT_am	F	Inversion burnoff temperature. Temperature required to dissipate nocturnal inversion in
	dayRH_am	pct	Daytime average humidity
	HAI_am	-	Haines Stability Index. Atmospheric stability index for large fire growth.
	BLH_am	ft	Top of the atmospheric boundary layer that is well mixed and in which smoke disperses best during midday.
	TWS_am	mph	Transport wind speed as average wind speed between surface and BLH
	TSC_am	-1 to 1	N-S wind component (1 is from the south)
	TEC_pm	-1 to 1	E-W wind component (1 is from the west)
	TWD_am	23-360	Transport wind direction as average wind direction between surface and BLH
	VR_am	ftmph	Ventialation Rate = BLH*TWS

The spatial correlation was subdivided into four data sets for four main air mass transport patterns; i.e. upwind (wind from NE), downwind (wind from SW), cross-wind from the ocean (sea breeze from SE), and cross-wind carrying continental air mass (land breeze from NW).

6.3.1. Sensitivity Analysis using Principal Components Regression (PCR)

PCR is a statistical analysis technique that combines principal components analysis (PCA) with multivariate regression [*Fekedulegn et al.*, 2002]. PCA is often used i) to remove multi-collinear effects of the original data (input variables), and ii) to reduce dimensionality of large data sets. Details of PCA are widely available in the literature but briefly, the first step in PCA is to normalize a given data matrix, X_{org}, so that all variables have a mean of 0 and standard deviation of 1 (called X_{std}, or just X for simplicity). Next, singular value decomposition is used to determine the principal components, which are the matrix of eigenvectors, V, of the dispersion matrix, $X^T X$. The relative strength, or scores, Z, of each component, for each day are therefore, a rotation of the data matrix X (Eq. 1).

$$Z = X * V \qquad \text{(Equation 1)}$$

PCA results in the same number of eigenvectors as variables in the data matrix, X. The eigenvectors, V, are orthonormal, and the resulting scores, Z are orthogonal, which has the net effect of removing collinearity within the data matrix, X. Also, the components are ordered by their eigenvalues. The higher the eigenvalue for a particular component, the more variability in X that component explains. Typically, PCA is used for exploratory analysis of a data set and is often used to identify variables that vary together and qualitatively to identify components which explain most of the variability. There are several different ways of determining the most important components; two often-used methods include choosing components that explain a certain percentage of the

total variability or only choosing components with eigenvalues greater than 1. In this work, we use the method of choosing components that together explain at least 80% of variability.

In standard multivariate regression, the dependent variable, *Y*, is regressed against the independent variable matrix, *X* (Eq. 2). In PCR, the dependent variable is regressed against the scores, *Z*, from the PCA analysis (Eq. 3). Note that *Y* in both equations represents $PM_{2.5}$ (either observed or interpolated) and that the scores are only from the components determined to be most important via the manner explained above.

$$Y = X * A + \varepsilon \qquad \text{(Equation 2)}$$

$$Y = Z * B + \varepsilon \qquad \text{(Equation 3)}$$

Since $V^TV = VV^T = I$, the identity matrix, due to orthonormality, we can also derive Eq. 3 from Eq. 2 via the following:

$$Y = X * A + \varepsilon = X * V * V^T * A + \varepsilon = Z * V^T * A + \varepsilon = Z * B + \varepsilon .$$

Thus, we can derive the relationship between *A* and *B* (Eq. 4).

$$A = V * B \qquad \text{(Equation 4)}$$

where *A* represents the unit-less sensitivity of $PM_{2.5}$ to the standardized variables, and *B* is the vector of regression coefficients from the PCR analysis. Further, only the coefficients in B that are less than a predetermined p-value are used. As explained later, we used a p-value of 0.08 (92% confidence). Physical units can be applied back to the unitless sensitivities by applying Eq. 5, where $D_{t,k}$ is the unit-based sensitivity of $PM_{2.5}$ to parameter *k*, $s_{PM_{2.5}}$ is the unit-based standard deviation of $PM_{2.5}$, s_k is the unit-based standard deviation of parameter *k*, and A_k is the unit-less sensitivity of $PM_{2.5}$ to parameter *k*:

$$D_{t,k} = \frac{s_{PM_{2.5}}}{s_k} A_k \qquad \text{(Equation 5)}$$

The full data set contained 635 days of observed and modeled (forecast) meteorological parameters and fire data (i.e. acres burned and KBDI). $PM_{2.5}$ data was not included in determining principal components (PCs) as this is the dependent variable in the PCR. The 635 days spanned a period from December 1, 2002 to March 15, 2007, largely determined by the occurrence of PB days, and thus was limited to days during the PB season (winter and spring). During this 4.5 year period, PB was conducted on 201 occasions (days) with areas burnt between 2 and 3800 acres, the daily average being 513 ±575 acres (single std.dev.) and median being 323 acres. We also gathered data for two days prior to and five days after a single burn day. The data from two days before a fire would capture the $PM_{2.5}$ conditions prior to the PB, and the five days after would allow the capture of potential effects from smoldering. PCA was applied to several data sets that represented a combination of exploratory variables. PCA was initially run on two data sets: i) observed meteorological conditions with AM-reported meteorological forecast (PC-AM) and ii) observed meteorological conditions with PM-reported meteorological forecast (PC-PM). Here, AM means the forecast in the morning of the performed PB and PM refers to the forecast in the evening prior to the PB. Subsequently, three more data sets were analyzed using PCA. They include meteorological observations with fire data (OBS only), AM forecast with fire data (AM ONLY) and PM forecast with fire data (PM ONLY).

Due to the specific setting of the source (MCBCL) relative to receptor (Jacksonville proper) in our case (see Figure 6.1), the PCR analysis is expected to be particularly sensitive to wind direction. Wind direction is reported in degrees for only the observed data and in main sectors (e.g. N, NNE, NE etc.) for the forecasted data, which were converted into degrees for the PCR analysis. We tested the sensitivity of the method to this parameter by running PCA on all five data sets with both observed and

forecast wind direction into to splitting the wind direction into its north-south and east-west vector components (Table 6.3) with the mathematical convention of southerly and westerly component flows being positive, respectively. Thus, there were a total of 10 data sets, each with 635 days of data, analyzed by PCA.

Table 6.3: List of data sets subject to PCA. The "X" indicates if the data subtype was used.

Name of PCA Run	PB Data	Observed Meteorology	AM Forecast	PM Forecast	Wind Direction
PCA-AM	X	X	X		Degrees
PCA-PM	X	X		X	Degrees
PCA-OBS ONLY	X	X			Degrees
PCA-AM ONLY	X		X		Degrees
PCA-PM ONLY	X			X	Degrees
PCA-AM WD	X	X	X		N-S/E-W Components
PCA-PM WD	X	X		X	N-S/E-W Components
PCA-OBS ONLY WD	X	X			N-S/E-W Components
PCA-AM ONLY WD	X		X		N-S/E-W Components
PCA-PM ONLY WD	X			X	N-S/E-W Components

Scores from the principal components which explained ~80% of the variance (the first seven components) were regressed against $PM_{2.5}$ on days that met two conditions: first, there was both $PM_{2.5}$ data and PB activity at MCBCL and second, there was no PB activity the previous two days. This allowed us to quantify same day effects (lag 0) of PB on $PM_{2.5}$. The importance of one and two-day lag on sensitivities was also examined. Understanding lag is important because the smoldering stages of a prescribed fire can last for days, and in effect turn it into a continuous source; also, potentially long transport times (e.g. under stagnant conditions) may result in impacts at a receptor location days after the actual PB conduct. We reduced the data set so that days with multiple lag effects were removed. For lag 1, regression was conducted only for days for which there

was PB activity on day *n* and $PM_{2.5}$ measurement on day *n+1* and no PB activity on days *n-1* and *n+1*. Similarly for lag 2, regression was conducted only for days for which there was PB activity on day *n* and $PM_{2.5}$ measurement on day *n+2* and no PB activity on days *n+1* and *n+2*. The reduced data set, with no multiple lag effects, resulted in data sets of 32 days for lag 0, 16 days for lag 1 and 32 days for lag 2. The regression yielded a standardized coefficient vector *B*, which was converted to unit-less sensitivities, *A*, via Equation 4, and to unit-based sensitivities, $D_{t,k}$, via Equation 5. Only the components that had regression coefficients that were significant at $p<0.08$ were used to calculate unit-less sensitivities (Tables 4a and b).

6.4. RESULTS

6.4.1. Principal Components and Regression

For all 10 PCA runs, the first seven principal components (V in Eq. 1) cumulatively explained at least 80 % of the total variance. Each of the seven PCs, with PC1 explaining the largest amount of variability in the $PM_{2.5}$ data set and PC7 the least, for each PCA run, is dominated by a few important parameters (Table 6.4). While there are some significant differences between these 10 cases, several of the principal components share similarities. For example, the first PC is dominated by relative humidity and temperature, indicating their importance to the overall variability of $PM_{2.5}$. The second component is typically loaded by a combination of temperature and dispersive parameters, including wind (WS, WD) and atmospheric stability (HAI). Prescribed burning (PB) is not prominent until the PC4 or later; this is to be expected because its sample size (i.e. 201 occurrences with values greater than zero of total 635 records) is far smaller than any meteorological parameter. Further, $PM_{2.5}$ is associated with meteorological parameters even if PBs are not present.

Table 6.4: Main parameters of the first seven principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
PCA AM	Relative Humidity	Temperature, WS, HAI	VR, WS	WD, Temp Diff., Visibility	Temp Diff., WD, KBDI, VIS	PB	Precipitation, BLH, HAI,
PCA PM	Relative Humidity	Temperature, WS	VR, WS	WD, Temp Diff.,	Temp Diff., Precipitation, WD, KBDI, VIS	PB	Precipitation, Temp Diff, BLH, PB
PCA AM WD	Relative Humidity, Temperature	Temperature, BLH, WS (N-S), HAI, KBDI	WS, VR	KBDI, BLH, VR and both Wind directions	PB, E-W wind direction	T Diff, mcaPCP, , mcaVIS, PB (neg),	PB, mcaPCP
PCA PM WD	Relative Humidity, Temperature	Temperature, BLH, WS (N-S), HAI	VR, WS, EW wind direction, T diff PM	KBDI, BLH, T diff, WS – both directions	EW wind dir., PB, VR, WS	PB, VIS, mcaPCP, HAI, BLH	Precipitation, PB, BLH, VIS
PCA Obs only	Relative Humidity, Temperature	Wind direction, Temperature, KBDI, mcaPCP	WS, VIS, KBDI, Temperature, T Diff, RH	PB, mcaPCP, T diff, VIS	mcaPCP, T diff, VIS, PB	mcaPCP, VIS, KBDI, WS	VIS, KBDI, RH
PCA AM only	Relative Humidity, POP, IBT,HAI,	BLH, Temp, KBDI, HAI, POP	WS, VR, POP	PB, WD, T Diff	PB, WD, HAI	T, WD, HAI, IBT	BLH, KBDI, POP, WS
PCA PM only	Relative Humidity, Temperature, POP, HAI, WD	BLH, Temp, HAI, IBT, KBDI, WS, POP	VR, WS, T Diff, BLH	WD, PB, T Diff	PB, WD, KBDI, HAI	KBDI, HAI< T diff	BLH, HAI, WS, WD, Temperature
PCA Obs Only WD	Relative Humidity, Temperature, RH, KBDI, VIS	NS Wind direction, Temperature, KBDI, WS, mcaPCP	WS, EW Direction, KBDI, Temperature, T Diff	KBDI, mcaPCP, Visibility	PB	mcaPCP, Temperature, E-W Wind,	WS, Both wind direction, mcaPCP
PCA AM Only WD	Relative Humidity, Temperature, POP, various others	BLH, Temp, KBDI	VR, WS, POP	PB	EW wind dir., PB, NS Wind direction, VR, BLH	EW wind dir., KBDI, HAI, N-S WD	Temp, PB, both Wind Direction
PCA PM Only WD	Relative Humidity, Temperature, POP, KBDI, Both WD, HAI	BLH, Temp, HAI, IBT, POP, E-W WD	VR, WS, both directions	NS Wind direction, PB, Temperature, KBDI, BLH	PB, E-W WD, KBDI	EW wind dir., PB, KBDI, T diff, NS WD, HAI	HAI, BLH, KBDI, WS

6.4.2. Unit-less sensitivities

Table 6.5: Regression p-values for the 10 cases of PCA runs. Numbers in bold highlight p-vaules ≤ 0.08.

	Intercept	PC1	PC2	PC3	PC4	PC5	PC6	PC7	R^2
PCA AM	0.57	**0.08**	0.52	0.13	0.19	0.44	**0.04**	0.67	0.28
PCA PM	0.93	**0.03**	0.23	0.1	0.14	0.98	**0.03**	0.26	0.29
PCA Obs only	0.65	**0.03**	0.75	0.14	0.16	**0.05**	0.3	**0.04**	0.37
PCA AM only	0.3	0.18	0.84	0.51	0.59	**0.02**	0.37	0.94	0.3
PCA PM only	0.23	0.43	0.95	0.45	0.81	**0.01**	0.28	0.7	0.29
PCA AM WD	0.79	**0.03**	0.15	0.27	0.28	0.23	0.25	**0.01**	0.34
PCA PM WD	0.99	**0.03**	0.12	0.18	0.45	0.16	0.11	**0.03**	0.32
PCA Obs WD	0.59	**0.05**	0.5	0.23	0.62	**0.08**	0.47	**0.04**	0.28
PCA AM Only WD	0.31	0.25	0.28	0.73	0.57	0.56	0.14	**0.03**	0.37
PCA PM Only WD	0.22	0.54	0.77	0.25	0.86	0.12	0.11	0.31	0.29

Although the regression's correlation coefficients were low (R^2 = 0.28 to 0.37) and only one to three components had p values less than 0.08 (Table 6.5), indicating large

variability in the original data, some important empirical relationships were found using the principal component vectors, V. Unitless sensitivities are used to directly compare how much relative influence individual parameters have on $PM_{2.5}$. Positive values indicate they contribute to an increase in $PM_{2.5}$, while negative values point to a decreasing effect on $PM_{2.5}$. The association of input parameters with $PM_{2.5}$ is investigated for the five sets where wind direction is used in degrees (Figure 6.2) and when it is split in to N-S and E-W components (Figure 6.3). In both cases, forecast parameters characterizing atmospheric stability (e.g. Haines Index, ventilation rate, wind speed and mixing height) have similar importance in keeping ambient $PM_{2.5}$ levels low, contributing to a decrease in $PM_{2.5}$, thus exhibiting a negative sensitivity. In contrast, PB and KBDI are important when $PM_{2.5}$ concentration increases. Sensitivity to PB is positive and consistently appears to be an important contributor to $PM_{2.5}$, and is similar in all four cases where forecast meteorology is used. However, $PM_{2.5}$ has a lower unit-less sensitivity to PB while KBDI is more strongly associated with $PM_{2.5}$ in the PC-OBS ONLY case. The forecast parameters IBT and dayT are consistently positive, indicating clear sky conditions under which a preceding strong nocturnal inversion allows $PM_{2.5}$ emissions to accumulate. Dispersive parameters TWS and VR and mcaWS are consistently negative, pointing to their diluting effects on $PM_{2.5}$ concentrations.

Splitting the wind direction into N-S and W-E vector components created some notable differences in unitless sensitivities. First, the PC-PM ONLY case with wind directions split resulted in no statistically significant regression coefficients. Second, splitting wind direction caused $PM_{2.5}$ sensitivity to be largest to measured precipitation. A similar effect can be seen for the forecast temperature difference parameter, diffT. In addition, when wind direction was used in degrees, it had a strong association with $PM_{2.5,}$ suggesting that continental air masses carry more $PM_{2.5}$ than maritime air from easterly directions. When wind direction is split into components, $PM_{2.5}$ shows a slight positive sensitivity to the E-W components when both observed and forecast parameters are used

and negative for N-S components, indicating additional $PM_{2.5}$ being transported to the receptor location from westerly and northerly directions, respectively, both of which can be considered continental air masses.

A number of parameters changed from a positive to a negative sensitivity, depending on if only observations, forecast parameters, or both were used. For example, in the case with wind direction in degrees, the forecast wind direction, TWD exhibits a large positive sensitivity (higher $PM_{2.5}$ in continental air masses) when only forecast parameters are used, whereas it appears to have a small negative sensitivity when observations and forecast parameters are used together (low $PM_{2.5}$ loadings in maritime air or sea breeze). To a lesser extent, mcaPCP has a positive association with $PM_{2.5}$ when observations and AM forecast is used, but more plausible negative association otherwise (the reason for this apparently implausible connection is discussed below). Similarly, diffT has a positive impact when only the AM forecast is used, but negative otherwise. Large temperature differences are an indicator for clear sky conditions causing large difference between daytime high and nighttime low temperatures prone for shallow nocturnal inversion layers near the ground, but can also indicate conditions for greater afternoon dispersion.

The wind direction split has similar effects only for the forecast parameters, BLH, TWS, and E-W wind component, such that the BLH and E-W wind component positively impact $PM_{2.5}$ when observations are combined with forecasts. However, BLH yields negative impact (pointing to $PM_{2.5}$ accumulation in shrinking BL) when only the AM forecast is used. When observations are combined with forecasts, $PM_{2.5}$ has a negative sensitivity to TWS due to its diluting effect on the BL, while a less plausible positive association is obtained when only the AM forecast is used.

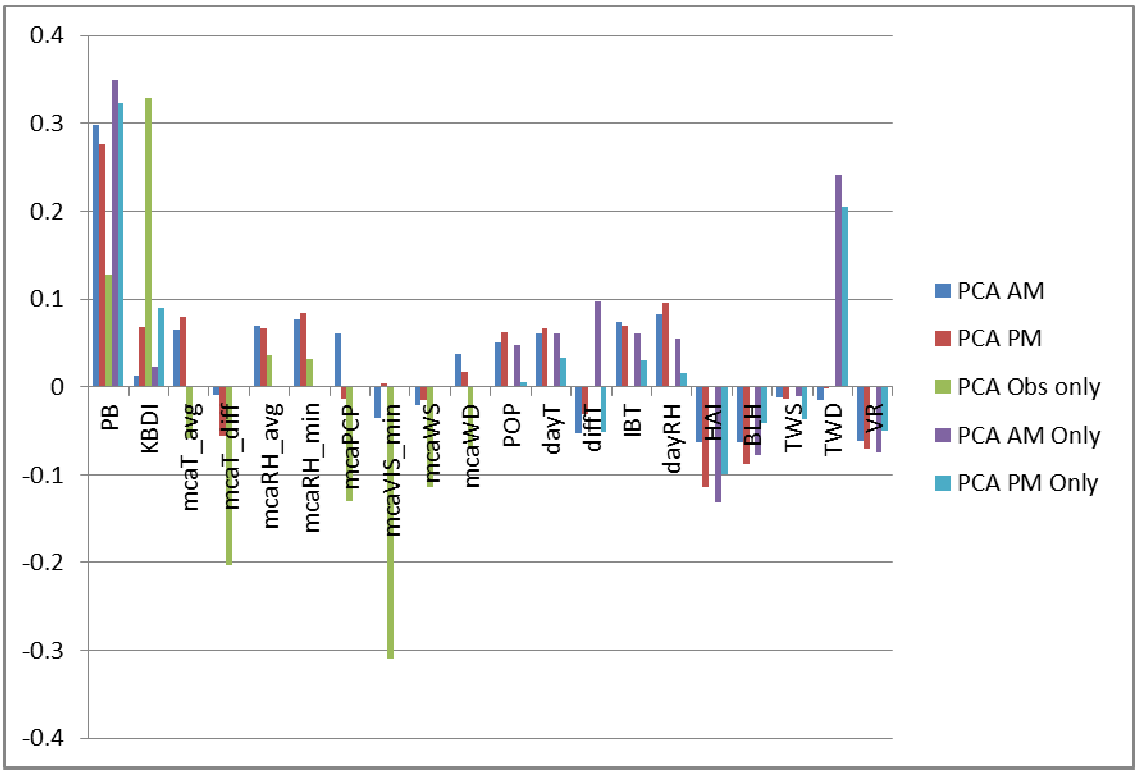


Figure 6.2: Unitless senstivities for the five cases with wind direction in degrees.

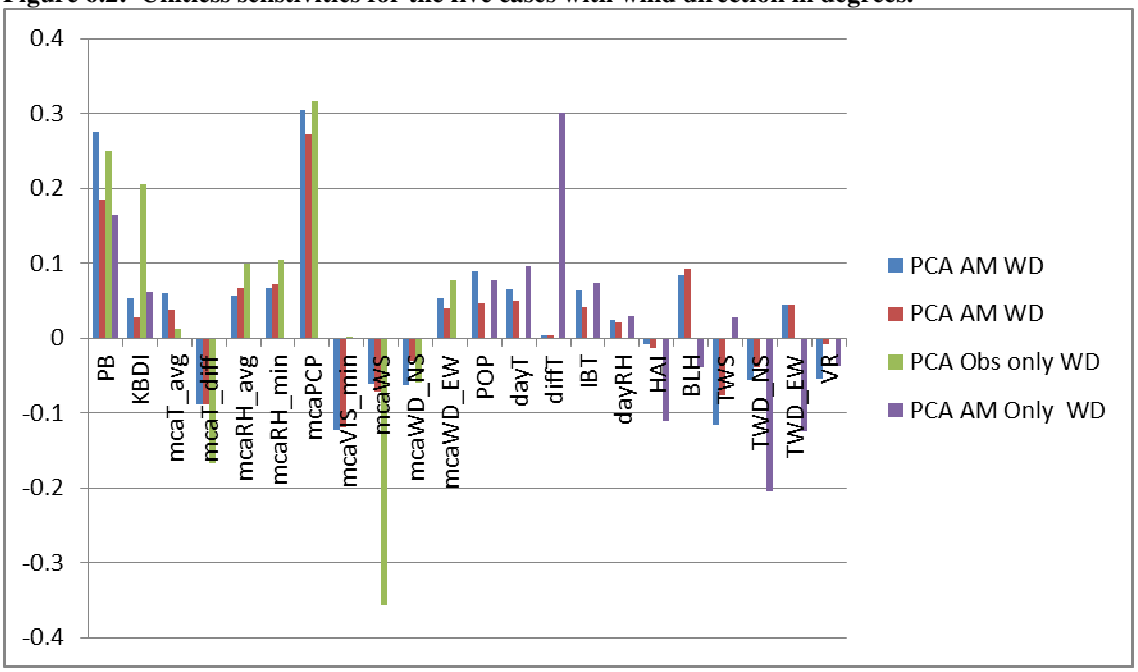


Figure 6.3: Unitless senstivities for the four cases with wind direction in N-S and E-W components. NOTE: PCA PM Only WD analysis did not result in any regression p values <0.08.

Taking the averages of the unitless sensitivities across all applicable runs, and taking the standard deviation as an estimate of uncertainty, indicates which parameters have the most significance to $PM_{2.5}$ (Figure 5.4). That is, unitless senstivities whose standard deviation cross the 0 of the y-axis, can be viewed as statistically not significant (at ~68% confidence). Thus, of the fire data parameter, senstivity of $PM_{2.5}$ to PB is significant, but not KBDI. For the measured meteorology, relative humidity and wind direction splits are significant. Positive $PM_{2.5}$ senstivities to E-W winds indicate importance of winds coming from the west while negative $PM_{2.5}$ senstivities to N-S winds indicate importance of winds coming from the north, the origin of both can be considered continental. For the forecast parameters, POP, dayT, IBT, and RH have significant postive sensitivities while HAI and VR have significant negative senstitivities. The postitive senstivity associated with POP was not expected but reflects the fire managers' actual positive consideration of POP in their decision to start PB on days when precepitation might be expected in late afternoons, because it acts as a natural fire break preventing uncontrolled fire spread. Similar to measurements, the forecast southerly wind component carried by the sea breeze leads to a decrease in $PM_{2.5}$.

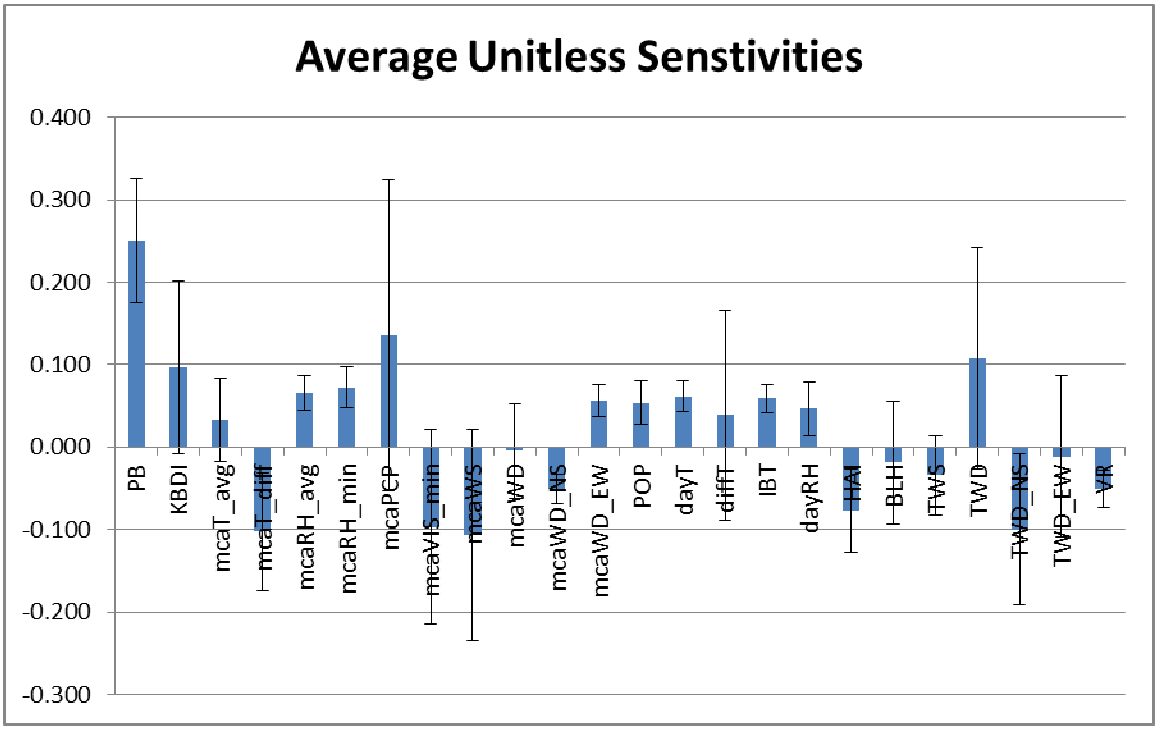


Figure 6.4: Average unitless senstitives. Error bars indicate standard deviation over all cases where parameter was used in the analysis.

6.4.3. Differences in AM versus PM forecasts based on unitless sensitivities

Fire managers use both AM and PM weather forecasts to determine if a particular day will have ideal conditions for PB. When observations are included in the PCA data set, splitting the wind direction into NS and EW components leads to a higher correlation between unitless sensitivities (Figure E.1), indicating that AM or PM forecasts are similarly associated with $PM_{2.5}$. When the PCA data sets include only fire data and forecasts (observations excluded), splitting the wind direction into NS and EW components leads to statistically significant unitless sensitivities for the AM forecast only (Figure E.2), indicating that AM forecasts are more stable to PCR analyses.

6.4.4. Unit-based Sensitivities

Unit-less sensitivities are used for direct comparison of the relative importance of different forecast parameters on the local $PM_{2.5}$ burden. However, it is of interest to air quality managers to know the unit-based sensitivity of $PM_{2.5}$ to parameters of interest, especially PB. Across all 10 cases, we find that at a p-value of 0.08 (92% confidence), there is consistency across all 10 test cases, resulting in an average sensitivity of $PM_{2.5}$ of 3.2 ± 1.0 $\mu g\ m^{-3}$ per 1000 acres burned (Table 6.6). A p-value of 0.08 was chosen because the PB based components in the PC-OBS ONLY case had p-values slightly higher than 0.05. However, at p-value of 0.08, $PM_{2.5}$ sensitivities are consistent across all cases, indicating that the PCR method employed in this work is stable.

An increase in daily average temperature of 1 °C (mcaT_avg) would add 0.02 $\mu g\ m^{-3}$ at MCBCL, which suggests that contribution from atmospheric SOA during the PB season is negligible compared to the spatial source-receptor relationship at this location. Larger temperature differences between daytime high and nighttime low have a decreasing effect on $PM_{2.5}$. For example, an increase of 1 K in observed daily max-min temperature difference (mcaT_diff) leads to a decrease of $PM_{2.5}$ by 0.14 ± 0.10 $\mu g\ m^{-3}$, pointing to the effect of increased dilution with larger mixing heights that result from solar heating. $PM_{2.5}$ sensitivity to the forecast parameter, diffT, is slightly positive (an increase of $PM_{2.5}$ by 0.033 ± 0.10 $\mu g\ m^{-3}$ for every increase of 1 K in modeled daily max-min temperature difference); however, the high standard deviation indicates that the predicted temperature difference has no significant relationship with measured PM2.5.

Table 6.6: Unit-based sensitivities of $PM_{2.5}$ to various parameters. NOTE: "Number of cases" refers to the number of datasets from Table 6.3 that a particular parameter is included in a PCA run.

Parameter	Average Unit-based Sensitivities	St. Dev	Number of Cases
PB	0.0032	0.001	9
KBDI	0.006	0.006	9
mcaT_avg	0.020	0.032	6
mcaT_diff	-0.137	0.096	6
mcaRH_avg	0.023	0.007	6
mcaRH_min	0.020	0.007	6
mcaPCP	0.104	0.145	6
mcaVIS_min	-0.129	0.158	6
mcaWS	-0.369	0.444	6
mcaWD	-0.108	0.203	4
mcaWD_NS	-0.356	0.127	3
mcaWD_EW	0.455	0.157	2
POP	0.010	0.005	7
dayT	0.021	0.006	7
diffT	0.033	0.103	7
IBT	0.019	0.005	7
dayRH	0.016	0.011	7
HAI	-0.438	0.285	7
BLH	0.000	0.000	7
TWS	-0.021	0.031	7
TWD	0.006	0.007	4
TWD_NS	-0.723	0.661	3
TWD_EW	-0.090	0.799	3
VR	0.000	0.000	7

Since rain is the main sink for PM in the atmosphere, it might be expected that an increase measured precipitation, mcaPCP, and forecast probability of precipitation POP reduces $PM_{2.5}$ mass concentration. However, we find that $PM_{2.5}$ increases by 0.10 ± 0.145 $\mu g\ m^{-3}$ for every mm increase in measured precipitation; albeit the high standard deviation indicates that there is significant variability across the different cases. $PM_{2.5}$

sensitivity to forecast POP, 0.01 ± 0.005 μg m^{-3} for every percent increase in probability of precipitation, is less variable than measured precipitation. This positive sensitivity is most likely due to fire managers' decision to start fires on days with forecast precipitation later in the day, after the fires are completed [*Becker*, 2013]. This is common practice because afternoon rain is considered a welcome fire break preventing unwanted spread of fires and naturally extinguishing smoldering fires.

The observed wind speed, mcaWS, has a more significant influence on [$PM_{2.5}$] than the modeled TWS, which may be due to the greater uncertainty of TWS, but fine PM levels decrease as either increases due to dilution. An increase of 1 m s^{-1} measured near the surface is associated with a reduction in $PM_{2.5}$ levels by ~0.37 μg m^{-3} whereas the same increase predicted for the entire BL would result in an 0.021 μg m^{-3} reduction on average only. Splitting the wind direction into NS and EW components showed that directionality is important. Our analysis shows that $PM_{2.5}$ is most sensitive to winds coming from the west (positive E-W sensitivity) and north (negative N-S sensitivity) (Table 6.5). Although the $PM_{2.5}$ monitor is located north from the MCBCL, forested lands ca. 20 km to the north-northeast (Hoffman Forest), 40-50 km to the east (Croatan National Forest) and 40-50 km to the west/southwest (Holly Shelter and Angola Bay game lands) receive PB treatment similar to MCBCL; i.e. foresters there employ similar criteria in their PB decision process. Wind rose plots of the lag 0 data (32 days) reveal that the highest $PM_{2.5}$ occurs when winds occur from the west/southwest and south/southeast (Figure 6.5a). When the winds are from the north, $PM_{2.5}$, although not as high as from the west/southwest and south/southeast, is moderate and greater the 8 μg m^{-3}. Easterly component winds are rare potentially influencing the directional sensitivity of $PM_{2.5}$.

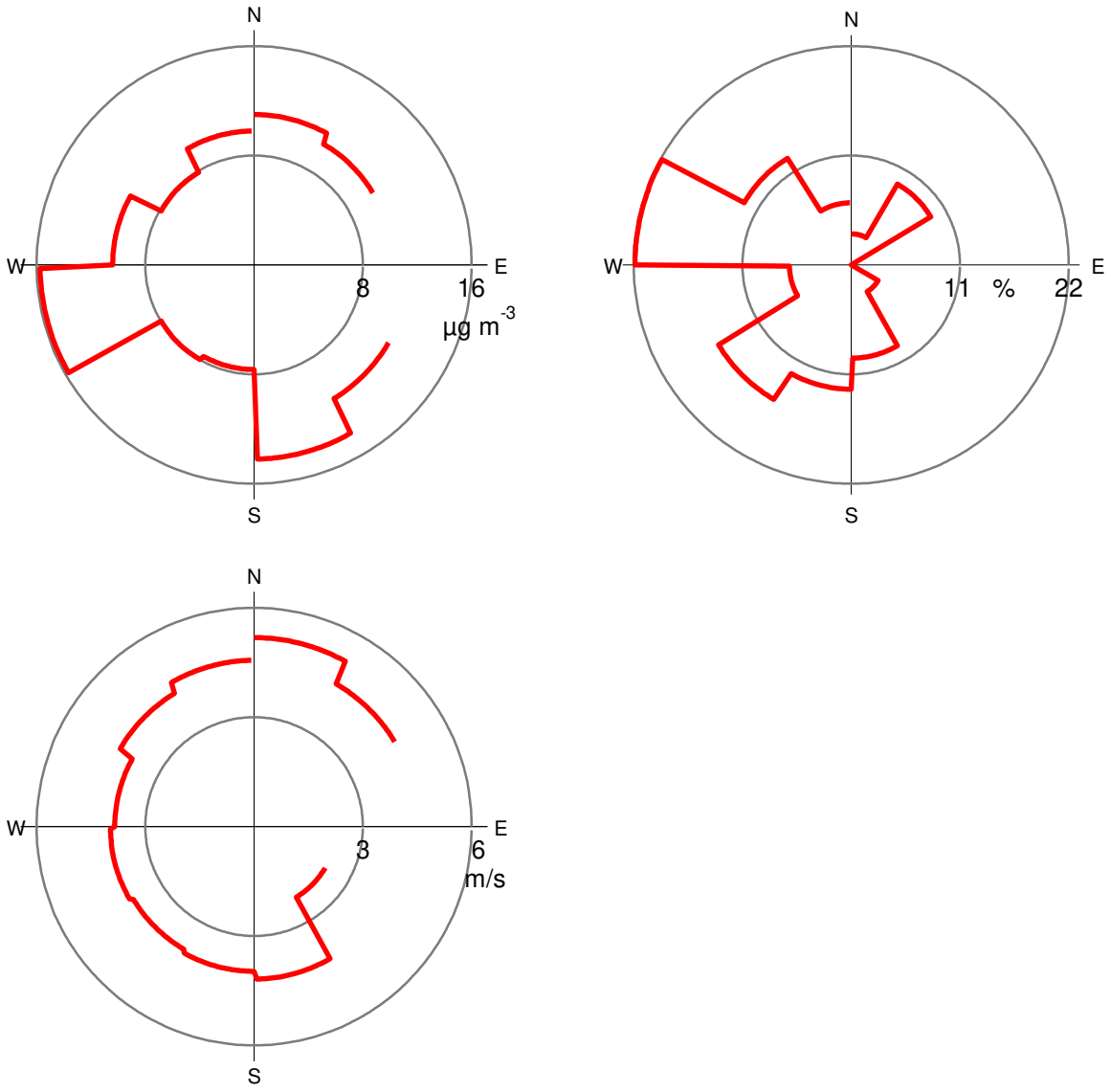


Figure 6.5: 24 hour average wind rose plots for lag 0 days of (a) $PM_{2.5}$, b) wind frequency and (c) wind speed. Gaps indicate no data.

The wind rose of $PM_{2.5}$ corroborates our analysis that $PM_{2.5}$ at the monitoring site is impacted by PB at MCBCL. It should be noted that we used only MCBCL's PB activity records because this is the most readily available PB data. However, all forested land in the greater area surrounding MCBCL proper is managed by PB. Therefore, the MCBCL records are potentially a surrogate measure for more wide spread PB activity in the region. Nevertheless, is the $PM_{2.5}$ monitor in Jacksonville directly adjacent to the

northern border of MCBCL, making this specific setting the most direct source-receptor relationship. Even if there was significant amounts of acres burned in regional forests, impact from those PB activities would not be as direct and far more dilute into the regional background level of $PM_{2.5}$. Thus, the PCR method employed in this work provides insight into local impacts from PB.

We also assessed the feasibility of using spatially versus temporally interpolated $PM_{2.5}$ data. Preliminary results with both spatially and temporally interpolated $PM_{2.5}$ leads to increased variability in sensitivities. This is most likely because the interpolation scheme introduces noise that is propagated into the regression results. Alternate spatial interpolation schemes, including fusing observations with chemical transport model results may be useful to reduce interpolation errors, leading to more days that can be used in the regression analysis.

6.5. Conclusions

Fire weather forecasts are used by wildlife managers in determining when PB activities are to occur. In this work, we explored differences in AM and PM forecasts, and impacts to air quality of PB by using PCR to estimate sensitivity of $PM_{2.5}$ to PB activity and meteorological parameters. We ran PCA on 10 data sets that included PB activity data along

with meteorological parameters of interest; the meteorological parameters included either observational data only, forecast data only or a combination of observations and forecasts. PCR was performed on the scores from the first seven components, which explained greater than 80% in all 10 data sets, and $PM_{2.5}$, to estimate sensitivities of $PM_{2.5}$ to all parameters of interest. $PM_{2.5}$ showed a significant association to PB, with a unit-based sensitivity of 3.2 ± 1 μg m^{-3} $PM_{2.5}$ per 1000 acres burned. $PM_{2.5}$ had a negative sensitivity dispersive parameters and was senstivite to winds coming from the west and the north, the origin of both can be considered continental.

6.6. References

1. Wade, D. D.; Brock, B. L.; Brose, P. H.; Grace, J. B.; Hoch, G. A.; Patterson III, W. A., Chapter 4: Fire in Eastern Ecosystems. In *Wildland fire in ecosystems: effects of fire on flora*, Smith, J. K.; Lyon, L. J., Eds. 240 West Prospect Road, Fort Collins, CO 80526-2098): U.S. Department of Agriculture Forest Service Rocky Mountain Research Station, 2000.
2. USCensus http://www.census.gov/popest/data/metro/totals/2012/index.html (March 19, 2013),
3. Davidson, C. I.; Phalen, R. F.; Solomon, P. A., Airborne particulate matter and human health: A review. *Aerosol Sci. Technol.* **2005,** *39*, (8), 737-749.
4. Lee, S.; Russell, A. G.; Baumann, K., Source apportionment of fine particulate matter in the southeastern united states. *Journal of the Air & Waste Management Association* **2007,** *57*, (9), 1123-1135.
5. Hu, Y.; Odman, M. T.; Chang, M. E.; Jackson, W.; Lee, S.; Edgerton, E. S.; Baumann, K.; Russell, A. G., Simulation of air quality impacts from prescribed fires on an urban area. *Environmental Science & Technology* **2008,** *42*, (10), 3676-3682.
6. Lee, S.; Baumann, K.; Schauer, J. J.; Sheesley, R. J.; Naeher, L. P.; Meinardi, S.; Blake, D. R.; Edgerton, E. S.; Russell, A. G.; Clements, M., Gaseous and particulate emissions from prescribed burning in Georgia. *Environmental Science & Technology* **2005,** *39*, (23), 9049-9056.
7. U.S.EPA, Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standards. In U.S. EPA Office of Air Quality Planning and Standards: 2011.
8. Baumann, K. *Study of Air Quality Impacts Resulting from Prescribed Burning, Final Report*; IMA-SERO and USAIC: 2005.
9. Fekedulegn, D.; Colbert, J. J.; Schuckers, M. E.; Hicks Jr., R. R., Coping with multicollinearity: An example on theory and application of principal components regression in dendroecology. In Service, U. S. F., Ed. US Department of Agriculture: 2002.

APPENDIX E: SUPPLEMENTAL INFORMATION FOR CHAPTER 6

Table E.1: Unitless Sensitivities.

	PCA AM	PCA PM	PCA Obs only	PCA AM Only	PCA PM Only	PCA AM WD	PCA PM WD	PCA Obs only WD	PCA AM Only WD
PB	0.298	0.276	0.128	0.349	0.322	0.275	0.184	0.251	0.164
KBDI	0.013	0.068	0.329	0.023	0.090	0.053	0.028	0.206	0.062
mcaT_avg	0.065	0.080	-0.059			0.059	0.038	0.012	
mcaT_diff	-0.009	-0.056	-0.203			-0.089	-0.088	-0.167	
mcaRH_avg	0.070	0.067	0.036			0.056	0.067	0.099	
mcaRH_min	0.077	0.084	0.031			0.068	0.071	0.105	
mcaPCP	0.061	-0.014	-0.130			0.304	0.272	0.317	
mcaVIS_min	-0.035	0.004	-0.310			-0.122	-0.120	0.002	
mcaWS	-0.021	-0.015	-0.114			-0.061	-0.071	-0.357	
mcaWD	0.038	0.018	-0.070						
mcaWD_NS						-0.064	-0.030	-0.059	
mcaWD_EW						0.054	0.039	0.078	
POP	0.051	0.062		0.048	0.005	0.089	0.047		0.076
dayT	0.060	0.066		0.060	0.033	0.065	0.050		0.095
diffT	-0.053	-0.031		0.097	-0.051	0.004	0.003		0.302
IBT	0.074	0.070		0.060	0.030	0.064	0.041		0.073
dayRH	0.083	0.095		0.055	0.016	0.025	0.021		0.029
HAI	-0.063	-0.114		-0.131	-0.100	-0.008	-0.013		-0.112
BLH	-0.063	-0.088		-0.077	-0.041	0.083	0.092		-0.039
TWS	-0.012	-0.013		-0.010	-0.037	-0.116	-0.077		0.027
TWD	-0.015	-0.002		0.242	0.205				
TWD_NS						-0.056	-0.039		-0.206
TWD_EW						0.045	0.045		-0.125
VR	-0.061	-0.071		-0.075	-0.050	-0.054	-0.009		-0.037

Table E.2: Unit-based Sensitivities.

	PCA AM	PCA PM	PCA Obs only	PCA AM Only	PCA PM Only	PCA AM WD	PCA PM WD	PCA Obs only WD	PCA AM Only WD
PB	0.004	0.004	0.002	0.004	0.004	0.004	0.002	0.003	0.002
KBDI	0.001	0.004	0.019	0.001	0.005	0.003	0.002	0.012	0.004
mcaT_avg	0.041	0.050	-0.037			0.037	0.024	0.008	
mcaT_diff	-0.013	-0.075	-0.272			-0.119	-0.118	-0.224	
mcaRH_avg	0.025	0.024	0.013			0.020	0.024	0.035	
mcaRH_min	0.022	0.023	0.009			0.019	0.020	0.029	
mcaPCP	0.047	-0.011	-0.100			0.233	0.209	0.244	
mcaVIS_min	-0.046	0.006	-0.413			-0.163	-0.159	0.003	
mcaWS	-0.073	-0.051	-0.394			-0.210	-0.246	-1.239	
mcaWD	0.002	0.001	-0.024						
mcaWD_NS						-0.445	-0.210	-0.412	
mcaWD_EW						0.432	0.311	0.622	
POP	0.009	0.012		0.009	0.001	0.016	0.009		0.014
dayT	0.021	0.023		0.021	0.012	0.022	0.017		0.033
diffT	-0.044	-0.022		0.079	-0.037	0.003	0.002		0.247
IBT	0.023	0.023		0.019	0.010	0.020	0.013		0.023
dayRH	0.029	0.034		0.019	0.006	0.009	0.008		0.010
HAI	-0.358	-0.643		-0.748	-0.563	-0.044	-0.073		-0.638
BLH	0.000	0.000		0.000	0.000	0.000	0.000		0.000
TWS	-0.008	-0.008		-0.007	-0.022	-0.076	-0.047		0.018
TWD	-0.001	0.000		0.013	0.011				
TWD_NS						-0.406	-0.281		-1.483
TWD_EW						0.361	0.381		-1.012
VR	0.000	0.000		0.000	0.000	0.000	0.000		0.000

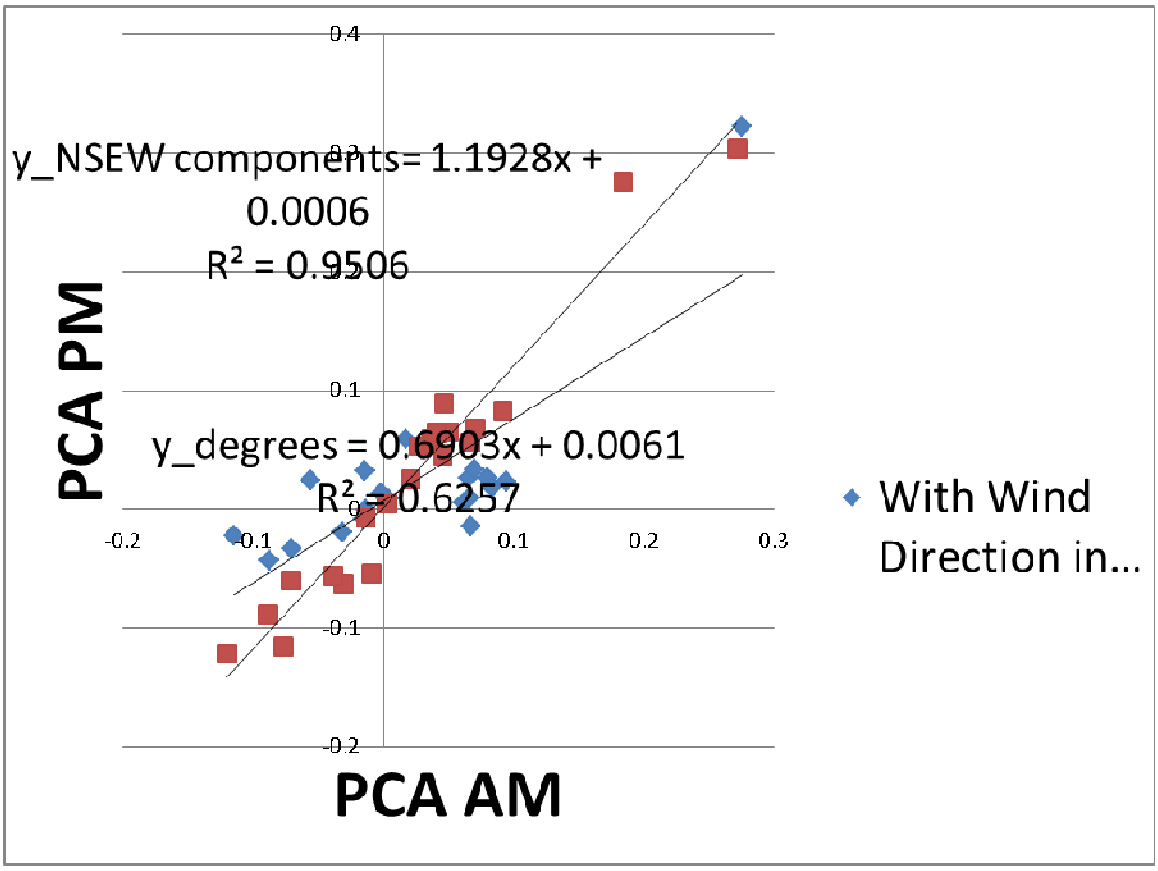


Figure E.1: Correlation of PCA_AM and PCA PM (i.e. PCA data comprised of fire data, observations and forecasts).

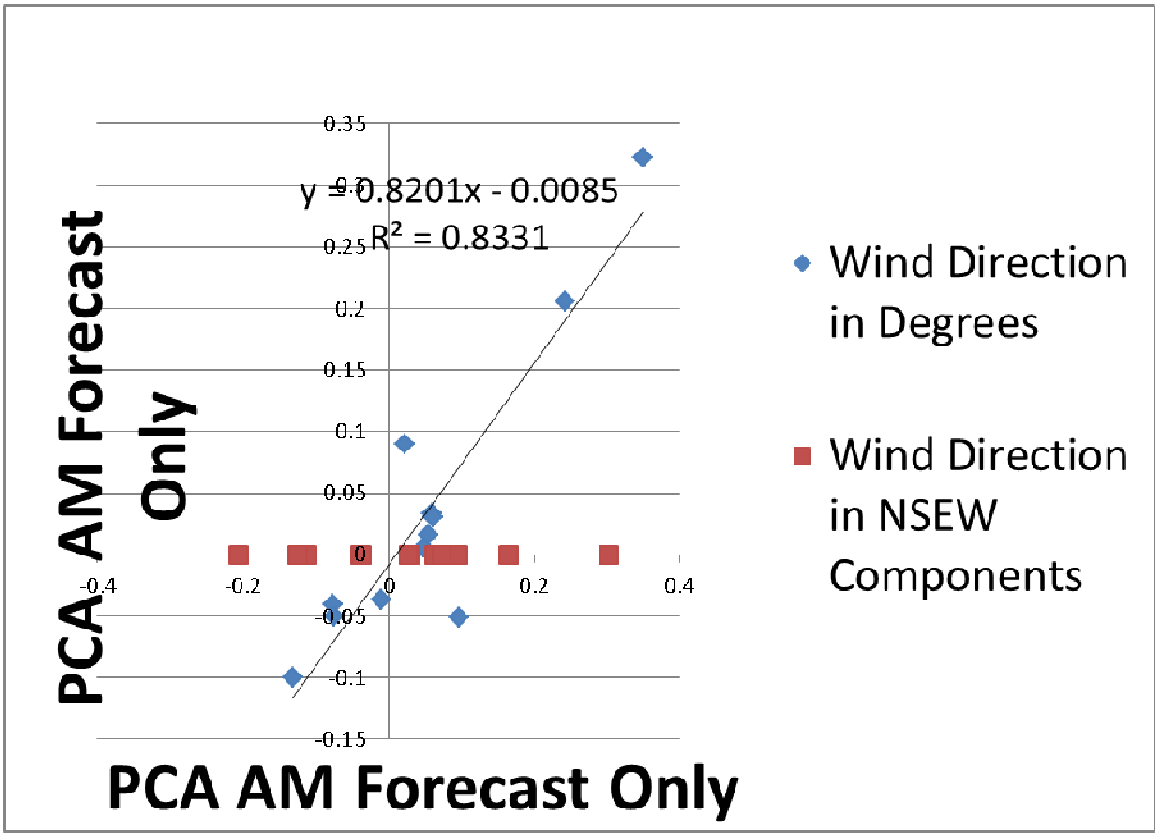


Figure E.2: Correlation of PCA_AM only and PCA PM only (i.e. PCA data comprised of fire data and forecasts).

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

7.1. Conclusions

In this dissertation, a number of inconsistencies and limitations of various source apportionment techniques are addressed by ensemble-averaging results from a short-term application of three receptor-based models and one emissions-based model. Individual SA methods were evaluated for how much they should weigh in the calculation of an ensemble average by exploration of how these methods calculate uncertainties. The method has a number of benefits over using one model exclusively. The method provides a way to evaluate different source apportionment (SA) models, including estimating uncertainties in a consistent manner. Highlights of this research work include:

Chapter 2: Ensemble-Trained Source Apportionment of Fine Particulate Matter and Method Uncertainty Analysis. Ensemble averaging results in updated estimates of source impacts with lower uncertainties than individual SA methods. Overall uncertainties for ensemble-averaged source impacts were ~45 - 74%. Calculated positive matrix factorization (PMF) uncertainties increased from ~40% to ~70-150%. Calculated chemical mass balance (CMB) with molecular markers and Community Multiscale Air Quality (CMAQ) model uncertainties decreased to ~70 - 90% in the summer. One use of these updated uncertainties is that they can be incorporated into epidemiologic studies, which can ultimately lead to improving our understanding of the relationships between $PM_{2.5}$ sources and health outcomes. Further, they can be used to inform policy makers of the effectiveness of control measures.

Chapter 3: Bayesian–Based Ensemble Source Apportionment of PM2.5. We extend the ensemble method by developing a Bayesian-based ensemble averaging technique. The Bayesian-based source impacts for biomass burning correlate better with observed levoglucosan (R^2=0.66) and water soluble potassium (R^2=0.63) than source

impacts estimated using more traditional methods, and more closely agreed with observed total mass. The Bayesian approach also captures the expected seasonal variation of biomass burning and secondary impacts. Sensitivity analysis found that using non-informative prior weighting performed better than using weighting based on method-derived uncertainties.

Chapter 4: Spectral Analysis of $PM_{2.5}$ Source Apportionment Methods. All power spectra derived using the Lomb-Scargle periodogram method (LSPM) show a strong peak at one year, independent of SA methods, species and source profiles/factors. Statistically significant peaks (α =0.05) are found for the frequency associated with one week for GV and DV at JST and SDK for most methods, but not at the rural YRK. BURN spectra have the greatest variation intra and inter-method, with low frequency signals at JST and SDK and YRK having both low frequency and weekly signals. Biomass burning profiles/factors have the greatest variability across methods and locations, especially with BBSPs and PMF factors. OC to EC ratios vary from 3 - 5 in EBSPs, to 3.9 - 17.6 with BBSPs and 3.1 - 10.8 in PMF, suggesting that biomass burning emissions have increased spatial variability as compared to other sources.

Chapter 5: Particulate and Gas Sampling of Prescribed Fires in South Georgia, USA. Major $PM_{2.5}$ components included OC (~57%), EC (~10 %), chloride (~1.6%), potassium (~0.7%) and nitrate (~0.9%). Major gaseous species include carbon dioxide, carbon monoxide, methane, ethane, methanol and ethylene. Particulate organic tracers of biomass burning, such as levoglucosan, dehydroabietic acid and retene, increased significantly during the burns. Water soluble organic carbon (WSOC) also increased significantly during the fire and levels are highly correlated with potassium (K) (R^2=.93) and levoglucosan (R^2=0.98). The average WSOC/OC ratio was 0.51 ± 0.03 and did not change significantly from background levels. Thus, the WSOC/OC ratio may not be a good indicator of secondary organic aerosol (SOA) in regions that are expected to be impacted by biomass burning. Results using a biomass burning source profile derived

from this work further indicate that source apportionment is sensitive to levels of potassium in biomass burning source profiles. This underscores the importance of quantifying local biomass burning source profiles.

Chapter 6: Verification of Fire Weather Forecasts Using PM2.5 Sensitivity Analysis We ran PCA on 10 data sets that included PB activity data along with meteorological parameters of interest; the meteorological parameters included either observational data only, forecast data only or a combination of observations and forecasts. For each data set, we regressed PCA scores from the first seven principal components against observed $PM_{2.5}$. $PM_{2.5}$ showed significant sensitivity to PB, with a unit-based sensitivity of 3.2±1 $\mu g\ m^{-3}$ $PM_{2.5}$ per 1000 acres burned. $PM_{2.5}$ had a negative sensitivity to dispersive parameters such as wind speed and had positive senstivity to winds coming from the west and the north, the origin of both can be considered continental. It is expected that fire managers will be able to utilize this information to determine if conditions are optimal for minimizing impacts to PM levels in surrounding communities.

7.2. Future Work

The ensemble method was developed with the goal of providing source impacts that can be included in health studies. Therefore, a future study should look at how health models are affected by ensemble-based source apportionment. With the use of Bayesian-based source profiles (BBSPs), multiple SA results are realized for each day. These sets of SA results can be easily incorporated into multiple health model runs. One aspect of this work will be to understand if and how health impact risk ratios change depending on the SA model used in health assessment. Another important aspect of this work would be to incorporate uncertainties into the health models. Because SA is conducted 10 times for each day, 10 different health assessments can be conducted. The variability of these

10 health assessments can be used as an estimate of how SA uncertainty propagates into health studies.

The Bayesian ensemble method currently uses a fully analytical framework, with an inverse gamma prior distribution and normal likelihood function. A next step would be to use Markov Chain Monte-Carlo (MCMC) framework to estimate the posterior distribution of weights. This would obviate the need for conjugate priors and as a result, more appropriate priors could be used. This could address the need for modeling lognormal or other right-tailed skewed distributions in source impacts and source profiles. In MCMC Bayesian analysis, a selection heuristic must be provided. Because SA models result in source impacts that have an autocorrelation structure, selection rules may need to account for this. For example, we can say that day-to-day changes in sources impacts must fall within a given range of autocorrelation. Therefore previous day source impacts can act as prior information that could guide the estimate of the posterior distribution.

Ideally, this would lead to a broad realization of source profiles that can be selected for source apportionment of a long-term data set. In this dissertation, source profiles were randomly sampled 10 times for each day in the long-term data set. In addition, seasonal profiles were developed for summer and winter. It would be of interest to develop source profiles for the fall and spring seasons or even profiles for each month. These profiles could also be binned according to meteorological conditions. Therefore, instead of sampling from one of two seasonal sets of profiles, we could sample from sets of profiles that reflect emissions given certain meteorological conditions. These conditions could be based on temperature, wind and/or season. In addition, this information could also guide SA results by including extra sources (e.g. a point source that could impact a receptor site a given wind direction).

Source apportionment techniques can be extended to measurements beyond $PM_{2.5}$ speciation. For example, results from measurement techniques such as aerosol mass

spectrometer (AMS), can be input into PMF. This can provide an additional way to quantify contributions from mobile, biogenic and secondary sources to organic aerosol. It would be of to compare how CMB-GC with BBSPs compare with AMS-based estimates of these source categories. In addition, the AMS-based SA results could also be used as an input into the ensemble-averaging for development of new BBSPs. These newer BBSPs can be compared against the older BBSPs.

Similar ensemble methods have been applied at receptor locations in St, Louis, MO and Dallas, TX; however, these studies did not use a Bayesian formulation for the ensemble. It would be of interest to compare of SA results in St. Louis, Dallas and Atlanta. In addition, major reasons for differences in SA results are the source profiles/factors that are input into the SA model. It would also be of interest to assess the differences in regional source profiles derived by the ensemble method. This could help shed light on regional variability of emissions.

One aspect of the ensemble method developed in this dissertation is a framework for estimating uncertainties in chemical transport models (CTMs). In this work, 32 source categories used in the CTM were binned to match the nine source categories used in CMB. CTM uncertainties were calculated for these nine binned source categories and used in both the standard and Bayesian ensemble. Propagation of errors can used to estimate uncertainties of the 32 CTM source categories. These estimated uncertainties can then be compared with other efforts to estimate CTM uncertainties. These include efforts to incorporate receptor models within the chemical transport model CTM framework, as well as efforts to use interpolated results CTM output for SA at receptor locations that do not have $PM_{2.5}$ speciation data. This work would fall into a broader category of comparison and evaluation of SA methods that would guide air quality policy development.