IMPACT AND SENSITIVITY ANALYSES OF ENERGY SECTOR EMISSIONS: AIR QUALITY MODELING OF THE PJM REGION

By

CAROLINE M. FARKAS

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Atmospheric Sciences

Written under the direction of

Ann Marie Carlton

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

MAY 2016

ABSTRACT OF THE DISSERTATION

Impact and Sensitivity Analyses of Energy Sector Emissions:

Air Quality Modeling of the PJM Region

by CAROLINE M. FARKAS

Dissertation Director:

Ann Marie Carlton

One in eight deaths globally is due to air pollution. Exposure to high concentrations of atmospheric fine particulate matter ($PM_{2.5}$) has negative health consequences. Air quality models, such as the Community Multiscale Air Quality (CMAQ) model, are employed to evaluate effectiveness of air pollution abatement strategies partly designed to minimize $PM_{2.5}$ exposure and protect human health. Energy production and consumption is the largest controllable source sector contributing to ambient $PM_{2.5}$ mass. The highest electricity sector (energy subdivision) emissions occur on hot, stagnant summer days, when energy demand is highest and the atmosphere is most conducive to photochemical production and $PM_{2.5}$ accumulation. Electricity generation is positively correlated with peak $PM_{2.5}$ concentrations. CMAQ consistently underpredicts these peak values. Accurate prediction of peak pollutant concentrations is critical to develop strategies that protect human health.

This dissertation works to reduce underprediction of peak $PM_{2.5}$ concentrations from an energy sector and heat wave event perspective in the Northeast U.S., where PJM Interconnection governs the electricity transmission for 61 million people. Temporal

representation of electricity sector emissions is improved in CMAQ during a heat wave, and episodic increases in peak $PM_{2.5}$ at the surface and aloft are predicted. PJM EGU emissions, especially sulfate, impact not only the PJM region, but also outlying areas. Monitored and controlled peaking units, EGUs used during highest electricity demand, contribute up to 87% of maximum hourly $PM_{2.5}$ concentrations. Urban areas experience the highest potential exposure (calculated as population-weighted concentrations (PWCs)) from peaking unit emissions, regardless of the location of predicted peak ambient concentrations. Peaking units contribute substantially to exposure potential on the worse air quality days, but are historically exempt from Federal air quality rules. Eight sensitivity experiments indicate CMAQ-predicted $PM_{2.5}$ PWCs are most sensitive to uncertainty in onroad primary organic carbon emissions, while ambient $PM_{2.5}$ concentrations are most sensitive to planetary boundary layer height. Model development strategies optimized to protect health may look different than traditional evaluation-focused strategies optimized to match annual averages in measured $PM_{2.5}$ mass. This dissertation provides issues to consider for prioritizing model development to address peak air quality events that drive non-attainment and threaten human health.

ACKNOWLEDGEMENTS

I first would like to thank my advisor, Dr. Ann Marie Carlton, for her continued support and invaluable guidance, and for being a wonderful mentor on all skills, professionally and personally, needed to become a successful scientist.

To the other members of my dissertation committee: Dr. Frank Felder showed incredible patience and provided immeasurable insight on all of my projects. Dr. Barbara Turpin was a supportive "grand-advisor" and her encouragement to think outside my comfort zone during joint group meetings is an invaluable skill to have learned. Finally, thank you to Dr. Enrique Curchitser for his encouragement and wisdom, especially in large-scale dynamics and on my final project.

My research was generously funded by the SEBS Excellence Fellowship, teaching assistantships from the Departments of Environmental Sciences and Chemistry, U.S. EPA Science To Achieve Results (STAR) Program, Grant #83504101, and *Climate to Humans: A Study of Urbanized Coastal Environments, Their Economics and Vulnerability to Climate Change*, National Science Foundation, OCE1049088.

Thank you to Dr. Tony Broccoli and Dr. Steven Decker for choosing me as a TA and giving me an outlet for my love of weather while researching air quality. Thank you to Dr. Mark Miller, Dr. Benjamin Lintner, Dr. Alan Robock, and the entire faculty of the Environmental Sciences Department for their expertise in the classroom and their helpful advice. I would also like to thank Melissa Arnesen, Bryan Raney, and the DES staff for helping me with every logistic and for keeping the program running smoothly.

Whether as a sounding board, a number cruncher, or just support while we watched the model produce error after error, Michael Moeller deserves a large amount of

credit for his continued work on these projects. I will be eternally grateful to my lab group, Dr. Neha Sareen, Khoi Nguyen, and Michael Moeller for their support and friendship. I have never been a member of a more fun group. I also thank Kirk Baker for all of his helpful contributions and advice through every step of this dissertation.

I thank the Adiabats softball team for allowing me to be your captain for 3 years. Thank you to Nicholas Pollock, Megan Francis, Dr. Bradford Greening, Arielle Alpert, Jacqueline McSweeney, Evan Buckland, Natalie Lemanski, Greg Seroka, and Dr. Ilissa Seroka for being not only outstanding colleagues but also true friends. To Dr. Amanda Luther, Jennifer Therkorn, Max Pike, Juan Perez Arango, Sara Duncan, Jenny Kafka, Matt Drews, Dr. Jessie Sagona, Zhongyu Kuang and all the DES graduate students – it is a special honor to have been apart of such a supportive graduate student community.

My family and friends have always provided me with unwavering support and love throughout my life. Dr. Christine Farkas not only gave me a goal to strive for but more importantly, she always reminds me not to take the little things so seriously and to find humor in even the worst situations. Thank you to Ryan Romanowski for seeing past the dissertation stress and being a pillar of strength and support in my craziest times; and to the entire Romanowski family for giving me family in New Jersey. Basia Koszalka is a source of infinite wisdom and kindness and she has been a large force behind my success. Thank you to my grandparents and the Stiles and Farkas families for giving me cheerleaders all around the country. Last, but absolutely not least, I owe everything to my parents, Michael and Judy Farkas, to whom I dedicate this dissertation. Thank you for encouraging me to always be curious and to chase my dreams, no matter how far away they may seem.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. CHAPTER 1

INTRODUCTION

1.1. Motivation

One in every 8 deaths globally is linked to air pollution exposure, with 3.7 million deaths annually attributable to ambient concentrations.[1] Exposure to fine particulate matter ($PM_{2.5}$), even over hourly time scales, contributes substantially to the impacts[2] and has most recently been linked to an increased risk of autism spectrum disorders.[3, 4] Ground level ozone is also dangerous to human health, negatively affecting the respiratory system especially in those who suffer from asthma, emphysema, and chronic obstructive pulmonary disease (COPD).[5] It is estimated that in the U.S., PM-related health impacts (including increased cancer risk,[6, 7] cardiovascular damage,[8] pulmonary disorders,[9] premature death[10, 11] and autism[3]) are a factor of 30 more than ozone-related impacts, measured in deaths and life years lost.[12] In the U.S, energy consumption and production is the largest controllable source of emissions that impact fine particle mass concentrations through direct emissions of $PM_{2.5}$ and of precursor gases that react to form particles in the atmosphere (e.g., sulfur dioxide (SO_2), nitrogen oxides (NO_X) → sulfate (SO_4), nitrate (NO_3), respectively).[13]

Aside from the health effects of $PM_{2.5}$ mass, elemental carbon (EC) exhibits a warming effect on Earth through absorption of sunlight and reduction of the albedo of ice and snow. Particulate sulfate negatively effects ecosystems and damages agricultural

crops through dry deposition and acid rain[14] and induces local cooling effects, a near-term climate forcer (NTCF), opposite to EC climate impacts. Over the Central and Eastern U.S., a decrease in surface temperatures, especially during heat waves, can be attributed to direct and indirect forcings of anthropogenic aerosols, largely due to electricity generation.[15, 16]

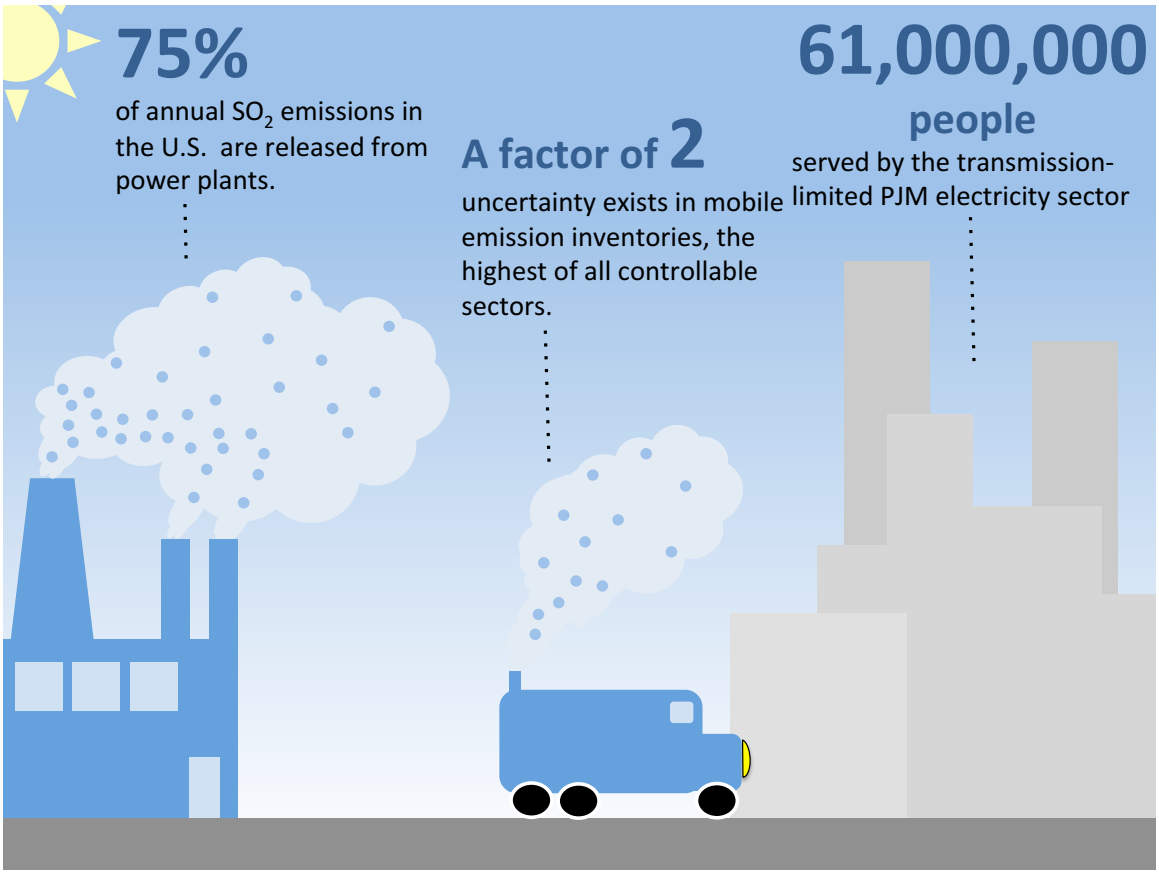

Figure 1-1: Overview of dissertation motivation. Data from (left to right) U.S. EPA, Hanna et al. (2001), and PJM Interconnection.

In the majority of the U.S., peak $PM_{2.5}$ mass concentrations are often observed during heat wave, stagnation events,[17-20] which are increasing in frequency and intensity in the Northeastern U.S. in the changing global climate.[20-26] These events frequently

correspond to periods of high electricity demand[27, 28] and increased energy-sector emissions. Electricity generation is positively associated with measured hourly peak concentrations of $PM_{2.5}$ and ozone in the Northeast U.S. region served by the regional transport organization (RTO), PJM Interconnection (Figure 1-2). EGU emissions of primary PM and the formation of secondary particulates through oxidation of SO_2 and NO_x emissions contribute to total $PM_{2.5}$ mass concentrations. Emissions of NOx combine with volatile organic compounds (VOCs) in the presence of sunlight to form ozone. EGU emissions of precursor gases and primary PM impact climate through the release or formation of greenhouse gases (GHGs) (e.g., N_2O, CO, CH_4, O_3) and NTCFs, such as particulate sulfate and black carbon.[15, 29, 30]

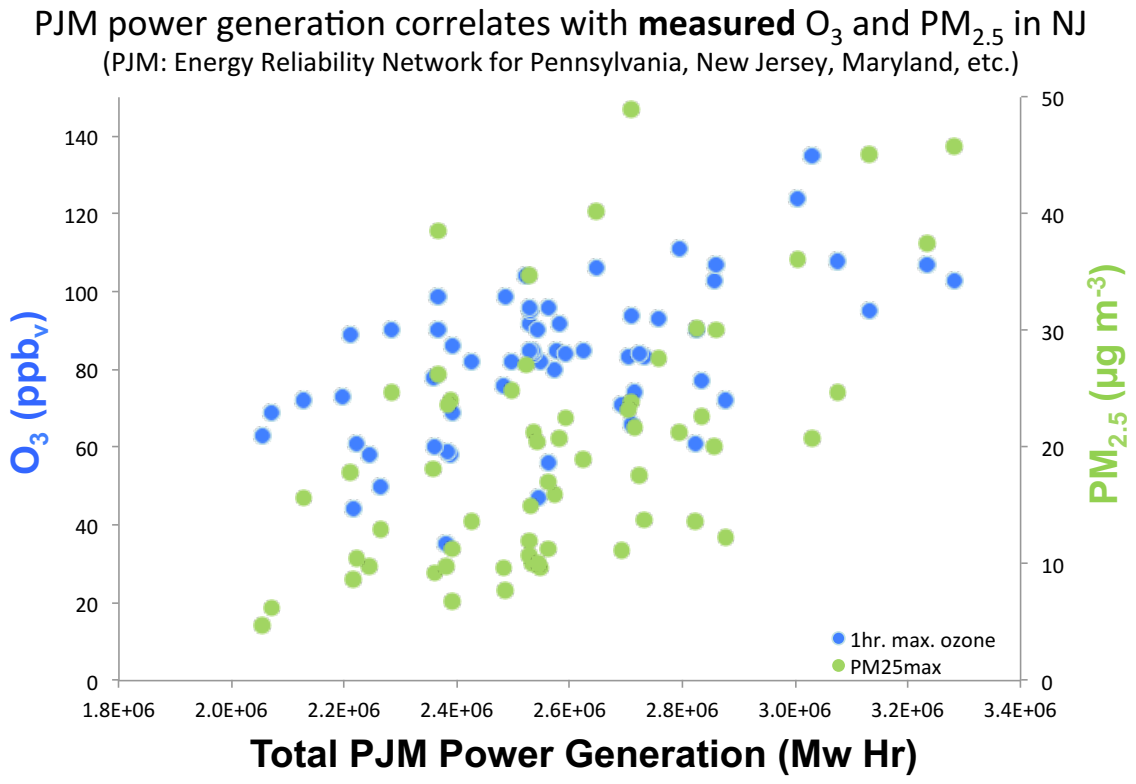

Figure 1-2: Total power generation in the PJM reliability network with measured ambient 1-hour maximum ozone (blue) and max daily $PM_{2.5}$ (green) in New Jersey.

Federal air quality initiatives and policies[31, 32] often target areas that are in non-attainment of National Ambient Air Quality Standards (NAAQS).[33] States develop State Implementation Plans (SIPs) to attain the NAAQS to ultimately safeguard human health. Because of the regional nature of air pollution due to chemical transformation during transport, individual states are often not solely responsible for local air quality. Recent Federal policies, such as the Cross-State Air Pollution Rule (CSAPR), employ a regional approach and require states to reduce emissions from electric generation units (EGUs) that contribute at least 1% of the $PM_{2.5}$ or ozone NAAQS in a downwind non-attainment area.[34] This rule is particularly important where the largest regional transmission organization in the nation, PJM, supplies electrical power to approximately 61 million people in the densely populated Northeast and mid-Atlantic regions.[35]

Peak concentrations drive non-attainment and potential adverse societal impacts are large. This is particularly true in densely populated urban areas and during the summer months when the atmosphere is primed for photochemical reactions that form $PM_{2.5}$ and ozone and people are engaged in outdoor activities. Accurate prediction of regional peak air quality events is essential to develop effective cross-state strategies that protect human health, but air quality models consistently underpredict peak $PM_{2.5}$ mass.[36, 37] To better predict peak air quality events, it is important to identify possible sources of inaccuracy within photochemical transport models that are used to develop air quality management strategies. A key knowledge gap exists in understanding the reason behind the routine underprediction of peak ozone and $PM_{2.5}$ by the U.S. Environmental Protection Agency's (EPA) Community Multiscale Air Quality (CMAQ) model[36, 37] used

to evaluate state implementation plan (SIP) effectiveness and develop regional strategies, such as those developed for CSAPR.

1.2. Background

1.2.1. Electricity Generation Emissions

The U.S. energy sector is the largest anthropogenic source of annual $PM_{2.5}$, SO_2, and NO_x emissions (Figure 1-3).[38] In particular, in 2008 electricity generation was the largest single source of primary $PM_{2.5}$ and was responsible for ~20% of annual PJM NO_x emissions.[38] Using the Comprehensive Air Quality Model with Extensions (CAMx) photochemical model, Fann et al.[39] found that in 2005, EGU point source emissions were the largest contributor to annually averaged $PM_{2.5}$ concentrations in the U.S., while the largest contributor to daily 8-hour maximum ozone concentrations in 2005 was the onroad mobile sector. EGU point sources were the 3rd highest contributor to ozone of the seven source sector categories investigated. Focus on annual average concentrations of $PM_{2.5}$ and daily 8-hour maximum concentrations of ozone occurs to match the time-frame definitions used to assess NAAQS attainment for these pollutants, however model evaluation focused on peak concentrations is warranted and necessary.

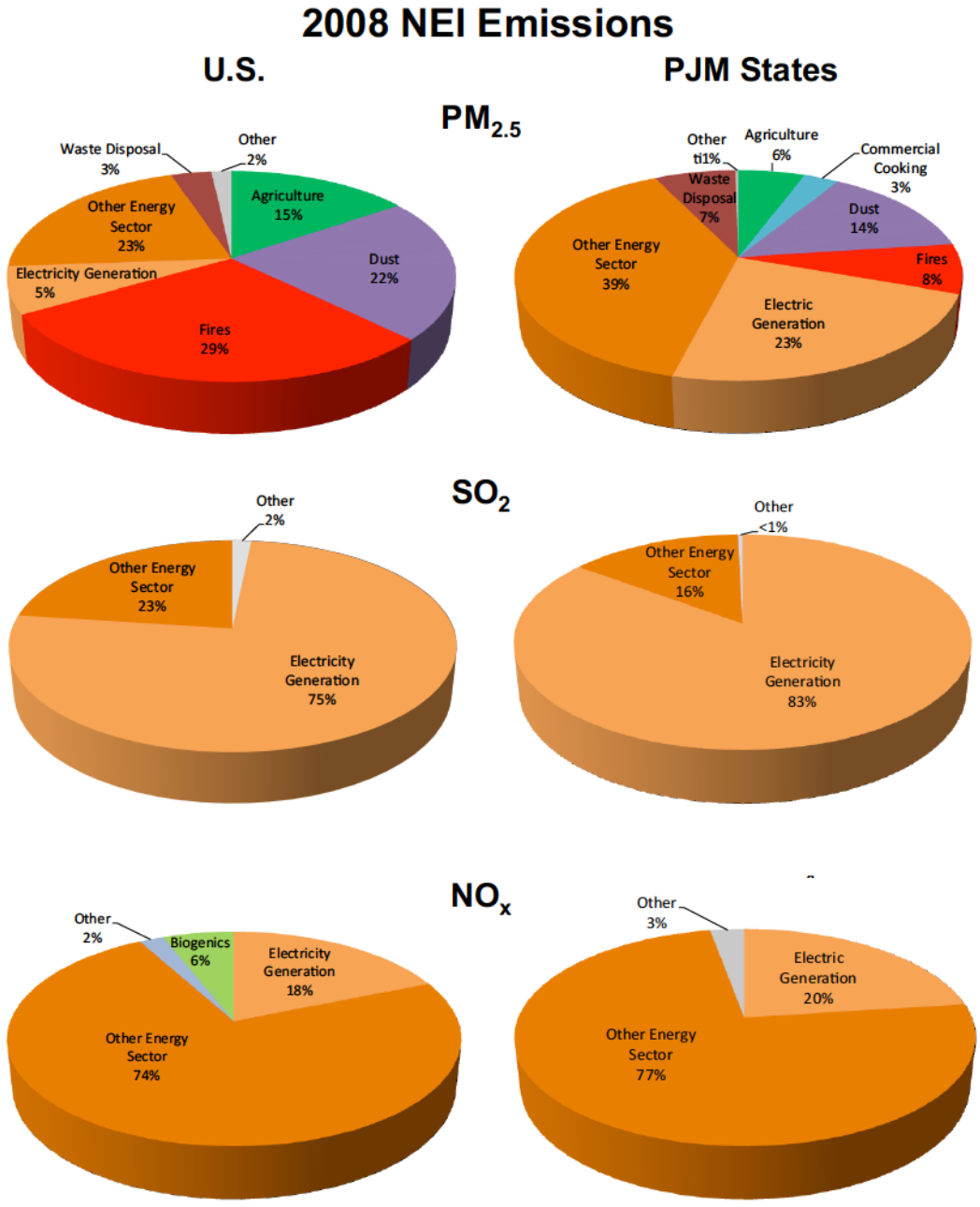

Figure 1-3: Sources of $PM_{2.5}$, SO_2, and NO_x emissions in the U.S. (left) and PJM states (right) as detailed in the 2008 NEI. "Other energy sector" includes emissions from mobile sources, industrial processes, and fuel combustion.

EGU emissions contribute to regional ambient concentrations of pollutants including the formation of secondary atmospheric pollutants. Primary emissions of $PM_{2.5}$ add to total ambient $PM_{2.5}$ mass. Emitted precursor gases (e.g. SO_2 and NO_x) can experience gas-phase photochemical or aqueous-phase reactions (i.e., within cloud and fog droplets) to form secondary $PM_{2.5}$ (e.g. sulfate, nitrate). Further, in the presence of sunlight these precursor gases react in the atmosphere to form tropospheric ozone.

Federal and state regulations[31, 32, 40] have been enacted to reduce emissions from the energy sector (including power plant and motor vehicle emissions) in a changing climate and NAAQS for $PM_{2.5}$ (daily: 35 $\mu g/m^3$; annually: 12 $\mu g/m^3$)[33] and ozone (8-hour: 0.075 ppm)[33] are in place. After the implementation of the Acid Rain Program and the NO_x Budget Program, Frost et al.[41] note a 49% decrease in summertime power plant NO_x emissions between 1999 and 2003 from 53 power plants in the Eastern U.S. He et al.[42] also observed overall decreases in summertime NO_x emissions after 2002 in 5 Eastern U.S. states, however they found a 2.5-4% increase in NO_x emissions per degree increase in temperature (°C) between 1997 and 2011. With expected emission decreases due to the enactment of CSAPR, new EGU and boiler maximum achievable technology (MACT) standards, and the NO_x SIP call, a ~60% decrease in the contribution of the EGU sector to total annual $PM_{2.5}$ mass concentrations is predicted for 2016 relative to 2005 when using the CAMx air quality model.[39] However, climate change has the potential to exacerbate peak concentrations of ozone and particulates.[43]

With global temperatures and U.S. heat wave stagnation events increasing and the subsequent increased likelihood of peak pollution events, accurate representation of air

pollution known to induce adverse health and welfare effects with tools that accurately predict peak concentrations in a changing climate is imperative.

1.2.2. Numerical Models

The work described in this dissertation utilizes a combination of four numerical models to simulate the emissions from the energy sector and determine the location and impact of the resulting pollutants. Figure 1-4 illustrates the connection and usage of the models described here.

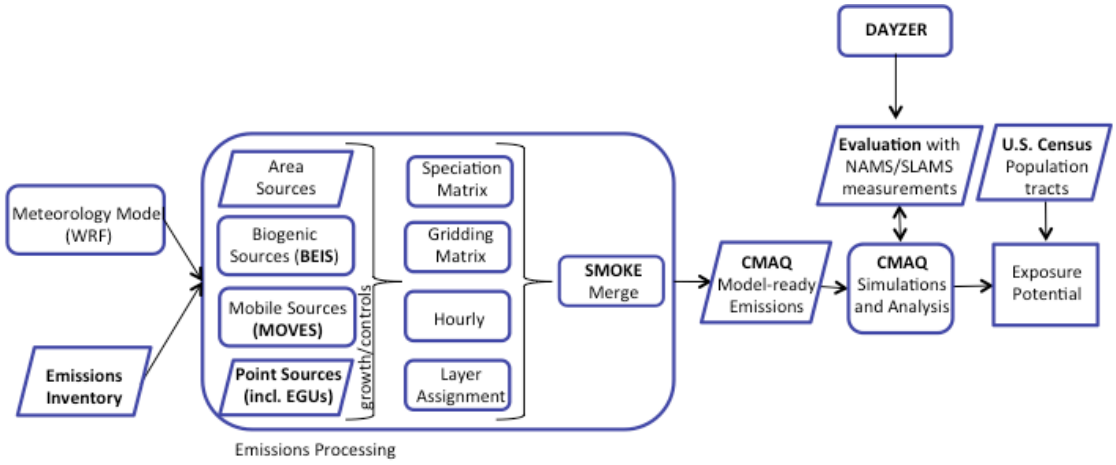

Figure 1-4: Diagram of project models (rectangle with rounded edges), datasets (rhombus), and calculations (rectangle). The changes made to the emission inventory in all three projects are processed in SMOKE and then simulated in CMAQ.

1.2.2.1. Sparse Matrix Operator Kernel Emissions (SMOKE) Preprocessor

The SMOKE model employs the EPA's National Emissions Inventory (NEI), a database of criteria and hazardous air emissions reported by state, local, and tribal agencies, to designate the type, quantity and location of anthropogenic air pollutants in the model domain. The emissions are reported by source sector at varying frequencies (annually, daily, hourly).[44] SMOKE converts emission data from the NEI into the temporal and spatial resolution needed by an air quality model. Emissions are chemically speciated and temporally and spatially allocated into hourly time steps in grid cells that comprise a gridded dataset for input into the air quality model.[45] Temporal profiles are designed to represent typical, repeatable patterns, not peak events, and are assigned to each pollution source based on its source classification code (SCC). In some cases day- or hour-specific emission data from the study year are reported and can be used in place of annual emissions, improving the temporal profile.

1.2.2.2. Community Multiscale Air Quality (CMAQ) Model

CMAQ is a three-dimensional photochemical model that simulates the relationships between meteorology, emissions, and chemical transport to produce 3D-gridded concentrations of atmospheric pollutants to understand air quality simulations on multiple scales.[46] The chemistry-transport portion of the model includes processes such as gas- and aqueous-phase reactions and deposition, photolytic rate computation, and

horizontal advection and vertical diffusion.[46] The meteorology is provided by a mesoscale numerical weather model, the Weather Research and Forecasting (WRF) model Advanced Research WRF core.[47] CMAQ is used by EPA, air quality planning agencies, and states to assess which state or regional plans will attain the NAAQS, predict effective national regulations and policies, and to produce air quality forecasts for the National Weather Service.[48]

1.2.2.3. Day-Ahead Locational Market Clearing Prices Analyzer (DAYZER)

DAYZER[49] simulates the operation of select electricity markets to forecast the hourly location market-clearing prices (LMPs), congestion costs, and emissions using the most recently available data on fuel prices, demand forecast, unit and transmission line outages, and emission permit costs. The electric load is forecasted based on historical load shape and forecasted peak demands, fuel prices from the New York Mercantile Exchange (NYMEX), random outage using the Bernoulli probability model, maintenance scheduling, and imports/exports. DAYZER is used in this work to compare electricity generation (actual values are proprietary) with observed pollutant concentrations.

1.2.2.4. Measurement Data Networks

National air quality is monitored through air measurement site networks such as the State and Local Air Monitoring Stations (SLAMS) and subsets including National Air Monitoring Stations (NAMS), and Photochemical Assessment Monitoring Stations

(PAMS).[50, 51] Criteria air pollutants (carbon monoxide, lead, NO_x, ozone, PM, and SO_2) are measured on varying time scales (from once every six days to hourly) depending on site requirements. The Chemical Speciation Network (CSN)[52] and Interagency Monitoring of Protected Visual Environments (IMPROVE)[53] are SLAMS subsets implemented specifically to measure species that contribute to total $PM_{2.5}$ mass. Data for these networks is available for download[54] and is used in this dissertation for comparison to model predictions.

1.3. Dissertation Overview

This work aims to improve the representation of anthropogenic emissions of compounds from EGU point sources for the PJM electricity sector and identify possible causes of the disagreement between CMAQ-predicted and measured values of $PM_{2.5}$ and ozone under peak conditions. Chapter 2 hypothesizes that inaccurate temporalization of a portion of EGU emissions in the PJM region contributes to underprediction of peak $PM_{2.5}$ concentrations. Continuous emission monitors (CEMS) placed on EGUs in compliance with Federal regulations (e.g., NO_x SIP call, Acid Rain Program) record hourly heat input and emission rates for some pollutants. Due to a source identification issue, SMOKE does not utilize all available hourly CEM data that could more accurately represent EGU emissions during peak air quality episodes and instead replaces actual measured data with default temporal profiles. Michael Moeller and I matched missing hourly CEM data from 267 EGUs (approximately half of unmatched CEMs in PJM) with their corresponding annual NEI records in SMOKE. With meteorology, biogenic and mobile emission inputs

developed and provided by Kirk Baker at the U.S. EPA, I employed CMAQ to simulate the air quality impacts of the additional CEM data. I found an increase in surface maximum hourly ambient $PM_{2.5}$ mass of up to 7.4 $\mu g/m^3$ with the addition of data from 267 CEMs with the majority of the maxima due to sulfate both at the surface and aloft. This is the first time the air quality impacts of missing CEM hourly data in the CMAQ model have been quantified.

Chapter 3 builds on the findings of Chapter 2 and investigates the impact of less frequently used EGUs known as peaking units to peak concentrations of $PM_{2.5}$ and ozone in population centers. Using the EPA peaking unit definition, Michael Moeller and I identified 544 EGUs in the PJM region as peaking units. I employed CMAQ with the same meteorology, biogenic and mobile emissions from Kirk Baker in Chapter 2 to simulate the air quality effects of peaking units. I found that peaking units contribute up to 110 $\mu g/m^3$ of maximum hourly $PM_{2.5}$ mass, but their impacts are localized. Primary unspeciated $PM_{2.5}$ contributes the most to these maxima in the PJM region. This is the first time air quality impacts of peaking units have been quantified and compared to those of EGUs from its own RTO and other surrounding RTOs.

Using population data I downloaded from the U.S. Census, a population-weighting formula from Carlton et al. (2010) and python code written by Barron Henderson, I calculated population-weighted concentrations from the CMAQ predictions to analyze the locations with the highest potential for human exposure to air pollution for peaking units. Highly populated urban areas both in and upwind from the PJM region have the highest exposure potential due to PJM EGUs (peaking and non-peaking). This is

the first time peaking unit air quality impacts have been population-weighted to analyze for potential exposure impacts.

Chapter 4 propagates uncertainties in controllable energy sector emissions and meteorology through the air quality prediction system. This chapter identifies differences in the predominant drivers of uncertainty in ambient predictions versus those for potential exposure estimates. I simulated perturbations of planetary boundary layer (PBL) height and onroad mobile source emissions of NOx (NO+NO2), primary organic carbon (POC), and unspeciated $PM_{2.5}$ (PMother) in SMOKE and CMAQ. Using the equation and python code from Chapter 3, population weights are applied to the predicted concentrations to determine which variables have the largest impact on potential exposure. I found that while CMAQ-predicted ambient concentrations of $PM_{2.5}$ mass were most sensitive (of the four studied variables) to PBL height perturbations, the majority of the impacts occurred over the ocean. Population-weighted $PM_{2.5}$ mass concentrations were most sensitive to fluctuations of onroad POC. To my knowledge, this is the first time population weights have been applied to pollutant concentration fluctuations due to uncertainty in an air quality model. Finally, Chapter 5 concludes this dissertation and proposes future directions.

1.4. References

1. World Health Organization. Air Pollution Estimates. http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf?ua=1 (August 30, 2015).

2. Robert D. Brook, J. R. B., Bruce Urch, Renaud Vincent, Sanjay Rajagopalan, Frances Silverman, Inhalation of Fine Particulate Air Pollution and Ozone Causes Acute

Arterial Vasoconstriction in Healthy Adults. *Circulation* **2002,** *105*, (13), 1534-1536.

3. Raz, R.; Roberts, A. L.; Lyall, K.; Hart, J. E.; Just, A. C.; Laden, F.; Weisskopf, M. G., Autism Spectrum Disorder and Particulate Matter Air Pollution before, during, and after Pregnancy: A Nested Case–Control Analysis within the Nurses' Health Study II Cohort. *Environ. Health Persp.* **2014**, 123, (3), 264-270.

4. Talbott, E. O.; Arena, V. C.; Rager, J. R.; Clougherty, J. E.; Michanowicz, D. R.; Sharma, R. K.; Stacy, S. L., Fine particulate matter and the risk of autism spectrum disorder. *Environ. Res.* **2015,** *140*, 414-420.

5. U.S. EPA, Ground Level Ozone Health Effects. http://www.epa.gov/ozonepollution/health.html (August 30, 2015).

6. Turner, M. C.; Krewski, D.; Pope, C. A.; Chen, Y.; Gapstur, S. M.; Thun, M. J., Long-term Ambient Fine Particulate Matter Air Pollution and Lung Cancer in a Large Cohort of Never-Smokers. *Amer. J. Resp. Crit. Care* **2011,** *184*, (12), 1374-1381.

7. Pope, C. A.; Burnett, R. T.; Turner, M. C.; Cohen, A.; Krewski, D.; Jerrett, M.; Gapstur, S. M.; Thun, M. J., Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure-Response Relationships. *Environ. Health Persp.* **2011,** *119*, (11), 1616-1621.

8. Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y. L.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.; Smith, S. C.; Whitsel, L.; Kaufman, J. D.; Epidemiol, A. H. A. C.; Dis, C. K. C.; Metab, C. N. P. A., Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association. *Circulation* **2010,** *121*, (21), 2331-2378.

9. Burnett, R. T.; Pope, C. A.; Ezzati, M.; Olives, C.; Lim, S. S.; Mehta, S.; Shin, H. H.; Singh, G.; Hubbell, B.; Brauer, M.; Anderson, H. R.; Smith, K. R.; Balmes, J. R.; Bruce, N. G.; Kan, H. D.; Laden, F.; Pruss-Ustun, A.; Michelle, C. T.; Gapstur, S. M.; Diver, W. R.; Cohen, A., An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environ. Health Persp.* **2014,** *122*, (4), 397-403.

10. Fang, Y.; Naik, V.; Horowitz, L. W.; Mauzerall, D. L., Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present. *Atmos. Chem. Phys.* **2013,** *13*, (3), 1377-1394.

11. Lepeule, J.; Laden, F.; Dockery, D.; Schwartz, J., Chronic Exposure to Fine Particles and Mortality: An Extended Follow-up of the Harvard Six Cities Study from 1974 to 2009. *Environ. Health Persp.* **2012,** *120*, (7), 965-970.

12. Fann, N.; Lamson, A. D.; Anenberg, S. C.; Wesson, K.; Risley, D.; Hubbell, B. J., Estimating the National Public Health Burden Associated with Exposure to Ambient $PM_{2.5}$ and Ozone. *Risk Anal.* **2012,** *32*, (1), 81-95.

13. U.S. EPA. The 2011 National Emissions Inventory v.1. http://www.epa.gov/ttnchie1/net/2011inventory.html (August 30, 2015).

14. U.S. EPA. *National Air Quality and Emissions Trends Report, Chapter 7*; 1999; pp 115-126.

15. Leibensperger, E. M.; Mickley, L. J.; Jacob, D. J.; Chen, W. T.; Seinfeld, J. H.; Nenes, A.; Adams, P. J.; Streets, D. G.; Kumar, N.; Rind, D., Climatic effects of 1950-2050 changes in US anthropogenic aerosols - Part 1: Aerosol trends and radiative forcing. *Atmos. Chem. Phys.* **2012,** *12*, (7), 3333-3348.

16. Leibensperger, E. M.; Mickley, L. J.; Jacob, D. J.; Chen, W. T.; Seinfeld, J. H.; Nenes, A.; Adams, P. J.; Streets, D. G.; Kumar, N.; Rind, D., Climatic effects of 1950-2050 changes in US anthropogenic aerosols - Part 2: Climate response. *Atmos. Chem. Phys.* **2012,** *12*, (7), 3349-3362.

17. Stagnation When Heat Waves Exist - Summer, 1950-2007. In NOAA/NCDC, Ed. United States Global Change Research Program: 2009.

18. Tai, A. P. K.; Mickley, L. J.; Jacob, D. J.; Leibensperger, E. M.; Zhang, L.; Fisher, J. A.; Pye, H. O. T., Meteorological modes of variability for fine particulate matter (PM2.5) air quality in the United States: implications for $PM_{2.5}$ sensitivity to climate change. *Atmos. Chem. Phys.* **2012,** *12*, (6), 3131-3145.

19. Mahmud, A.; Hixson, M.; Kleeman, M. J., Quantifying population exposure to airborne particulate matter during extreme events in California due to climate change. *Atmos. Chem. Phys.* **2012,** *12*, (16), 7453-7463.

20. Mickley, L. J.; Jacob, D. J.; Field, B. D.; Rind, D., Effects of future climate change on regional air pollution episodes in the United States. *Geophys. Res. Lett.* **2004,** *31*, (24).

21. Mwaniki, G. R.; Rosenkrance, C.; Wallace, H. W.; Jobson, B. T.; Erickson, M. H.; Lamb, B. K.; Hardy, R. J.; Zalakeviciute, R.; VanReken, T. M., Factors contributing to elevated concentrations of $PM_{2.5}$ during wintertime near Boise, Idaho. *Atmos. Pol. Res.* **2014,** *5*, 96-103.

22. Meehl, G.A.; Tebaldi, C., More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science* **2004,** *305*, (5686), 994-997.

23. Jacob, D. J.; Winner, D. A., Effect of climate change on air quality. *Atmos. Environ.* **2009,** *43*, (1), 51-63.

24. Horton, D. E.; Harshvardhan; Diffenbaugh, N. S., Response of air stagnation frequency to anthropogenically enhanced radiative forcing. *Environ. Res. Lett.* **2012,** *7*, (4).

25. National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Air Stagnation Index. http://www.ncdc.noaa.gov/societal-impacts/air-stagnation/ (August 30, 2015).

26. Trail, M.; Tsimpidi, A. P.; Liu, P.; Tsigaridis, K.; Hu, Y.; Nenes, A.; Russell, A. G., Downscaling a global climate model to simulate climate change over the US and the implication on regional and urban air quality. *Geosci. Model Dev.* **2013,** *6*, (5), 1429-1445.

27. Amato, A. D. R., M.; Kirshen, P.; Horwitz, J., Regional Energy Demand Responses To Climate Change: Methodology And Application To The Commonwealth Of Massachusetts. *Climatic Change* **2005,** *71*, (1-2), 175-201.

28. Miller, N. L.; Hayhoe, K.; Jin, J.; Auffhammer, M., Climate, Extreme Heat, and Electricity Demand in California. *J. Appl. Meteorol. and Clim.* **2008,** *47*, (6), 1834-1844.

29. Twomey, S., The Influence of Pollution on the Shortwave Albedo of Clouds. *J. Atmos. Sci.* **1977,** *34*, (7), 1149-1152.

30. Jacobson, M. Z., Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols. *Nature* **2001,** *409*, (6821), 695-697.

31. U.S. EPA. Acid Rain Program. http://www.epa.gov/airmarkets/progsregs/arp/basic.html (August 30, 2015).

32. U.S. EPA. NO_x Budget Trading Program. http://www.epa.gov/airmarkets/programs/nox/ (August 30, 2015).

33. U.S. EPA. National Ambient Air Quality Standards (NAAQS). http://www.epa.gov/air/criteria.html (August 30, 2015).

34. U.S. EPA, *EPA-HQ-OAR-2009-0491: Federal Implementation Plans to Reduce Interstate Transport of Fine Particulate Matter and Ozone in 27 States; Correction of SIP Approvals for 22 States*; 2011.

35. PJM Who We Are. http://www.pjm.com/about-pjm/who-we-are.aspx (August 30, 2015).

36. Foley, K. M.; Roselle, S. J.; Appel, K. W.; Bhave, P. V.; Pleim, J. E.; Otte, T. L.; Mathur, R.; Sarwar, G.; Young, J. O.; Gilliam, R. C.; Nolte, C. G.; Kelly, J. T.; Gilliland, A. B.; Bash, J. O., Incremental testing of the Community Multiscale Air Quality (CMAQ) modeling system version 4.7. *Geosci. Model Dev.* **2010,** *3*, (1), 205-226.

37. Zhang, H.; Chen, G.; Hu, J.; Chen, S.-H.; Wiedinmyer, C.; Kleeman, M.; Ying, Q., Evaluation of a seven-year air quality simulation using the Weather Research and Forecasting (WRF)/Community Multiscale Air Quality (CMAQ) models in the eastern United States. *Sci Total Environ.* **2014,** *473–474*, (0), 275-285.

38. U.S. EPA. The 2008 National Emissions Inventory. http://www.epa.gov/ttnchie1/net/2008inventory.html (September 2, 2015).

39. Fann, N.; Fulcher, C. M.; Baker, K., The Recent and Future Health Burden of Air Pollution Apportioned Across U.S. Sectors. *Environ. Sci. Tech.* **2013,** *47*, (8), 3580-3589.

40. U.S. EPA. Tier 3 Vehicle Emissions and Fuel Standards Program. http://www3.epa.gov/otaq/tier3.htm (October 8, 2015).

41. Frost, G. J.; McKeen, S. A.; Trainer, M.; Ryerson, T. B.; Neuman, J. A.; Roberts, J. M.; Swanson, A.; Holloway, J. S.; Sueper, D. T.; Fortin, T.; Parrish, D. D.; Fehsenfeld, F. C.; Flocke, F.; Peckham, S. E.; Grell, G. A.; Kowal, D.; Cartwright, J.; Auerbach, N.; Habermann, T., Effects of changing power plant NOx emissions on ozone in the eastern United States: Proof of concept. *J. Geophys. Res.-Atmos.* **2006,** *111*, (D12).

42. He, H.; Hembeck, L.; Hosley, K. M.; Canty, T. P.; Salawitch, R. J.; Dickerson, R. R., High ozone concentrations on hot days: The role of electric power demand and NOx emissions. *Geophys. Res Lett.* **2013,** *40*, (19), 5291–5294.

43. U.S. EPA. Our Nation's Air - Status and Trends through 2008. http://www3.epa.gov/airquality/airtrends/2010 (August 30, 2015).

44. U.S. EPA. The 2008 National Emissions Inventory. http://www.epa.gov/ttnchie1/net/2008inventory.html (August 30, 2015).

45. The Institute for the Environment -The University of North Carolina Chapel Hill. SMOKE v3.5 User's Manual. http://www.cmascenter.org/smoke/documentation/3.5/manual_smokev35.pdf (August 30, 2015).

46. 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*, (2), 51-77.

47. Skamarock, W. C. K., J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. *A description of the Advanced Research WRF version 3. NCAR Technical Note NCAR/TN-475+STR*; 2008.

48. U.S. EPA Community Multiscale Air Quality Model. http://www.epa.gov/amad/Research/RIA/cmaq.html (August 30, 2015).

49. Cambridge Energy Solutions. DAYZER. http://www.ces-us.com/product-dayzer.asp (August 30, 2015).

50. EPA, U.S. SLAMS Networks. http://www3.epa.gov/ttn/amtic/slams.html (August 30, 2015).

51. EPA, U. S. *SLAMS/NAMS/PAMS Network Review Guidance*; 1998.

52. EPA, U. S. Chemical Speciation Network. http://www3.epa.gov/ttnamti1/speciepg.html (August 30, 2015).

53. IMPROVE Interagency Monitoring of Protected Visual Environments. http://vista.cira.colostate.edu/improve/ (August 30, 2015).

54. EPA, U. S. Download Detailed AQS Data. http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm (August 30, 2015).

2. CHAPTER 2

TEMPORALIZATION OF PEAK ELECTRIC GENERATION PM EMISSIONS DURING HIGH ENERGY DEMAND DAYS

Material in this chapter is adapted from:

Farkas, C.M.; Moeller, M.D.; Felder, F.A.; Baker, K.A.; Rodgers, M.; Carlton, A.G., Temporalization of Peak Electric Generation PM Emissions During High Energy Demand Days. *Environ. Sci. Tech*. **2015**, 49, (7), 4696-4704.

2.1. Abstract

Underprediction of peak ambient pollution by air quality models hinders development of effective strategies to protect health and welfare. EPA's Community Multiscale Air Quality (CMAQ) model routinely underpredicts peak ozone and fine particulate matter ($PM_{2.5}$) concentrations. Temporal mis-allocation of electricity sector emissions contributes to this modeling deficiency. Hourly emissions are created for CMAQ using temporal profiles applied to annual emission totals unless a source is matched to a continuous emissions monitor (CEM) in the National Emissions Inventory (NEI). More than 53% of CEMs in the Pennsylvania-New Jersey-Maryland (PJM) electricity market and 45% nationally are unmatched in the 2008 NEI. For July 2006, a U.S. heat wave with high electricity demand, peak electric sector emissions, and elevated ambient $PM_{2.5}$ mass, we match hourly emissions for 267 CEM/NEI pairs in PJM

(approximately 49% and 12% of unmatched CEMs in PJM and nationwide) using state permits, electricity dispatch modeling and CEMs. Hourly emissions for individual facilities can differ up to 154% during the simulation when measurement data is used rather than default temporalization values. Maximum CMAQ $PM_{2.5}$ mass, sulfate, and elemental carbon predictions increase up to 83%, 103%, and 310%, at the surface and 51%, 75%, and 38% aloft (800 mb), respectively.

2.2. Introduction

Photochemical transport models are integral tools used to help develop air quality management strategies to decrease ambient pollution concentrations and minimize human exposure to harmful pollution. While Electric Generating Units (EGU) can exacerbate ambient ozone through nitrogen oxide (NO_x) emissions,[1-3] they are the largest source of primary particulate matter (PM) in the Northeast[4] and fine mode ($PM_{2.5}$) particulate sulfate in the continental U.S.[5] A variety of adverse PM-related health consequences negatively impact human health, including increased cancer risk,[6, 7] cardiovascular damage,[8] pulmonary disorders,[9] premature death.[10, 11] and autism.[12] For example, exposure to peak concentrations of $PM_{2.5}$ over just 2 hours leads to the constriction of blood vessels in arteries, which can cause acute cardiac events.[13] Fann et al. (2012) estimate that costs associated with pre-mature deaths and life years lost due to PM-related health impacts are approximately a factor of 30 more than ozone-related impacts in the U.S.[14] In the Eastern U.S., peak $PM_{2.5}$ mass concentrations are often observed during heat waves, typically characterized by a period of prolonged stagnation.[15-18] The frequency,

duration and intensity of stagnation events are projected to increase due to climate change[18-22] and observations from the National Climatic Data Center (NCDC)[23] (see Figure A-1) and independent modeling[24] indicate stagnation in the Northeastern U.S. is increasing. During heat waves and summertime stagnation in the Eastern U.S., electricity demand is highest due to increased air conditioning loads.[25-27] These high electric demand days (HEDDs) are characterized by maximum emissions from the electric generation sector[2, 3] and poor surface air quality. It follows from the above trends that peak $PM_{2.5}$ episodes over the heavily populated Eastern U.S. could also increase. Despite the critical importance of peak air quality episodes, accurate characterization by photochemical transport models of these events is elusive. In particular, the Community Multi-scale Air Quality model (CMAQ) underpredicts peak summertime $PM_{2.5}$ concentrations in the Eastern U.S.[28, 29] CMAQ is often used to assess the effectiveness of rules created to reduce the highest ambient concentrations of $PM_{2.5}$ and other pollutants in order to achieve National Ambient Air Quality Standards (NAAQS). Extreme events frequently drive non-attainment of $PM_{2.5}$ NAAQS and are not the result of average ambient concentrations, but rather exacerbated conditions like HEDDs and stagnation that result in peak pollution episodes. Therefore, it is especially important to accurately quantify emissions during these events.

In addition to health effects associated with surface level $PM_{2.5}$ mass concentrations, aloft particulate sulfate and black carbon are near-term climate forcers (NTCFs) that can impact surface and cloud albedos and have strong radiative forcing properties, in particular over the Eastern U.S.[30-32] Both the direct and indirect forcings of anthropogenic aerosols largely due to electricity generation in the U.S. peaked between

1980-1990 at more than -2.0 W m^{-2} each and are associated with a decrease in surface temperatures over the Central Eastern U.S. of 0.5-1.0 °C annually. The strongest cooling often occurs during heat waves.[32, 33] The magnitudes of these forcings have decreased to approximately -1.0 W m^{-2} as sulfur dioxide (SO_2) emissions have decreased and are projected to continue decreasing but at a slower rate.[32, 33] Future energy policy and associated environmental regulations are key uncertainties, but with respect to future climate warming, sulfate PM mass may increase with rising temperatures due to faster SO_2 oxidation[34, 35] and increased power plant emissions on hot days[2, 3]. EGUs are also a large source of primary $PM_{2.5}$, elemental (or black) carbon (EC), and NO_x emissions, which can modulate ambient PM concentrations (e.g., nitrate, secondary organic aerosol (SOA)[36]).

Continuous emission monitors (CEMs) installed on EGUs and industrial facilities with a capacity above 25 MW in response to Federal or State regulations such as the Acid Rain program[37] or NO_x State Implementation Plan (SIP) Call[1, 38] provide direct hourly measurement of NO_x and SO_2 emissions from stationary point sources. Other regulated pollutants for which CEMs are not required or are not readily available, such as PM, are reported to the NEI on an annual basis. In addition to containing emissions not measured with CEMs, the NEI contains critically important source information for air quality modeling such as location, stack height, and other stack parameters such as exit temperature and velocity. NEI emissions are gridded, speciated, and temporalized for input into photochemical models, like CMAQ, using a pre-processor such as the Sparse Matrix Operator Kernel Emissions modeling system (SMOKE).[39] CEMs are identified and paired to their corresponding NEI annual records by their Office of Regulatory

Information Systems identifiers (ORIS IDs) and corresponding boiler codes recorded in the NEI. NEI facility and CEM unit identifiers are established by different federal agencies (U.S. EPA and U.S. Energy Information Administration (EIA))[40, 41], and it is not required that plants report CEM identifiers to the NEI with their annual emission reports. EGU emissions in the NEI without both the corresponding ORIS ID and boiler code result in hourly CEM-measured data not being input to CMAQ. Instead, SMOKE assigns a source classification code (SCC)-specific *a priori* temporalization to the NEI-reported annual emission total.

In this work we investigate the degree to which EGU sector emissions are accurately represented at peak energy demand during a heat wave stagnation event in an eastern area of the U.S. served by the Pennsylvania-New Jersey-Maryland (PJM) Interconnection, the largest regional transmission organization (RTO) in the U.S.[42] We explore the hypothesis that underprediction of peak $PM_{2.5}$ concentrations in a portion of the PJM energy sector is due, in part, to inaccurate temporalization of EGU sector emissions for which measurement data is available. Further, inaccurate temporalization of EGU sector emissions may also provide a plausible explanation for overprediction of low PM concentrations.[28] Because the generic SMOKE profile is biased low for peak generation, lower than average generation during other time periods may be overpredicted. To test this hypothesis, 267 units from 91 stationary sources (EGUs and industrial facilities of varying load capacities) from 5 states within the PJM network with measured but unmatched CEM data are paired with their corresponding NEI annual records and reprocessed through the SMOKE model to remove the generic temporal profiles and replace default patterns with actual measurements. We conduct two

simulations with CMAQv5.01 to analyze the effect of the improved hourly temporalization of point source emissions. CMAQ predictions are evaluated with Interagency Monitoring of Protected Visual Environments (IMPROVE)[43] and EPA's Air Quality System (AQS) measurement data.

2.3. Methods

To simulate air quality for the Northeastern U.S., we apply CMAQ to the heat wave of 2006 using emissions from the 2008 NEI with 2006 CEM data. Estimated electricity dispatch from the Day-Ahead Locational Market Clearing Prices Analyzer (DAYZER), $PM_{2.5}$ mass from AQS and chemically speciated PM measurements from the IMPROVE network are employed to provide context for CMAQ-predicted changes in ambient concentrations of $PM_{2.5}$ mass, sulfate, and EC from the improved temporal emission assignments during the heat wave.

July and August 2006 are characterized by a widespread heat wave that produced higher-than-normal daily maximum and minimum temperatures throughout the U.S. as well as below-average precipitation.[44] The nationally averaged temperature was 77.2ºF, and precipitation averaged less than 2.6 inches for the U.S. July 2006 is the third warmest July on record nationally.[45] These conditions were accompanied by higher than normal consumption of electricity over the majority of the nation, leading to increased electricity sector emissions. Peak summertime sunlight and hot temperatures with little precipitation or wind coupled with increased emissions are conducive to photochemical reactions and poor air quality.

2.3.1. Air Quality Modeling

We apply CMAQv5.0.1 with 12 km by 12 km grid cells (109 rows by 144 columns) and 34 vertical layers up to 50 millibars. The Carbon Bond 05 chemical mechanism with toluene chemistry extensions[46] is used in the simulations to describe gas phase photochemistry and the AERO5 aerosol module which includes secondary organic aerosol formed through semi-volatile partitioning[47] and cloud chemistry.[48] The height of point source emission stacks in the five study states ranged from 0 to 320 meters (approximately CMAQ layer 1 to 6) with an average stack height of approximately 60 meters (layer 2). Initial and boundary conditions for CMAQ are extracted from a continental scale simulation that matched the time period of this application. Details related to inputs, application, and evaluation of the larger continental scale CMAQ simulation are described elsewhere.[49] The Weather Research and Forecasting model (WRF), Advanced Research WRF core version 3.1[50] is used to generate gridded meteorological data that generates emissions and drives transport in the photochemical modeling.

Two CMAQ simulations are conducted for June 28-July 31, 2006, which includes a portion of the heat wave. The first three simulation days are excluded from the analysis for model spin up to minimize the impact of initial conditions.[51] Day and hour specific WRF-estimated solar radiation and temperatures are input to the Biogenic Emission Inventory (BEIS) version 3.14 model to generate biogenic emission estimates.[52] Anthropogenic emissions originating in the United States are based on the 2008 NEI version 2.[53] Stationary point sources reporting CEM data (http://ampd.epa.gov/ampd/) are

modeled with day and hour specific emissions matching the simulation period. Onroad mobile emissions are estimated with the Motor Vehicle Emissions Simulator (MOVES) model 2010b.[54] Canadian emissions are based on a 2002 inventory and not projected to 2006.[55] All emissions are processed for input to CMAQ using SMOKE. Both CMAQ simulations include identical meteorology, chemistry and emissions, with the exception that the sensitivity study incorporates hourly emissions measurements for 267 units that employed default temporalizations in the base case simulation.

2.3.2. EGU Emissions Processing

Prior to the 2008 NEI, facility and emission information was collected under the Consolidated Emissions Reporting Rule (CERR) using the NEI Input Format (NIF), which does not require ORIS ID submission.[56] EPA promulgated the Air Emissions Reporting Requirements (AERR) in December 2008, updating the previous CERR.[57] The AERR requires state, local, and tribal agencies to submit emission inventories to the Emission Inventory System (EIS) in the Consolidated Emissions Reporting Schema (CERS) XML reporting format. ORIS IDs are stored in EIS as an alternative identifier when available, but CERS does not have an explicit reportable ORIS ID data element.[58]

To account for this deficiency, we utilize the Mid-Atlantic Regional Air Management Association (MARAMA) crosswalk[59], a document that matches the different identifiers for a facility or unit across inventories, to match 267 units from 91 stationary sources with registered but unmatched CEM data and identification numbers for five states in the Northeast: Delaware, New Jersey, Pennsylvania, Virginia, and

Maryland (Figure 2-1a). These units represent 49% of the missing matches in the PJM region, and 12% of the total missing units throughout the U.S. (Table 2-1).

To more accurately represent the actual temporal profile of peak emissions during the heat wave, 2006 CEM data is used instead of the default 2008 NEI-included CEM data. Matching the ORIS IDs/boiler codes involves manual insertion of the correct codes into the corresponding columns in the NEI Integrated Planning Model (IPM) point-source sector (ptipm) inventory. CEMs are smoke stack specific, however, a single stack can be associated with multiple units (e.g. boilers). The NEI is reported on the unit level, not on the stack level. The SMOKE utility program CEMScan[39] is used to sum annual totals of NO_x and SO_2 emissions, heat input (heat energy input in mmBTU), steam load (steam generated), and gross load (net energy MW output) from hourly data at the CEM monitoring location. When CEM and NEI ORIS ID and boiler codes match, SMOKE uses the output from CEMScan to allocate hourly emissions for NO_x and SO_2 as well as all other pollutants (e.g. PM, mercury) with an annual emission total for each point source using Equation 2-1[39]. The hourly emissions for each source (i) for each emitted pollutant (j) are calculated as:

$$Hourly\ emissions_{i,j} = annual\ factor_i \times annual\ emissions_{i,j} \quad \text{[Equation 2-1]}$$

$$\text{where, } Annual\ Factor_i = \frac{hourly\ heat\ input\ for\ ORIS/boiler}{annual\ summed\ heat\ input\ for\ ORIS/boiler} \quad \text{[Equation 2-2]}$$

To quality assure emission changes using this method, we manually calculate the hourly emissions using Equation 2-1 for primary $PM_{2.5}$ and SO_2. We compare these calculations

of hourly emissions from CEM heat inputs to that of the default temporal profile assigned by SMOKE based on the unit's SCC[39].

2.3.3. Electricity Modeling

The Day-Ahead Locational Market Clearing Prices Analyzer (DAYZER)[60] simulates the operation of select electricity markets using the most recently available data on fuel prices, electricity demand, unit and transmission line outages, and emission permit costs. Detailed information about DAYZER and its required data input can be found elsewhere[60]. We employ DAYZER in these retrospective simulations to evaluate power generation during this heat wave stagnation event and relate electricity dispatch on HEDD to measured peak $PM_{2.5}$ mass, sulfate, and EC measurements.

2.3.4. Ambient Evaluation

CMAQ predictions and DAYZER dispatch data are compared with ambient measurements from the nine IMPROVE monitoring sites and 13 AQS locations in the modeling domain (Figure 2-1b). IMPROVE and AQS measurement data is downloaded from the Technology Transfer Network (TTN) Air Quality System (AQS)[61] and is paired in space and time with CMAQ results using the site compare (sitecmp) tool.[62] The data is analyzed using the Visualization Environment for Rich Data Interpretation (VERDIv1.4)[63] as well as the R software environment[64].

2.4. Results and Discussion

2.4.1. Air Quality and Electricity Dispatch

During the simulation time period, trends in $PM_{2.5}$ and sulfate mass concentrations are positively correlated with electricity dispatch in PJM. The correlation coefficient between daily DAYZER-predicted electricity dispatch and the daily average maximum $PM_{2.5}$ mass at all AQS monitoring sites is r =0.6 (Figure A-2) and that for daily maximum sulfate at IMPROVE locations is r =0.5 (Figure 2-2). On July 31, when the $PM_{2.5}$ mass concentration is highest and the fractional contribution of sulfate is greatest (EGUs are the largest source of continental sulfate), there is a corresponding peak in PJM electricity generation. Conversely, on July 8-9, when electricity demand is lowest, regional ambient $PM_{2.5}$ concentrations are lowest and the fractional contribution of sulfate is smallest. Note, however, that there is an offset because electricity generation data is hourly, while speciated $PM_{2.5}$ measurements are available every third day.

Similar to previous studies[28], the CMAQ model indicates a positive model bias (overprediction) at low ambient mass concentrations and a negative model bias (underprediction) when measurement values are high (Figure 2-3). Linear model relationships between model bias and ambient measurements for the two species are robust and statistically significant, (sulfate: r =-0.4 and $p = 3.1 \times 10^{-7}$; $PM_{2.5}$: r = -0.6 and $p = 4 \times 10^{-15}$). This is consistent with temporal mis-allocation of EGU emissions where default profiles are designed to predict averages and typical conditions, not peak or minimum events. In the sensitivity simulation where 267 CEM matches are added, 27%

of CEMS in PJM (From Table 2-1: (541-267)/1006) and 40% of CEMS nationally that could be improved remain unmatched, and hourly emissions for these sources remain temporally mis-allocated. These percentages are calculated using values from the Clean Air Markets Division (CAMD) database.

Actual utilization of hourly CEM emission data from the additional 267 CEMs in SMOKE increases CMAQ-predicted peak surface mass concentrations of hourly ambient $PM_{2.5}$ up to 7.4 $\mu g/m^3$, an 83% increase, during the heat wave studied here. Maximum hourly ambient $PM_{2.5}$ mass increases are primarily a consequence of increases in sulfate (from SO_2 precursor emissions) and EC mass (Figure 2-5, Figure A-3). At the surface, increases of sulfate are highest in the Virginia Beach/Norfolk, VA area (2.5 $\mu g/m^3$) (103%) near coal plants (Figure 2-1a). A smaller but noticeable increase is estimated near Blacksburg, VA (1-1.6 $\mu g/m^3$). Increases in maximum hourly modeled EC concentrations in specific grid cells are up to 5.3 $\mu g/m^3$ near Southwestern NJ and Northeastern DE, a 310% increase. This is consistent with locations of coal fueled plants, the largest point sources of EC emissions[4], (Figure 2-1) near Virginia Beach/Norfolk, VA, in the suburbs of Wilmington, DE, and Blacksburg, VA. The largest $PM_{2.5}$ mass concentration increases occur near population centers: Blacksburg, VA (1-1.5 $\mu g/m^3$), Virginia Beach/Norfolk, VA area (1.5-2.5 $\mu g/m^3$), Pittsburgh, PA (1.5-2.5 $\mu g/m^3$) and in the suburbs of Wilmington, DE (4.5-7.4 $\mu g/m^3$). Aloft, at 800mb, an approximation for the daytime boundary layer height, ambient $PM_{2.5}$ concentrations increase up to 1.5 $\mu g/m^3$ with sulfate comprising nearly all of the change. These concentrations represent increases of 51% ($PM_{2.5}$) and 75% (sulfate) aloft. Changes in predicted ambient EC concentrations are small aloft (0.17 $\mu g/m^3$), but represent a 38% increase.

CMAQ predictions of other PM species are also impacted by the emissions changes. Particulate ammonium increases up to 0.5 $\mu g/m^3$ at the surface and 0.2 $\mu g/m^3$ aloft, with the largest differences occurring near Newark, NJ, the Virginia Beach/Norfolk, VA area and Blacksburg, VA. In Pennsylvania, the increases in ammonium are spatially consistent with nitrate increases. Ammonium increases in the Virginia appear to be driven by sulfate, consistent with coal plant locations (Figure 2-1a). Nitrate exhibits increases up to 0.8 $\mu g/m^3$ at the surface with the largest concentration differences occurring near East-Central Pennsylvania. Aloft, changes in predicted nitrate are similar in amount to changes at the surface (up to 0.7 $\mu g/m^3$) however, geospatially they are different. Maximum aloft increases are located near Newark, NJ and New York City, NY. Mass concentrations of organic carbon (OC) increase up to 1.6 $\mu g/m^3$ in North-Central VA, Southern NJ and Pittsburgh, PA, driven by increased emissions of primary organic carbon. Aloft, OC has a small signature, with increases of up to 0.1 $\mu g/m^3$ near Pittsburgh, PA and New York City, NY (Figure A-4).

Changes in ambient ozone mixing ratios in these simulations at the surface are less than 2%, however percent differences aloft are larger because absolute concentrations are smaller (Figure A-5). In the region modeled in this work, 60% of NO_x emissions stem from mobile sources, with only 20% resulting from electric generation (Figure A-6). The small difference in CMAQ predicted ozone between simulations is consistent with the electricity sector contributing a relatively small amount to the domain total NO_x emissions.

2.4.2. Matching Facility IDs and Improved Temporalization of Emissions

These changes in CMAQ-predicted ambient concentrations are a consequence of correcting a portion of the temporally mis-allocated SO_2, primary $PM_{2.5}$ and EC emissions. The generic SCC-based temporal profiles that are assigned to facilities are designed to represent average and typical conditions, not peak events. Whereas these profiles capture pattern differences between day and night, weekday and weekend, and seasons, they are limited in their ability to account for discrepancies due to year-specific events (e.g. heat waves). EPA and MARAMA have addressed this issue to an extent with the MARAMA-developed "CAMD to NIF (National Emissions Inventory Input Format) Crosswalk" to match CEM and NEI identifiers in the Northeastern and Mid-Atlantic regions of the U.S.[65] Still, more than 40% of the national 2006 CEM measurement data goes unused in simulations that employ the 2008 NEI. For example, the Hopewell, VA coal plant CEM data indicates the unit operated only in May through October of 2006. The 2008 NEI is missing the ORIS ID and boiler codes associated with this unit's CEM, therefore SMOKE temporalizes Hopewell emissions over the entire year (Figure 2-6a). During January – May and September – December, periods of lower than average electricity demand, SO_2, EC, $PM_{2.5}$ emissions are largely overpredicted. During higher summertime peak demand, emissions can be underpredicted by more than a factor of three. In the portion of the heat wave simulated here, use of CEM data for this coal plant results in an average hourly increase of $PM_{2.5}$ and SO_2 emissions of 154% over the default SMOKE emissions when no match occurs (Figure 2-6b). The relative increases for each pollutant are the same for an individual unit using this method because emissions

are heat input dependent (Equation 2-1). A similar trend in both $PM_{2.5}$ and SO_2 emissions is observed at a New Jersey gas plant, where the CEM estimate represents an increase of 131% in $PM_{2.5}$ and SO_2 emissions (Figures 2-6c and 2-6d). It should be noted that the annual averages and total of emissions for both the default temporal profile and matched CEM profile are the same, however temporal allocation results in hourly, daily and monthly emission disparities.

Similar to emissions processing programs, routine air quality monitoring networks, in particular for chemically speciated PM, are designed to characterize annual percentiles and not peak episodes. Episodic peak events occur on short time scales (~hours) and negatively impact health[7, 13], but are not adequately characterized by conventional air quality surface networks that make measurements every three days or even daily. Therefore, comparison of hourly model predictions to measurements collected every 3rd day (e.g. IMPROVE) to evaluate changes in model performance induced by improvements to peak hourly emissions shows little to no improvement in model predictive skill. However, using the non-reference method hourly $PM_{2.5}$, episodic improvements in prediction of peak concentrations are observed (Figure 2-4). During a period of higher electricity demand, July 16-18, 2006, predicted $PM_{2.5}$ concentrations are closer to measured values (maximum: up to 4 $\mu g/m^3$ closer to observed values; average: 0.1 $\mu g/m^3$ closer to observed values over entire domain) for the added CEM sensitivity simulation for a few hours each day. The majority of these slight improvements occur when observed $PM_{2.5}$ mass concentrations exceed 50 $\mu g/m^3$ (i.e. peak concentrations). Speciated hourly measurements of sulfate and EC would more enable more accurate assessment the effectiveness of the temporalization scheme.

CMAQ predictions of a variety of pollutants increase, indicative of the peak 2006 summer heat wave emissions that occur when the matched EGU emissions are temporally reassigned based on their hourly CEM data. A similar study for an annual simulation is unlikely to perturb annual $PM_{2.5}$ mass concentration averages. However, episodic increases in $PM_{2.5}$ mass predictions are expected, as observed during the heat wave modeled here. Decreases in predicted $PM_{2.5}$ mass during time periods of low electricity demand due to hourly and daily allocation differences are also expected. This suggests that better temporalization of EGU emissions may improve CMAQ's ability to reproduce PM variability at the surface and aloft. While this paper focuses on temporalization for better prediction of $PM_{2.5}$, sulfate, and EC ambient mass concentrations, this analysis could be applied to any number of pollutants emitted from point sources, for example mercury, whose largest emission source is power plants[66]. Furthermore, these results represent units from a single RTO and only 12% of the total unmatched CEM units in the U.S. A study examining this issue in other regions of the U.S. electricity grid would provide an interesting comparison to the results observed here. For example, a region in which power plants are the dominant source of NO_x would provide a different perspective of ozone sensitivity to temporalization of EGU sector emissions. Finally, the MARAMA crosswalk is essential to this work, a national crosswalk, or combination of available RTO crosswalks, would aide in a solution to this problem.

2.5. References

1. Frost, G. J.; McKeen, S. A.; Trainer, M.; Ryerson, T. B.; Neuman, J. A.; Roberts, J. M.; Swanson, A.; Holloway, J. S.; Sueper, D. T.; Fortin, T.; Parrish, D. D.; Fehsenfeld, F. C.; Flocke, F.; Peckham, S. E.; Grell, G. A.; Kowal, D.; Cartwright, J.; Auerbach, N.; Habermann, T., Effects of changing power plant NO_x emissions on ozone in the eastern United States: Proof of concept. *J. Geophys. Res.-Atmos.* **2006,** *111*, (D12).

2. He, H.; Loughner, C. P.; Stehr, J. W.; Arkinson, H. L.; Brent, L. C.; Follette-Cook, M. B.; Tzortziou, M. A.; Pickering, K. E.; Thompson, A. M.; Martins, D. K.; Diskin, G. S.; Anderson, B. E.; Crawford, J. H.; Weinheimer, A. J.; Lee, P.; Hains, J. C.; Dickerson, R. R., An elevated reservoir of air pollutants over the Mid-Atlantic States during the 2011 DISCOVER-AQ campaign: Airborne measurements and numerical simulations. *Atmos. Environ.* **2014,** *85*, (0), 18-30.

3. He, H.; Hembeck, L.; Hosley, K. M.; Canty, T. P.; Salawitch, R. J.; Dickerson, R. R., High ozone concentrations on hot days: The role of electric power demand and NO_x emissions. *Geophys. Res. Lett.* **2013,** *40*, (19) 5291-5294.

4. U.S. EPA. The 2011 National Emissions Inventory v.1. http://www.epa.gov/ttnchie1/net/2011inventory.html (August 30, 2015).

5. Appel, K. W.; Foley, K. M.; Bash, J. O.; Pinder, R. W.; Dennis, R. L.; Allen, D. J.; Pickering, K., A multi-resolution assessment of the Community Multiscale Air Quality (CMAQ) model v4.7 wet deposition estimates for 2002–2006. *Geosci. Model Dev.* **2011,** *4*, (2), 357-371.

6. Turner, M. C.; Krewski, D.; Pope, C. A.; Chen, Y.; Gapstur, S. M.; Thun, M. J., Long-term Ambient Fine Particulate Matter Air Pollution and Lung Cancer in a Large Cohort of Never-Smokers. *Am. J. Resp. Crit. Care* **2011,** *184*, (12), 1374-1381.

7. Pope, C. A.; Burnett, R. T.; Turner, M. C.; Cohen, A.; Krewski, D.; Jerrett, M.; Gapstur, S. M.; Thun, M. J., Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure-Response Relationships. *Environ. Health Persp.* **2011,** *119*, (11), 1616-1621.

8. Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y. L.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.; Smith, S. C.; Whitsel, L.; Kaufman, J. D.; Epidemiol, A. H. A. C.; Dis, C. K. C.; Metab, C. N. P. A., Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association. *Circulation* **2010,** *121*, (21), 2331-2378.

9. Burnett, R. T.; Pope, C. A.; Ezzati, M.; Olives, C.; Lim, S. S.; Mehta, S.; Shin, H. H.; Singh, G.; Hubbell, B.; Brauer, M.; Anderson, H. R.; Smith, K. R.; Balmes, J. R.; Bruce, N. G.; Kan, H. D.; Laden, F.; Pruss-Ustun, A.; Michelle, C. T.; Gapstur, S. M.; Diver, W. R.; Cohen, A., An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environ. Health Persp.* **2014,** *122*, (4), 397-403.

10. Fang, Y.; Naik, V.; Horowitz, L. W.; Mauzerall, D. L., Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present. *Atmos. Chem. Phys.* **2013,** *13*, (3), 1377-1394.

11. Lepeule, J.; Laden, F.; Dockery, D.; Schwartz, J., Chronic Exposure to Fine Particles and Mortality: An Extended Follow-up of the Harvard Six Cities Study from 1974 to 2009. *Environ. Health Persp.* **2012,** *120*, (7), 965-970.

12. Raz, R.; Roberts, A. L.; Lyall, K.; Hart, J. E.; Just, A. C.; Laden, F.; Weisskopf, M. G., Autism Spectrum Disorder and Particulate Matter Air Pollution before, during, and after Pregnancy: A Nested Case–Control Analysis within the Nurses' Health Study II Cohort. *Environ. Health Persp.* **2015,** 123, 264-270.

13. Robert D. Brook, J. R. B., Bruce Urch, Renaud Vincent, Sanjay Rajagopalan, Frances Silverman, Inhalation of Fine Particulate Air Pollution and Ozone Causes Acute Arterial Vasoconstriction in Healthy Adults. *Circulation* **2002,** *105*, (13).

14. Fann, N.; Lamson, A. D.; Anenberg, S. C.; Wesson, K.; Risley, D.; Hubbell, B. J., Estimating the National Public Health Burden Associated with Exposure to Ambient $PM_{2.5}$ and Ozone. *Risk Anal.* **2012,** *32*, (1), 81-95.

15. Stagnation When Heat Waves Exist - Summer, 1950-2007. In NOAA/NCDC, Ed. United States Global Change Research Program: 2009.

16. Tai, A. P. K.; Mickley, L. J.; Jacob, D. J.; Leibensperger, E. M.; Zhang, L.; Fisher, J. A.; Pye, H. O. T., Meteorological modes of variability for fine particulate matter ($PM_{2.5}$) air quality in the United States: implications for $PM_{2.5}$ sensitivity to climate change. *Atmos. Chem. Phys.* **2012,** *12*, (6), 3131-3145.

17. Mahmud, A.; Hixson, M.; Kleeman, M. J., Quantifying population exposure to airborne particulate matter during extreme events in California due to climate change. *Atmos. Chem. Phys.* **2012,** *12*, (16), 7453-7463.

18. Mickley, L. J.; Jacob, D. J.; Field, B. D.; Rind, D., Effects of future climate change on regional air pollution episodes in the United States. *Geophys. Res. Lett.* **2004,** *31*, (24).

19. Mwaniki, G. R.; Rosenkrance, C.; Wallace, H. W.; Jobson, B. T.; Erickson, M. H.; Lamb, B. K.; Hardy, R. J.; Zalakeviciute, R.; VanReken, T. M., Factors contributing to elevated concentrations of $PM_{2.5}$ during wintertime near Boise, Idaho. *Atmos. Pol. Res.* **2014,** *5*, 96-103.

20. Gerald A. Meehl, C. T., More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science* **2004,** *305*, (5686), 994-997.

21. Jacob, D. J.; Winner, D. A., Effect of climate change on air quality. *Atmos. Environ.* **2009,** *43*, (1), 51-63.

22. Horton, D. E.; Harshvardhan; Diffenbaugh, N. S., Response of air stagnation frequency to anthropogenically enhanced radiative forcing. *Environ. Res. Lett.* **2012,** *7*, (4).

23. National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Air Stagnation Index. http://www.ncdc.noaa.gov/societal-impacts/air-stagnation/ (August 30, 2015).

24. Trail, M.; Tsimpidi, A. P.; Liu, P.; Tsigaridis, K.; Hu, Y.; Nenes, A.; Russell, A. G., Downscaling a global climate model to simulate climate change over the US and the implication on regional and urban air quality. *Geosci. Model Dev.* **2013,** *6*, (5), 1429-1445.

25. Amato, A. D. R., M.; Kirshen, P.; Horwitz, J., Regional Energy Demand Responses To Climate Change: Methodology And Application To The Commonwealth Of Massachusetts. *Climatic Change* **2005,** *71*, (1-2), 175-201.

26. Miller, N. L.; Hayhoe, K.; Jin, J.; Auffhammer, M., Climate, Extreme Heat, and Electricity Demand in California. *J. Appl. Meteorol. Clim.* **2008,** *47*, (6), 1834-1844.

27. ICF International. *Clean Energy Options for Addressing High Electic Demand Days*; 2008. http://www.epa.gov/statelocalclimate/documents/pdf/hedd_clean_energy_options.pdf (August 30, 2015).

28. Zhang, H.; Chen, G.; Hu, J.; Chen, S.-H.; Wiedinmyer, C.; Kleeman, M.; Ying, Q., Evaluation of a seven-year air quality simulation using the Weather Research and Forecasting (WRF)/Community Multiscale Air Quality (CMAQ) models in the eastern United States. *Sci. Total Environ.* **2014,** *473–474*, (0), 275-285.

29. Foley, K. M.; Roselle, S. J.; Appel, K. W.; Bhave, P. V.; Pleim, J. E.; Otte, T. L.; Mathur, R.; Sarwar, G.; Young, J. O.; Gilliam, R. C.; Nolte, C. G.; Kelly, J. T.; Gilliland, A. B.; Bash, J. O., Incremental testing of the Community Multiscale Air

Quality (CMAQ) modeling system version 4.7. *Geosci. Model Dev.* **2010,** *3*, (1), 205-226.

30. Twomey, S., The Influence of Pollution on the Shortwave Albedo of Clouds. *J. Atmos. Sci.* **1977,** *34*, (7), 1149-1152.

31. Jacobson, M. Z., Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols. *Nature* **2001,** *409*, (6821), 695-697.

32. Leibensperger, E. M.; Mickley, L. J.; Jacob, D. J.; Chen, W. T.; Seinfeld, J. H.; Nenes, A.; Adams, P. J.; Streets, D. G.; Kumar, N.; Rind, D., Climatic effects of 1950-2050 changes in US anthropogenic aerosols - Part 1: Aerosol trends and radiative forcing. *Atmos. Chem. Phys.* **2012,** *12*, (7), 3333-3348.

33. Leibensperger, E. M.; Mickley, L. J.; Jacob, D. J.; Chen, W. T.; Seinfeld, J. H.; Nenes, A.; Adams, P. J.; Streets, D. G.; Kumar, N.; Rind, D., Climatic effects of 1950-2050 changes in US anthropogenic aerosols - Part 2: Climate response. *Atmos. Chem. Phys.* **2012,** *12*, (7), 3349-3362.

34. Dawson, J. P.; Adams, P. J.; Pandis, S. N., Sensitivity of $PM_{2.5}$ to climate in the Eastern US: a modeling case study. *Atmos. Chem. Phys.* **2007,** *7*, (16), 4295-4309.

35. Kleeman, M. J., A preliminary assessment of the sensitivity of air quality in California to global change. *Climatic Change* **2008,** *87*, S273-S292.

36. Fry, J. L.; Kiendler-Scharr, A.; Rollins, A. W.; Wooldridge, P. J.; Brown, S. S.; Fuchs, H.; Dubé, W.; Mensah, A.; dal Maso, M.; Tillmann, R.; Dorn, H. P.; Brauers, T.; Cohen, R. C., Organic nitrate and secondary organic aerosol yield from NO_3 oxidation of β-pinene evaluated using a gas-phase kinetics/aerosol partitioning model. *Atmos. Chem. Phys.* **2009,** *9*, (4), 1431-1449.

37. U.S. EPA, Acid Rain Program. http://www.epa.gov/airmarkets/progsregs/arp/basic.html (August 30, 2015).

38. U.S. EPA, Finding of Significant Contribution and Rulemakings for Certain States in the Ozone Transport Assessment Group Region. http://www.epa.gov/ttn/naaqs/ozone/rto/sip/index.html (August 30, 2015).

39. U.S. EPA, SMOKE v3.5 User's Manual. http://www.cmascenter.org/smoke/documentation/3.5/manual_smokev35.pdf (August 30, 2015).

40. U.S. EPA, *NEI Input Format (NIF) Version 3.0 User's Guide Instructions and Conventions of Use*. U.S. EPA Emission Factor & Inventory Group: Research Triangle Park, NC, 2003.

41. E.H. Pechan & Associates, I. *The Emissions & Generation Resource Integrated Database For 2010 (eGRID) Technical Support Document*; 2010.

42. PJM. Who We Are. http://www.pjm.com/about-pjm/who-we-are.aspx (August 30, 2015).

43. IMPROVE Interagency Monitoring of Protected Visual Environments. http://vista.cira.colostate.edu/improve/ (August 30, 2015).

44. National Centers for Environmental Information, National Oceanic and Atmospheric Administration. State of the Climate: National Overview for July 2006. http://www.ncdc.noaa.gov/sotc/national/2006/7 (August 30, 2015).

45. National Centers for Environmental Information, National Oceanic and Atmospheric Administration. Climatological Rankings. http://www.ncdc.noaa.gov/temp-and-precip/climatological-rankings/index.php (August 30, 2015).

46. Sarwar, G.; Luecken, D.; Yarwood, G.; Whitten, G. Z.; Carter, W. P. L., Impact of an Updated Carbon Bond Mechanism on Predictions from the CMAQ Modeling System: Preliminary Assessment. *J. Appl. Meteorol. Clim.* **2008,** *47*, (1), 3-14.

47. Carlton, A. G.; Bhave, P. V.; Napelenok, S. L.; Edney, E. D.; Sarwar, G.; Pinder, R. W.; Pouliot, G. A.; Houyoux, M., Model Representation of Secondary Organic Aerosol in CMAQv4.7. *Environ. Sci. Techn.* **2010,** *44*, (22), 8553-8560.

48. Carlton, A. G.; Turpin, B. J.; Altieri, K. E.; Seitzinger, S. P.; Mathur, R.; Roselle, S. J.; Weber, R. J., CMAQ Model Performance Enhanced When In-Cloud Secondary Organic Aerosol is Included: Comparisons of Organic Carbon Predictions with Measurements. *Environ. Sci. Tech.* **2008,** *42*, (23), 8798-8802.

49. Appel, K. W.; Chemel, C.; Roselle, S. J.; Francis, X. V.; Hu, R. M.; Sokhi, R. S.; Rao, S. T.; Galmarini, S., Examination of the Community Multiscale Air Quality (CMAQ) model performance over the North American and European domains. *Atmos. Environ.* **2012,** *53*, 142-155.

50. Skamarock, W. C. K., J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. *A description of the Advanced Research WRF version 3. NCAR Technical Note NCAR/TN-475+STR*; 2008.

51. Kwok, R. H. F.; Baker, K. R.; Napelenok, S. L.; Tonnesen, G. S., Photochemical grid model implementation of VOC, NO_x, and O_3 source apportionment. *Geosci. Model Dev. Discuss.* **2014,** *7*, (5), 5791-5829.

52. Carlton, A. G.; Baker, K. R., Photochemical Modeling of the Ozark Isoprene Volcano: MEGAN, BEIS, and Their Impacts on Air Quality Predictions. *Environ. Sci. Tech.* **2011,** *45*, (10), 4438-4445.

53. U.S. EPA. The 2008 National Emissions Inventory. http://www.epa.gov/ttnchie1/net/2008inventory.html (August 30, 2015).

54. U.S. EPA, Motor Vehicle Emission Simulator (MOVES) User Guide for MOVES2010b. EPA-420-B-12-001b. **2012**.

55. U.S. EPA, North American Emissions Inventories - Canada. **2014**.

56. U.S. EPA, *NEI Input Format (NIF) Version 3.0 User's Guide Instructions and Conventions of Use*; 2003.

57. U.S. EPA, *FACT SHEET Revisions to the Air Emissions Reporting Requirements: Revisions to Lead (Pb) Reporting Threshold and Clarifications to Technical Reporting Details*; 2013.

58. U.S. EPA, *2008 National Emissions Inventory: Emissions Inventory System Implementation Plan Appendix 3*; 2008.

59. Mid Atlantic Regional Air Mangement Administration (MARAMA), *Development and Analysis of 2007 Hourly Point Source Emissions in the Northeast/Mid-Atlantic Region: Appendix A*; 2010.

60. Cambridge Energy Solutions. DAYZER. http://www.ces-us.com/product-dayzer.asp (August 30, 2015).

61. U.S. EPA, Download Detailed AQS Data. http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm (August 30, 2015).

62. U.S. EPA, Community Modeling and Analysis System, Institute for the Environment – University of North Carolina Chapel Hill. Operational Guidance for the Community Multiscale Air Quality (CMAQ) Modeling System. http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_(February_2010_release)_OGD&CFID=1566640&CFTOKEN=15aa7e2b6d5df141-36D372BC-DCE3-B3AE-7D6D6D8A88E1C553 (August 30, 2015).

63. Institute for the Environment - University of North Carolina at Chapel Hill, U.S. EPA, *Visualization Environment for Rich Data Interpretation (VERDI 1.4) User's Manual*. 2011.

64. R Core Team. *A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.

65. AMEC Environment & Infrastructure, S. I., Inc., MARAMA *Technical Support Document for the Development of the 2007 Emission Inventory for Regional Air Quality Modeling in the Northeast/Mid-Atlantic Region*; 2012; p Appendix A.

66. U.S. EPA, Cleaner Power Plants. http://www.epa.gov/mats/powerplants.html (August 30, 2015).

	Base Simulation	Sensitivity Simulation
PJM Region		
Total Units with CEMs	1006	1006
Matched ORIS ID/boiler in NEI	465	732 (465 + 267 new matches)
Unmatched ORIS ID/boiler in NEI	541	274 (541 – 267 new matches)
United States		
Total Units with CEMs	4830	4830
Matched ORIS ID/boiler in NEI	2644	2911 (2644 + 267 new matches)
Unmatched ORIS ID/boiler in NEI	2186	1919 (2186 – 267 new matches)

Table 2-1. Continuous emission monitor matched and unmatched unit data.

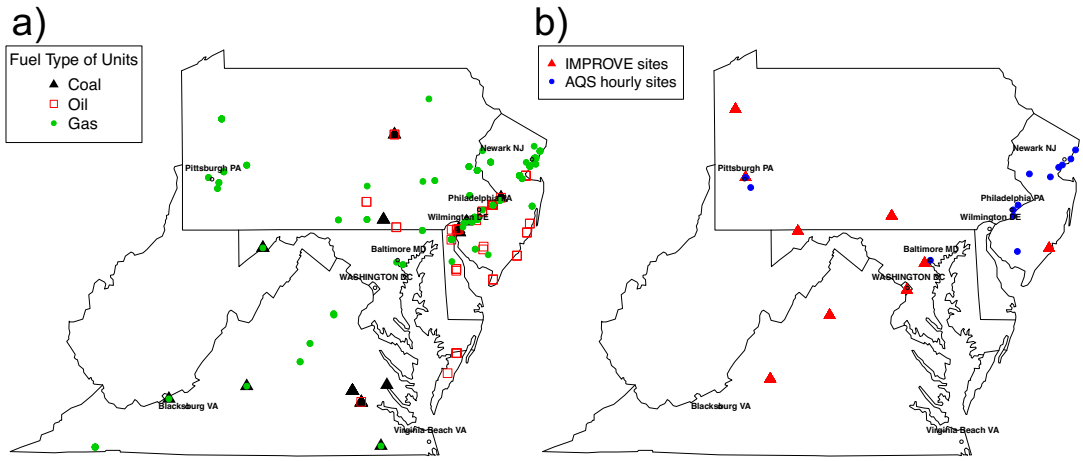

Figure 2-1. a) The location of 267 units by fuel type, whose identifying numbers are added to the 2008 NEI for this study and b). The locations of the IMPROVE (red triangles) and EPA AQS (blue circles) observation stations in the study domain.

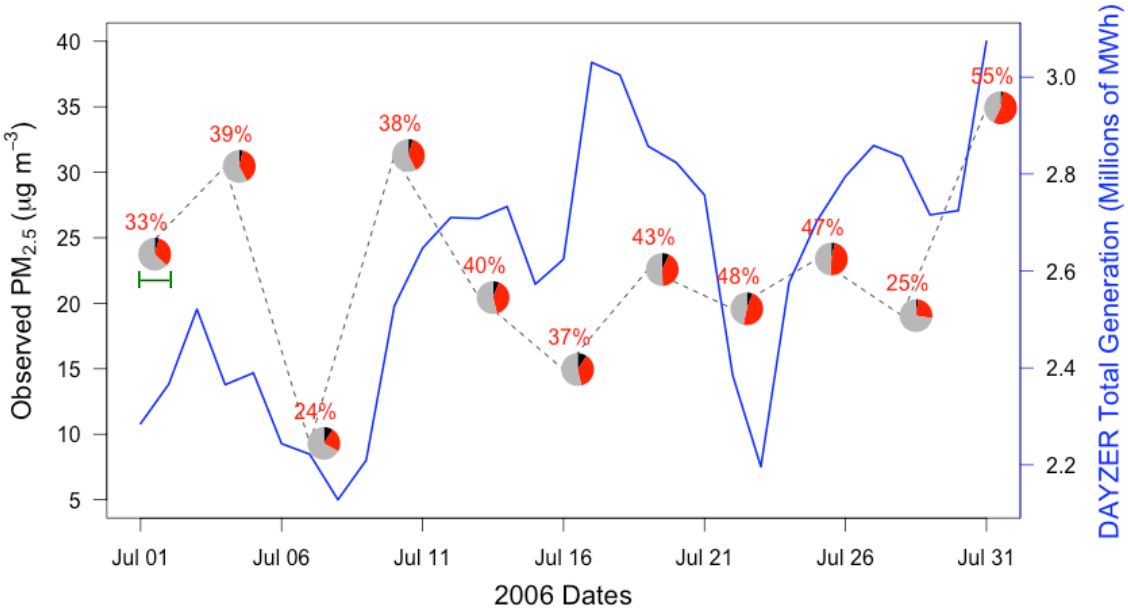

Figure 2-2. IMPROVE network daily maximum $PM_{2.5}$ observations from July 2006 (dotted line) measured every third day, divided into categories of sulfate (red), EC (black), and all other species (gray) for the 5 studied states: (Delaware, Maryland, New Jersey, Pennsylvania, Virginia). The percentages represent the fraction of $PM_{2.5}$ that is sulfate at each observation. The daily DAYZER total electricity generation for the entire PJM region during the same time period is in blue. Note: The width of each pie chart is equal to the 24-hour period over which that measurement was made, as demonstrated by the green bracket under the first pie chart.

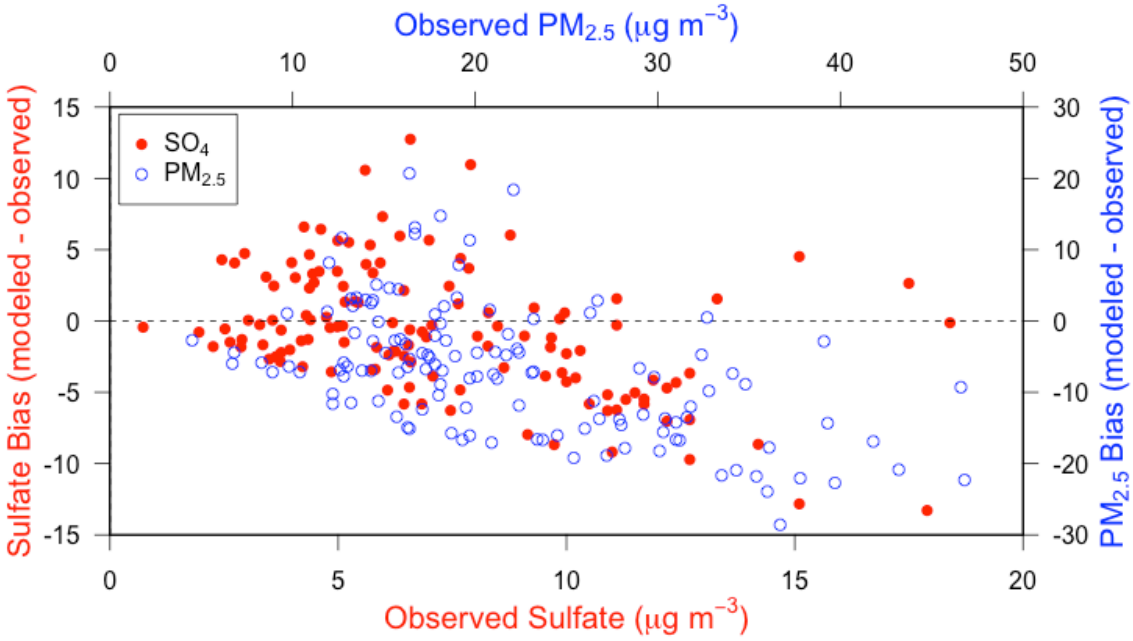

Figure 2-3. The base case model bias of sulfate (red) and $PM_{2.5}$ (blue). The observation data are IMPROVE sites that lie within the modeling domain for all measured days in July 2006.

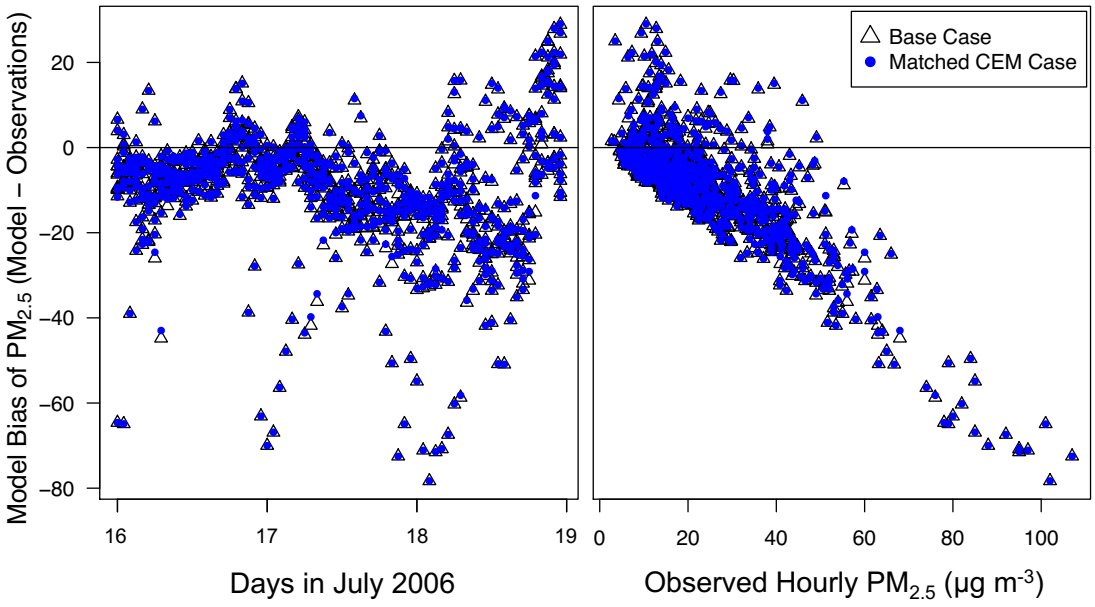

Figure 2-4. CMAQ model bias of $PM_{2.5}$ ambient concentrations at the surface between July 16-18, 2006 as a time series (left) and compared to hourly observed values (right) for both the base case (unfilled triangles) and the case with additional CEM matches (blue filled circles).

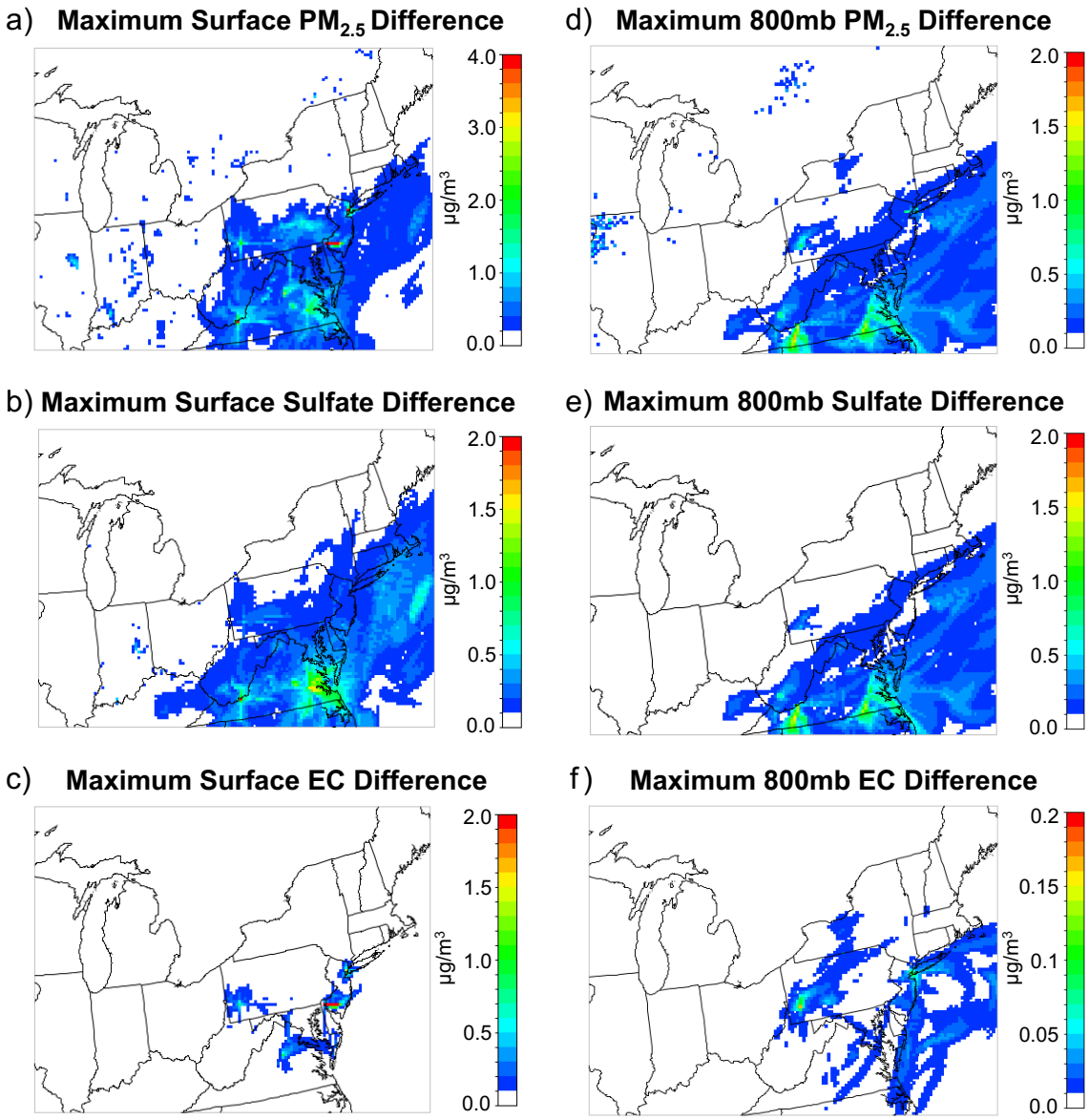

Figure 2-5. CMAQ-predicted maximum hourly ambient concentration differences (matched CEM simulation – base simulation) at the surface and the 800mb level between July 1-31, 2006 of $PM_{2.5}$ (a and d), sulfate (b and e) and EC (c and f). NOTE: not all concentration scales are the same.

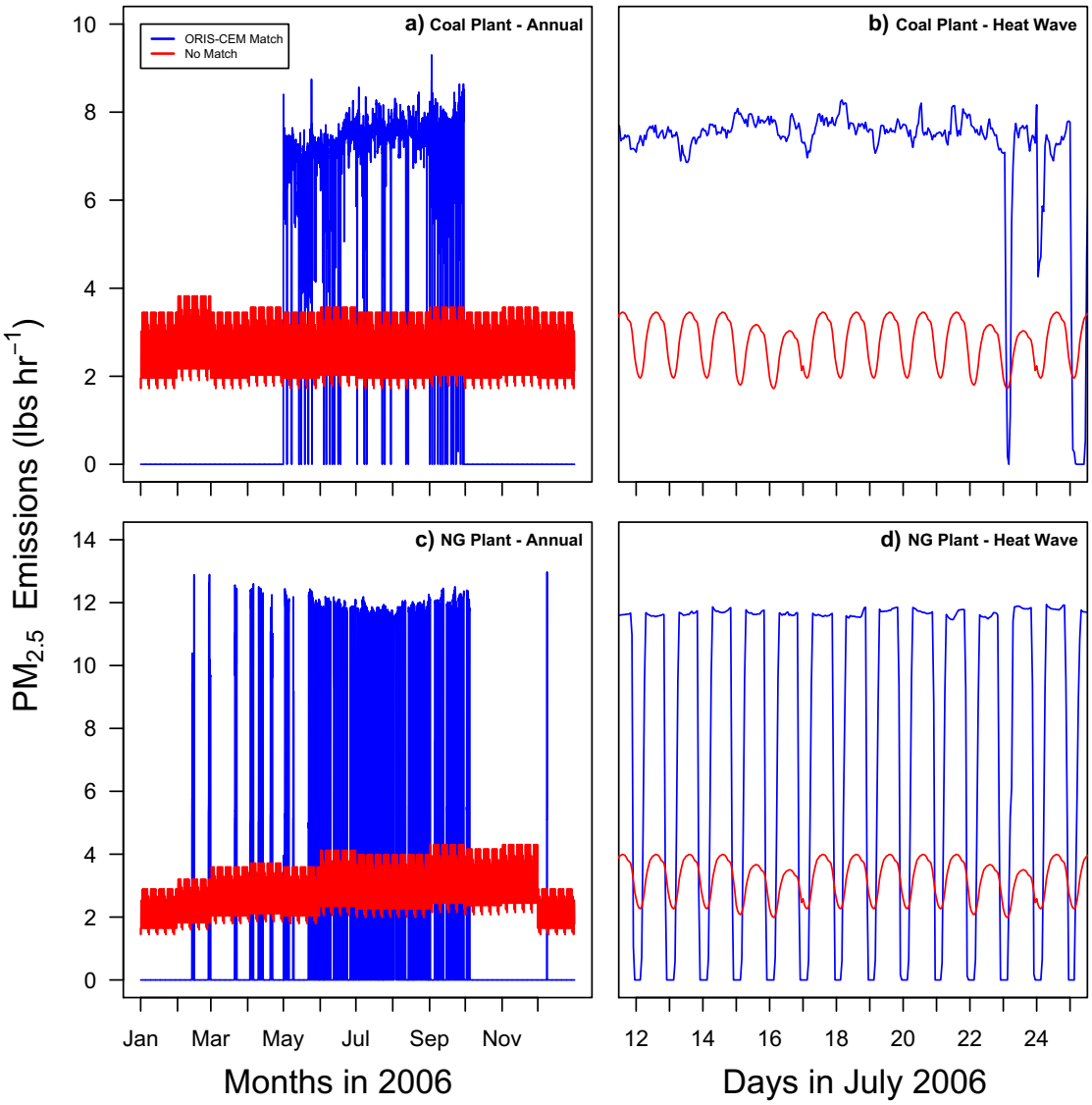

Figure 2-6. A manual comparison of the default temporalization in the SMOKE model without matched CEM hourly data (red) and with the temporalization from the hourly CEM data (blue) using Equation 1. The top two plots are data from the Cogentrix-Hopewell coal plant in Virginia and the bottom two plots are data from the AES Red Oak plant in New Jersey. The left plots show the temporalization over the entire year; the right plots show July 12-26, 2006, outlining a portion of the heat wave.

3. CHAPTER 3

HIGH ELECTRICITY DEMAND IN PJM: ELECTRICTY SECTOR CONTRIBUTION TO AIR QUALITY AND PEAKING UNIT POTENTIAL EXPOSURE

Material in this chapter to be submitted as:

Farkas, C.M.; Moeller, M.D.; Felder, F.A.; Henderson, B.H.; Carlton, A.G., High Electricity Demand in PJM: Electricity Sector Contribution to Air Quality and Peaking Unit Potential Exposure. *Environ. Sci. & Tech. (in prep)*

3.1. Abstract

On high electricity demand days, when air quality is often poor, regional transmission organizations (RTOs), such as PJM Interconnection, ensure reliability of the electric grid by employing peak-use electric generating units (EGUs). These "peaking units" are exempt from some federal and state air quality rules that traditionally target larger (e.g., >25 MW) facilities. We identify RTO assignment and peaking unit classification for EGUs in the Eastern U.S. The Community Multiscale Air Quality (CMAQ) model estimates fine particulate matter ($PM_{2.5}$) mass and ozone concentrations, and we population-weight ambient values as a surrogate for potential exposure during a heat wave in July 2006. Non-PJM EGU emissions produce high hourly maximum $PM_{2.5}$ concentrations (up to 140 $\mu g/m^3$), predominantly impacting areas outside PJM. Monitored and controlled PJM peaking units contribute up to 87% of hourly maximum

$PM_{2.5}$ mass locally. Potential exposure to peaking unit $PM_{2.5}$ mass is highest in the domain's most populated cities. Average daily temperature and national Gross Domestic Product (GDP) drive peaking unit heat input. Air quality planning that capitalizes on the understanding of electricity demand and economics may provide a holistic approach to protect human health within the context of energy needs in a changing world.

3.2. Introduction

In the Northeast U.S., high-energy demand days (HEDDs) are typically hot, stagnant summer days when air conditioning loads are high and air quality is poor. Electricity usage is linearly correlated with daily maximum temperatures above 23˚C,[1-5] because generation rises to meet increased demand. Peak generation is met by employing "peaking units", often older, less efficient, less regulated[6, 7] electric generating units (EGUs), to ensure reliability of the grid. In many regions of the U.S., dispatch within and among electric grids is managed by regional transmission organizations (RTOs), which administer the transmission of electricity in a particular region. PJM Interconnection, the RTO governing the electricity transmission for all or parts of Illinois, Indiana, Michigan, Ohio, Kentucky, Virginia, West Virginia, Pennsylvania, Delaware, Maryland, New Jersey, North Carolina, Tennessee, and the District of Columbia, is the largest U.S. RTO, managing roughly 20% of electricity generation in the U.S.[8] PJM is transmission limited and employs many peaking units that are often located in close proximity to population-dense urban centers (Figure 3-1) to provide extra electricity quickly to areas of highest demand. Meteorological conditions that trigger high demand episodes (e.g. heat waves,

stagnation events) have increased in frequency and duration and are projected to continue this trend with the changing climate.[9-12] Further, higher air temperatures reduce the efficiency of electricity generators and transmission lines.[13] In the coming decade, summer peak load in PJM is forecast to increase by an average of 1% yearly.[14]

The electricity sector and fossil-fueled EGUs, in particular, are the largest controllable source of primary particulate matter[15] and are responsible for 75% of annual sulfur dioxide (SO_2) emissions in the U.S.[16] EGUs are the second largest controllable source of nitrogen oxide (NO_x) emissions in the U.S.[17] These emissions and secondary pollutants formed from them in the atmosphere, such as sulfate, nitrate, and ozone cause cardio and pulmonary diseases[18-21] and premature death[22, 23] in humans. EGU-derived particles (e.g. black carbon (BC) and sulfate) act as near term climate forcers (NTCFs) and can modulate regional surface temperatures, in particular in the Eastern U.S.[24-27] EGU emission rates of individual pollutants[28] vary (Table B-1) dependent on fuel type, generation amount, and operating control equipment. Emission factors in PJM are lower than the national averages for all categories. In 2006, of the peaking units in PJM for which data is available, (i.e., monitored and controlled EGUs with continuous emission monitors (CEMs)), 1% used coal, 24% used oil, and 75% used gas. Many peaking units however are not fitted with CEMs and their emissions, even in high population density areas, are not measured and difficult to quantify.

Despite the close proximity of some peaking units to high population areas, and their projected increase in fractional contribution to total electricity generation,[14] these facilities are not well characterized in the emissions inventory and often fall outside criteria for inclusion in emissions control programs. The exact number of unmonitored

peaking units is difficult to quantify because when peaking units fall below a regulatory threshold nameplate generating capacity, CEMs and national emissions inventory (NEI) reporting are not required. For example, if a peaking unit has a nameplate generation capacity below the policy-implemented threshold (25 MW)[6, 7] of the Acid Rain Program, the SO_2 emissions may not be regulated unless the unit's state requires it. Peaking units not located in an ozone or PM non-attainment region or with low annual or seasonal emissions, regardless of peak hourly emission rates, can be exempt from federal NO_x control and monitoring rules.[29]

It is well-established in electricity markets, that dispatch and short term generation planning is required for Day-Ahead market based demand, transmission capacity and fuel price.[30] In PJM, as in many parts of the U.S., the dispatch of generation units is a function of demand, generation offers (primarily fuel costs), transmission constraints, and reliability requirements. Long term energy planning (e.g. new facility construction) is based on long-term forecasts of fuel price and demand.[31, 32] Short term fuel price fluctuations largely control electricity dispatch, in particular, in the Northeast U.S., and in particular at peak demand time periods.[30] However, short-term spot market economics are absent in air quality forecasts, meant to alert and protect the public. This represents a key knowledge gap.

Holistic and regional regulation and analyses of air pollution, including extreme events, such as heat waves and peak electricity demand episodes, are needed because of the interconnected nature of the electricity system (e.g. electricity demand from Philadelphia may require fossil fuel burning in Ohio) and the transport of pollution across state boundaries. EGU emissions impact air quality locally and regionally because they

directly emit pollution and precursors that form secondary fine particles ($PM_{2.5}$) and ozone. Regional regulations such as the Cross-State Air Pollution Rule (CSAPR) aim to reduce emissions from EGUs that contribute to regional $PM_{2.5}$ and ozone, and this is critical. Local rules aimed at small generators are also helpful to curtail pollution events and avert adverse health impacts. In this work we investigate the air quality impacts of emissions from 544 monitored and controlled peaking units during a heat wave stagnation event during July 2006 in the Eastern U.S. This period is characterized by peak electricity demand that broke demand records set one year prior.[33] We compare the air quality impacts of PJM peaking units to those associated with all EGUs from PJM and EGUs from surrounding RTOs that fall within the model domain. We employ the Community Multiscale Air Quality (CMAQ) model to simulate air quality and population-weight the hourly concentrations as an index for exposure potential. We explore associations among fuel price, electricity demand and air quality in PJM during the heat wave. To our knowledge, this is the first discussion of the air quality impacts of emissions from an individual RTO on its own region and on surrounding regions in the literature. We focus on PM because health impacts and the associated costs are larger than ozone impacts.[34]

3.3. Methods

We investigate air quality specific to an extreme event, namely peak electricity generation in PJM from July 1-31, 2006 during a heat wave stagnation event. Higher-than-normal temperatures and below-average precipitation produced the third warmest

July on record[35, 36] and resulted in above-average electricity consumption and poor air quality in many cities.[37] CMAQ is employed to simulate air quality and we assess differences in ambient and population-weighted concentrations of total $PM_{2.5}$ mass, primary unspeciated $PM_{2.5}$, sulfate, elemental carbon (EC), and ozone from perturbations in EGU-sector emissions.

3.3.1. Air Quality Modeling

We conduct four CMAQ simulations incorporating emission scenarios described below. The simulations here employ the same meteorology, biogenic emissions and chemistry as described in detail in Farkas et al., 2015.[38] The Base Case simulation in this work is the sensitivity simulation in Farkas et al., 2015. Briefly, CMAQv5.0.1 is applied to a domain covering the majority of the PJM region on a 12 km by 12 km grid with 34 vertical layers up to 50 mb for the heat wave of June 28-July 31, 2006. The first three days are excluded from the analysis for model spin up. Air quality impacts from three sensitivity analyses are estimated via difference from the Base Case, 1.) PJM peaking unit emissions, 2.) All EGU emissions from PJM, and 3.) EGU emissions from RTOs other than PJM (Table 3-1). Emission sectors (biogenic, mobile, etc) remain constant across the simulations.

3.3.2. Emissions Processing

Four different emission scenarios (Table 3-1) are created and processed in the Sparse Matrix Operator Kernel Emissions (SMOKE) model (https://www.cmascenter.org/smoke/). The base case scenario contains U.S. anthropogenic emissions obtained from the 2008 NEI (v2) and July 2006 hour-specific CEM data. From Farkas et al. (2015),[38] 267 Office of Regulatory Information Systems (ORIS) identifiers are added to the point source sector emissions for more accurate temporal representation of EGU emissions. Biogenic emissions are generated from the Biogenic Emissions Inventory model (BEISv3.14)[39] with day- and hour-specific solar radiation and temperatures from the Weather Research and Forecasting (WRF) model. The Motor Vehicle Emission Simulator (MOVESv2010b) is employed to estimate onroad mobile emissions. A 2002 inventory for Canadian emissions are used and are not projected to 2006.

We identify EGUs as peaking units by applying the Environmental Protection Agency's (EPA) peaking unit definition to the point source sector of the 2008 NEI with 2006 CEM data. The EPA categorizes an EGU as a peaking unit if it meets two criteria: 1) the unit's average annual capacity factor is less than 10% over a three-year period and 2) the unit's annual capacity factor is less than 20% in each of the three years. The capacity factor of a unit is defined as "either 1) the ratio of the unit's actual electrical output to the nameplate capacity times 8,760 or 2) the ratio of the unit's actual annual heat input to the maximum design heat input times 8,760"[40] (Figure B-1). Employing the EPA definition with maximum-rated hourly heat input data over a three-year period (2006-2008) we identify

544 point-sector (Integrated Planning Model (IPM) sector and non-IPM sector) peaking units in states that are members of PJM from the 2008 NEI. The 544 units represent 10-20% of total electricity generation in PJM on the highest demand days (Figure B-2). These units represent a lower bound for peak electricity generation and air quality impacts, as they are monitored and controlled peaking units fitted with CEMs and represent a subset of EGUs used for peak demand (e.g. behind-the-meter (BTM) generation is not considered and impacts air quality).[41] The second scenario, hereafter the "No Peak" emission scenario, contains the same emissions as the Base Case scenario with the exception that air pollution emissions from 544 peaking units are removed from the annual and hourly SMOKE emission inputs.

We employ the EPA's Clean Air Markets Division (CAMD)[28] database to separate EGUs into PJM and non-PJM categories. Of note, North Carolina and Michigan are considered non-PJM states in this analysis because a small number of facilities in these states are part of PJM. For the "No PJM" scenario, we remove 1,017 PJM EGUs from the annual and hourly SMOKE emission inputs. The "Only PJM" scenario removes all EGUs except the 1,017 identified PJM EGUs from the annual and hourly SMOKE emissions inputs. To identify and remove individual EGUs (No Peak, No PJM, and Only PJM scenarios) we use ORIS identifying numbers. Although we employ all publically available data, a key limitation is that not all facilities report ORIS ID numbers to the NEI and therefore some qualifying facilities may not properly identified.[38] This impairs the accuracy of air quality predictions. However the same method and data is used for all RTO emission scenarios and while exact air quality predictions are uncertain, the relative impact is more reliable.

3.3.3. Emissions, Temperature, and GDP Correlation

To investigate and identify factors that influence peaking unit usage and ultimately help derive predictive indicators, we compare both annual PJM and U.S. peaking unit heat input with PJM and U.S. average daily summer temperature and national gross domestic product (GDP) growth rate from 2004-2014. A yearly list of PJM and U.S. peaking units is complied to account for the variability in peaking unit usage. Annual PJM and national EPA-defined peaking unit heat input is obtained from CAMD.[28] PJM and U.S. summer temperatures are calculated as the average daily temperature in the months of June, July, and August from the National Centers for Environmental Information's climatological rankings database.[36] The U.S. GDP annual growth rate for 2004-2014 is obtained from The World Bank.[42]

3.3.4. Population-Weighted Concentration Analysis

To estimate locations with the highest potential for human exposure due to peaking units, we calculate population-weighted concentrations. We assign populations from the 2010 National Census Tracts Gazetteer,[43] based on fractional area overlap, and match with CMAQ concentrations of total $PM_{2.5}$ mass, primary unspeciated $PM_{2.5}$, sulfate, EC, and ozone from the simulations described above using the PseudoNetCDF, custom-developed software that is freely available

(http://www.github.com/barronh/pseudonetcdf). Equation 1[44] is used to calculate population-weighted concentrations in the study domain:

$$\frac{\sum(P_i \times C_i)}{\sum P_i} \text{ x } 1000\ people \qquad \text{[Equation 1]}$$

where P_i is the population of the grid cell and C_i is the concentration of the grid cell.[44] The resulting population-weighted figures are multiplied by 1000 people for scaling purposes and are subtracted from the population-weighted Base Case simulation to estimate the impact due only to the specific sector.

County-level total population data from the 2010 American Community Survey (spanning 2006-2010) from the U.S. Census Bureau (https://www.census.gov/programs-surveys/acs/) was downloaded and mapped using the "acs"[45] and "choroplethr"[46] packages in the R statistical language.[47] County-level median annual household income data from the 2010 American Community Survey (spanning 2006-2010) from the U.S. Census Bureau was also downloaded using the "acs" package, and was adjusted by 2015 cost of living state data created by the Council for Community and Economic Research (https://www.c2er.org) and download from the Missouri Economic Research and Information Center (https://www.missourieconomy.org/indicators/cost_of_living/). Normalized median annual household income by county was mapped using the "choroplethr" R package.

3.4. Results and Conclusions

3.4.1. Peaking Unit Usage

Between 2004-2014, PJM peaking unit heat inputs from CEM data vary annually and are dependent on meteorology and economic indicators (Figure 3-2). A statistically robust positive correlation ($r = 0.8$, $p = 3\times10^{-3}$) exists between average summer (e.g., when peaking units are typically the most employed: June, July, August) daily temperatures in the PJM states and annual peaking unit heat input. Repeating the analysis for all U.S. peaking unit heat inputs and U.S. average summer daily temperature yields the same statistically discernible positive correlation ($r = 0.8$, $p = 3\times10^{-3}$, Figure B-3a). National GDP growth rate has a weaker positive correlation ($r = 0.36$, $p < 0.3$) over the 10-year period. However, during the most recent U.S. recession (2007-2011), the national GDP growth rate was 2.5% or less. The positive correlation between PJM peaking unit heat inputs and national GDP growth rate for this time period is strong ($r=0.9$, $p < 0.05$). A stronger positive correlation exists between all U.S. peaking unit heat inputs and national GDP growth rate during the recession ($r = 0.94$, $p < 0.02$). This suggests that GDP growth rate affects electricity usage below an economic threshold and may imply that GDP growth is linked to disposable income available for air conditioning up to certain economic and temperature limits.

Among EGU fuel types, the price of coal (shipment price)[48] has the highest correlation with PJM peaking unit heat input ($r=0.32$, $p=0.5$). Though this association is suggestive of increased usage of peaking units when coal prices are above a threshold

price the relationship is not statistically discernable. The average annual prices of oil (first purchase price)[49] and gas (wellhead price)[50] are not correlated with PJM peaking unit usage ($r = 0.05$, $p = 0.87$ and $r = 0.17$, $p = 0.65$, respectively, Figure B-3b). The majority of PJM peaking units use gas as a fuel, only 1% use coal. Base and intermediate load facilities (i.e., non-peaking units) employ coal to a greater extent (55%) as a primary fuel. At times when coal price is high (relative to gas prices) increased use of peaking units may be more economically preferable. An important caveat is that these relationships are based on those peaking units in PJM that were identified through the facility ID matching process and not inclusive of all peaking units. Further, this analysis is limited to annual and not short-term (e.g., daily, weekly) price data, which is proprietary. The higher resolution data and the distribution of fuel usage and other factors in the unknown subset of peaking units could impact these relationships.

3.4.2. CMAQ-Simulated Air Quality Impacts

The maximum hourly $PM_{2.5}$ concentration is highest in the Other RTOs simulation (Figure 3-3c), and most widespread in the all PJM simulation (Figure 3-3b). Spatial distributions of hourly maximum concentrations differ for each simulation. For example, hourly maximum $PM_{2.5}$ concentrations from EGUs in RTOs other than PJM are relatively low in PJM states, however PJM EGUs result in areas of high (>25 $\mu g/m^3$) maximum hourly $PM_{2.5}$ mass (Figure 3-3, dark orange shading) throughout PJM states, as well as New York and Vermont. This finding suggests that not only do EGUs impact air quality in their own RTOs but also the air quality of states in other RTOs. Monitored and

controlled PJM peaking unit emissions contribute to high hourly maximum $PM_{2.5}$ concentrations. CMAQ-predicted hourly maximum surface $PM_{2.5}$ mass concentrations due to PJM peaking units are as high as 110 $\mu g/m^3$ in specific grid cells (Figure 3-3a), but the impacts are more localized than the other two scenarios. These peaking unit $PM_{2.5}$ maxima represent up to 87% of the total maximum hourly $PM_{2.5}$ ambient concentrations of the Base Case (Figure B-4a). While these maxima are spatially sporadic and temporally episodic, nine of the thirty-one days studied exhibit peaking unit contributions to total hourly $PM_{2.5}$ mass in at least one grid cell higher than 50% (Figure B-5).

The All PJM Case impacts sulfate mass concentrations throughout the modeling domain (> 6 $\mu g/m^3$ except in Michigan, Wisconsin, and Canada, Figure 3-3e). The Other RTO Case impacts maximum sulfate concentrations less (Figure 3-3f). Simulation analysis of the PJM Peaking Units Case indicates sporadic high maximum hourly sulfate concentrations in Northwestern WV, Virginia Beach, VA, and the Baltimore, MD/Washington, DC area (Figure 3-3d). These results highlight the localized impacts of peaking units. In small areas, primary unspeciated $PM_{2.5}$ contributes the most to peak concentrations of total $PM_{2.5}$ mass in two of three of the cases, i.e., PJM Peaking Unit Case and All PJM Case, both with maximum hourly concentrations in Northwestern West Virginia/Southern Pennsylvania and surrounding Baltimore, Maryland (up to 85 $\mu g/m^3$, Figures 3-4d and 3-4e). This implies that the highest values of primary $PM_{2.5}$ in the PJM area cannot be chemically identified. The maxima in both the PJM Peaking Unit and All PJM Cases on the Ohio/West Virginia border and Pennsylvania/West Virginia border correspond with the locations of PJM peaking units that use coal as the primary fuel (Figure 3-1a). Ambient concentrations of primary unspeciated $PM_{2.5}$ for the Other

RTOs Case are highest southeast of Chicago, Illinois, near Bay City, Michigan, and in Southern Indiana/Northern Kentucky. The maximum in Southern Indiana/Northern Kentucky corresponds with EGU locations in PJM (Figure 3-1b). This may indicate incorrect ORIS ID reporting to the NEI. Some highly populated areas (e.g., Baltimore, MD, Washington, DC) have localized high concentrations of EC attributable to PJM peaking units (Figure 3-4a), however maximum hourly EC from any of the simulations does not exceed 5 μg/m^3. It would be useful to evaluate these high primary unspeciated $PM_{2.5}$ concentrations to determine the validity of this result; however due to the episodic nature of these events[38] and the infrequency of chemically-speciated ambient measurements,[51, 52] this type of comparison for the studied heat wave is difficult with existing data sets.

3.4.3. Potential Exposure to Peaking Unit Incremental Exposure

PJM peaking unit emissions do not impact domain-wide air quality to the extent of all EGUs in a particular RTO, but have potential to adversely affect human health in major cities with disproportionate lower income levels. Coal PJM peaking plants are located in the Ohio Valley and on the West Virginia-Pennsylvania border (Figure 3-1a) and this is where their impacts on ambient concentrations are greatest. However, Washington, DC, Baltimore, MD, and New York City, NY have the highest population-weighted CMAQ-predicted hourly maximum concentrations (surrogate for exposure) for total $PM_{2.5}$ mass (Figure 3-5a), particulate sulfate (Figure 3-5d), elemental carbon (Figure B-9a) and primary unspeciated $PM_{2.5}$ (Figure B-9d). Implications for human health and

cross-state exposure to pollution from rural areas to highly populated cities, even for peaking units.

Approximately 60% of PJM Peaking units studied here are located in counties with median annual household incomes (adjusted by cost of living) between $40,000-$60,000 (Figures 3-1a and B-10). This income level represents the lower end of the income spectrum and more than half (60%) of the population in the domain live in counties characterized by this median income range. No identified peaking units are located in counties below $30,000 or above $90,000. A proprietary county-level cost of living index exists and may result in stronger associations, particularly in states with large ranges of cost of living (e.g., New York).

PJM peaking unit heat inputs positively correlate with average summer daily temperatures and national GDP growth rate and air quality managers can capitalize on that knowledge to develop strategies to protect human health. Emissions from EGUs in the RTOs studied here impact regional air quality in geographic regions outside of their electricity network. PJM peaking units modeled in this study impact local air quality largely due to emissions of primary unspeciated $PM_{2.5}$. Further, peaking units perturb air quality in highly populated areas of this domain and in lower income areas, even when located in rural areas. Peaking units not accounted for in this work because they do not employ CEMs, also impact air quality and health and further effort to include these units in air quality modeling is needed. Holistic energy planning with regional and local analyses is needed to determine the largest sources of air pollution-induced health impacts in order to maintain sustainable and healthy communities.

3.5. References

1. Miller, N. L.; Hayhoe, K.; Jin, J.; Auffhammer, M., Climate, Extreme Heat, and Electricity Demand in California. *J. Appl. Meteorol. Clim.* **2008,** *47*, (6), 1834-1844.

2. Amato, A. D. R., M.; Kirshen, P.; Horwitz, J., Regional Energy Demand Responses To Climate Change: Methodology And Application To The Commonwealth Of Massachusetts. *Climatic Change* **2005,** *71*, (1-2), 175-201.

3. Ali, M.; Iqbal, M. J.; Sharif, M., Relationship between extreme temperature and electricity demand in Pakistan. *Int. J. Energ. Environ. Eng.* **2013,** *4*, (1).

4. Colombo, A. F.; Etkin, D.; Karney, B. W., Climate Variability and the Frequency of Extreme Temperature Events for Nine Sites across Canada: Implications for Power Usage. *J. Climate* **1999,** *12*, (8), 2490-2502.

5. Gastli, A.; Charabi, Y.; Alammari, R. A.; Al-Ali, A. M. In *Correlation between climate data and maximum electricity demand in Qatar*, 2013, IEEE: pp 565-570.

6. Protection of the Environment Title 40 Subchapter C Part 64: Compliance Assurance Monitoring. Federal Code of Regulations, 2002. http://www.gpo.gov/fdsys/pkg/CFR-2002-title40-vol12/pdf/CFR-2002-title40-vol12-part64.pdf (August 30, 2015).

7. Protection of the Environment Title 40 Subchapter C Part 74: Sulfur Dioxide Opt-Ins. Federal Code of Regulations, 2002. http://www.gpo.gov/fdsys/pkg/CFR-2002-title40-vol13/pdf/CFR-2002-title40-vol13-part74.pdf (August 30, 2015).

8. PJM, Who We Are. http://www.pjm.com/about-pjm/who-we-are.aspx (October 8, 2015).

9. Jacob, D. J.; Winner, D. A., Effect of climate change on air quality. *Atmos. Environ.* **2009,** *43*, (1), 51-63.

10. Mwaniki, G. R.; Rosenkrance, C.; Wallace, H. W.; Jobson, B. T.; Erickson, M. H.; Lamb, B. K.; Hardy, R. J.; Zalakeviciute, R.; VanReken, T. M., Factors contributing to elevated concentrations of PM2.5 during wintertime near Boise, Idaho. *Atmos. Pol. Res.* **2014,** *5*, (1), 96-103.

11. Mickley, L. J.; Jacob, D. J.; Field, B. D.; Rind, D., Effects of future climate change on regional air pollution episodes in the United States. *Geophys. Res. Lett.* **2004,** *31*, (24).

12. Meehl, G. A.; Tebaldi, C., More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **2004,** *305*, (5686), 994-997.

13. Chandramowli, S. N.; Felder, F. A., Impact of climate change on electricity systems and markets – A review of models and forecasts. *Sustain. Energ. Tech. A.* **2014,** *5*, 62-74.

14. PJM Resource Adequacy Planning Department, *PJM Load Forecast Report*; 2015.

15. U.S. EPA, The 2011 National Emissions Inventory v.1. http://www.epa.gov/ttnchie1/net/2011inventory.html (September 2, 2015).

16. U.S. EPA, National Summary of Sulfur Dioxide Emissions. http://www.epa.gov/cgi-bin/broker?_service=data&_debug=0&_program=dataprog.national_1.sas&polchoice=SO2 (September 2, 2015).

17. U.S. EPA, National Summary of Nitrogen Oxides Emissions. http://www.epa.gov/cgi-bin/broker?_service=data&_debug=0&_program=dataprog.national_1.sas&polchoice=NOX (September 2, 2015).

18. Burnett, R. T.; Pope, C. A.; Ezzati, M.; Olives, C.; Lim, S. S.; Mehta, S.; Shin, H. H.; Singh, G.; Hubbell, B.; Brauer, M.; Anderson, H. R.; Smith, K. R.; Balmes, J. R.; Bruce, N. G.; Kan, H. D.; Laden, F.; Pruss-Ustun, A.; Michelle, C. T.; Gapstur, S. M.; Diver, W. R.; Cohen, A., An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environ. Health Persp.* **2014,** *122*, (4), 397-403.

19. Turner, M. C.; Krewski, D.; Pope, C. A.; Chen, Y.; Gapstur, S. M.; Thun, M. J., Long-term Ambient Fine Particulate Matter Air Pollution and Lung Cancer in a Large Cohort of Never-Smokers. *Am. J. of Resp. and Crit. Care* **2011,** *184*, (12), 1374-1381.

20. Pope, C. A.; Burnett, R. T.; Turner, M. C.; Cohen, A.; Krewski, D.; Jerrett, M.; Gapstur, S. M.; Thun, M. J., Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure-Response Relationships. *Environ. Health Persp.* **2011,** *119*, (11), 1616-1621.

21. Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y. L.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.; Smith, S. C.; Whitsel, L.; Kaufman, J. D.; Epidemiol, A. H. A. C.; Dis, C. K. C.; Metab, C. N. P. A., Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association. *Circulation* **2010,** *121*, (21), 2331-2378.

22. Fang, Y.; Naik, V.; Horowitz, L. W.; Mauzerall, D. L., Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present. *Atmos. Chem. Phys.* **2013,** *13*, (3), 1377-1394.

23. Lepeule, J.; Laden, F.; Dockery, D.; Schwartz, J., Chronic Exposure to Fine Particles and Mortality: An Extended Follow-up of the Harvard Six Cities Study from 1974 to 2009. *Environ. Health Persp.* **2012,** *120*, (7), 965-970.

24. Twomey, S., The Influence of Pollution on the Shortwave Albedo of Clouds. *J. Atmos. Sci.* **1977,** *34*, (7), 1149-1152.

25. Jacobson, M. Z., Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols. *Nature* **2001,** *409*, (6821), 695-697.

26. Leibensperger, E. M.; Mickley, L. J.; Jacob, D. J.; Chen, W. T.; Seinfeld, J. H.; Nenes, A.; Adams, P. J.; Streets, D. G.; Kumar, N.; Rind, D., Climatic effects of 1950-2050 changes in US anthropogenic aerosols - Part 1: Aerosol trends and radiative forcing. *Atmos. Chem. Phys.* **2012,** *12*, (7), 3333-3348.

27. Leibensperger, E. M.; Mickley, L. J.; Jacob, D. J.; Chen, W. T.; Seinfeld, J. H.; Nenes, A.; Adams, P. J.; Streets, D. G.; Kumar, N.; Rind, D., Climatic effects of 1950-2050 changes in US anthropogenic aerosols - Part 2: Climate response. *Atmos. Chem. Phys.* **2012,** *12*, (7), 3349-3362.

28. U.S. EPA, Clean Air Markets Division (CAMD), Air Markets Program Data. Ed. 2015.

29. EPA *Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans. Appendix I: Methods for Quantifying Energy Efficiency and Renewable Energy Emission Reductions*; 2012.

30. Creswell, J.; Gebeloff, R., Traders Profit as Power Grid Is Overworked. *New York Times* August 14, 2014.

31. N.Y. State, *Energy Price and Demand Annual Long-Term Forecast: 2009-2028. New York State Energy Plan 2009*; 2009.

32. U.S. Energy Information Agency (EIA), *Assumptions to the Annual Energy Outlook 2014: Electricity Market Module*; 2014.

33. Federal Energy Regulatory Commission (FERC), *Electric Power Market Summary: Summer 2006*; 2006.

34. Fann, N.; Lamson, A.D.; Anenberg, S.C.; Wesson, K.; Risley, D.; Hubbell, B.J., Estimating the National Public Health Burden Associated with Exposure to Ambient PM2.5 and Ozone. *Risk Anal.* **2012**, *32*, (1), 81-95.

35. National Centers for Environmental Information, National Oceanic and Atmospheric Administration, State of the Climate: National Overview for July 2006. http://www.ncdc.noaa.gov/sotc/national/2006/7 (August 30, 2015).

36. National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Climatological Rankings. http://www.ncdc.noaa.gov/temp-and-precip/climatological-rankings/index.php (August 30, 2015).

37. U.S. EPA, Download Detailed AQS Data. http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm (October 8, 2015).

38. Farkas, C. M.; Moeller, M. D.; Felder, F. A.; Baker, K. R.; Rodgers, M.; Carlton, A. G., Temporalization of Peak Electric Generation Particulate Matter Emissions during High Energy Demand Days. *Environ. Sci. Tech.* **2015,** *49*, (7), 4696-4704.

39. Carlton, A. G.; Baker, K. R., Photochemical Modeling of the Ozark Isoprene Volcano: MEGAN, BEIS, and Their Impacts on Air Quality Predictions. *Environ. Sci. Tech.* **2011,** *45*, (10), 4438-4445.

40. U.S. EPA, Plain English Guide to the Part 75 Rule, Appendix E: Methodology for Gas-Fired and Oil-Fired Peaking Units. In Vol. Part 75.

41. Zhang, X.; Zhang, M. K., Demand Response, Behind-the-Meter Generation and Air Quality. *Environ. Sci. Tech.* **2015,** *49*, 1260-1267.

42. The World Bank, GDP Growth Rate. In 2015.

43. 2010 National Census Tracts Gazetteer File. In U.S. Census Bureau, Ed. 2012.

44. Carlton, A. G.; Pinder, R. W.; Bhave, P. V.; Pouliot, G. A., To What Extent Can Biogenic SOA be Controlled? *Environ. Sci. & Tech.* **2010,** *44*, (9), 3376-3380.

45. Glenn, E. H. *acs: Download, manipulate, and present data from the US Census American Community Survey*; 2014.

46. Lamstein, A.; Johnson, B. P. *choroplethr: Simplify the Creation of Choropleth Maps in R*; 2016.

47. Team, R. C. *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

48. U.S. Energy Information Agency (EIA), Coal Shipments to the Electric Power Sector: Price, by plant state. In 2015.

49. U.S. Energy Information Agency (EIA), Domestic Crude Oil First Purchase Prices by Area. In 2015.

50. U.S. Energy Information Agency (EIA), Natural Gas Prices. In 2015.

51. IMPROVE Interagency Monitoring of Protected Visual Environments. http://vista.cira.colostate.edu/improve/ (August 30, 2015).

52. EPA, U. S. Chemical Speciation - General Information. http://www.epa.gov/ttnamti1/specgen.html (August 30, 2015).

Emission Scenarios	
Name	**Description**
Base Case	Base Case emissions as described in 3.3.1
No Peak	Base emissions scenario with 544 PJM peaking units removed from the annual NEI and CEM hourly data
No PJM	Base emissions scenario with all PJM EGUs identified by ORIS numbers removed from the annual NEI and CEM hourly data (1,017 EGUs removed)
Only PJM	Base emissions scenario with EGUs in other RTOs (i.e., not in PJM) identified by ORIS numbers removed from the annual NEI and CEM hourly data (only 1,017 EGUs remain)
CMAQ Output Analyses	
Name	**Formula**
PJM Peaking Units	Base Case scenario – No Peak scenario
All PJM	Base Case scenario – No PJM scenario
Other RTOs	Base Case scenario – Only PJM scenario

Table 3-1: Emission scenarios and CMAQ output analyses and descriptions

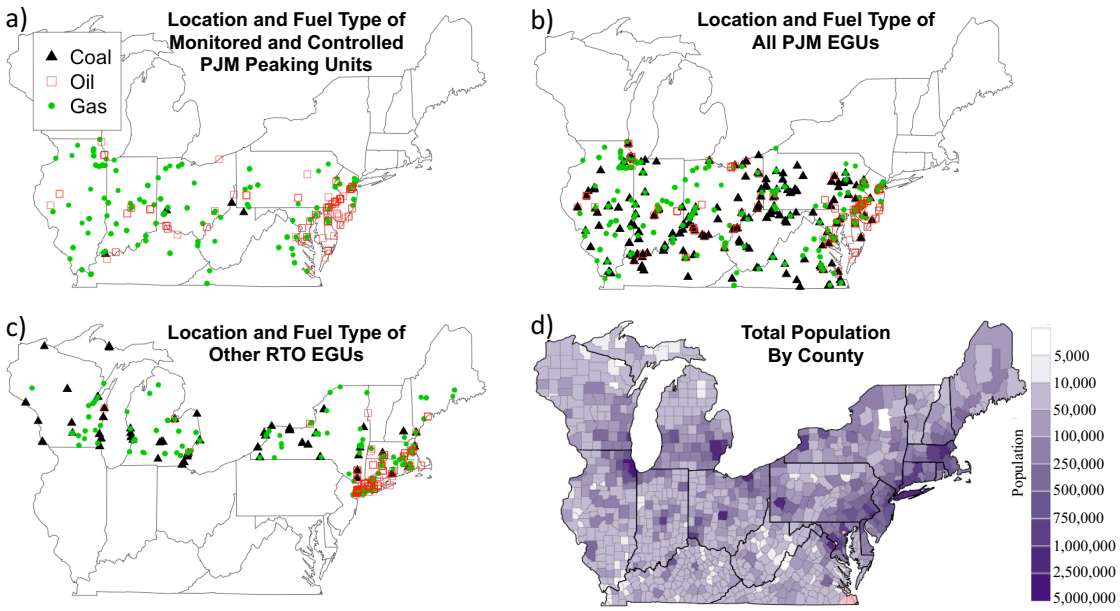

Figure 3-1: Locations and fuel types of EGUs from the PJM Peaking Unit Case (a), All PJM Case (b), and Other RTOs Case (c). Total population by county from the American Community Survey from 2006-2010 (d). The pink coloring in Virginia represents areas with missing data.

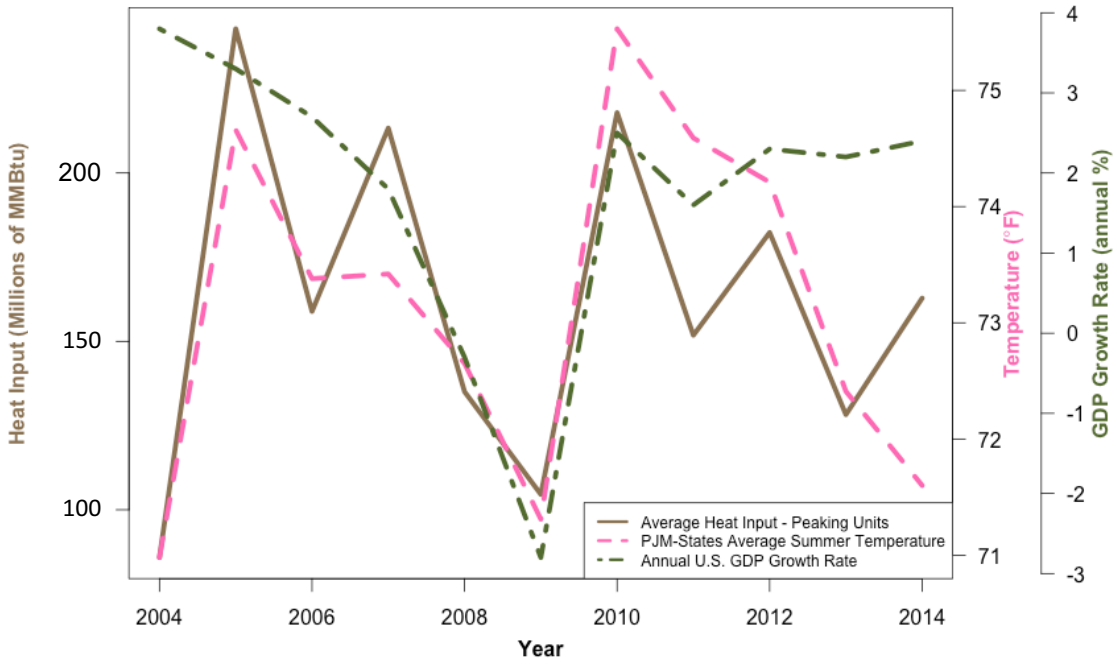

Figure 3-2: PJM annual heat input (millions of MMBtu) of the 544 peaking units studied here (brown) compared to average summer (June, July, August) temperature (pink dashed) and annual U.S. gross domestic product (GDP) growth rate (green dash-dot). Sources: EPA clean air markets division (CAMD - heat input), NCDC (temperatures), The World Bank (U.S. GDP growth rate).

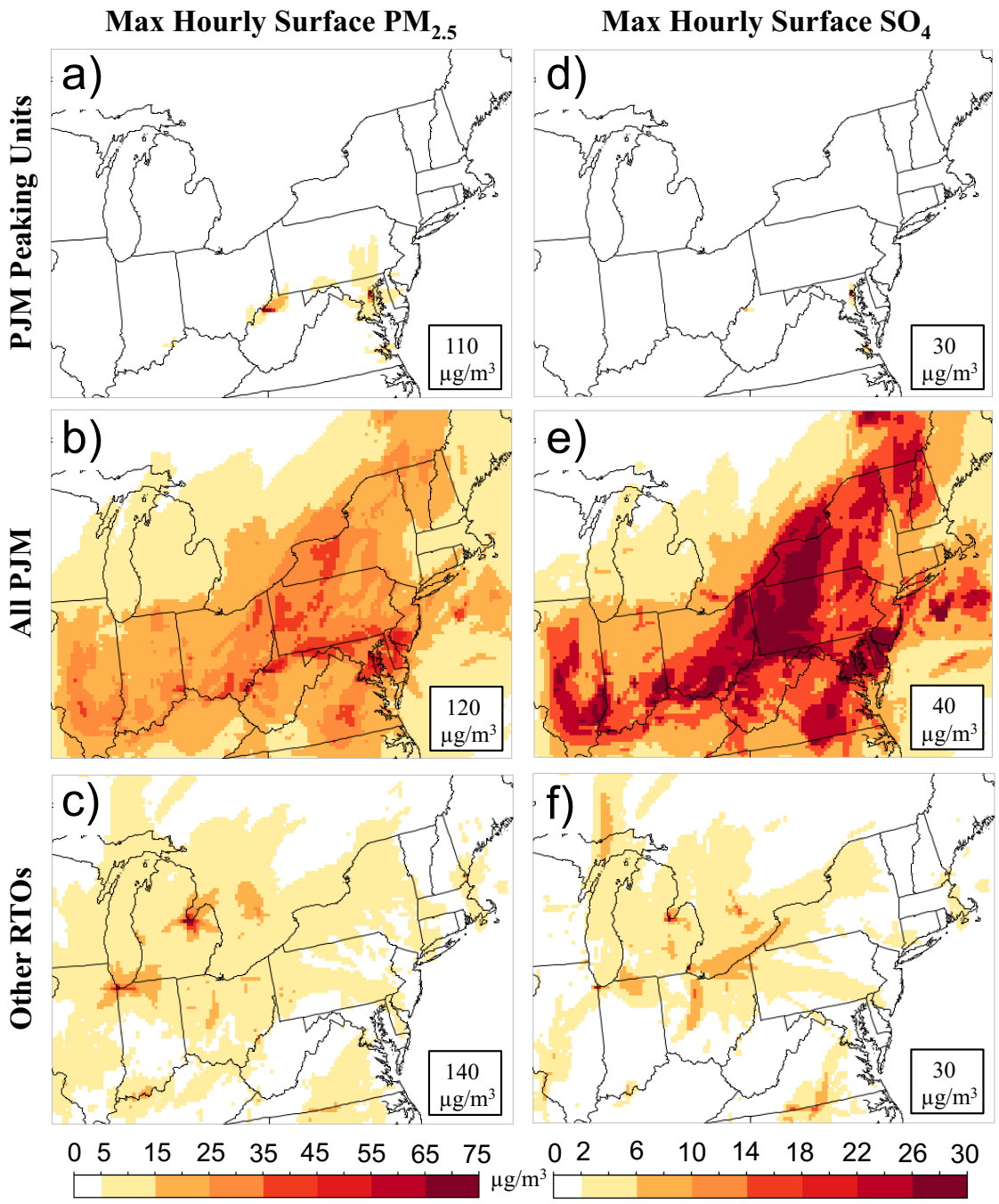

Figure 3-3: CMAQ-predicted maximum hourly ambient concentrations of $PM_{2.5}$ and sulfate mass for PJM Peaking Units, All PJM, and Other RTOs at the surface from July 1-31, 2006. The boxed numbers represent the highest hourly maximum concentration that occurred over the month of July 2006.

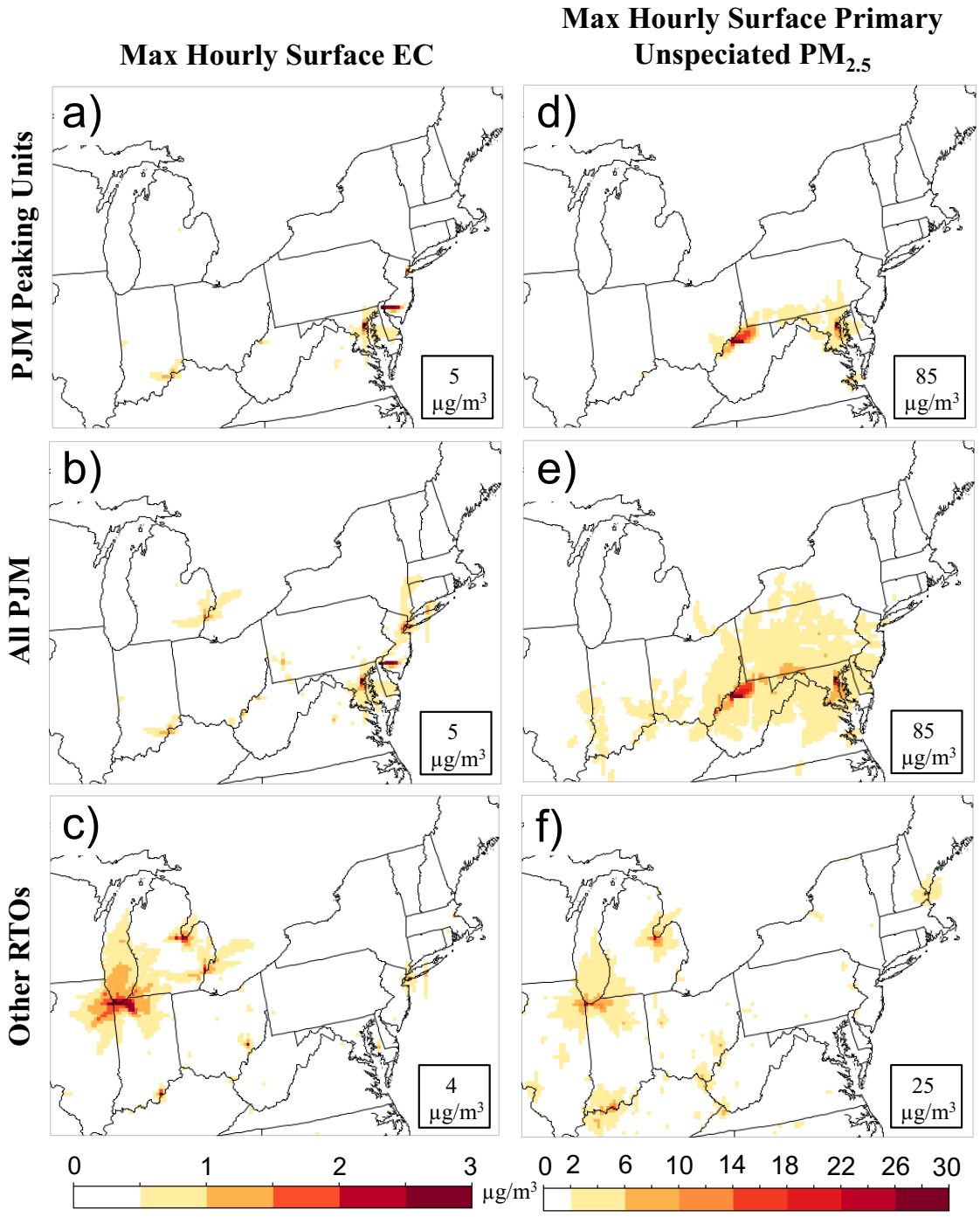

Figure 3-4: CMAQ-predicted maximum hourly ambient concentrations of EC and primary unspeciated $PM_{2.5}$ mass for PJM Peaking Units, All PJM, and Other RTOs at the surface from July 1-31, 2006. The boxed numbers represent the highest hourly maximum concentration that occurred over the month of July 2006.

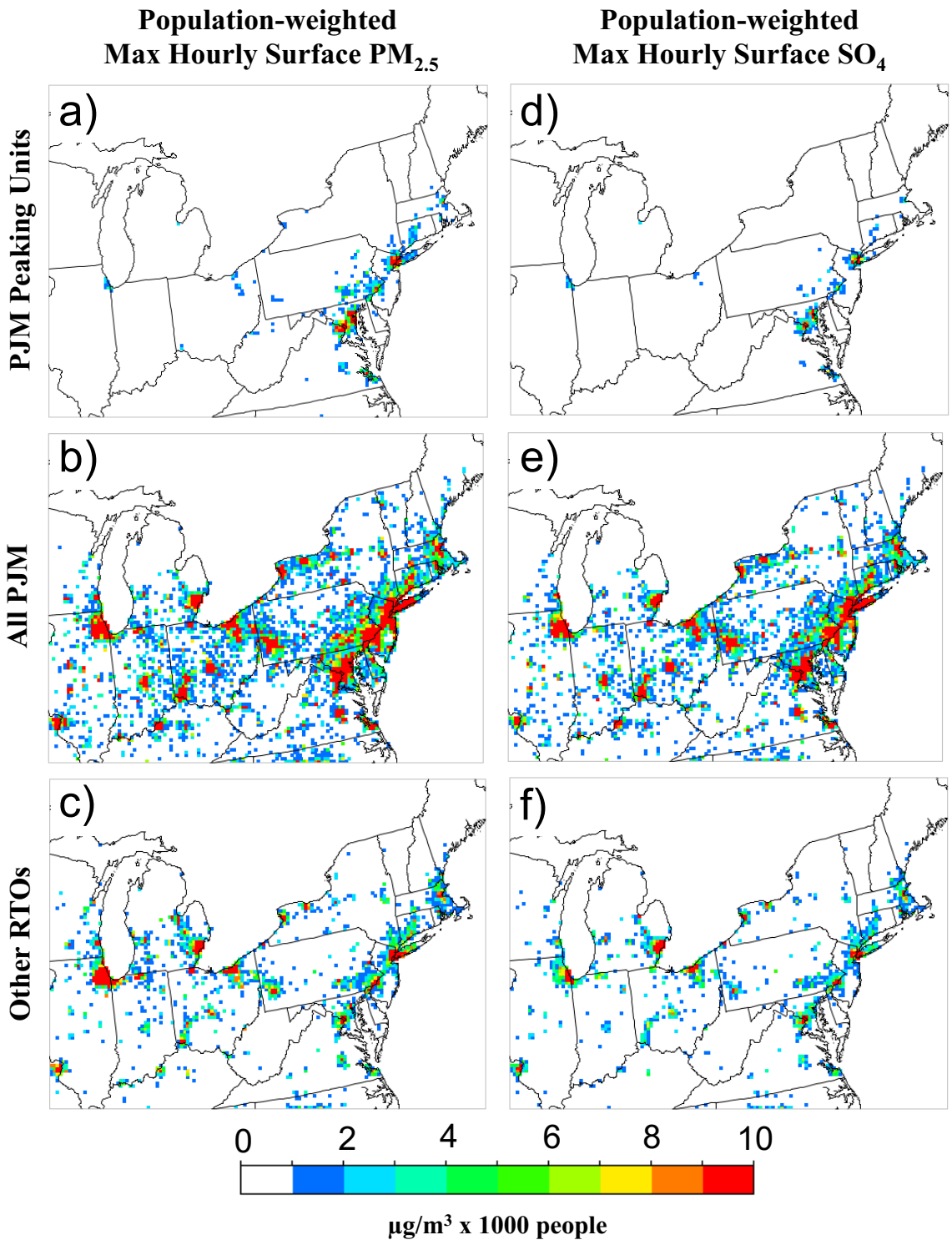

Figure 3-5: CMAQ-predicted maximum hourly population-weighted ambient concentrations of $PM_{2.5}$ and sulfate mass for PJM Peaking Units, All PJM, and Other RTOs at the surface from July 1-31, 2006.

4. CHAPTER 4

SENSITIVITY ANALYSIS OF CMAQ ESTIMATION OF POTENTIAL EXPOSURE TO $PM_{2.5}$ AND OZONE CONCENTRATIONS

Material in this chapter to be submitted as:

Farkas, C.M.; Carlton, A.G., Sensitivity Analysis of CMAQ Estimation of Potential Exposure to $PM_{2.5}$ and Ozone Concentrations. *Atmos. Environ.* *(in prep)*

4.1. Abstract

Identifying and reducing sources of uncertainty in photochemical models improves air quality predictions used to influence air quality management policies for the protection of human health and welfare. Inaccuracies in three-dimensional photochemical transport models employed to predict air quality could arise from a range of sources, including meteorological inputs, chemical processes, and emission inventories. The planetary boundary layer (PBL) height and the mobile emission sector have been identified as leading causes of uncertainty in predictions in surface level concentrations of fine particle ($PM_{2.5}$) and ozone. Exploring the impacts of uncertainty in these variables on population-weighted concentrations may provide insight into the most important factors that confound potential exposure estimates. In this study we employ the

Community Multiscale Air Quality (CMAQ) model to perturb variables of PBL height and onroad mobile emissions of nitrogen oxides ($NO + NO_2 = NO_X$), primary organic carbon (POC), and unspeciated $PM_{2.5}$ (PM_{other}) in eight separate simulations during the 2006 U.S. heat wave in the northeast quarter of the U.S. We population-weight the resulting concentrations as a potential exposure index to identify the controlling uncertainties potentially most impactful to CMAQ-estimated human exposure estimates that would be calculated using CMAQ predictions of ambient concentrations. This is in contrast to existing CMAQ sensitivity studies that evaluate controlling uncertainties for average domain-wide ambient concentration bias. Of the studied variables, average and maximum hourly $PM_{2.5}$ predicted ambient concentrations are most sensitive to PBL height perturbations in locations over the ocean. In contrast, changes in onroad POC emissions result in the largest impact on population-weighted concentrations. Both predicted ambient and population-weighted ozone mixing ratios are most sensitive to onroad NO_x emissions. The findings suggest efforts to improve CMAQ simulations for development of effective air quality management policies to protect human health should concentrate on reducing uncertainties in onroad mobile POC and NO_x emissions.

4.2. Introduction

Air quality models, such as the Community Mutli-scale Air Quality (CMAQ) model, provide an approximation of the atmosphere and the physical and chemical interactions with air pollutant emissions that impact air quality. When utilizing model results to develop plans for air quality management, it is important to consider that air

quality models, such as CMAQ, are estimations of a physical reality that simulate the relationships between chemical reactions and transport with measured or estimated inputs (e.g. meteorological inputs, emission inventories), all of which include uncertainty.[1, 2] For instance, a single model variable in the domain used here (e.g. WRF-predicted air temperature) represents 10^7 uncertain CMAQ input values for a one month model simulation,[3] generating many possible points of inaccuracy. Identifying the sources of largest uncertainty in CMAQ and their impacts is critical for accurate modeling of pollutants, such as fine particulate matter ($PM_{2.5}$) and tropospheric ozone that are linked to acute and chronic cardiovascular and pulmonary diseases and autism.[4-13] Focusing, however, on domain-wide average concentration and error relative to measurement data,[14] may or may not identify the most critical variables for predictive skill (i.e., the most harmful pollutants in the highest populated areas).

Estimating human exposure to $PM_{2.5}$ or ozone whether outdoors, indoors, or in-vehicle can have many sources of uncertainty. In-vehicle $PM_{2.5}$ exposure estimates can be highly variable depending on in-vehicle to ambient ratios and have been estimated to represent up between 6%-57% of total $PM_{2.5}$ exposure.[15] Further, concentration gradients of NO_x close to roadways (~100-400m) have been found to change on hourly time-scales and depend on variations in boundary layer height, traffic flow rate and wind speed.[16, 17] Estimation of the transport of outdoor pollutants indoors and their impacts on residential indoor air quality is dependent on chemical composition and can be simulated with outdoor-to-indoor transport models.[18] Most uncertainty analyses employing CMAQ have identified the largest contributors to propagated uncertainty on domain-wide ozone concentrations (i.e., chemical mechanism, turbulent closure, NO_2 photolysis rate).[3, 19-22]

Previous research has estimated uncertainty of $PM_{2.5}$ concentrations that propagates from coupling models to a source-to-dose modeling system to understand exposure.[23] I am not aware of any CMAQ uncertainty analysis focused not on domain averages, but rather which factor creates the largest uncertainty in an exposure potential to $PM_{2.5}$ and ozone calculated based on CMAQ ambient predictions. This represents a key knowledge gap that hinders development of effective air quality strategies to protect human health.

Uncertainty analyses use a variety of techniques to test the sensitivity of pollutant concentrations to model uncertainties. Monte Carlo analyses have been used to evaluate several model inputs at once;[21, 22] however because of the large number of simulations and computation time required, a reduced form of the numerical photochemical model is often used, introducing additional uncertainties.[1] The complex non-linear system of chemical reactions that are CMAQ's chief strength and focus here are removed in such approaches. Brute force analysis is a technique of testing the sensitivity of model predictions to a single variable by running a new model simulation after each change to determine its impact on predictions. In the past, brute force techniques have been used to apportion sources of $PM_{2.5}$[24] and to define the controllability of biogenic secondary organic aerosol (SOA)[25] in CMAQ. The decoupled direct method (DDM) uses the 1st and 2nd derivatives of the chief equations of the model to create sensitivity coefficients for clear sky conditions.[3, 26-28] Three-dimensional DDM analyses of inorganic sulfate PM mass changes due to emissions of SO_2 have been shown to compare well with results of brute force approaches, but more in-direct relationships between some gaseous and aerosol species do not compare as well to brute force results.[29] Further, evaluation during extreme events (e.g. heat waves and maximum $PM_{2.5}$ concentrations) is needed because

they can drive adverse human health impacts but are not described well in approaches that employ average or typical conditions.

There exist many avenues through which inaccuracies can lead to uncertainty in air quality predictions. While the resolution of grid cells chosen for an air quality simulation can change the meteorological and transport parameters used within the model, several studies have found that resolutions finer than 12 km x 12km grid cells (used in this work) are unlikely to significantly alter predictions of pollutant concentrations and health impacts.[30-32] Meteorology, in particular, the planetary boundary layer (PBL) height modulates the diurnal pattern observed in pollutant concentrations and is found to have an effect on $PM_{2.5}$ predictions.[33] The PBL height is subject to measurement limitations with current technologies. Observation networks are sparse[34] with approximately 3% of weather stations that estimate PBL,[35] leading to uncertainties in PBL height and therefore air quality.[36] Balzarini et al. (2014)[37] analyzed five PBL schemes in the Weather Research and Forecasting (WRF) model over Northern Italy and found mean biases ranging anywhere from -43.82 m to +248.05 m, depending on the scheme. In Southeast Texas, Kolling et al. (2012) found the use of the asymmetric convective model version 2 (ACM2) PBL scheme in the WRF model yielded modeled-observed PBL differences from -275m to 260m.[38]Uncertainty in cloud modeling within CMAQ has been shown to greatly impact air quality simulations through limiting of ozone formation[39] and gas-to-cloud ice partitioning that can impact aloft pollutant concentrations.[40] Further, biogenic emission inaccuracies have led to differences in predictions of up to 5% of the national Ambient Air Quality Standard (NAAQS) for ozone and up to 2% for $PM_{2.5}$ mass.[41] Federal and state air quality regulations aim to

reduce emissions from the controllable sectors and improving inaccuracies within these sectors is critical to designing effective air quality management strategies.

This study employs the brute force sensitivity analysis to investigate several sources of uncertainty, previously identified (Table 4-1), that propagate errors in the CMAQ model and affect air quality predictions of $PM_{2.5}$ and ozone. Propagating these uncertainties through an air quality model may help to determine if model-predicted ambient concentrations in simulation results fall outside a range of uncertainty. I focus on uncertainty in controllable anthropogenic emissions and PBL height.

The energy sector, including motor vehicles and electric generating unit (EGU) sectors, are the largest contributors to controllable emissions that degrade air quality and impact climate. EGU emissions of sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and primary $PM_{2.5}$ impact regional ambient ozone and $PM_{2.5}$ concentrations.[42-45] Point source emissions are the least uncertain because exact locations are known and unchanging and annual total emissions are reported, though errors exist.[46-48] I focus on mobile emissions, the largest controllable source of primary $PM_{2.5}$ and nitrogen oxides (NO_x, where $NO_x = NO + NO_2$),[49] which contributes to ozone, particulate organics and nitrate in close proximity to people. Due to the inherently stochastic nature of mobile emissions (i.e. large diversity in fleet model years, traffic patterns, etc.) emission estimates remain uncertain, especially in urban areas at street level,[22, 50, 51] where potential exposure is high due to close proximity of sources (motor vehicles) and receptors (humans). In Atlanta, uncertainties in mobile NO_x emissions impact predicted ozone concentrations to the largest extent relative to all anthropogenic sectors.[22] Zavala et al. (2009)[50] found NO overpredictions of 14-20% and more than a factor of 4 underestimation of $PM_{2.5}$ in

mobile emissions in Mexico City. Primary organic carbon (POC), a chemically specific subset of $PM_{2.5}$ emissions, may be the most uncertain due to additional emission factor errors (e.g., intermediate-volatility organic compounds (IVOCs)[52]).[53] During the spring and summer 2006-2008, CMAQ-estimated total particulate carbon (organic carbon (OC) + elemental carbon (EC)) was biased low relative to surface mass measurements due to underprediction of organic carbon;[33] however this could also be a consequence of errors in the model description of secondary formation as well.[54] The same study found an underprediction in "other" $PM_{2.5}$ species (the sum of $PM_{2.5}$ species without sulfate, nitrate, ammonium, and total carbon) during the summer in the areas included in our study domain, which has been shown to cause underpredictions in total $PM_{2.5}$ mass.[55] Nationally, approximately 30% of all primary $PM_{2.5}$ mass emitted in 2005 was POC.[56] The largest national contributor, biomass burning, occurs mostly in the Western U.S., while effective public policy to abate air quality issues reduces controllable emissions. Therefore, I focus on uncertainty of POC in the largest controllable emission sources, which originate from the transportation sector.[56]

4.3. Methods

I perform eight CMAQ brute force perturbations of PBL height, NO_x, primary organic carbon (POC), and PM_{other} ($PM_{2.5}$ not categorized as organic carbon (OC), elemental carbon (EC), sulfate, nitrate, H_2O, Na^+, Cl^-, NH_4^+, non-carbon organic matter (NCOM), Al, Ca^{2+}, Fe, Si, Ti, Mg^{2+}, K^+, or Mn) in the onroad sector of the 2008 National emissions inventory (NEI) in the Northeast U.S. from July 1, 2006 to July 31, 2006. This

period is characteristic of Northeast stagnation heat wave events when ambient concentrations of ozone and $PM_{2.5}$ are often high. We analyze sensitivities of $PM_{2.5}$ and ozone concentrations in the first model layer because that is where people live and are exposed to pollution. Further, the concentrations are population-weighted using the 2010 national census to determine which variables have the largest effect on potential human exposure.

4.3.1. Sensitivity Simulations

The Base Case simulation used to compare to all perturbations is consistent with the sensitivity simulation described in detail in Farkas et al. (2015).[47] Briefly, a 12 km by 12 km CMAQ grid with 34 vertical layers up to 50 mb is applied to anthropogenic emissions from the 2008 NEI version 2 with continuous emission monitor (CEM) data from 2006 and additional identifiers in the point source sector emissions to incorporate additional CEM data. The Advanced Research WRF core model (version 3.1)[57] is employed to generate the gridded meteorological data for the study period. The Biogenic Emissions Inventory (BEISv3.14)[58] is used to generate biogenic emissions with day- and hour-specific solar radiation and temperatures from the WRF model. The Motor Vehicle Emission Simulator (MOVESv2010b) estimates the onroad mobile emissions for the study period. The model is simulated from June 28, 2006 through July 31, 2006, the first three model days are excluded from the analysis to allow for model spin up.

Each sensitivity case contains one or two modifications (Table 4-1); all other variables remain consistent with the Base Case. Case 1 and Case 2 perturb the WRF

gridded meteorological data to increase and decrease the PBL by 100 m, respectively. Full analyses of different WRF physics options have been discussed in detail,[57, 59] I focus here on the meteorological variable shown to have the largest uncertainty similar to previous work. Case 3 doubles the emissions of NO and NO_2 in the MOVES-generated onroad mobile sector, while Case 4 decreases both by half. Case 5 and 6 alter the onroad mobile sector emissions of POC also by a factor of two, respectively. Case 7 doubles the PM_{other} emissions from the onroad mobile sector, while Case 8 divides the variable in half. It is important to note that this designation is different than the "other" $PM_{2.5}$ described in previous literature above[33, 55] due to a recent development of the CMAQ model. Each respective input file was altered using the combine tool included with the CMAQ source code. Cases 3-8 are processed through the Sparse Matrix Operator Kernel Emissions (SMOKEv3.5) modeling system[60] after each alteration to incorporate the changes of each case into the mobile sector and prepare CMAQ-ready gridded emissions.

4.3.2. Population-weighted Concentrations

The population-weighted concentrations are calculated using census tract data from the 2010 U.S. National Census.[61] The centroid of each census tract is matched to its corresponding CMAQ grid cell, and weighted pollutant concentrations are calculated with the equation below[25] using the Python language (www.python.org).

$$\frac{\sum(P_i \times C_i)}{\sum P_i}, \qquad \text{[Equation 1]}$$

where P_i is the population of the grid cell and C_i is the CMAQ concentration of the grid cell.

4.4. Results

4.4.1. Predicted $PM_{2.5}$ Concentration Sensitivity

Of the studied variables, POC emission changes (Cases 5 and 6) in the onroad mobile sector have the largest impact on population-weighted hourly average $PM_{2.5}$ concentrations (Figures 4-1a and 4-1b). In nearly all-major cities in the domain, increases in onroad POC emissions result in increases in population-weighted hourly average $PM_{2.5}$ up to 13 $\mu g/m^3$ for every 1000 people when compared to the base case. Decreases in onroad POC emissions have the opposite effect in the same areas, with decreases in population-weighted hourly average $PM_{2.5}$ up to -6.8 $\mu g/m^3$ for every 1000 people. Conversely, changes in domain-wide hourly average ambient $PM_{2.5}$ concentrations are less substantial. When onroad mobile POC emissions are doubled (Case 5), $PM_{2.5}$ increases up to 7% (1.1 $\mu g/m^3$, Figures 4-2a, 4-3a), and reducing POC by half (Case 6) decreases average ambient $PM_{2.5}$ concentrations up to 3.5% (< 1$\mu g/m^3$, Figures 4-2b, 4-3b). Maximum hourly $PM_{2.5}$ concentrations increase by up to 8 $\mu g/m^3$ when the onroad mobile POC is doubled and up to 0.9 $\mu g/m^3$ when POC is halved (Figures 4-4a and 4-4b, respectively). The unbalanced difference in the addition/reduction of POC emissions to potential ambient $PM_{2.5}$ exposure highlights the complex non-linear chemistry associated with air quality. This demonstrates the usefulness of a brute force technique for the first

analysis of this type, as an inherent limitation of reduced form model may not represent the non-linear relationship well.

Whereas population-weighted hourly average $PM_{2.5}$ concentrations are most sensitive to perturbations of onroad mobile POC, domain-wide ambient $PM_{2.5}$ concentrations are most sensitive to changes in PBL height (Table 4-2). Increasing the PBL height by 100 m (Case 1) produces hourly average ambient $PM_{2.5}$ concentration increases of up to 2.1 $\mu g/m^3$ (Figure 4-5a), an increase of 106% (Figure 4-6a), in the eastern portion of the domain, with maximum hourly increases in $PM_{2.5}$ concentrations up to 44 $\mu g/m^3$ (Figure 4-7a). However, the majority of the highest hourly $PM_{2.5}$ concentrations occur over the ocean, greatly reducing the population-weighted impact (Figures 4-8a, up to 5.9 $\mu g/m^3$ for every 1000 people). Case 1 produces hourly average decreases east of Chicago, IL and north of New York, New Hampshire, and Vermont in Canada (up to -1.8 $\mu g/m^3$ or -8.6%) and population-weighted $PM_{2.5}$ concentrations decreases in Chicago, IL and Detroit, MI (up to -3.0 $\mu g/m^3$ for every 1000 people). Decreasing the PBL by 100 m (Case 2) results in decreases in hourly average $PM_{2.5}$ concentrations of up to -0.9 $\mu g/m^3$ (-10%) over the ocean and increases of up to 1.2 $\mu g/m^3$ (+10%) in Canada (Figures 4-5c and 4-6c). Case 2 produces increases in population-weighted $PM_{2.5}$ concentrations of up to 4 $\mu g/m3$ for every 1000 people (Figure 4-8c) in the urban areas in the western portion of the domain and decreases in the eastern urban areas (up to -1 $\mu g/m^3$ for every 1000 people). The exception to this is New York City, NY, which displays an increase in population-weighted $PM_{2.5}$ concentrations. Maximum hourly $PM_{2.5}$ concentrations in Case 2 increase from the base case up to 56 $\mu g/m^3$ (Figure 4-7c).

An increase in onroad mobile NO_x emissions (Case 3) results in increases of CMAQ-predicted $PM_{2.5}$ hourly average concentrations up to 2.3 $\mu g/m^3$ (Figure 4-5e), a 13.7% increase (Figure 4-6e). The locations of the maxima are not along major interstates or in highly populated urban areas, as may be expected. These increases are a consequence of increases in nitrate (Figure C-1a). Rural areas where the higher concentrations are noted are agricultural regions of the U.S. with a large presence of livestock and fertilizer, the largest sources of ammonia emissions.[62] Ammonia is the most abundant atmospheric alkaline gas and an aerosol free acidity analysis[63] demonstrates low concentrations of free H+ in these areas (Figure C-1c). Although ammonia typically reacts with sulfuric acid to form ammonium-sulfate aerosol, high concentrations of ammonia in the presence of HNO_3 can result in formation of aerosol ammonium nitrate as observed in these simulations.[64] This demonstrates the possibility that highway and road emissions which are highest in urban areas can impact $PM_{2.5}$ in rural, agricultural areas should ammonia emissions be high. Decreasing onroad mobile NO_x emissions by half (Case 4) produces decreases in hourly average $PM_{2.5}$ concentrations up to -1.3 $\mu g/m^3$ (-7.3%) in the same geographic locations as the increase in Case 3 (Figures 4-5g and 4-6g) also due to nitrate. Maximum hourly increases in $PM_{2.5}$ concentrations from the base case for Cases 3 and 4 are 22 $\mu g/m^3$ and 6 $\mu g/m^3$, respectively (Figures 4-7e and 4-7g). $PM_{2.5}$ sensitivity to changes in onroad NO_x emissions (Cases 3 and 4) has population-weighted impacts in urban areas (Case 3: up to 2.6 $\mu g/m^3$ for every 1000 people, Case 4: up to 2 $\mu g/m^3$ for every 1000 people). However, there are also maxima in less populated areas of Pennsylvania and Ohio (Figures 4-8e and 4-8g), which correspond with the rural

locations observed above again, highlighting complex non-linearity, in this case due to inorganic thermodynamic phase partitioning of nitrogen.

Perturbations of onroad mobile PM_{other} (Cases 7 and 8) do not have large impacts on average hourly $PM_{2.5}$ concentrations, with the majority of differences from the base case <1% (Figures 4-2c, 4-2d, 4-3c, and 4-3d). Increases of up to 2.2 $\mu g/m^3$ for every 1000 people in population-weighted hourly average $PM_{2.5}$ are observed in Case 7 (Figure 4-1c), while Case 8 results in decreases of hourly average population-weighted $PM_{2.5}$ concentration (Figure 4-1d, up to -1.1 $\mu g/m^3$). Maximum hourly increases of $PM_{2.5}$ for Cases 7 and 8 are 1.4 $\mu g/m^3$ and 1.1 $\mu g/m^3$, respectively (Figures 4-4c and 4-4d). The smaller impacts from Cases 7 and 8 indicate that $PM_{2.5}$ concentrations are not very sensitive to changes in PM_{other}, most likely a consequence of improvements in $PM_{2.5}$ speciation.

4.4.2. Predicted Ozone Concentration Sensitivity

Ambient and population-weighted ozone mixing ratios are most sensitive to changes in onroad mobile NO_x emission changes. Doubling NO_x emissions (Case 3) results in average hourly ozone increases across a majority of the domain up to 4.5 ppb_v (an increase of 8.2%, Figures 4-5f and 4-6f), with maximum increase of up to 32 ppb_v. Yet, in urban areas and along the I-95 corridor, we observe decreases of up to -8.5 ppb_v (a decrease of -55%) in average hourly ozone. Ozone rapidly reacts with NO near the roadway in Case 3, titrating to NO_2 and decreasing ozone mixing ratios.[65] Case 3 produces the largest decreases of hourly average population-weighted ozone mixing

ratios (up to -118 ppb_v for every 1000 people, Figure 4-8f). Decreasing onroad mobile NO_x emissions by half (Case 4) results in average decreases of ozone mixing ratios in the majority of the domain. However, along major highways and in cities, average hourly ozone concentrations increases up to 6.3 ppb_v (Figure 4-5h), with hourly maximum ozone increases up to 72 ppbv (Figure 4-7h). This is counter-intuitive to the ozone reduction goals of NO_x regulations. Increases in population-weighted hourly average ozone mixing ratios (up to 79 ppb_v for every 1000 people) are observed when halving NO_x onroad emissions (Figure 4-8h).

When PBL height is increased by 100 m (Case 1), average hourly ozone mixing ratios increase over land (Figure 4-5b, up to 3 ppb_v) and decrease over the ocean (up to 2.5 ppb_v). The largest percent increases (>100%) occur in large urban areas (e.g. New York, NY, Philadelphia, PA, Chicago, IL, Figure 4-6b). Maximum hourly ozone mixing ratios are as high as 72 ppb_v (Figure 4-7b). When PBL height is decreased by 100 m (Case 2), ozone mixing ratios appear to decrease over the domain by as much as -3.3 ppb_v (Figure 4-5d). These decreases are probably due to the decrease in the volume of the air column, increasing the rate at which NO_x titration occurs. The only exception to these results is an increase of up to 1 ppbv near Philadelphia, PA, representing a 120% increase in ozone (Figure 4-6d). Elsewhere, average hourly ozone percent changes are +/- 5% sporadically. Population-weighted hourly average ozone is only sensitive to perturbations of the PBL height in major cities in the domain (New York City, Philadelphia, and Chicago). Case 1 produces population-weighted average hourly ozone increases of up to 23.6 ppbv (Figure 4-8b), while Case 2 results in decreases of population-weighted average hourly ozone by as much as -18 ppb_v (Figure 4-8d). Recent coupling of the

WRF-CMAQ system has displayed changes in PBL height relative to the one-way coupling due to direct aerosol feedback[66, 67] and repeating this analysis with the coupled model is a necessary next step. As expected, Cases 5-8 show negligible changes in average or maximum hourly ozone mixing ratios, since POC and PM_{other} do not impact ozone mixing ratios.

4.5. References

1. Bridges, A.; Felder, F. A.; McKelvey, K.; Niyogi, I., Uncertainty in energy planning: Estimating the health impacts of air pollution from fossil fuel electricity generation. *Energ. Res. Soc. Sci.* **2015,** *6*, 74-77.

2. Cohan, D. S.; Koo, B.; Yarwood, G., Influence of uncertain reaction rates on ozone sensitivity to emissions. *Atmos. Environ.* **2010,** *44*, (26), 3101-3109.

3. Pinder, R. W.; Gilliam, R. C.; Appel, K. W.; Napelenok, S. L.; Foley, K. M.; Gilliland, A. B., Efficient Probabilistic Estimates of Surface Ozone Concentration Using an Ensemble of Model Configurations and Direct Sensitivity Calculations. *Environ. Sci. Tech.* **2009,** *43*, (7), 2388-2393.

4. Talbott, E. O.; Arena, V. C.; Rager, J. R.; Clougherty, J. E.; Michanowicz, D. R.; Sharma, R. K.; Stacy, S. L., Fine particulate matter and the risk of autism spectrum disorder. *Environ. Res.* **2015,** *140*, 414-420.

5. Volk, H. E.; Lurmann, F.; Penfold, B.; Hertz-Picciotto, I.; McConnell, R., TRaffic-related air pollution, particulate matter, and autism. *JAMA Psychiatry* **2013,** *70*, (1), 71-77.

6. Volk, H. E.; Hertz-Picciotto, I.; Delwiche, L.; Lurmann, F.; McConnell, R., Residential Proximity to Freeways and Autism in the CHARGE Study. *Environ. Health Persp.* **2011,** *119*, (6), 873-877.

7. Cohen, A.J.; Anderson, H.R.; Ostra, B.; Pandey, K. D.; Krzyzanowski, M.; Kunzli, N.; Gutschmidt, K.; Pope, A.; Romieu, I.; Samet, J. M.; Smith, K., The Global Burden of Disease Due to Outdoor Air Pollution. *J. Tox. and Environ. Health, Part A* **2005**, (68), 1-7.

8. Alexis, N. E.; Huang, Y. C. T.; Rappold, A. G.; Kehrl, H.; Devlin, R.; Peden, D. B., Patients with Asthma Demonstrate Airway Inflammation after Exposure to Concentrated Ambient Particulate Matter. *Am. J. Resp. Crit. Care* **2014,** *190*, (2), 235-237.

9. Brook, R. D.; Brook, J. R.; Urch, B.; Vincent, R.; Rajagopalan, S.; Silverman, F., Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation* **2002,** *105*, (13), 1534-1536.

10. Brook, R. D.; Rajagopalan, S.; Pope, C. A.; Brook, J. R.; Bhatnagar, A.; Diez-Roux, A. V.; Holguin, F.; Hong, Y. L.; Luepker, R. V.; Mittleman, M. A.; Peters, A.; Siscovick, D.; Smith, S. C.; Whitsel, L.; Kaufman, J. D.; Epidemiol, A. H. A. C.; Dis, C. K. C.; Metab, C. N. P. A., Particulate Matter Air Pollution and Cardiovascular Disease An Update to the Scientific Statement From the American Heart Association. *Circulation* **2010,** *121*, (21), 2331-2378.

11. Fang, Y.; Naik, V.; Horowitz, L. W.; Mauzerall, D. L., Air pollution and associated human mortality: the role of air pollutant emissions, climate change and methane concentration increases from the preindustrial period to present. *Atmos. Chem. Phys.* **2013,** *13*, (3), 1377-1394.

12. Raz, R.; Roberts, A. L.; Lyall, K.; Hart, J. E.; Just, A. C.; Laden, F.; Weisskopf, M. G., Autism Spectrum Disorder and Particulate Matter Air Pollution before, during, and after Pregnancy: A Nested Case–Control Analysis within the Nurses' Health Study II Cohort. *Environ. Health Persp.* **2014,** *123*, (3), 264-270.

13. Valavanidis, A.; Vlachogianni, T.; Fiotakis, K., Airborne Particulate Matter in Urban Areas and Risk for Cardiopulmonary Mortality and Lung Cancer. Dietary antioxidants and supplementation for prevention of adverse health effects. *Pharmakeftiki* **2014,** *26*, (IV), 139-156.

14. Appel, K. W.; Foley, K. M.; Bash, J. O.; Pinder, R. W.; Dennis, R. L.; Allen, D. J.; Pickering, K., A multi-resolution assessment of the Community Multiscale Air Quality (CMAQ) model v4.7 wet deposition estimates for 2002–2006. *Geosci. Model Dev.* **2011,** *4*, (2), 357-371.

15. Liu, X.; Frey, H. C.; Cao, Y.; Deshpande, B., Modeling of In-vehicle PM(2.5) Exposure Using the Stochastic Human Exposure and Dose Simulation Model. *Annual meeting & exhibition proceedings CD-ROM Air & Waste Management Association. Meeting* **2009,** *2*, (102), 1087-1100.

16. Durant, J. L.; Ash, C. A.; Wood, E. C.; Herndon, S. C.; Jayne, J. T.; Knighton, W. B.; Canagaratna, M. R.; Trull, J. B.; Brugge, D.; Zamore, W.; Kolb, C. E., Short-term variation in near-highway air pollutant gradients on a winter morning. *Atmos. Chem. Phys.* **2010,** *10*, (17), 8341-8352.

17. Beckerman, B.; Jerrett, M.; Brook, J. R.; Verma, D. K.; Arain, M. A.; Finkelstein, M. M., Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. *Atmospheric Environment* **2008,** *42*, (2), 275-290.

18. Hodas, N.; Meng, Q.; Lunden, M. M.; Turpin, B. J., Toward refined estimates of ambient PM2.5 exposure: Evaluation of a physical outdoor-to-indoor transport model. *Atmospheric Environment* **2014,** *83*, 229-236.

19. Hanna, S. R.; Chang, J. C.; Fernau, M. E., Monte carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmos. Environ.* **1998,** *32*, (21), 3619-3628.

20. Mallet, V.; Sportisse, B., Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling. *J. of Geophys. Res. Atmospheres* **2006,** *111*, (D1).

21. Hanna, S. R.; Lu, Z.; Christopher Frey, H.; Wheeler, N.; Vukovich, J.; Arunachalam, S.; Fernau, M.; Alan Hansen, D., Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmos. Environ.* **2001,** *35*, (5), 891-903.

22. Tian, D.; Cohan, D. S.; Napelenok, S.; Bergin, M.; Hu, Y.; Chang, M.; Russell, A. G., Uncertainty Analysis of Ozone Formation and Response to Emission Controls Using Higher-Order Sensitivities. *J. Air Waste M. A.* **2010,** *60*, (7), 797-804.

23. Özkaynak, H.; Frey, H. C.; Burke, J.; Pinder, R. W., Analysis of coupled model uncertainties in source-to-dose modeling of human exposures to ambient air pollution: A PM(2.5) case study. *Atmos. Environ.* **2009,** *43*, (9), 1641-1649.

24. Burr, M. J.; Zhang, Y., Source apportionment of fine particulate matter over the Eastern U.S. Part I: source sensitivity simulations using CMAQ with the Brute Force method. *Atmos. Pol. Res.* **2011,** *2*, (3), 300-317.

25. Carlton, A. G.; Pinder, R. W.; Bhave, P. V.; Pouliot, G. A., To What Extent Can Biogenic SOA be Controlled? *Environ. Sci. Tech.* **2010,** *44*, (9), 3376-3380.

26. Cohan, D. S.; Hakami, A.; Hu, Y.; Russell, A. G., Nonlinear Response of Ozone to Emissions: Source Apportionment and Sensitivity Analysis. *Environ. Sci. Tech.* **2005,** *39*, (17), 6739-6748.

27. Hakami, A.; Odman, M. T.; Russell, A. G., High-Order, Direct Sensitivity Analysis of Multidimensional Air Quality Models. *Environ. Sci. Tech.* **2003,** *37*, (11), 2442-2452.

28. Dunker, A. M., The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. *J. Chem. Phys.* **1984,** *81*.

29. Napelenok, S. L.; Cohan, D. S.; Hu, Y.; Russell, A. G., Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM). *Atmos. Environ.* **2006,** *40*, (32), 6112-6121.

30. Thompson, T. M.; Selin, N. E., Influence of air quality model resolution on uncertainty associated with health impacts. *Atmos. Chem. Phys.* **2012,** *12*, (20), 9753-9762.

31. Valari, M.; Menut, L., Does an Increase in Air Quality Models' Resolution Bring Surface Ozone Concentrations Closer to Reality? *Journal of Atmospheric and Oceanic Technology* **2008,** *25*, (11), 1955-1968.

32. Vivanco, M.; Azula, O.; Palomino, I.; Martín, F., Evaluating the Impact of Resolution on the Predictions of an Air Quality Model over Madrid Area (Spain). In *Geocomputation, Sustainability and Environmental Planning*, Murgante, B.; Borruso, G.; Lapucci, A., Eds. Springer Berlin Heidelberg: 2011; Vol. 348, pp 145-162.

33. Doraiswamy, P.; Hogrefe C Fau - Hao, W.; Hao W Fau - Civerolo, K.; Civerolo K Fau - Ku, J.-Y.; Ku Jy Fau - Sistla, G.; Sistla, G., A retrospective comparison of model-based forecasted PM2.5 concentrations with measurements. *J. Air Waste M. A.* **2010**, *60*, 1293-1308.

34. Du, C. L.; Liu, S. Y.; Yu, X.; Li, X. M.; Chen, C.; Peng, Y.; Dong, Y.; Dong, Z. P.; Wang, F. Q., Urban Boundary Layer Height Characteristics and Relationship with Particulate Matter Mass Concentrations in Xi'an, Central China. *Aerosol Air Qual. Res.* **2013,** *13*, (5), 1598-1607.

35. Thompson, G. *Surface Station List – RAP Real-Time Weather*; 2015.

36. Coen, M. C.; Praz, C.; Haefele, A.; Ruffieux, D.; Kaufmann, P.; Calpini, B., Determination and climatology of the planetary boundary layer height above the Swiss plateau by in situ and remote sensing measurements as well as by the COSMO-2 model. *Atmos. Chem. Phys.* **2014,** *14*, (23), 13205-13221.

37. Balzarini, A.; Angelini, F.; Ferrero, L.; Moscatelli, M.; Perrone, M. G.; Pirovano, G.; Riva, G. M.; Sangiorgi, G.; Toppetti, A. M.; Gobbi, G. P.; Bolzacchini, E., Sensitivity analysis of PBL schemes by comparing WRF model and experimental data. *Geosci. Model Dev. Discuss.* **2014,** *7*, (5), 6133-6171.

38. Kolling, J. S.; Pleim, J. E.; Jeffries, H. E.; Vizuete, W., A multisensor evaluation of the Asymmetric Convective Model, Version 2, in Southeast Texas. *J. Air Waste Manage. Assoc.* **2012,** *63*, (1), 41-53.

39. Dolwick, P. In *The effects of cloud attenuation on air quality: A comparison of model treatments.*, Proc. 14th Joint Conference on the Applications of Air Pollution Meteorology with the Air and Waste Management Association, Atlanta, GA, 2006; American Meteorological Society: Atlanta, GA, 2006.

40. Marmo, B. P.; Carlton, A. G.; Henderson, B. H., Partitioning of HNO3, H2O2 and SO2 to cloud ice: Simulations with CMAQ. *Atmospheric Environment* **2014,** *88*, 239-246.

41. Hogrefe, C.; Isukapalli, S.S.; Tang, X.; Georgopoulos, P. G.; He, S.; Zalewsky, E. E.; Hao, W.; Ku, J.-Y.; Key, T.; Sistla, G., Impact of biogenic emission uncertainties on the simulated response of ozone and fine particulate matter to anthropogenic emission reductions. *J. Air Waste M. A.* **2011**, *61,* (1)*,* 92-108.

42. U.S. EPA, National Summary of Sulfur Dioxide Emissions. http://www.epa.gov/cgi-bin/broker?_service=data&_debug=0&_program=dataprog.national_1.sas&polchoice=SO2 (August 30, 2015).

43. He, H.; Hembeck, L.; Hosley, K. M.; Canty, T. P.; Salawitch, R. J.; Dickerson, R. R., High ozone concentrations on hot days: The role of electric power demand and NOx emissions. *Geophys. Res. Lett.* **2013,** *40*, (19), 2013.

44. Frost, G. J.; McKeen, S. A.; Trainer, M.; Ryerson, T. B.; Neuman, J. A.; Roberts, J. M.; Swanson, A.; Holloway, J. S.; Sueper, D. T.; Fortin, T.; Parrish, D. D.; Fehsenfeld, F. C.; Flocke, F.; Peckham, S. E.; Grell, G. A.; Kowal, D.; Cartwright, J.; Auerbach, N.; Habermann, T., Effects of changing power plant NOx emissions on ozone in the eastern United States: Proof of concept. *J. Geophys. Res.-Atmos.* **2006,** *111*, (D12).

45. U.S. EPA, The 2011 National Emissions Inventory v.1. http://www.epa.gov/ttnchie1/net/2011inventory.html (August 30, 2015).

46. Zhang, H.; Chen, G.; Hu, J.; Chen, S.-H.; Wiedinmyer, C.; Kleeman, M.; Ying, Q., Evaluation of a seven-year air quality simulation using the Weather Research and Forecasting (WRF)/Community Multiscale Air Quality (CMAQ) models in the eastern United States. *Sci. Tot. Environ.* **2014,** *473–474*, (0), 275-285.

47. Farkas, C. M.; Moeller, M. D.; Felder, F. A.; Baker, K. R.; Rodgers, M.; Carlton, A. G., Temporalization of Peak Electric Generation Particulate Matter Emissions during High Energy Demand Days. *Environ. Sci. & Tech.* **2015,** *49*, (7), 4696-4704.

48. Foley, K. M.; Roselle, S. J.; Appel, K. W.; Bhave, P. V.; Pleim, J. E.; Otte, T. L.; Mathur, R.; Sarwar, G.; Young, J. O.; Gilliam, R. C.; Nolte, C. G.; Kelly, J. T.; Gilliland, A. B.; Bash, J. O., Incremental testing of the Community Multiscale Air

Quality (CMAQ) modeling system version 4.7. *Geosci. Model Dev.* **2010,** *3*, (1), 205-226.

49. U.S. EPA, National Summary of Nitrogen Oxides Emissions. http://www.epa.gov/cgi-bin/broker?_service=data&_debug=0&_program=dataprog.national_1.sas&polchoice=NOX (August 30, 2015).

50. Zavala, M.; Herndon, S. C.; Wood, E. C.; Onasch, T. B.; Knighton, W. B.; Marr, L. C.; Kolb, C. E.; Molina, L. T., Evaluation of mobile emissions contributions to Mexico City's emissions inventory using on-road and cross-road emission measurements and ambient data. *Atmos. Chem. Phys.* **2009,** *9*, (17), 6305-6317.

51. Pan, L.; Tong, D.; Lee, P.; Kim, H. C.; Chai, T., Assessment of NOx and O3 forecasting performances in the U.S. National Air Quality Forecasting Capability before and after the 2012 major emissions updates. *Atmos. Environ.* **2014,** *95*, 610-619.

52. Zhao, Y.; Hennigan, C. J.; May, A. A.; Tkacik, D. S.; de Gouw, J. A.; Gilman, J. B.; Kuster, W. C.; Borbon, A.; Robinson, A. L., Intermediate-Volatility Organic Compounds: A Large Source of Secondary Organic Aerosol. *Environ. Sci. & Tech.* **2014,** *48*, (23), 13743-13750.

53. Kanakidou, M.; Seinfeld, J. H.; Pandis, S. N.; Barnes, I.; Dentener, F. J.; Facchini, M. C.; Van Dingenen, R.; Ervens, B.; Nenes, A.; Nielsen, C. J.; Swietlicki, E.; Putaud, J. P.; Balkanski, Y.; Fuzzi, S.; Horth, J.; Moortgat, G. K.; Winterhalter, R.; Myhre, C. E. L.; Tsigaridis, K.; Vignati, E.; Stephanou, E. G.; Wilson, J., Organic aerosol and global climate modelling: a review. *Atmos. Chem. Phys.* **2005,** *5*, (4), 1053-1123.

54. Carlton, A. G.; Bhave, P. V.; Napelenok, S. L.; Edney, E. D.; Sarwar, G.; Pinder, R. W.; Pouliot, G. A.; Houyoux, M., Model Representation of Secondary Organic Aerosol in CMAQv4.7. *Environ. Sci. Tech.* **2010,** *44*, (22), 8553-8560.

55. Appel, K.W.; Bhave, P. V.; Gilliland, A. B.; Sarwar, G.; Roselle, S. J., Evaluation of the community multiscale air quality (CMAQ) model version 4.5: Sensitivities impacting model performance; Part II—particulate matter. *Atmos. Environ.* **2008,** *42*, (24), 6057-6066.

56. U.S. EPA. *Report to Congress on Black Carbon*; 2012.

57. Skamarock, W. C. K., J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. *A description of the Advanced Research WRF version 3. NCAR Technical Note NCAR/TN-475+STR*; 2008.

58. Carlton, A. G.; Baker, K. R., Photochemical Modeling of the Ozark Isoprene Volcano: MEGAN, BEIS, and Their Impacts on Air Quality Predictions. *Environ. Sci. Tech.* **2011,** *45*, (10), 4438-4445.

59. Borge, R.; Alexandrov, V.; José del Vas, J.; Lumbreras, J.; Rodríguez, E., A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula. *Atmospheric Environment* **2008,** *42*, (37), 8560-8574.

60. The Institute for the Environment – Univeristy of North Carolina Chapel Hill, SMOKE v3.5 User's Manual. http://www.cmascenter.org/smoke/documentation/3.5/manual_smokev35.pdf (August 30, 2015).

61. 2010 National Census Tracts Gazetteer File. In U.S. Census Bureau, Ed. 2012.

62. Behera, S.; Sharma, M.; Aneja, V.; Balasubramanian, R., Ammonia in the atmosphere: a review on emission sources, atmospheric chemistry and deposition on terrestrial bodies. *Environ. Sci. Pollut. Res.* **2013,** *20*, (11), 8092-8131.

63. Carlton, A. G.; Turpin, B. J., Particle partitioning potential of organic compounds is highest in the Eastern US and driven by anthropogenic water. *Atmos. Chem. Phys.* **2013,** *13*, (20), 10203-10214.

64. CERNS Air Quality Research Subcommittee, *Atmospheric Ammonia: Sources and Fate - A Review of Ongoing Federal Research and Future Needs*; 2000.

65. Seinfeld, J. H.; Pandis, S. N., *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Second ed.; John Wiley & Sons, Inc.: Hoboken, New Jersey, 2006.

66. Wong, D. C.; Pleim, J.; Mathur, R.; Binkowski, F.; Otte, T.; Gilliam, R.; Pouliot, G.; Xiu, A.; Young, J. O.; Kang, D., WRF-CMAQ two-way coupled system with aerosol feedback: software development and preliminary results. *Geosci. Model Dev.* **2012,** *5*, (2), 299-312.

67. U.S. EPA, Coupled WRF-CMAQ Modeling System. http://www.epa.gov/amad/Research/Air/twoway.html (August 30, 2015).

Case	CMAQ variable	Alteration	References
1	PBL	+ 100 m	**Balzarini et al. (2014)**: PBL mean biases of -43.82m to 248.05m **Doraiswamy et al. (2010)**: sharp changes in PBL impact diurnal variation peaks of pollutants
2	PBL	- 100 m	
3	NO, NO_2 mobile emissions	doubled	**Hanna et al. (2001)**: factor of 2 uncertainty in area mobile emissions **Zavala et al. (2009)**: 14-20% overpredictions of NO emissions **Pan et al. (2014)**: 10-20% underestimation of mobile NO_x emissions **Tian et al. (2010):** factor of 2 uncertainty in mobile emissions has the largest impact on ozone predictions of 5 sectors
4	NO, NO_2 mobile emissions	halved	
5	POC mobile emissions	doubled	**Doraiswamy et al. (2010):** underpredictions of organic carbon in spring and summer up to ~50%
6	POC mobile emissions	halved	
7	PMOTHR mobile emissions	doubled	**Doraiswamy et al. (2010)**: underpredictions of "other" unspeciated $PM_{2.5}$ of ~50-60% in the summer **Appel et al. (2008)** find an underprediction in $PM_{2.5}$ mass due to an underprediction in PM_{other} by ~11-20%
8	PMOTHR mobile emissions	halved	

Table 4-1. Description of sensitivity simulations

Ambient Concentration Sensitivity Rank			Population-weighted Sensitivity Rank	
$PM_{2.5}$	**Ozone**		**$PM_{2.5}$**	**Ozone**
PBL height	onroad NO_x	most sensitive	onroad POC	onroad NO_x
onroad NO_x	PBL height		PBL height	PBL height
onroad POC	-------		onroad NO_x	-------
onroad PM_{other}	-------	least sensitive	onroad PM_{other}	-------

Table 4-2: CMAQ Simulation Sensitivity Ranks
CMAQ perturbation simulations ranked in order of $PM_{2.5}$ or ozone sensitivity (most sensitive to least sensitive)

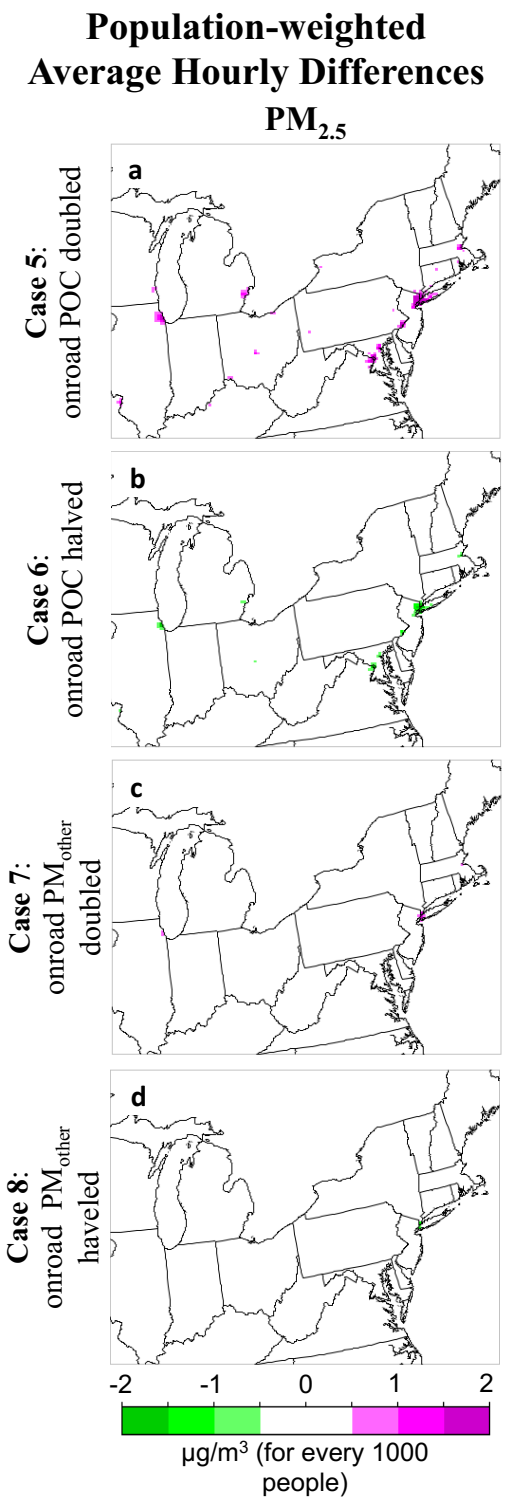

Figure 4-1: Sensitivity Cases 5-8, population-weighted CMAQ-predicted ambient concentration differences from the Base Case of average hourly $PM_{2.5}$ from July 1, 2006 to July 31, 2006. All concentrations are per 1000 people. Ozone is not included because Cases 5-8 did not have an effect on CMAQ-predicted ozone mixing ratios.

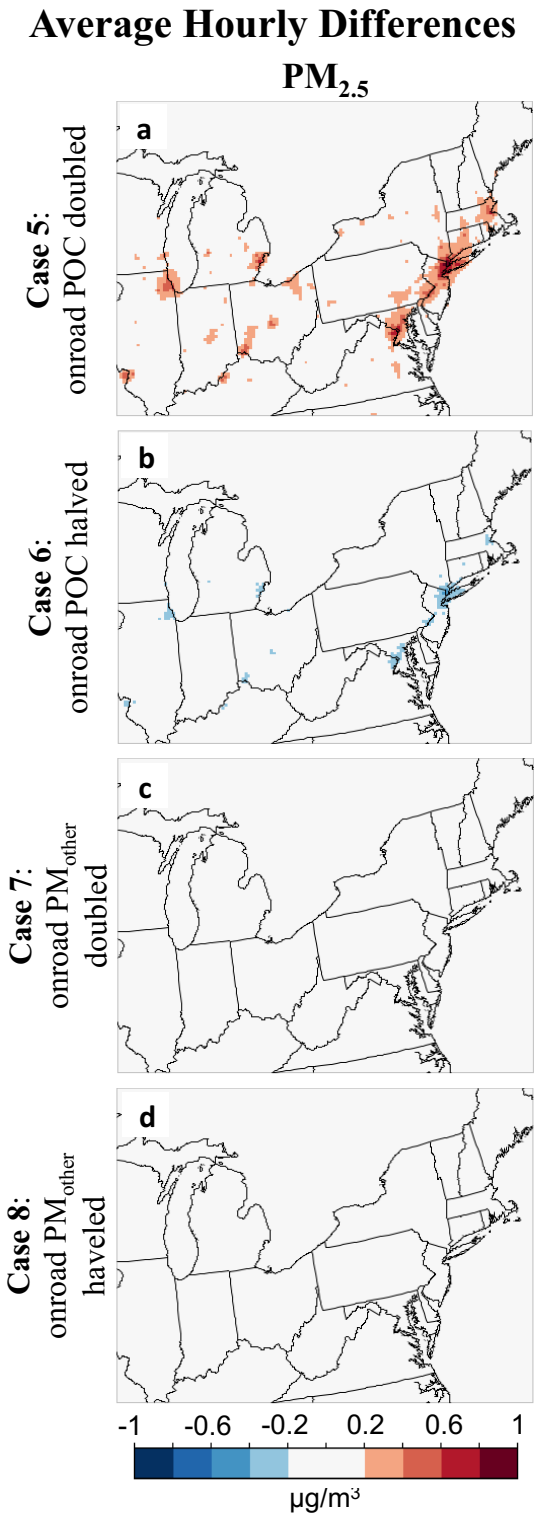

Figure 4-2: Sensitivity Cases 5-8, CMAQ-predicted ambient concentration differences from the Base Case of average hourly $PM_{2.5}$ from July 1, 2006 to July 31, 2006. Ozone is not included because Cases 5-8 did not have an effect on CMAQ-predicted ozone mixing ratios.

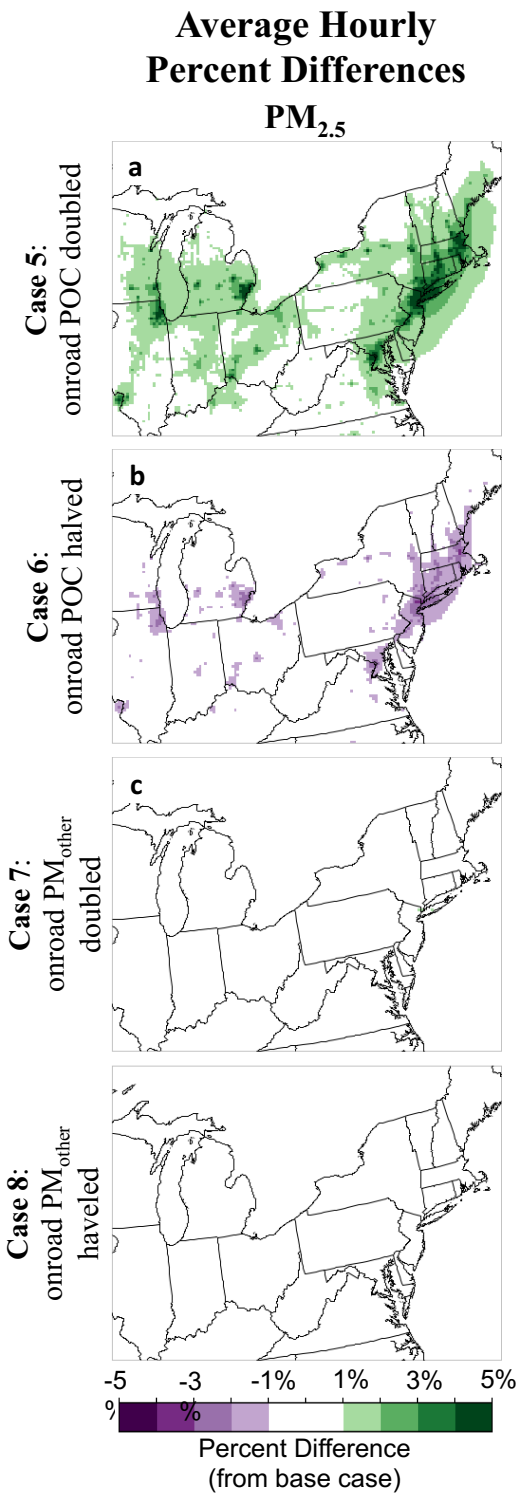

Figure 4-3: Sensitivity Cases 5-8, CMAQ-predicted ambient concentration percent differences from the Base Case of average hourly $PM_{2.5}$ from July 1, 2006 to July 31, 2006. Ozone is not included because Cases 5-8 did not have an effect on CMAQ-predicted ozone mixing ratios.

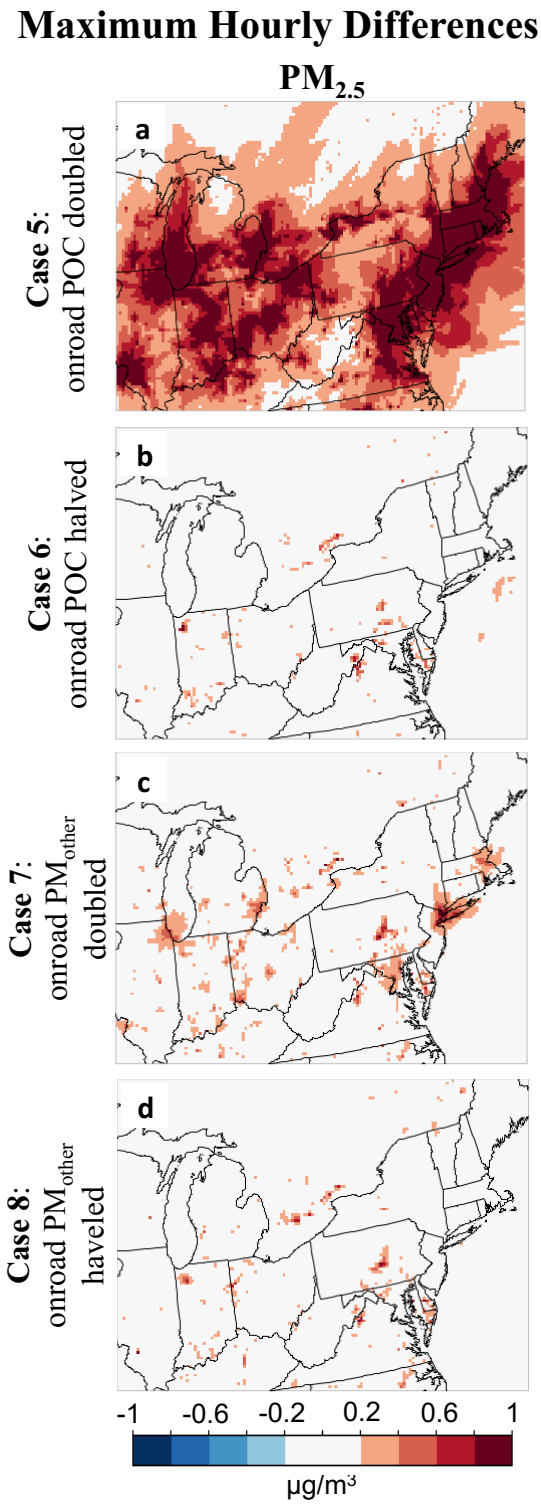

Figure 4-4: Sensitivity Cases 5-8, CMAQ-predicted maximum hourly ambient concentration differences from the Base Case of maximum hourly $PM_{2.5}$ (a,c,e,g) from July 1, 2006 to July 31, 2006. Ozone is not included because Cases 5-8 did not have an effect on CMAQ-predicted ozone mixing ratios.

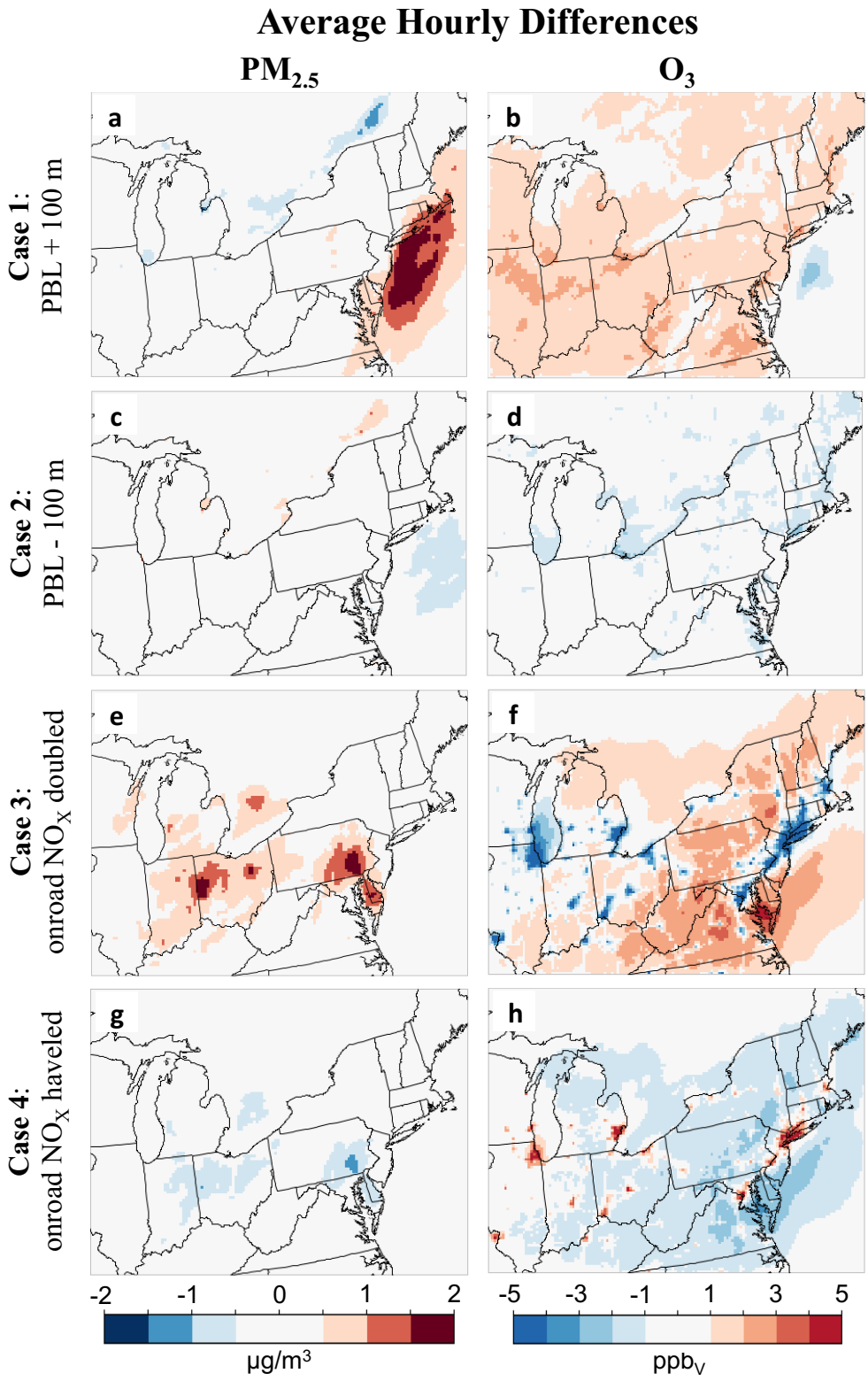

Figure 4-5: Sensitivity Cases 1-4, CMAQ-predicted ambient concentration differences from the Base Case of average hourly $PM_{2.5}$ (a,c,e,g) and average hourly ozone (b,d,f,h) from July 1, 2006 to July 31, 2006.

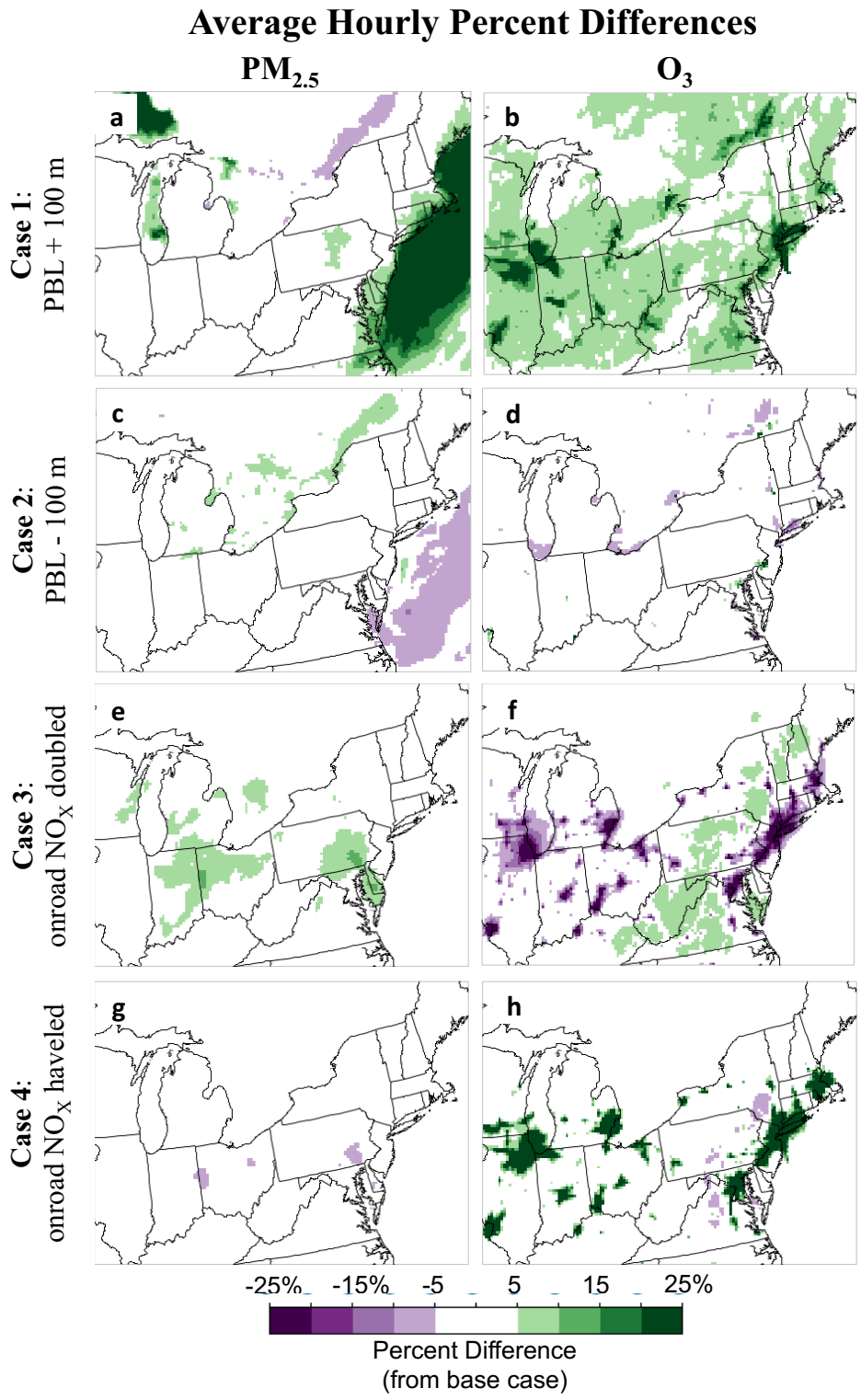

Figure 4-6: Sensitivity Cases 1-4, CMAQ-predicted ambient concentration percent differences from the Base Case of average hourly $PM_{2.5}$ (a,c,e,g) and average hourly ozone (b,d,f,h) from July 1, 2006 to July 31, 2006.

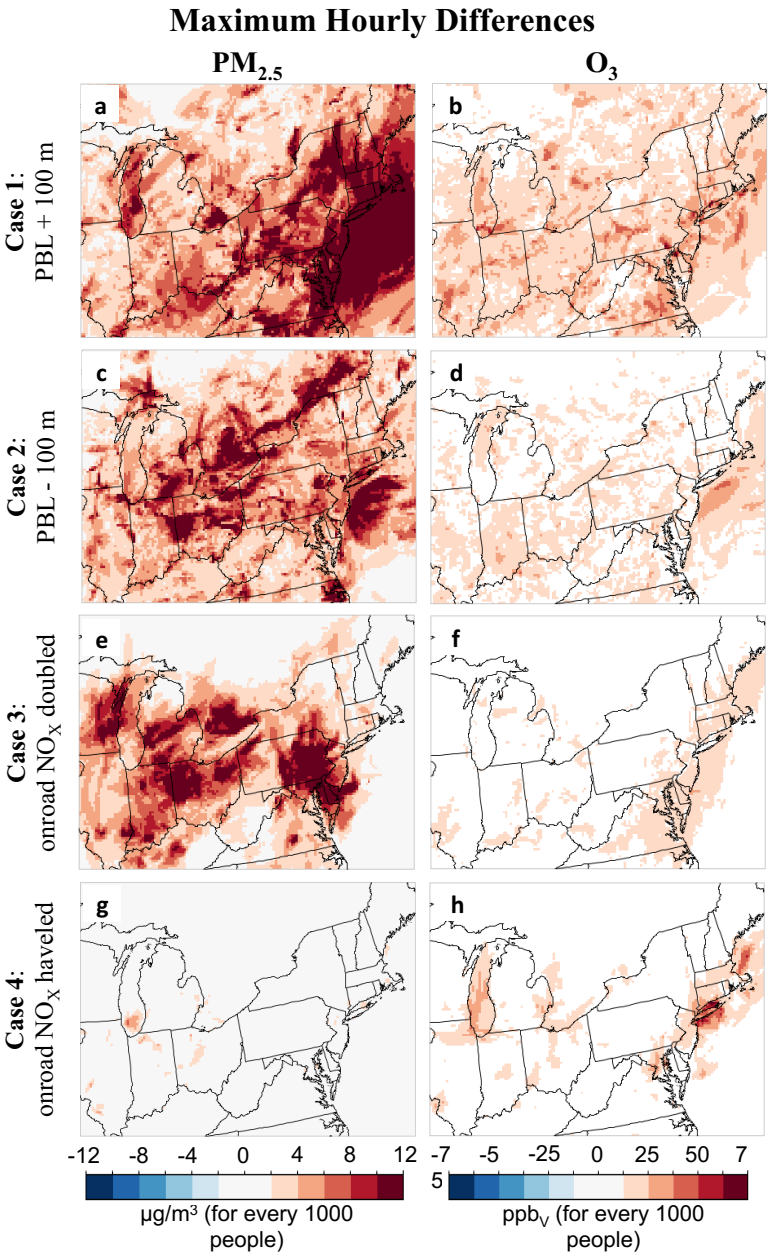

Figure 4-7: Sensitivity Cases 1-4, CMAQ-predicted ambient concentration differences from the Base Case of maximum hourly $PM_{2.5}$ (a,c,e,g) and maximum hourly ozone (b,d,f,h) from July 1, 2006 to July 31, 2006.

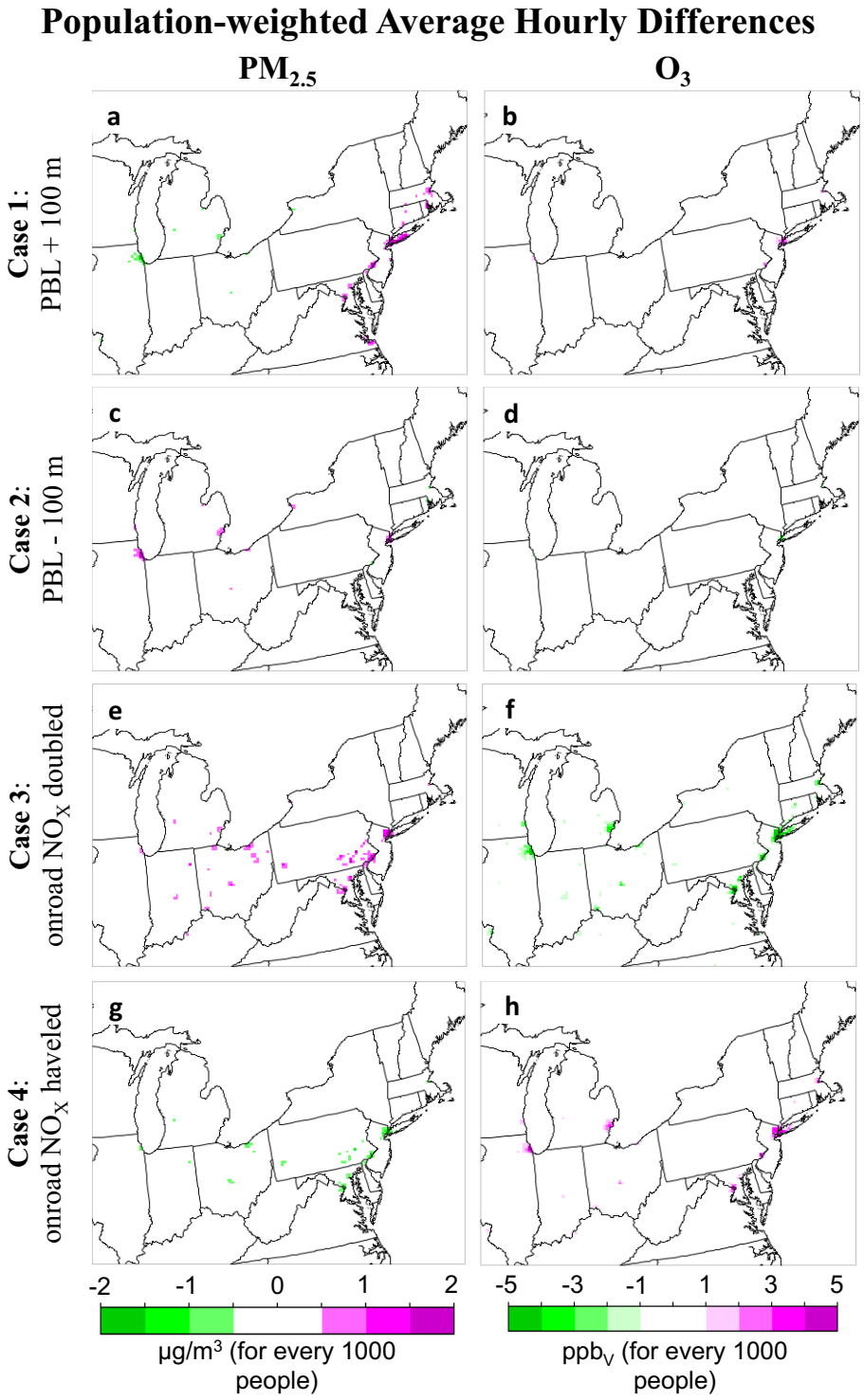

Figure 4-8: Sensitivity Cases 1-4, population-weighted CMAQ-predicted ambient concentration differences from the Base Case of average hourly $PM_{2.5}$ (a,c,e,g) and average hourly ozone (b,d,f,h) from July 1, 2006 to July 31, 2006. All concentrations are per 1000 people.

5. CHAPTER 5

SUMMARY, IMPLICATIONS, AND FUTURE DIRECTIONS

5.1. Summary

The work presented in this dissertation addresses an ongoing mission to improve agreement between ambient concentrations of $PM_{2.5}$ mass species and ozone measured in the atmosphere and those predicted by photochemical models for the right reasons. The first project focuses on improvement of point source emission estimates in a photochemical model by incorporating actual emission measurements to improve the temporal profile of electric generating unit (EGU) sector emissions. The second project focuses on the contribution of peaking unit emissions from the PJM Interconnection regional transmission organization (RTO) to model estimated $PM_{2.5}$ and ozone in comparison to those from all EGUs in PJM and other RTOs. The final project estimates sources of uncertainty for $PM_{2.5}$ and ozone CMAQ-predictions using a brute force analysis of meteorological and mobile emission variables that have been found to have high uncertainty in previous research. The resulting ambient concentrations are population-weighted as estimation for potential exposure.

Chapter 2 demonstrates that while improved temporalization of CEM emission data predicts substantial, episodic increases in surface concentrations of $PM_{2.5}$ mass and small increases in surface ozone mixing ratios, the model bias between CMAQ predicted concentrations and observed ambient concentrations of these pollutants showed small

improvements (up to 3 $\mu g/m^3$) in the case with additional ORIS IDs. These improvements generally occurred during the hottest days of the heat wave and at peak concentrations of $PM_{2.5}$ mass. Fine particulate matter is often studied in an annual context and measurements are designed for this as opposed to capturing peak concentrations.

The same heat wave period described in Chapter 2 is used in Chapter 3 to estimate the air quality impact due to peak-usage EGUs, also known as peaking units and compare them to the impacts from all EGUs in PJM and surrounding RTOs. Population-weighted concentrations, as a simplified index of potential human exposure, are employed to study impacts of ambient concentrations of $PM_{2.5}$ and ozone due to peaking units from a more health-centered perspective. Emissions from PJM EGUs not only impact the PJM region, but surrounding regions as well. Maximum ambient mass concentrations of $PM_{2.5}$ due to peaking units are predicted to be very high at peak hours during the heat wave period and these concentrations have the largest effect on the most populated cities in the Northeast U.S., regardless of the location of the maximum in ambient concentrations. Peaking unit usage over the past 10 years in the PJM region correlates well with average summer air temperature and with GDP when the growth rate is 2.5% or less, suggesting that the summer air temperature and the national economic strength, likely indicators of available disposable income to spend on air conditioning during these peak events, may strongly affect peaking unit usage. The most monitored, most regulated peaking units are studied here and represent a lower bound estimate. Peaking unit usage is projected to increase over the next ten years, as are heat waves and stagnation events, bringing attention to the likely air quality impact and the importance of identifying and quantifying their emissions.

Limitations of these studies exist and are controlled in large part by uncertainties of the models and their inputs. Small errors in a single input can propagate throughout the model simulation, affecting predicted concentrations. Chapter 4 uses a brute force sensitivity analysis to determine the resulting fluctuations in model-predicted concentrations when changes are made individually to single variables. Specifically, I investigate how mass concentrations of $PM_{2.5}$ species and ozone mixing ratios react to perturbations in planetary boundary layer (PBL) height and increases and decreases in mobile emissions of primary organic carbon (POC), nitrogen oxides ($NO + NO_2 = NO_x$), and unspeciated $PM_{2.5}$ (PM_{other}). I found that while ambient concentrations of $PM_{2.5}$ are most sensitive to changes in PBL height, the majority of the sensitivity occurs over the ocean. An assessment for potential exposure assessment shows that population-weighted $PM_{2.5}$ concentrations are most sensitive to onroad mobile primary organic carbon (POC) changes. Ambient and population-weighted ozone concentrations are most sensitive to onroad mobile NO_x increases and decreases. Model development intended to help develop air quality strategies to protect human health may be more effective if the focus is on motor vehicle emissions of POC and NO_x.

When considering the uncertainty findings of Chapter 4 in relation to Chapter 2, the additional hourly ambient $PM_{2.5}$ mass observed after 267 ORIS IDs were added to the simulation was as high as 7.4 $\mu g/m^3$. This value is above the range of uncertainty observed with onroad PMother, which had a maximum increase of 1.4 $\mu g/m^3$. The maximum findings of Chapter 2 lie within the range of uncertainty for onroad POC (maximum 8 $\mu g/m^3$), onroad NO_x (maximum difference of 22 $\mu g/m^3$) and PBL height (maximum uncertainty of 56 $\mu g/m^3$). Note that the majority of the PBL maxima were

located over the ocean. The additional ORIS matches from Chapter 2 represent 50% of missing ORIS IDs in the NEI during this simulation. If all 541 missing ORIS IDs were matched and simulated and we assume the impacts are linear, they lie outside the maximum uncertainty observed in Chaper 4 for onroad POC and onroad PM_{other}, but not PBL height or onroad NO_x This suggests that while onroad NO_x and PBL height have the potential to impact air quality forecasts in a larger way, correct hourly temporal allocation is necessary for accurate short-term forecasting, especially within individual $PM_{2.5}$ species.

When considering the uncertainty findings of Chapter 4 in relation to Chapter 3, the maximum ambient $PM_{2.5}$ mass observed due solely to 544 PJM peaking units is 110 $\mu g/m^3$. This concentration lies outside the uncertainty due to any of the studied variables in the uncertainty analysis, implying that these estimated impacts are not due to model error and could possibly benefit from a policy that aims to reduce emissions from peaking units on the hottest days.

5.2. Implications

This work has implications both in the modeling community and the policy community. Chapter 2 identifies an area ripe for improvement with the use of existing higher temporal resolution data. In this study I manually paired the data using identifying numbers and electricity permits. This problem is not due to a lack of science or engineering understanding, rather a policy fix, such as a permit requirement mandating a single identifier or submission of the Office of Regulatory and Information Systems

(ORIS) identifier across all point source sectors and emission reporting procedures to the national emissions inventory (NEI) would facilitate use of measured data in atmospheric models. The air quality impacts due to peaking units presented in Chapter 3 are limited to regulated peaking units with CEMs, suggesting that unmonitored facilities and behind-the-meter generation units could have an equal or worse impact on air quality and that the results shown here are a lower bound. This has even larger implications for areas of high-population. Chapter 4 contributes to the modeling community by identifying model uncertainties most impactful to potential human exposure to ambient air pollution. Ranking sensitivities by population-weighted concentrations can highlight the model inaccuracies most critical to improving accuracy of air quality simulations used to affect public health and welfare policies.

5.3. Future Directions

5.3.1. Reducing Uncertainty in Motor Vehicle Emissions

Large contributors to uncertainty in mobile emissions are traffic patterns, vehicle fleets, and miles driven. Often vehicle fleets are characterized by state registrations. Occasionally, two simultaneous video cameras are used for eight hours a day to capture license plates at a specific location. The license plates are usually recorded by hand; however license plate identification software does exist.[1] This method only collects information on the vehicle, not the miles driven nor the time spent in traffic. I propose utilizing existing highway cameras as a more efficient and accurate method of collecting

traffic data. Automated vehicle identification software has been used in the past by toll agencies, traffic monitoring agencies, and law enforcement offices.[2, 3] Using a combination of highway cameras and vehicle identification software, the vehicle fleet, traffic patterns, and corresponding emissions can be estimated. This method could be used to improve hourly motor vehicle emission data for the major highways in the U.S.

5.3.2. Detailed Demographic Health Impact Analysis

Population-weighted concentrations can be used not only to detail the areas most affected by air pollutants but also in interdisciplinary studies to investigate correlations between air pollution and human characteristics, such as socio-economic status, race, age, and other variables accounted for in the national census. This could help to answer interesting science-sociology and human ecology questions about the location of point source emissions and the populations they affect the most, as well as introducing the need for policies protecting those most affected by air pollution.

5.3.3. Coupled Modeling

The DAYZER model used in Chapter 2 can be a powerful tool in hypothetical studies involving changes in future energy choices or equipment. One study that is already in progress from this work is investigating the air quality implications of an offshore wind farm in New Jersey. Wind data from the Weather Research and Forecasting (WRF) model and DAYZER-generated hypothetical electricity market

simulations are employed to observe fluctuations in air emissions, and the resulting air quality, due to the inclusion of three different sized wind farms into the power grid. With the results, an environmental economic-based study can be performed to quantify the regional health costs or savings associated with an offshore wind farm.

The work presented in this dissertation highlights the human dimension to air quality modeling and the exposure potential of air pollution on different regions of the Northeast U.S. The aim of the U.S. EPA is to protect human health and welfare. While public policy and energy planning measures can work to bring regions into attainment with NAAQS, emission models that do not account for economic or potential human exposure impacts (e.g. population-weighted concentrations) will not be the most effective at protecting health-reducing exposure. By performing sensitivity analyses of the CMAQ model and population-weighting the results, we gain a better understanding of which variables have the greatest impact on the model results with the highest exposure potential. The newly coupled CMAQ-WRF model is a two-way model in which predicted concentrations can impact meteorological variables (e.g. direct aerosol feedbacks). The sensitivity analysis performed here should be repeated with the coupled model to evaluate the impact of this improvement relative to the one-way coupled model.

5.3.4. Incorporating Environmental Externalities into Electricity Price

Currently, electricity grids are optimized for reliability. Using the DAYZER-SMOKE-CMAQ modeling system, we could investigate impacts of optimizing the electricity grid for air quality or human health impacts. These simulations could be

compared with health impacts of current air quality simulations and answer questions such as 1) When do electricity costs outweigh health costs if the electricity grid was optimized for air quality?, 2) Would blackout or brownout events increase?, 3) How much would ambient concentrations of $PM_{2.5}$ and ozone change?

5.4. References

1. University of California at Riverside – Center for Environmental Research and Technology, *Improving Vehicle Fleet, Activity, and Emissions Data for On-Road Mobile Sources Emissions Inventories*; 2011.

2. 3M, Automatic Vehicle Classification Software. http://solutions.3m.com/wps/portal/3M/en_US/NA_Motor_Vehicle_Services_Systems/Motor_Vehicle_Industry_Solutions/product_catalog/parking-tolling-and-dmv-vehicle-software-3m-motor-vehicle-systems/vehicle-identification-systems-classification-software/ (September 10, 2015).

3. Eyedea, Vehicle Type Recognition. http://www.eyedea.cz/vehicle-type-recognition (September 10, 2015).

A. APPENDIX A

SUPPORTING INFORMATION FOR CHAPTER 2

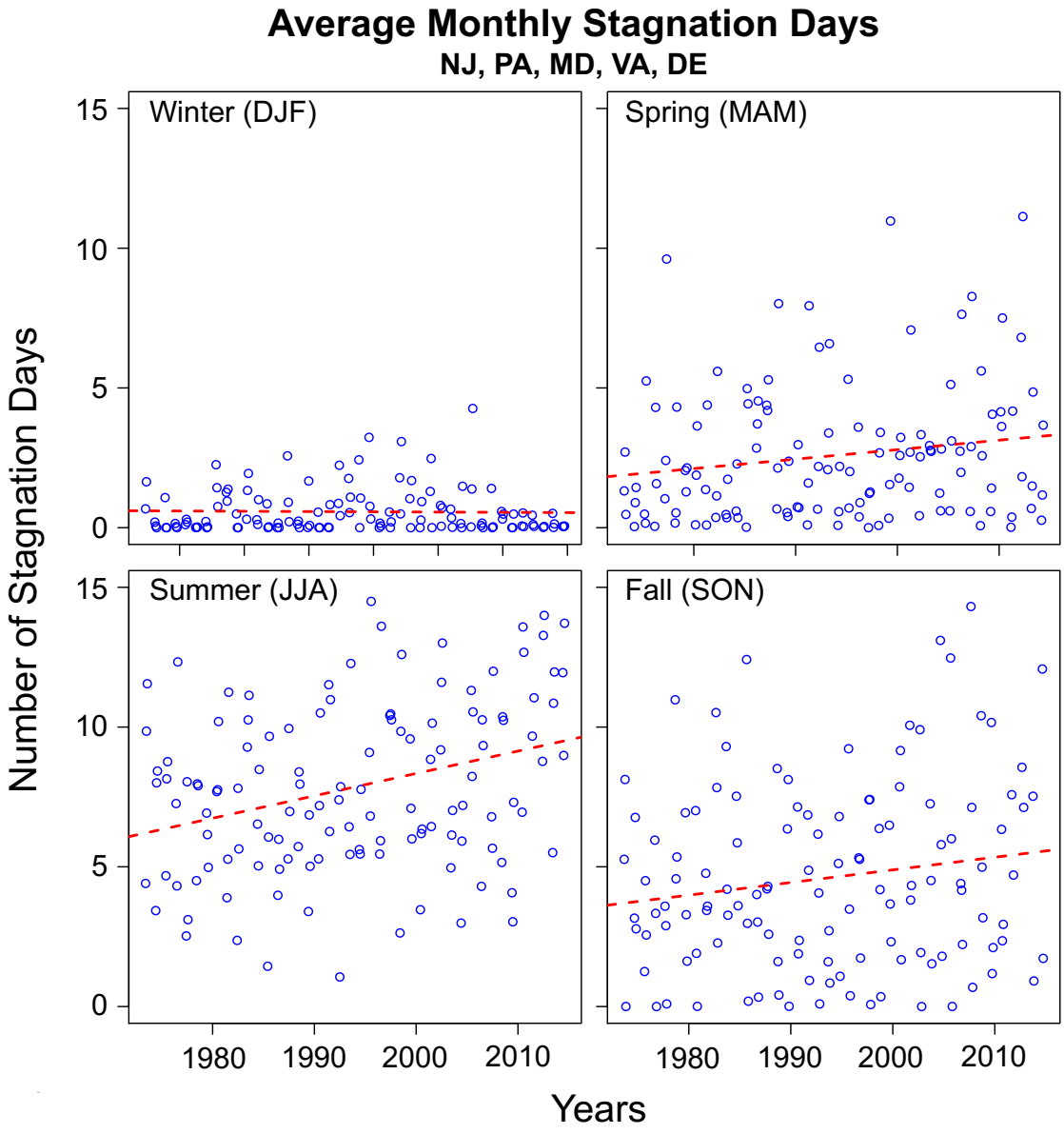

Figure A-1. Average monthly stagnation days by season from the National Climatic Data Center (NCDC) for the 5 studied states (Delaware, Maryland, New Jersey, Pennsylvania, Virginia) are represented by blue open circles. The red dashed line is a linear trendline.

Electricity Modeling Details:

DAYZER simulations were created using the DAYZER tool in long-term form downloaded from Cambridge Energy Solutions (http://www.ces-us.com/product-dayzer.asp). The July 1-31, 2006 simulation used in this work was simulated using the "PJM RTO" pool with the corresponding PJM power flow data (PFD). The default user database (UserDB) was used as a baseline.

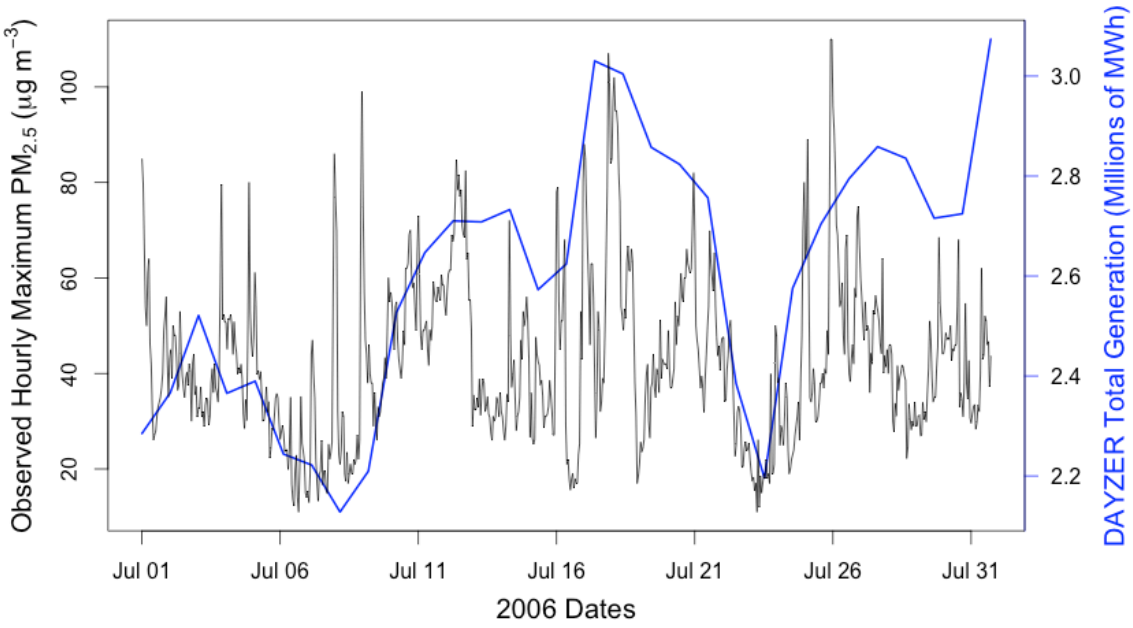

Figure A-2. CSN maximum hourly $PM_{2.5}$ observations from July 2006 for the study domain are represented by the thin black line. The daily DAYZER total electricity generation for the entire PJM region during the same time period is in blue.

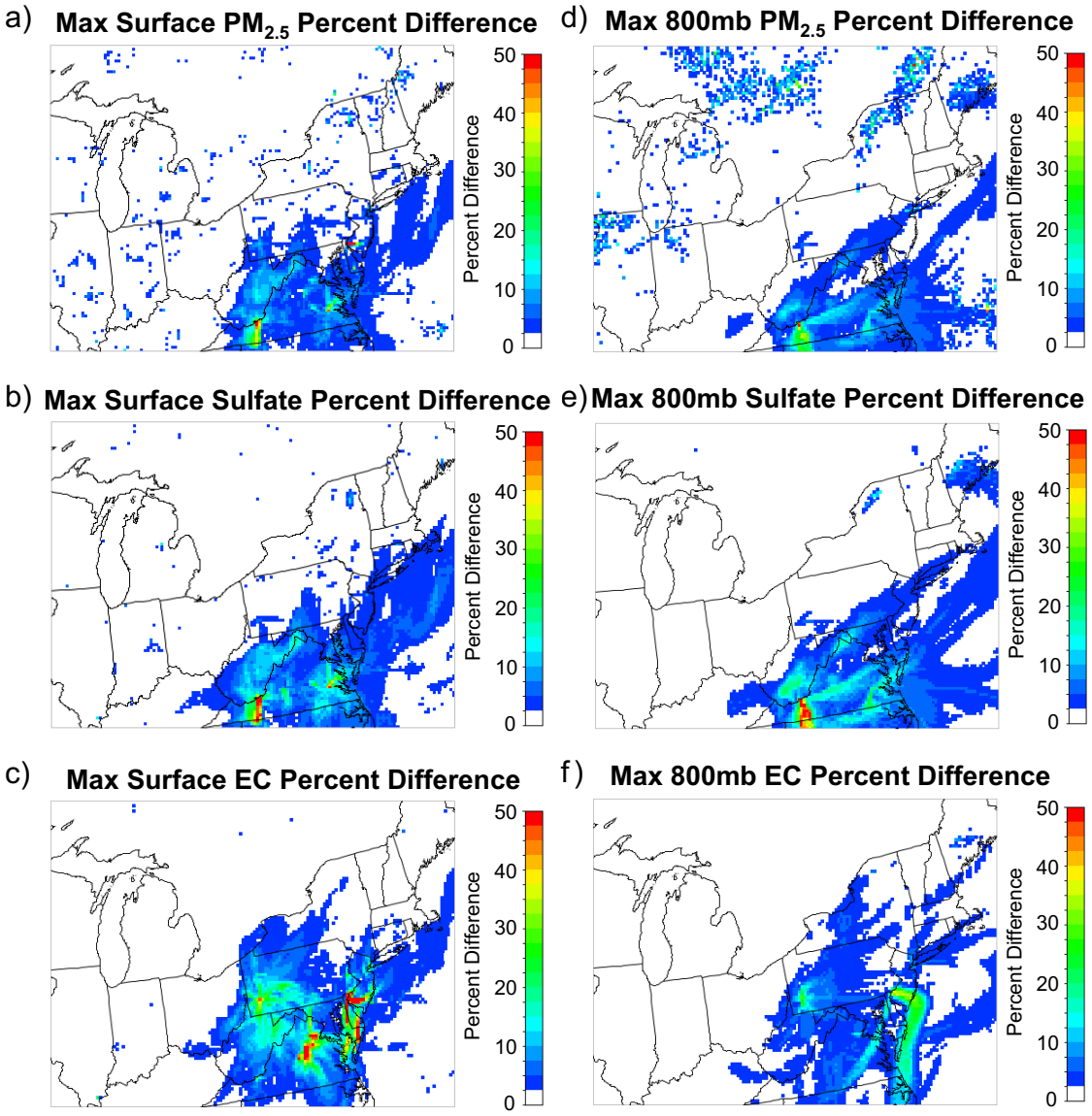

Figure A-3. CMAQ-predicted maximum percent differences of the two study simulations at the surface and 800mb between July 1-31, 2006 of ambient concentrations of $PM_{2.5}$ (a and d), sulfate (b and e) and EC (c and f).

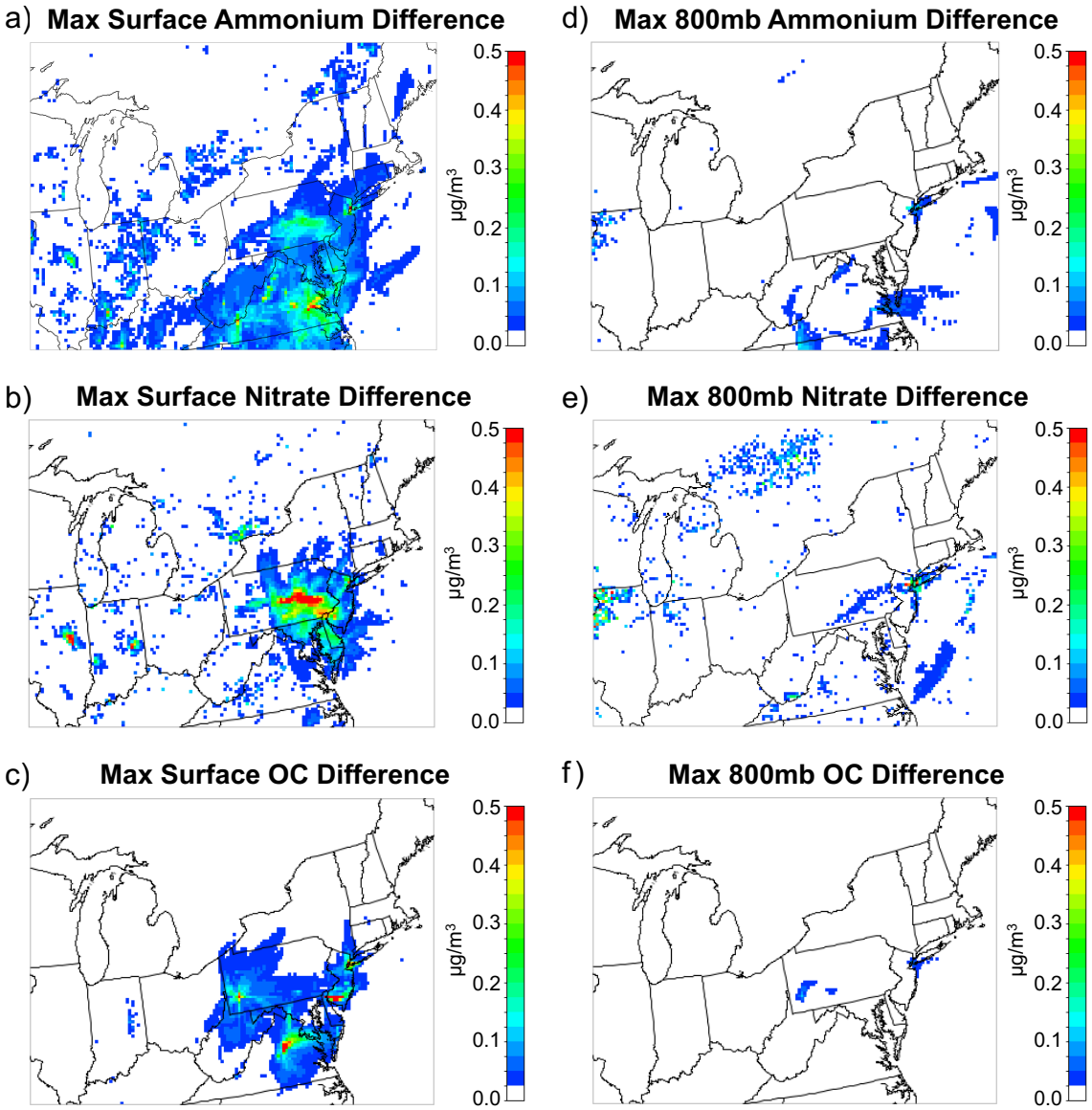

Figure A-4. CMAQ-predicted maximum hourly ambient concentration differences of the two study simulations at the surface and the 800mb level between July 1-31, 2006 of ambient concentrations of ammonium (a and d), nitrate (b and e) and OC (c and f).

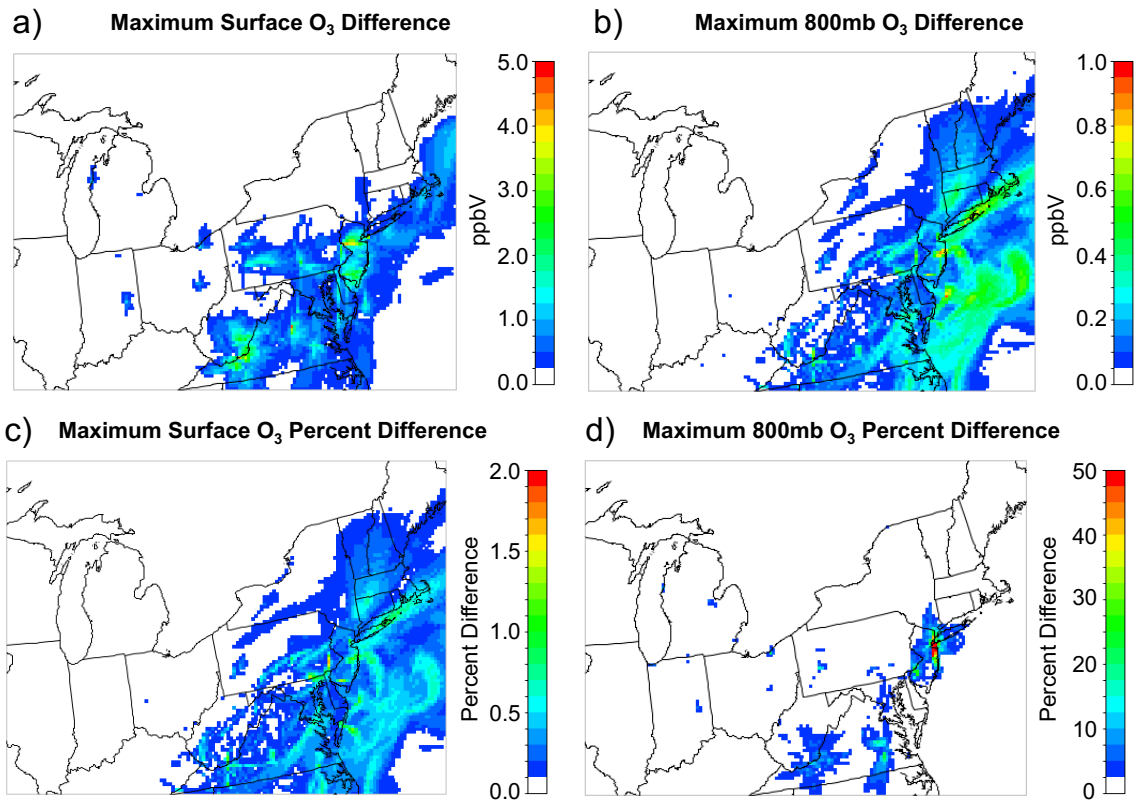

Figure A-5. CMAQ-predicted ppbV maximum hourly differences and maximum percent differences of the two study simulations at the surface (a,c) and 800mb (b,d) between July 1-31, 2006 of ambient mixing ratios of ozone.

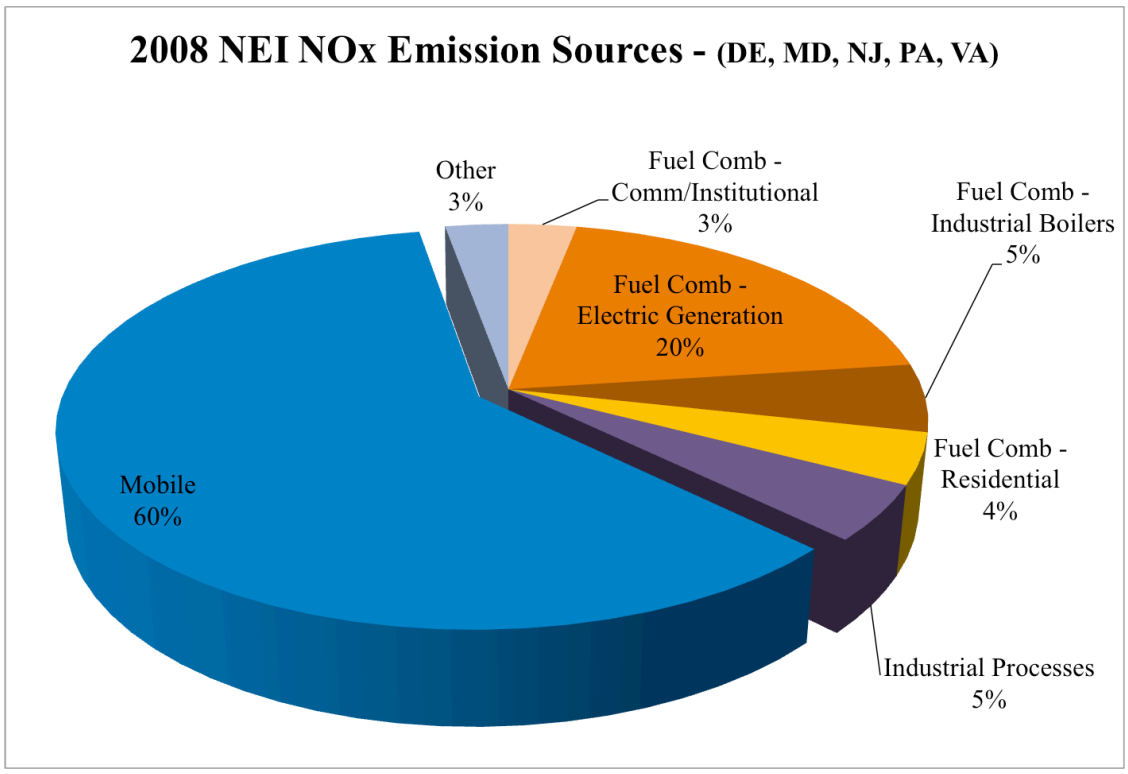

Figure A-6. Percentage of NO_x emissions emitted from each category of sources from the 2008 NEI for Delaware, Maryland, New Jersey, Pennsylvania, and Virginia.

B. APPENDIX B

SUPPORTING INFORMATION FOR CHAPTER 3

National		
Fuel Type	**SO_2** (lbs/MWh)	**NO_x** (lbs/MWh)
Coal	13	6
Oil	12	4
Gas	0.1	1.7
PJM		
Coal	5	0.9
Oil	1.1	1.7
Gas	0.03	1.3

Table B-1: National and PJM EGU Average Emission Rates of SO_2 and NO_x by Fuel Source: EPA peaking unit definition; 2006 continuous emission monitor (CEM) data from EPA's Clean Air Markets Division

EPA's peaking unit definition:
EGU generates < 10% of 3-year average capacity and < 20% each year

Plant 1
Capacity: 100,000 million British thermal units (MMBTU) per year

Year 1	Year 2	Year 3
16,000 MMBTU	5,000 MMBTU	8,000 MMBTU
(16%)	(5%)	(8%)

3-year average = 9.7%

Figure B-1: Example of the classification of an EGU as an EPA-defined peaking unit.

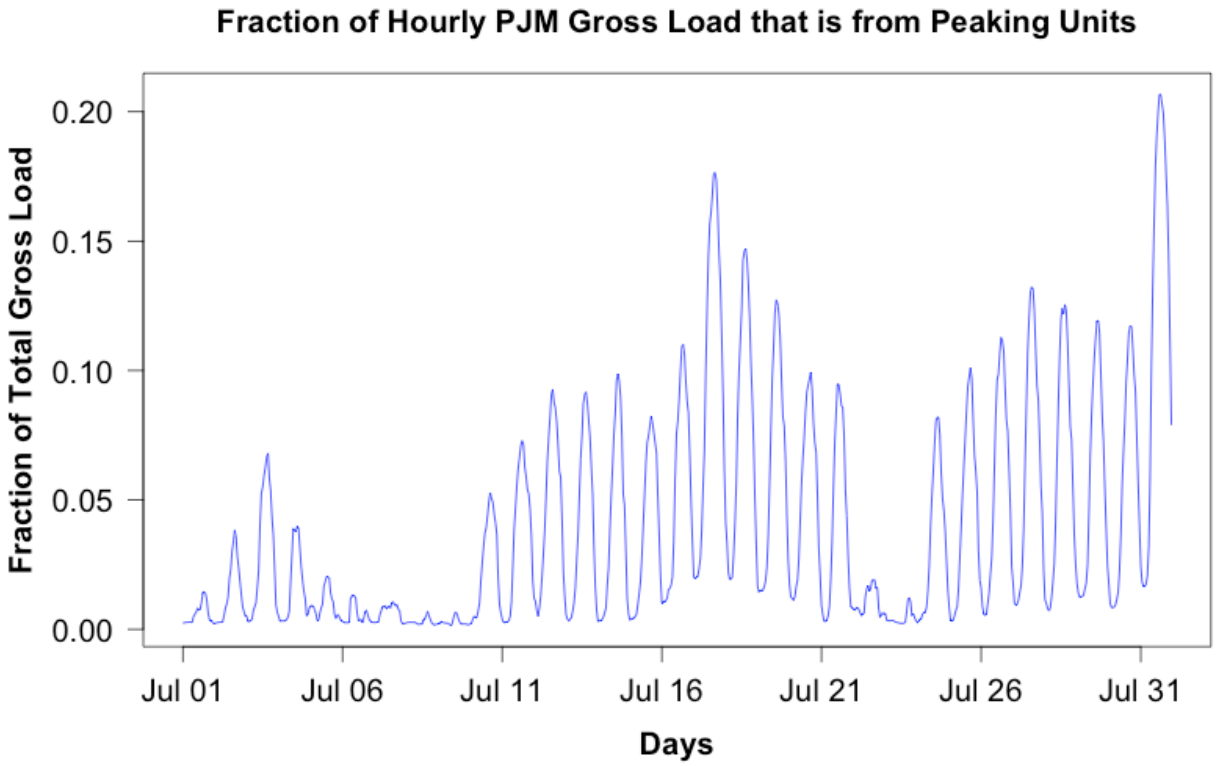

Figure B-2: The fraction of hourly total PJM gross load that is PJM peaking unit load during July1-31, 2006. Note: The hourly data is from PJM continuous emissions monitors and does not include units that are not fitted with monitors.

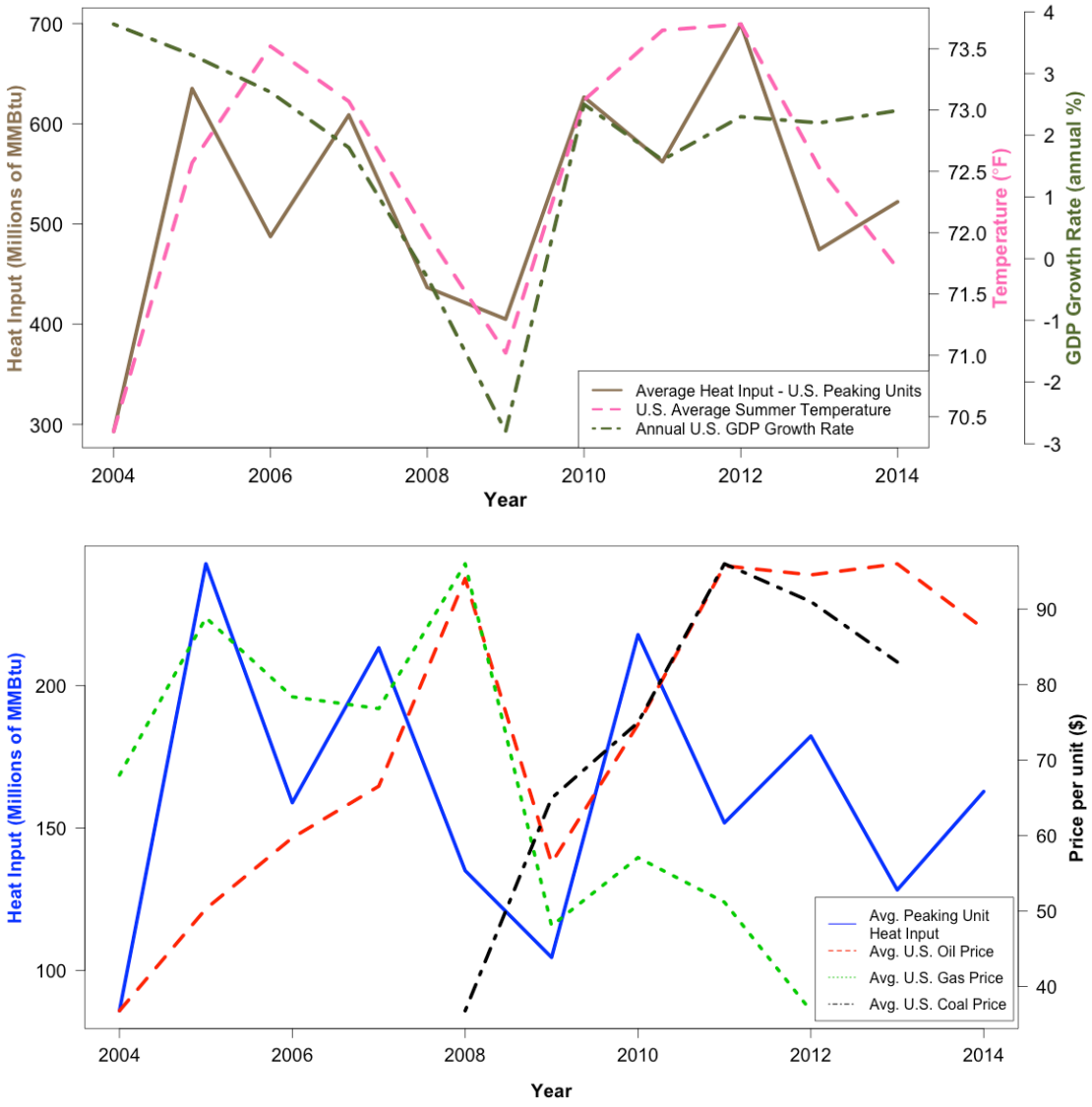

Figure B-3: a) U.S. annual heat input (millions of MMBtu) of peaking units (brown) compared to annual average temperature (pink dashed) and annual gross domestic product (GDP) growth rate (green dash-dot). The daily average temperature (Figure 2) has a slightly better correlation ($r = 0.8$, $p = 3\times10^{-3}$) than average summer high temperatures ($r = 0.79$, $p = 3\times10^{-3}$) partly due to the fact that overnight summer temperatures in addition to daily maximum also play a critical role in energy usage. b) PJM annual heat input (millions of MMBtu) of peaking units (blue) compared to annual average U.S. oil price per barrel (red dashed), annual average U.S. gas price per thousand cubic feet (green dotted) and annual average coal price per short ton (black dash-dot).

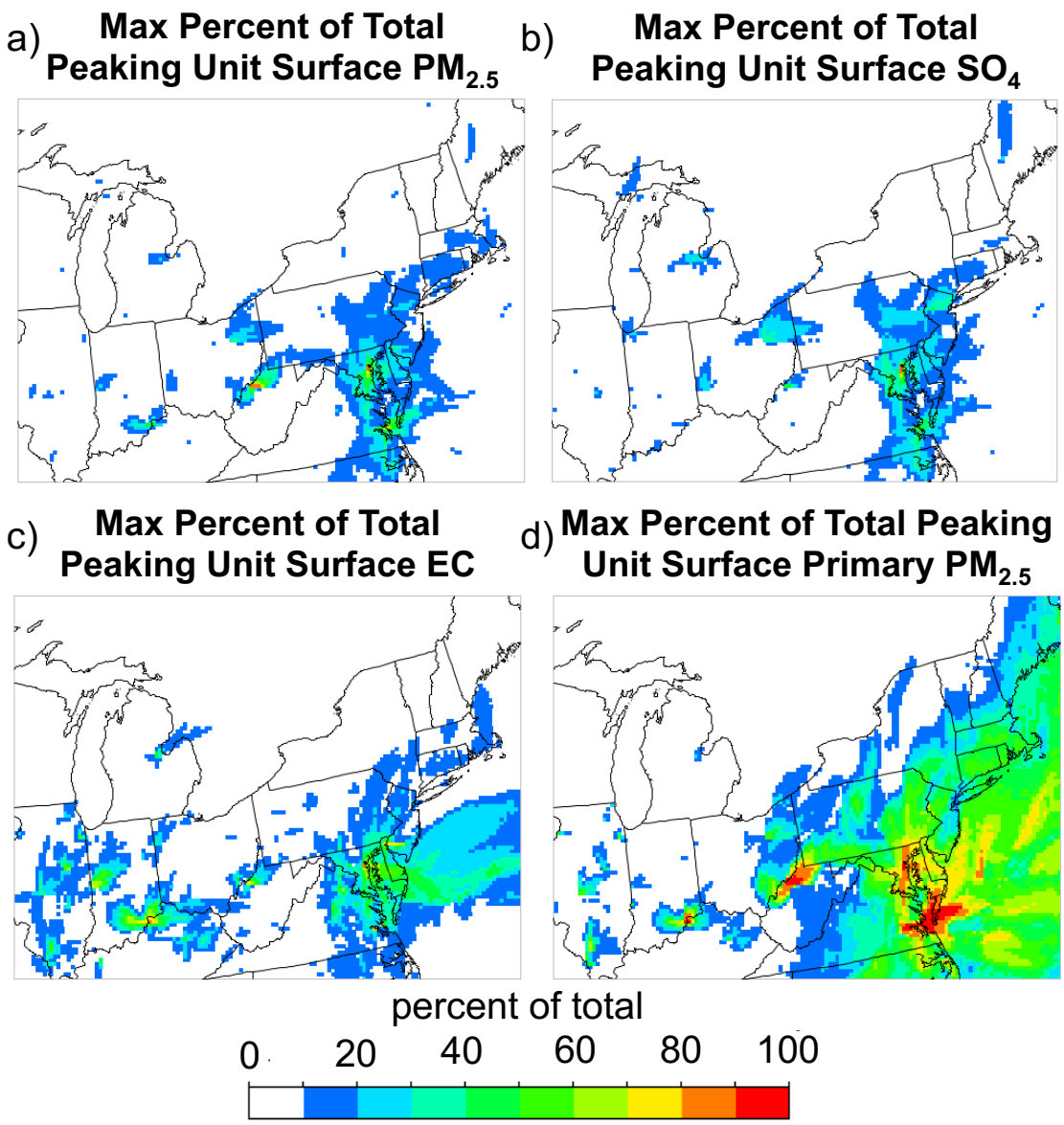

Figure B-4: Maximum percent of total from the 544 peaking units studied here to the Base Case concentrations of a) $PM_{2.5}$ mass b) particulate sulfate, c) elemental carbon, and d) primary unspeciated $PM_{2.5}$ from July1-31, 2006. This represents the maximum contribution of these peaking units to hourly total $PM_{2.5}$ during the studied time period.

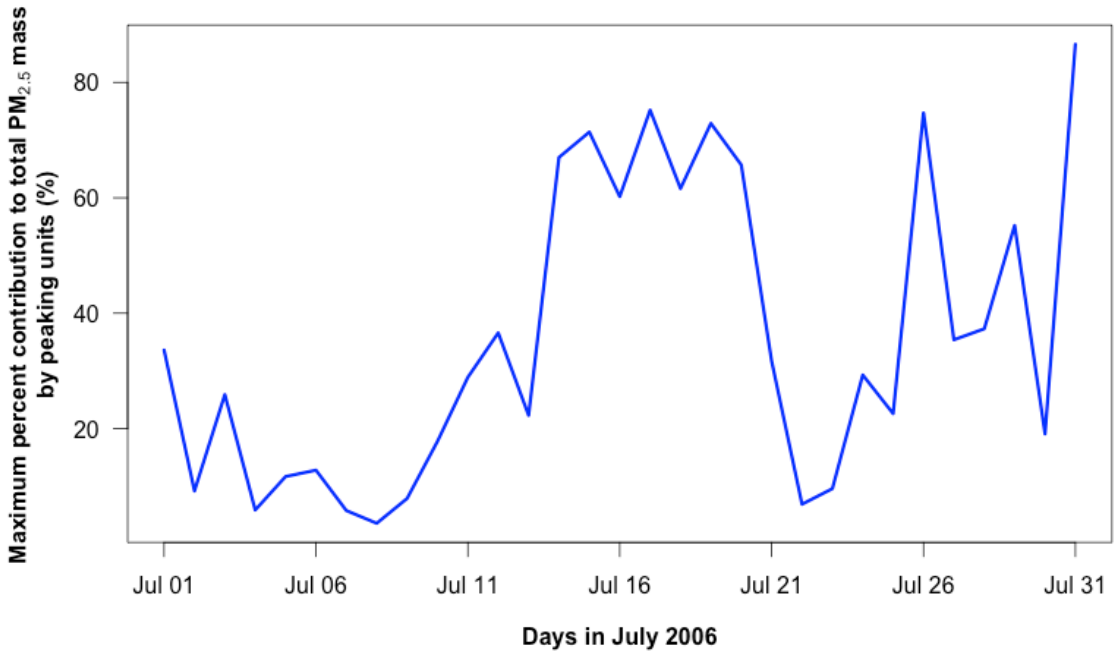

Figure B-5: Maximum daily percent contribution to CMAQ-predicted hourly ambient $PM_{2.5}$ mass by the 544 peaking units studied here from July 1-31, 2006.

Ozone Analysis

Ambient surface hourly maximum ozone mixing ratios are highest in the All PJM simulation throughout and outside the PJM region (Figure B-5b, up to 45 ppb_V). PJM Peaking Units result in hourly maximum ozone mixing ratios of up to14 ppb_V (Figure B-5a), 9% of the maximum hourly O_3 concentrations of the Base Case (Figure B-6), with multiple peak values occurring from Northern New Jersey extending south through the coast of Virginia.

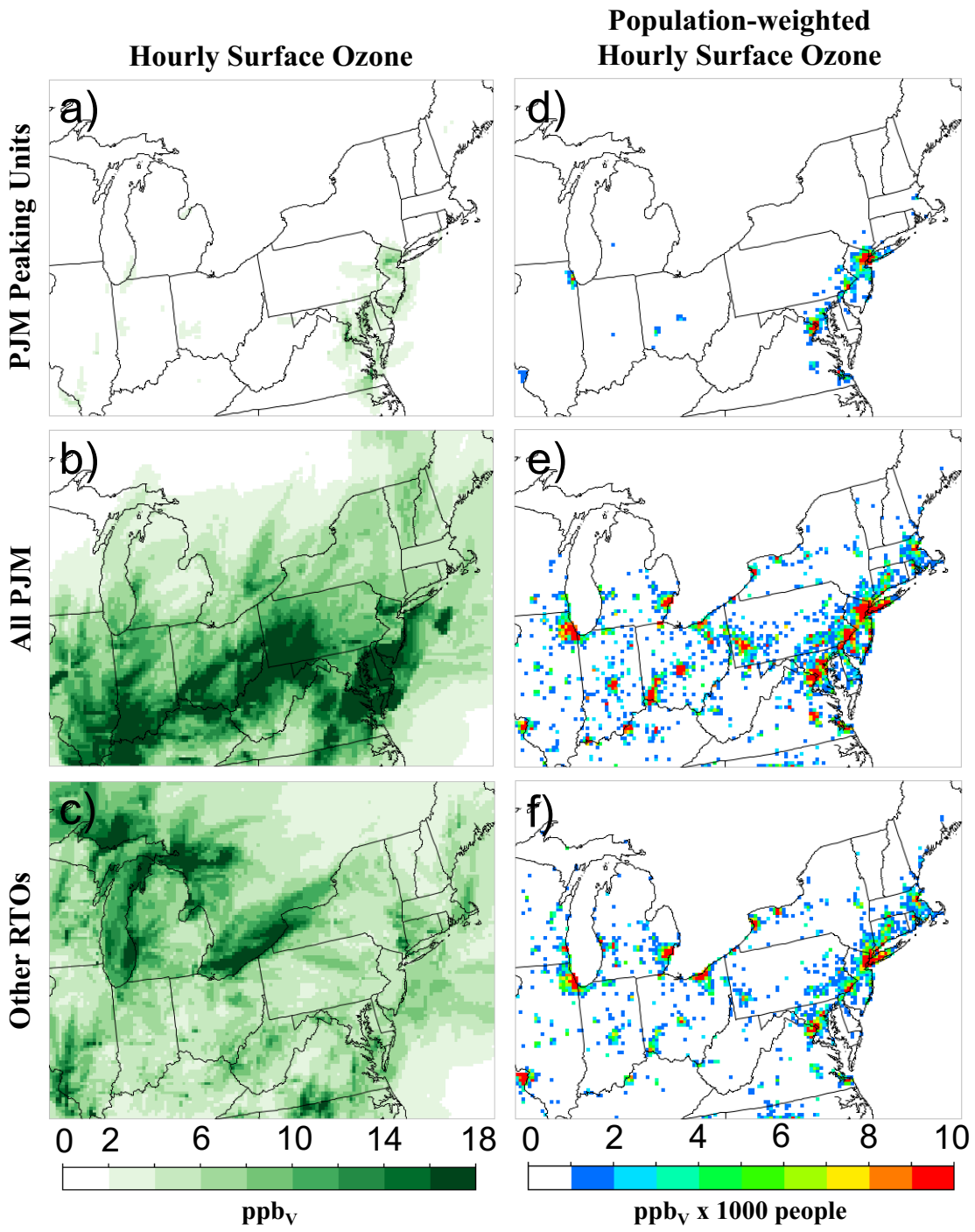

Figure B-6: CMAQ-predicted maximum hourly ambient concentrations of ozone for PJM Peaking Units, All PJM, and Other RTOs at the surface from July 1-31, 2006 (a-c) and CMAQ-predicted maximum hourly population-weighted ambient concentrations of ozone for PJM Peaking Units, All PJM, and Other RTOs at the surface from July 1-31, 2006 (d-f).

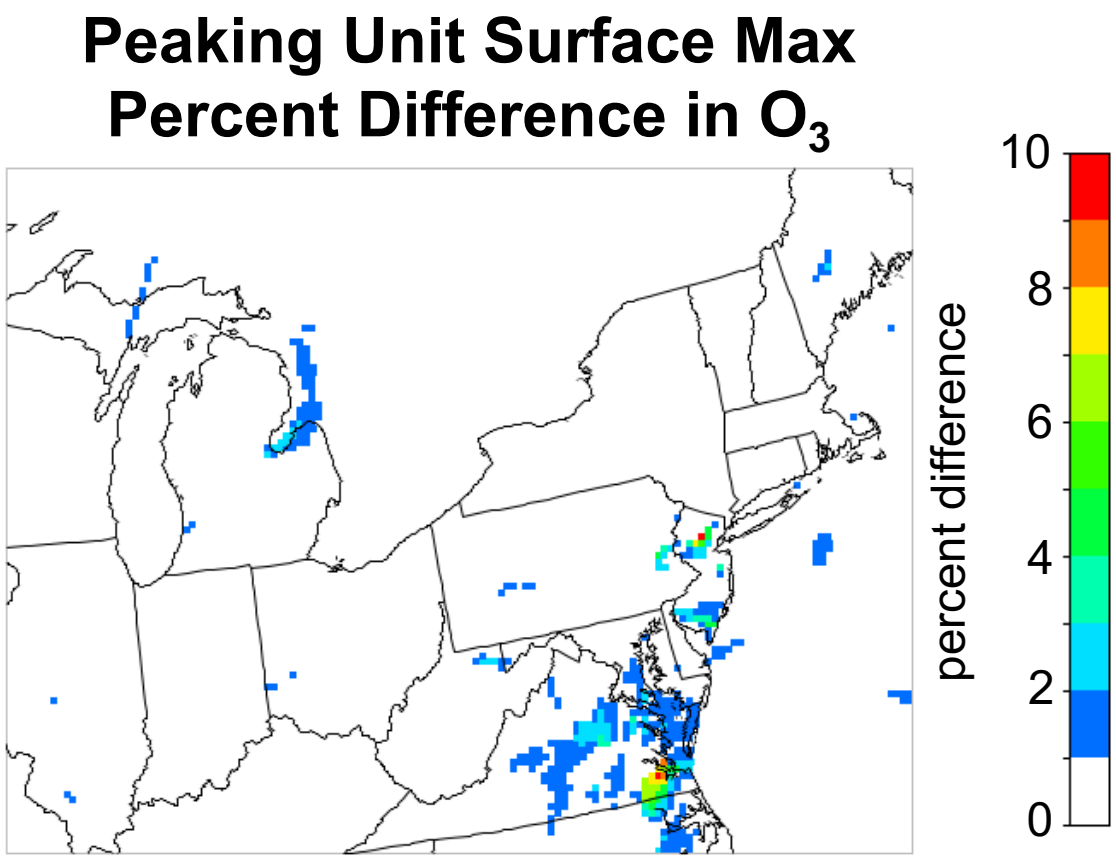

Figure B-7: Maximum percent difference between the 544 peaking units studied here and the Base Case for ozone from July1-31, 2006. This represents the maximum contribution of these peaking units to hourly total ozone during the studied time period.

Aloft analysis

Meteorology can exacerbate air quality episodes, often through stagnation that causes accumulation of trace species such as chemically-produced sulfate and $PM_{2.5}$ mass. Along coastal areas, land/sea breeze events can also exacerbate pollution events. Aloft, at 800 mb, CMAQ-predicted peaking unit ambient concentrations of $PM_{2.5}$ reach an hourly maximum up to 6.6 $\mu g/m^3$ with primary unspeciated $PM_{2.5}$ and sulfate species contributing the majority, up to ~3 $\mu g/m^3$ each (Figure B-7). This result is unique. Primary $PM_{2.5}$ is emitted at the surface, while sulfate forms aloft in the atmosphere, often in clouds, areas of convective mixing and vertical transport. Note the spatial distribution in maximum change aloft is different than at the surface, with the majority of the maxima over Southeastern Virginia, Western Maryland and Southern New Jersey.

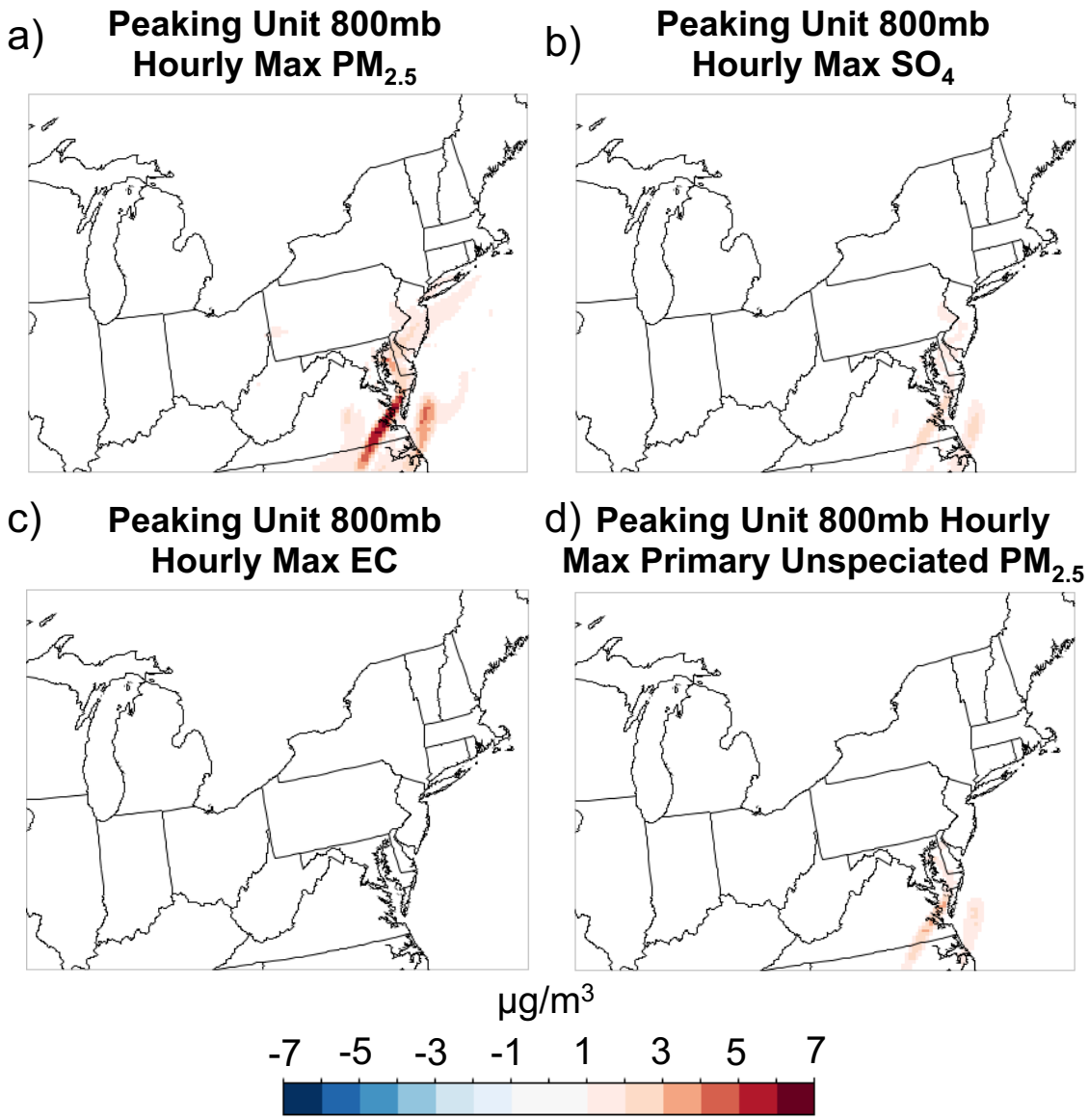

Figure B-8: CMAQ-predicted ambient hourly maximum concentrations due to peaking units at 800mb of total $PM_{2.5}$ mass (a), sulfate (b), EC (c) and primary unspeciated $PM_{2.5}$ (d) from July 1-31, 2006.

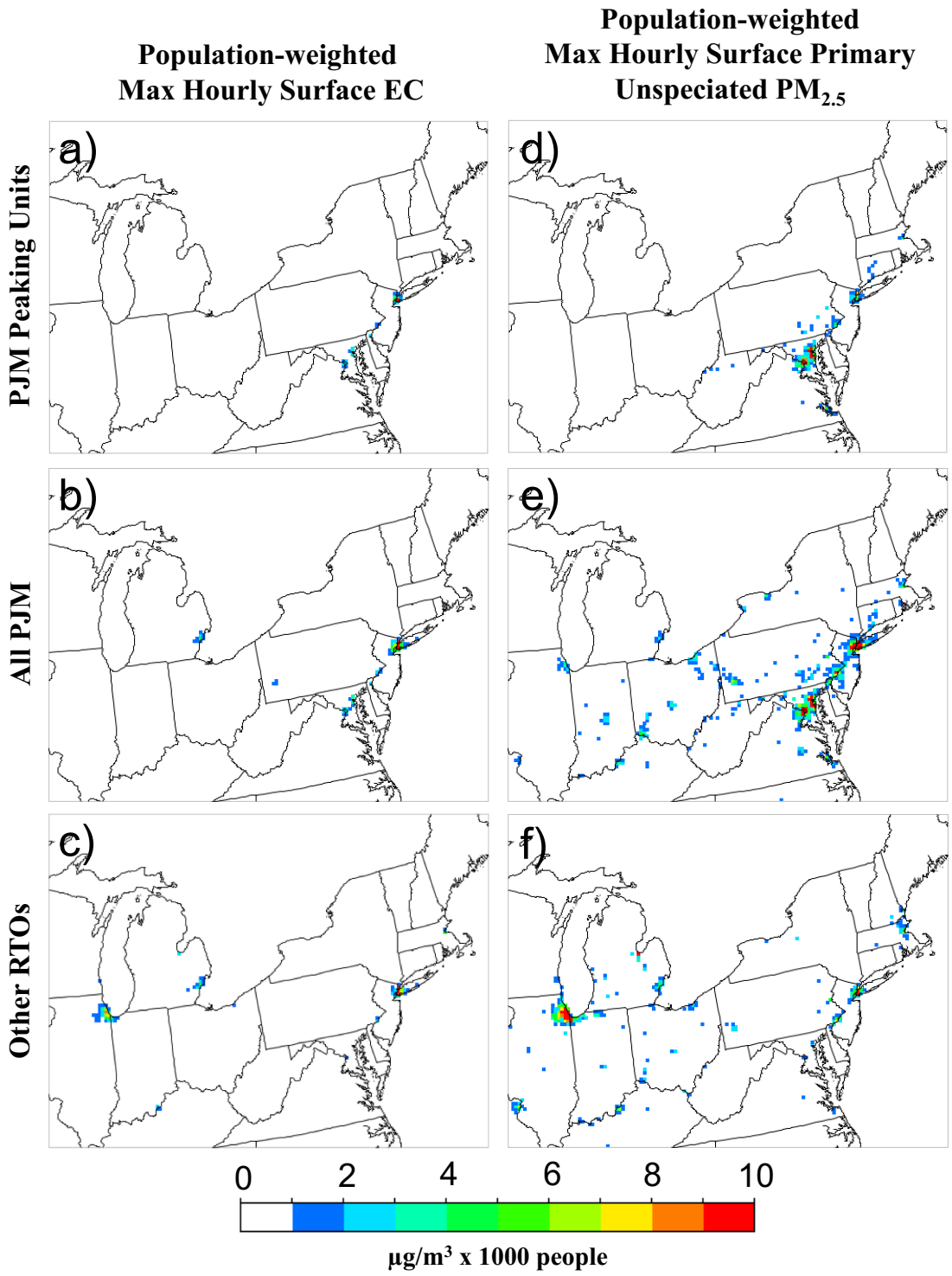

Figure B-9: CMAQ-predicted maximum hourly population-weighted ambient concentrations of EC and primary unspeciated $PM_{2.5}$ mass for PJM Peaking Units, All PJM, and Other RTOs at the surface from July 1-31, 2006.

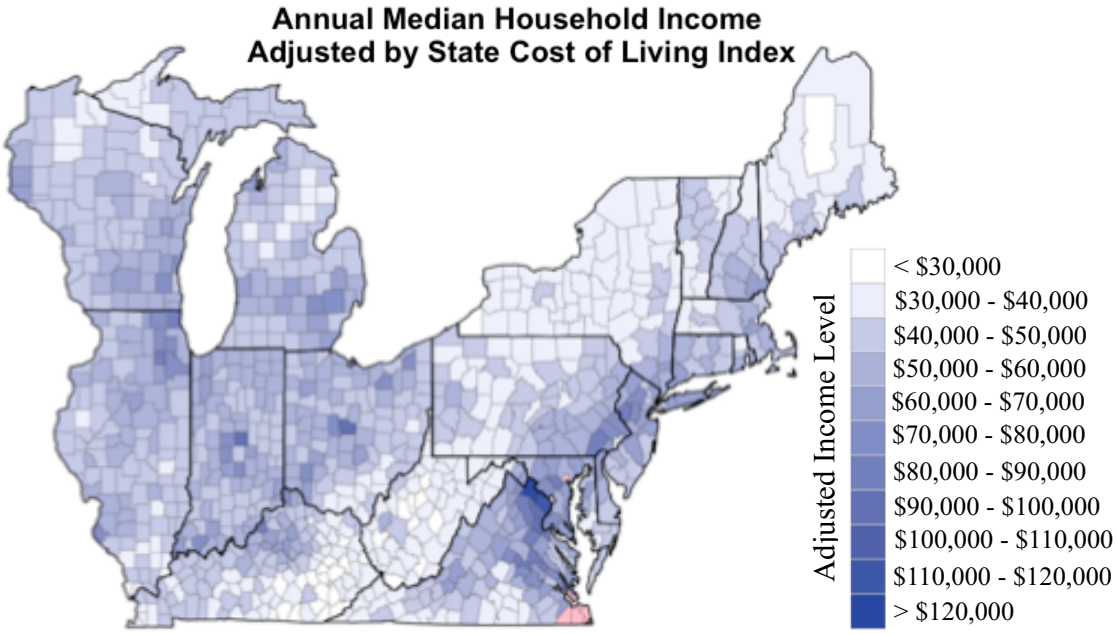

Figure B-10: Annual median household income for 2006-2010 as reported in the 2010 American Community Survey by the US Census Bureau adjusted by a state-level cost of living index. The pink coloring in Virginia represents areas with missing data.

C. APPENDIX C

SUPPORTING INFORMATION FOR CHAPTER 4

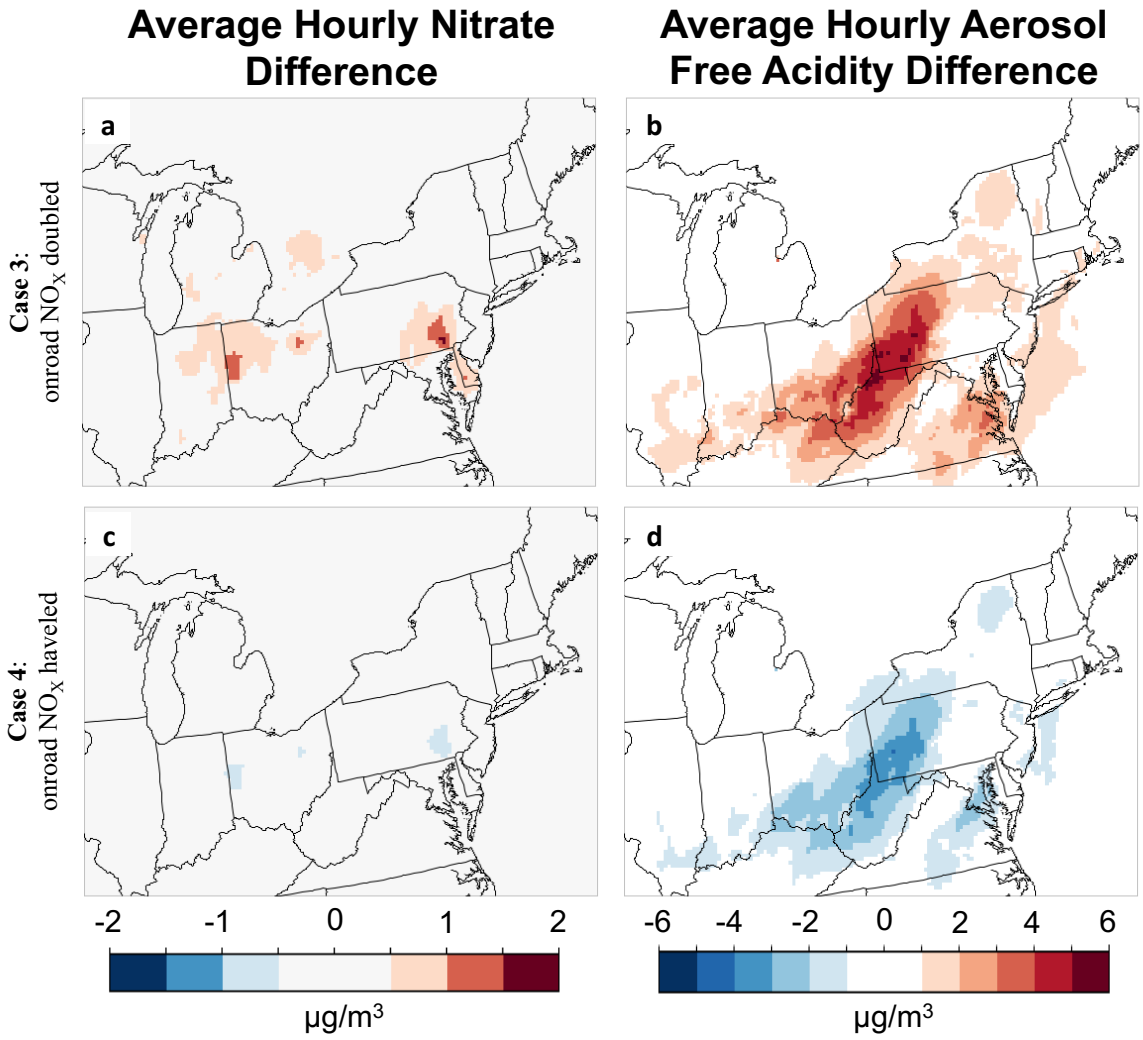

Figure C-1: Sensitivity Cases 3-4, CMAQ-predicted ambient concentration differences from the Base Case of average hourly nitrate (a,c) and hourly average aerosol free acidity (b,d) from July 1, 2006 to July 31, 2006.

D. APPENDIX D

EFFICIENCY OF BASE CASE MODEL RUNS ON PHOTON SERVER

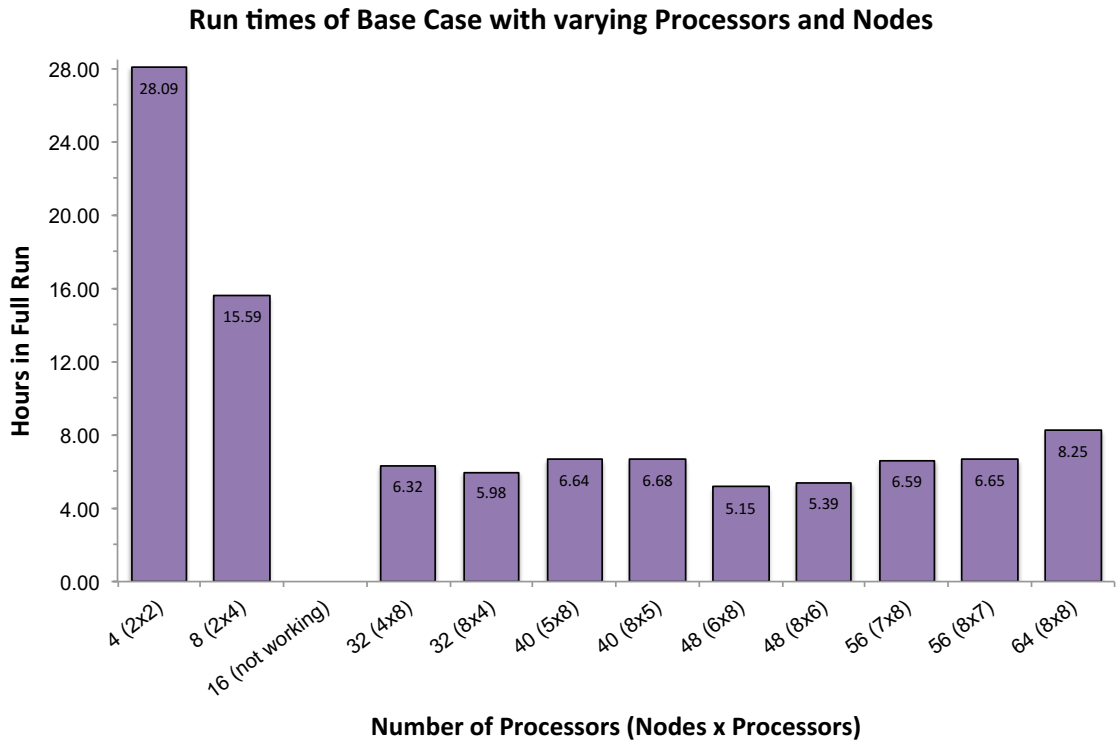

Figure D-1: Total run times of the Base Case simulation from Chapter 2 with varying configurations of nodes and processors to determine the optimal configuration that provides the fastest run time on the Rutgers' Department of Environmental Science's PHOTON server.

E. APPENDIX E

R SCRIPTS USED FOR THIS DISSERTATION

Appendix E Table of Contents

```
##############################################################################
# Figure 2-1b of Caroline Farkas Dissertation (2016)
# Script for Map of AQS and IMPROVE sites used
##############################################################################

#load libraries
> library(maps)
> library(mapproj)

#Read in IMPROVE data and get unique sites
> col_names5 <- c("Dataset", "Sitecode", "Date", "POC", "Lat", "Lon", "State", "ec",
        "pm25", "so4")
> col_classes5 <- c("character","character","character","numeric","numeric", "numeric",
        "character", rep("numeric",3))
> improve_sites<- read.csv(file=
        "/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_analaysis/IMPROVE
        _2006.txt", sep=",",  skip=1, header=F, col.names=c(col_names5),
        colClasses = c(col_classes5), na.strings=c("-999.0000","-999"))
> improve_unique <- unique(improve_sites$Sitecode)
> length(improve_unique)
#remove unneeded columns
> improve <- improve_sites[, c(2,5,6,7)]
> improve <- improve[!duplicated(improve),]
```

```
> nrow(improve) #should match length(improve_unique)


#Read in hourly PM2.5 data for AQS sites
> col_names <- c("siteID", "lat", "lon", "column", "row", "timeOn", "timeOff",
        "obPM25", "modPM25")
> col_classes <- c("numeric", "numeric", "numeric", "numeric", "numeric", "character",
        "character", rep("numeric",2))
> sitexAdded<- read.csv(file=
        "/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_analaysis/QA_2006_
        addto2008NEI_hourly.csv", sep=",", skip=6, header=F,col.names=c(col_names),
        colClasses=c(col_classes), na.strings=c("NA", "", "-999"))
> sites <- unique(sitexAdded$siteID)
#remove unneeded columns
> hourly <- sitexAdded[, c(1,2,3)]
> hourly <- hourly[!duplicated(hourly),]


#Plot AQS and IMPROVE onto map using different colors and print as PDF format
#begin PDF
> pdf("/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_analaysis/obs_
        location_map.pdf", width=8, height=7)
> map(database="state",regions=c("Virginia", "New Jersey", "Pennsylvania", "West
        Virginia", "Delaware", "Maryland", "DC"))
> points(improve$Lon,improve$Lat,cex=1.5,pch=17,col="red")
```

```
> points(hourly$lon,hourly$lat, cex=1.5,pch=20,col="blue")
> legend("topleft", c("IMPROVE sites", "AQS hourly sites"), pch=c(17,20),
        col=c("red","blue"), cex=0.65)
#add names of cities above 250k people for reference
> data(us.cities)
> map.cities(us.cities, country=c("NJ"), label=T, minpop=250000, font=2)
> map.cities(us.cities, country=c("PA"), label=T, minpop=250000, font=2)
> map.cities(us.cities, country=c("MD"), label=T, minpop=250000, font=2)
> map.cities(us.cities, country=c("DC"), label=T, minpop=250000, font=2)
> map.cities(us.cities, country=c("VA"), label=T, minpop=250000, font=2)
> map.cities(us.cities, country=c("VA"), label=T, minpop=40050, maxpop=41000,
        cex=1, font=2)
> map.cities(us.cities, country=c("DE"), label=T, font=2)
#end PDF
> dev.off()
```

```
##############################################################################
# Figure 2-2 of Caroline Farkas Dissertation (2016)
# Time series of July 2006 showing daily concentrations of PM2.5, measured every third
day (IMPROVE) versus DAYER daily concentrations of PM2.5. Additional pie plots
show the percentage of species that make up PM2.5.
##############################################################################
> library(mapplots)
> library(lubridate)
# Read in IMPROVE data, separated by pollutant
> col_names6 <-c("Dataset", "Sitecode", "Date", "POC", "Lat", "Lon", "State", "ec",
        "pm25", "so4")
> col_classes6 <- c("character","character","character","numeric","numeric", "numeric",
        "character",rep("numeric",3))
> improve<-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
        analaysis/IMPROVE_2006.txt",sep=",", skip=1, header=F,
        col.names=c(col_names6), colClasses=c(col_classes6), na.strings=c("-
        999.0000","-999"))
> improve$Dataset <- NULL
> improve$Date <- mdy(improve$Date)
> improve$other <- improve$pm25 - improve$so4 - improve$ec
#make data frame of maximum of all points on certain date
> max.improve <- aggregate(cbind(ec,pm25,so4,other)~Date, data=improve,
        FUN="max")
```

```
> write.csv(max.improve, "/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
        analaysis/max_improve.csv",row.names=FALSE)

#IMPROVE is measured every 3 days. So, in textedit, I added the missing dates with no
(or blank) data to more easily match up with the DAYZER data that had points for
everyday. Read back in, format date, merge with DAYZER data and remove unwanted
columns:
> col_names9 <- c("dateon","ec","pm25","so4","other")
> col_classes9 <- c("character",rep("numeric",4))
> max.improve.new<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/
        Manuscript_data_analaysis/max_improve.csv",sep=",", skip=1,header=F,
        col.names=c(col_names9), colClasses=c(col_classes9))
> max.improve.new$dateon <- ymd(max.improve.new$dateon)
> all_max.improve <- merge(max.improve.new,dayzer,by.x=("dateon"),by.y=("Date"))
> all_max.improve$Generation_MWh <- NULL
> all_max.improve$Spin_MWh <- NULL

#read in DAYZER generation
> dayzer.file <- "/Users/Carna/Downloads/Daily+Generation+from+Jul-Aug+2006-
        2.csv"
> col_names2 <- c("Date", "Generation_MWh", "Spin_MWh",
        "Total_Generation_MWh")
> col_classes2 <- c("character", "numeric", "numeric","numeric")
```

```
> dayzer <- read.csv(file=dayzer.file, skip=1, sep=c(","), header=F,
        col.names=c(col_names2), colClasses=c(col_classes2), na.strings=c("#N/A", "-
        999", "-9"))
#I made a vector called dates to subset out the dates in DAYZER that are for July 2006.
This could be done MUCH easier by just subsetting with less than and greater than
statements. This was early on in my R coding abilities….
> dates <- c("7/1/06", "7/2/06", "7/3/06", "7/4/06", "7/5/06", "7/6/06", "7/7/06", "7/8/06",
        "7/9/06", "7/10/06", "7/11/06", "7/12/06", "7/13/06", "7/14/06", "7/15/06",
        "7/16/06", "7/17/06", "7/18/06", "7/19/06", "7/20/06", "7/21/06", "7/22/06",
        "7/23/06", "7/24/06", "7/25/06", "7/26/06", "7/27/06", "7/28/06", "7/29/06",
        "7/30/06", "7/31/06")
> dayzer <- subset(dayzer, dayzer$Date%in%dates)
> dayzer$Date <- mdy(dayzer$Date)
> dayzer$Mil_MWh <- dayzer$Total_Generation_MWh/1000000


#Making the Graph
> par(mar=c(4,4,1,5))
> plot(max.improve$Date, max.improve$pm25, ylim=c(5,40),
        xlim=c(1151712000,1154304000), xlab="2006 Dates",
        ylab=expression(paste("Observed PM"[2.5], " (", mu, "g ",m^{-3},
        ")")),cex.lab=1.25,mgp=c(2.5,1,0), las=1, type="l",lty=2)
> par(new=T)
```

```
> plot(dayzer$Date,dayzer$Mil_MWh, type="l", xlim=c(1151712000,1154304000),
        col="blue",yaxt="n", ylab=NA, xlab=NA,axes=F, lwd=2)
> axis(4,at=c("2.0","2.2","2.4","2.6","2.8","3.0","3.2"),col="blue",las=1,col.ticks="blue")
> mtext(side=4,line=3,"DAYZER Total Generation (Millions of
        MWh)",col="blue",cex=1.25)
> par(new=T)
> plot(NA,NA, xlim=c(1151712000,1154304000), ylim=c(5,40) ,las=1, ylab="", xlab="",
        xaxt="n", yaxt="n",bty="n")
#make pie charts appear on graph
> a=1
> for (i in seq(1151712000,1154304000,86400)){
        add.pie(z=c(max.improve[a,2],max.improve[a,4],max.improve[a,5]),
                x=max.improve[a,1], y=max.improve[a,3], radius=1.25, labels="",
                col=c("black","red","gray73"), border=F)
         percent <-max.improve[a,4] / (max.improve[a,2] + max.improve[a,4] +
                max.improve[a,5])*100
        percent <- formatC(percent,format="f",digits=0)
        text(paste("  ",percent,"%",sep=""), x=max.improve[a,1], y=max.improve[a,3] +
                2.5, cex=1,col="red")
        a=a+1
        }
```

```
##############################################################################
# Figure 2-3 of Caroline Farkas Dissertation (2016)
# Scatter plot of CMAQ model bias using IMPROVE measured data for SO4 and PM2.5
##############################################################################
#read in IMPROVE data
> col_names <- c("siteID", "lat", "lon", "column", "row", "timeOn", "timeOff",
        "obASO4", "modASO4", "obAEC", "modAEC", "obPM25", "modPM25")
> col_classes <- c("numeric", "numeric", "numeric", "numeric", "numeric", "character",
        "character", rep("numeric",6))
> sitexBase<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
        analaysis/QA_stn_2006.csv",sep=",", skip=6, header=F,col.names=c(col_names),
        colClasses=c(col_classes))

#format time and calculate biases
sitexBase$timeOn <- mdy_hm(as.character(sitexBase$timeOn))
sitexBase$biasASO4 <- sitexBase$modASO4 - sitexBase$obASO4
sitexBase$biasPM25 <- sitexBase$modPM25 - sitexBase$obPM25

#Plot model bias of base case with PM2.5 and SO4 on the same plot
> par(mar=c(4,4,4,4))
> plot(sitexBase$obASO4,sitexBase$biasASO4, col = "red", xlab="", xlim=c(0,20),
        ylim=c(-15,15), xaxs="i", yaxs="i", ylab="", cex.lab=1.25, lab.col="red",
        mgp=c(2.75,1,0), las=1, pch=16, Family="Times")
```

```
> mtext(expression(paste("Observed Sulfate (", mu, "g ",m^{-3},  ")")),
        line=2.5,side=1,cex=1.25, col="red")
> mtext(expression(paste("Sulfate Bias (modeled - observed)")), line=2.25, side=2,
        cex=1.25, col="red", Family="Times")
> abline(v=0,h=0,lty=2,col="black")
> par(new=T)
> plot(sitexBase$obPM25, sitexBase$biasPM25, col="blue", xaxs="i", yaxs="i",
        xlab="", ylab="", xlim=c(0,50), ylim=c(-30,30), xaxt="n", yaxt="n",
        mgp=c(2.75,1,0), las=1)
> axis(3)
> mtext(expression(paste("Observed PM"[2.5], " (", mu, "g ",m^{-3}, ")")), line=2,
        side=3, cex=1.25, col="blue")
> axis(4,las=1)
> mtext(expression(paste("PM"[2.5]," Bias (modeled - observed)")), line=2.75, side=4,
        cex=1.25, col="blue")
> legend(0.5,29, c(expression(paste("SO"[4])), expression(paste("PM"[2.5]))),
        pch=c(16,1), col=c("red","blue"))
```

```
##############################################################################
# Figure 2-4 of Caroline Farkas Dissertation (2016)
# CMAQ model bias of PM2.5 during the month of July 2006 and compared to hourly measured values
##############################################################################
> library(lubridate)

#Read in the base case vs obs (hourly data downloaded from ttn - used 88502)
> col_names <- c("siteID", "lat", "lon", "column", "row", "timeOn", "timeOff",
        "obPM25", "modPM25")
> col_classes <- c("numeric", "numeric", "numeric", "numeric", "numeric", "character",
        "character", rep("numeric", 2))
> sitexBase<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
        analaysis/QA_2006_2008NEIbase_hourly.csv",sep=",", skip=6, header=F,
        col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA", "", "-
        999"))

> a <- unique(sitexBase$siteID)
> length(a)
> sitexBase$dateTime <- mdy_hm(sitexBase$timeOn)
```

```
#Read in the sensitivity simulation (added ORIS IDs case) vs obs
> col_names <- c("siteID", "lat", "lon", "column", "row", "timeOn", "timeOff",
        "obPM25", "modPM25")
col_classes <- c("numeric", "numeric", "numeric", "numeric", "numeric", "character",
        "character", rep("numeric", 2))
sitexAdded<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
        analaysis/QA_2006_addto2008NEI_hourly.csv",sep=",", skip=6, header=F,
        col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA", "", "-
        999"))

> b <- unique(sitexAdded$siteID)
> length(b)
> sitexAdded$dateTime <- mdy_hm(sitexAdded$timeOn)

#merge the two datasets
> all <- merge(sitexBase,sitexAdded, by.x=c("siteID", "dateTime"),
        by.y=c("siteID","dateTime"))
> all$basePM25 <- all$modPM25.x
> all$addedPM25 <- all$modPM25.y
> all$lat.y <- NULL
> all$lon.y <- NULL
> all$column.y <- NULL
> all$row.y <- NULL
```

```
> all$obPM25.y <- NULL

> all$modPM25.y <- NULL

> all$modPM25.x <- NULL


#compute difference between observed and modeled

> all$baseDiff <- all$basePM25-all$obPM25  #base - obs

> all$addedDiff <- all$addedPM25-all$obPM25  #added - obs


#fix the date

> all$timeOn <- mdy_hm(all$timeOn)


#2-panel plot of model bias

> par(mfrow=c(1,2), oma = c(3,3,1,0) + 0.1, mar = c(0,0,1,1) + 0.1)

#plot bias vs time

> par(mar=c(1.25,1,0,0.3))

> plot(sub$dateTime,sub$baseDiff, pch=2, mgp=c(2.5,1,0), cex.lab=1.25, las=1)

> points(sub$dateTime, sub$addedDiff, col="blue", pch=20)

> abline(h=0)

#plot bias vs obs

> par(mar=c(1.25,0.3,0,0.25))

> plot(sub$obPM25,sub$baseDiff, pch=2, yaxt='n')

> points(sub$obPM25,sub$addedDiff, col="blue", pch=20)

> abline(h=0)
```

```
> legend(57,32,c("Base Case","Matched CEM Case"),pch=c(2,20),col=c("black","blue"),
        cex=1, pt.cex=1.5)

> title(xlab = expression(paste("Dates                    Observed Hourly PM"[2.5], "
        (", mu, "g ", m^{-3}, ")")), ylab=expression(paste("Model Bias of PM"[2.5],"
        (Model - Observations)")), cex.lab=1.25, outer = TRUE, mgp=c(1.5,1,0),
        cex.main=2)
```

```
##############################################################################
# Figure 2-6 of Caroline Farkas Dissertation (2016)
# 4-panel comparison of default SMOKE temporalization versus matched CEM hourly
with ORIS IDs
##############################################################################
> library(lubridate)

#Read in Coal Plant data
> col_names2 <- c("State", "FacilityName", "FacilityID", "UnitID", "AssociatedStacks",
        "Year", "Date", "Hour", "HeatInput", "AnnFactor", "AnnSO2",
        "SMOKE_SO2_hourly", "CEMScan_SO2_hourly", "AnnPM25",
        "SMOKE_PM25_hourly", "CEMScan_PM25_hourly")
> col_classes2 <- c(rep("character",2),"numeric", "character", "character", "numeric",
        "character" ,rep("numeric",9))
> Coal<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
        analaysis/James_Cogen_Coal.csv", sep=",", skip=1, header=F,
        col.names=c(col_names2), colClasses=c(col_classes2))
> Coal$newdate <- mdy_hm(paste(Coal$Date, " ", Coal$Hour,":00",sep=""))

#Change tons to lbs
> Coal$CEMScan_PM25_hourly <- Coal$CEMScan_PM25_hourly*2000
> Coal$SMOKE_PM25_hourly <- Coal$SMOKE_PM25_hourly*2000
```

```
#Read in NG plant data
> col_names <- c("Date", "Hour", "HeatInput", "AnnFactor", "AnnPM25",
          "CEMScan_PM25_hourly", "SMOKE_PM25_hourly", "AnnSO2",
          "CEMScan_SO2_hourly", "SMOKE_SO2_hourly")
> col_classes <- c("character",rep("numeric",9))
> NG<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript_data_
          analaysis/AES_Red_Oak.csv", sep=",", skip=1, header=F,
          col.names=c(col_names), colClasses=c(col_classes))

> NG$newdate <- mdy_hm(paste(NG$Date, " ", NG$Hour,":00",sep=""))
> Dates <- unique(NG$Date)

#Change tons to lbs
> NG$CEMScan_PM25_hourly <- NG$CEMScan_PM25_hourly*2000
> NG$SMOKE_PM25_hourly <- NG$SMOKE_PM25_hourly*2000

#For chart labels - time goes by seconds (86,400 sec per day since 1970)
> months <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct",
          "Nov", "Dec")
> mo <- c(jan,feb,mar,apr,may,jun,jul,aug,sep,oct,nov,dec)
> jul_days <- c("12","13","14","15","16","17","18","19","20","21","22","23","24","25")
```

```
> jul_sec <- c(jul12, jul13, jul14, jul15, jul16, jul17, jul18, jul19, jul20, jul21, jul22,
jul23, jul24, jul25)

#Make 4 panel plots
> par(mfrow=c(2,2),oma = c(3,4,1,0) + 0.1,mar = c(0,0,1,1) + 0.1)
#1
> par(mar=c(1.25,1,0,0.5))
> plot(Coal$newdate,Coal$CEMScan_PM25_hourly, type = "l", col="blue", xlab="",
        ylab="", las=1, ylim=c(0,10), xaxt="n")
> axis(1,at=mo,labels=NA)
> lines(Coal$newdate,Coal$SMOKE_PM25_hourly, col="red")
> legend(1136073601,10, c("ORIS-CEM Match","No Match"), cex=0.55, lty=c(1,1),
        lwd=c(1.5,1.5), col=c("blue","red"))
> text(1162339201+324000,10, "Coal Plant - Annual", cex=0.75, font=2)
#2
> par(mar=c(1.25,0.5,0,1))
> plot(Coal$newdate,Coal$CEMScan_PM25_hourly, type = "l", col="blue", xlab="",
        ylab="", ylim=c(0,10), xlim=c(jul12,jul25), yaxt="n", las=1)
> axis(1,at=jul_sec,labels=NA)
> lines(Coal$newdate,Coal$SMOKE_PM25_hourly, col="red")
> text(jul22+43200,10, "Coal Plant - Heat Wave", cex=0.75, family="Arial",font=2)
#3
> par(mar=c(1,1,0.25,0.5))
```

```
> plot(NG$newdate,NG$CEMScan_PM25_hourly, type = "l",col="blue", xlab="",
        ylab="", ylim=c(0,14), xaxt="n",las=1)
> axis(1,at=mo,labels=months)
> lines(NG$newdate,NG$SMOKE_PM25_hourly, col="red")
> text(1162339201+700000,14, "NG Plant - Annual", cex=0.75, family="Arial",font=2)
#4
> par(mar=c(1,0.5,0.25,1))
> plot(NG$newdate,NG$CEMScan_PM25_hourly, type = "l",col="blue", ylab="",
        xlab="",ylim=c(0,14), xlim=c(jul12,jul25), las=1, yaxt="n")
> axis(1,at=jul_sec,labels=jul_days)
> lines(NG$newdate,NG$SMOKE_PM25_hourly, col="red")
> text(jul22+60200,14, "NG Plant - Heat Wave", cex=0.75, font=2)

> title(xlab = "Months in 2006                    Days in July 2006", ylab =
        expression(paste("PM"[2.5], "  Emissions (lbs hr"^-1,")")), cex.lab=1.5, outer =
        TRUE, mgp=c(1.75,1,0), cex.main=2, family="Arial")

#Calculating actual numbers for comparing the data
> July <- c("7/1/06", "7/2/06", "7/3/06", "7/4/06", "7/5/06", "7/6/06", "7/7/06", "7/8/06",
"7/9/06", "7/10/06", "7/11/06", "7/12/06", "7/13/06", "7/14/06", "7/15/06", "7/16/06",
"7/17/06", "7/18/06", "7/19/06", "7/20/06", "7/21/06", "7/22/06", "7/23/06", "7/24/06",
"7/25/06", "7/26/06", "7/27/06", "7/28/06", "7/29/06", "7/30/06", "7/31/06")
```

```
#Coal
> July_Coal <- subset(Coal, Coal$Date%in%July)
> a <- mean(July_Coal$SMOKE_PM25_hourly)
> b <- mean(July_Coal$CEMScan_PM25_hourly)
> c <- mean(July_Coal$SMOKE_SO2_hourly)
> d <- mean(July_Coal$CEMScan_SO2_hourly)

#percent increase of PM2.5 – coal plant
> e <- (b-a)/a *100

#percent increase of SO2 – coal plant
> f <- (d-c)/c*100 #the same as e

#Gas
> July_Gas <- subset(NG, NG$Date%in%July)
> g <- mean(July_Gas$SMOKE_PM25_hourly)
> h <- mean(July_Gas$CEMScan_PM25_hourly)
> i <- mean(July_Gas$SMOKE_SO2_hourly)
> j <- mean(July_Gas$CEMScan_SO2_hourly)

#percent increase of PM2.5 - gas plant
> k <- (h-g)/g *100
```

```
#percent increase of SO2 – gas plant

> l <- (j-i)/i*100 #the same as k


#Find max differences by looking at ratio of CEMScan/SMOKE values

> July_Coal$PM25diff <- July_Coal$CEMScan_PM25_hourly /
        July_Coal$SMOKE_PM25_hourly
> July_Coal$SO2diff <- July_Coal$CEMScan_SO2_hourly /
        July_Coal$SMOKE_SO2_hourly
> July_Gas$PM25diff <- July_Gas$CEMScan_PM25_hourly /
        July_Gas$SMOKE_PM25_hourly
```

```
##############################################################################
# Data Analysis for Chapter 3 of Caroline Farkas Dissertation (2016)
# This script determines which EGUs are peaking units based on the EPA definition (less
than 10% of capacity factor over 3-year average, and less than 20% of capacity factor in
each of those years). This approach was used to find PJM peaking units too by subsetting
the data by state (not shown here but very easy to do using this script)
##############################################################################
#Read in US CEM Capacity Factors
> col_names <- c("State", "FacilityName", "ORIS", "BLR", "Year", "MaxHourlyHIrate")
> col_classes <- c("character", "character","character", "character", rep("numeric",2))
> cap <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/US_CEM_CAPACITY_FACTORS_2002_2014.csv", sep=",", skip=1,
      header=F, col.names=c(col_names), colClasses=c(col_classes),
      na.strings=c("NA","-999", ""))

#Read in CEM data
> col_names1 <- c("State", "FacilityName", "ORIS", "BLR", "Year",
      "HeatInput_MMBTU", "GrossLoad_MWh")
> col_classes1 <- c("character", "character","character","character", rep("numeric",3))
> cem <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/US_CEM_2002_2014.csv", sep=",", skip=1, header=F,
      col.names=c(col_names1), colClasses=c(col_classes1), na.strings=c("NA","-
      999"))
```

```
#calculate maximum annual heat input by multiplying maximum hourly heat input by
number of hours in a year (8,760) and then merge the 2 data frames
> cap$maxHI <- cap$MaxHourlyHIrate*8760
> cem_cap <- merge(cem,cap,by.x=c("State", "FacilityName", "ORIS","BLR","Year"),
        by.y=c("State", "FacilityName","ORIS","BLR","Year"))
#calculate percentage use of max annual heat input
> cem_cap$pUse <- (cem_cap$HeatInput_MMBTU/cem_cap$maxHI)


#subset the years out
> y2002 <- subset(cem_cap, Year==2002)
> y2003 <- subset(cem_cap, Year==2003)
> y2004 <- subset(cem_cap, Year==2004)
> y2005 <- subset(cem_cap, Year==2005)
> y2006 <- subset(cem_cap, Year==2006)
> y2007 <- subset(cem_cap, Year==2007)
> y2008 <- subset(cem_cap, Year==2008)
> y2009 <- subset(cem_cap, Year==2009)
> y2010 <- subset(cem_cap, Year==2010)
> y2011 <- subset(cem_cap, Year==2011)
> y2012 <- subset(cem_cap, Year==2012)
> y2013 <- subset(cem_cap, Year==2013)
> y2014 <- subset(cem_cap, Year==2014)
```

```
# rename pUse for each year so I can keep them straight when I do the averages
> colnames(y2002)[10] <-"pUse2002"
> colnames(y2003)[10] <-"pUse2003"
> colnames(y2004)[10] <-"pUse2004"
> colnames(y2005)[10] <-"pUse2005"
> colnames(y2006)[10] <-"pUse2006"
> colnames(y2007)[10] <-"pUse2007"
> colnames(y2008)[10] <-"pUse2008"
> colnames(y2009)[10] <-"pUse2009"
> colnames(y2010)[10] <-"pUse2010"
> colnames(y2011)[10] <-"pUse2011"
> colnames(y2012)[10] <-"pUse2012"
> colnames(y2013)[10] <-"pUse2013"
> colnames(y2014)[10] <-"pUse2014"

> colnames(y2002)[6] <-"HeatInput2002_MMBTU"
> colnames(y2003)[6] <-"HeatInput2003_MMBTU"
> colnames(y2004)[6] <-"HeatInput2004_MMBTU"
> colnames(y2005)[6] <-"HeatInput2005_MMBTU"
> colnames(y2006)[6] <-"HeatInput2006_MMBTU"
> colnames(y2007)[6] <-"HeatInput2007_MMBTU"
> colnames(y2008)[6] <-"HeatInput2008_MMBTU"
> colnames(y2009)[6] <-"HeatInput2009_MMBTU"
```

```
> colnames(y2010)[6] <-"HeatInput2010_MMBTU"
> colnames(y2011)[6] <-"HeatInput2011_MMBTU"
> colnames(y2012)[6] <-"HeatInput2012_MMBTU"
> colnames(y2013)[6] <-"HeatInput2013_MMBTU"
> colnames(y2014)[6] <-"HeatInput2014_MMBTU"

#merge them into 3-year subsets for the rolling 3-year average (NOTE: do this for all 10
years, but to save space, I did not rewrite them all here)
#2004
> cem_2002_2004 <- merge(y2002[,c("State", "FacilityName", "ORIS","BLR",
        "pUse2002")], y2003[,c("State","FacilityName","ORIS","BLR","pUse2003")],by.
        x=c("State", "FacilityName", "ORIS","BLR"), by.y=c("State", "FacilityName",
        "ORIS", "BLR"))
> cem_2002_2004 <- merge(cem_2002_2004, y2004[,c("State", "FacilityName","ORIS",
        "BLR", "HeatInput2004_MMBTU","pUse2004")])
> cem_2002_2004$averagePUse <- rowMeans(subset(cem_2002_2004, select =
        c(pUse2002, pUse2003, pUse2004), na.rm=TRUE))

> peak_cem2004 <- subset(cem_2002_2004, averagePUse<=0.1 & pUse2002<=0.2 &
        pUse2003<=0.2 & pUse2004<=0.2)

#make data frame
> year <- c(2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014)
```

```
> heat_input <- c(sum(peak_cem2004$HeatInput2004_MMBTU),
        sum(peak_cem2005$HeatInput2005_MMBTU),
        sum(peak_cem2006$HeatInput2006_MMBTU),
        sum(peak_cem2007$HeatInput2007_MMBTU),
        sum(peak_cem2008$HeatInput2008_MMBTU),
        sum(peak_cem2009$HeatInput2009_MMBTU),
        sum(peak_cem2010$HeatInput2010_MMBTU),
        sum(peak_cem2011$HeatInput2011_MMBTU),
        sum(peak_cem2012$HeatInput2012_MMBTU),
        sum(peak_cem2013$HeatInput2013_MMBTU),
        sum(peak_cem2014$HeatInput2014_MMBTU))
> USpeak <- data.frame(year,heat_input)
```

```
##############################################################################
# Data analysis for Chapter 3 of Caroline Farkas Dissertation (2016)
# This script creates the “No PJM” and “Only PJM” emission scenarios by removing
PJM EGUs from CEM and NEI data that is used as inputs for SMOKE/CMAQ for 2006
##############################################################################
#read in PJM CEM data
> col_names <- c("State", "FacilityName", "ORISPL", "UnitID", "Month", "Year",
        "SO2", "AvgNOxRate", "NOx", "CO2", "HeatInput", "GrossLoad", "UnitType",
        "FuelType", "FuelTypeSec", "SO2Control", "NOxControl", "PMControl",
        "other")
> col_classes <- c(rep("character",18))
> PJM <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/July_2006_PJM_CEM.csv", sep=",", skip=1, header=F,
        col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA","-999",
        ""))
> PJM$pair <- paste(PJM$ORISPL,PJM$UnitID,sep=” “)

#read in 2008 NEI (ptipm is 1 file, ptnonipm is 6 files because of its size)
> col_names3 <- c("country_cd", "region_cd", "tribal_code", "facility_id", "unit_id",
        "rel_point_id", "process_id", "agy_facility_id", "agy_unit_id",
        "agy_rel_point_id", "agy_process_id", "SCC", "poll", "ann_value",
        "ann_pct_red", "facility_name", "erptype", "stkhgt", "stkdiam", "stktemp",
        "stkflow", "stkvel", "naics", "longitude", "latitude", "ll_datum",
```

```
        "horiz_coll_mthd", "design_capacity", "design_capacity_units", "reg_codes",
        "fac_source_type", "unit_type_code", "control_ids", "control_measures",
        "current_cost", "cumulative_cost", "projection_factor", "submitter_id",
        "calc_method", "data_set_id", "facil_category_code", "ORIS","BLR", "ipm_yn",
        "calc_year", "date_updated", "fug_height", "fug_width_ydim",
        "fug_length_xdim", "fug_angle", "zipcode", "annual_avg_hours_per_year",
        "jan_value", "feb_value", "mar_value", "apr_value", "may_value", "jun_value",
        "jul_value", "aug_value", "sep_value", "oct_value", "nov_value", "dec_value",
        "jan_pctred", "feb_pctred", "mar_pctred", "apr_pctred", "may_pctred",
        "jun_pctred", "jul_pctred", "aug_pctred", "sep_pctred", "oct_pctred",
        "nov_pctred", "dec_pctred", "comment")
> col_classes3 <- c(rep("character",77))
> NEI <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/2008PTIPMImprovedv2.txt", sep=",", skip=1, header=F,
        col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
        999"))

> NONIPM1 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/2008PTNONIPMImproved1v1.txt", sep=",", skip=1, header=F,
        col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
        999"))
```

```
> NONIPM2 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/2008PTNONIPMImproved2v3.txt", sep=",", skip=1, header=F,
      col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
      999"))

> NONIPM3 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/2008PTNONIPMImproved3v1.txt", sep=",", skip=1, header=F,
      col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
      999"))

> NONIPM4 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/2008PTNONIPMImproved4v2.txt", sep=",", skip=1, header=F,
      col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
      999"))

> NONIPM5 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/2008PTNONIPMImproved5v1.txt", sep=",", skip=1, header=F,
      col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
      999"))
```

```
> NONIPM6 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/2008PTNONIPMImproved6v1.txt", sep=",", skip=1, header=F,
        col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
        999"))

#make SCC codes numeric and pair ORIS and BLR identifiers into 1 column to make
matching and identification easier
> NEI$SCC <- as.numeric(NEI$SCC)
> NEI$pair <- paste(NEI$ORIS,NEI$BLR,sep=" ")

> NONIPM1$SCC <- as.numeric(NONIPM1$SCC)
> NONIPM1$pair <- paste(NONIPM1$ORIS,NONIPM1$BLR,sep=" ")

> NONIPM2$SCC <- as.numeric(NONIPM2$SCC)
> NONIPM2$pair <- paste(NONIPM2$ORIS,NONIPM2$BLR,sep=" ")

> NONIPM3$SCC <- as.numeric(NONIPM3$SCC)
> NONIPM3$pair <- paste(NONIPM3$ORIS,NONIPM3$BLR,sep=" ")

> NONIPM4$SCC <- as.numeric(NONIPM4$SCC)
> NONIPM4$pair <- paste(NONIPM4$ORIS,NONIPM4$BLR,sep=" ")
```

```
> NONIPM5$SCC <- as.numeric(NONIPM5$SCC)

> NONIPM5$pair <- paste(NONIPM5$ORIS,NONIPM5$BLR,sep=" ")


> NONIPM6$SCC <- as.numeric(NONIPM6$SCC)

> NONIPM6$pair <- paste(NONIPM6$ORIS,NONIPM6$BLR,sep=" ")


#read in CEM data for July 2006, make a single file

> col_names6 <- c("ORISID", "BLRID", "Date", "Hour", "NOXmass", "SO2mass",
        "NOXrate", "optime", "gload", "sload", "heatinput", "heatinput_measure",
        "so2measure", "noxmmeasure", "noxrmeasure", "unitflow")

> CEM1<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/HOUR_UNIT_2006_07_1-30jul.txt", sep=",", skip=1, header=F,
        col.names=c(col_names6), colClasses=c(col_classes6), na.strings=c("NA","-
        999"))


> CEM2 <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/HOUR_UNIT_2006_07_31jul.txt", sep=",", skip=1, header=F,
        col.names=c(col_names6), colClasses=c(col_classes6), na.strings=c("NA","-
        999"))


> CEM <- rbind(CEM1,CEM2)

> CEM$pair <- paste(CEM$ORISID,CEM$BLRID,sep=" ")
```

```
# match CEM and PJM ORIS/BLR pairs to make 2 dataframes (1. CEM PJM data and
2.CEM data without PJM units)
> a <- unique(PJM$pair)
> CEM_PJM <- subset(CEM, CEM$pair%in%PJM$pair)
#PJM units not in CEM data
> CEM_PJM_not <- subset(CEM,!(CEM$pair%in%PJM$pair))

# put in correct SMOKE input format and write to csv form
> CEM_PJM$ORISID <- as.numeric(CEM_PJM$ORISID)
> CEM_PJM$BLRID <- as.character(CEM_PJM$BLRID)
> CEM_PJM$Date <- as.character(CEM_PJM$Date)
> CEM_PJM$NOXrate <- as.numeric(CEM_PJM$NOXrate)
> CEM_PJM$heatinput = as.numeric(CEM_PJM$heatinput)
> CEM_PJM$heatinput_measure = as.numeric(CEM_PJM$heatinput_measure)
> CEM_PJM$so2measure = as.numeric(CEM_PJM$so2measure)
> CEM_PJM <- CEM_PJM[,1:16]
> write.csv(CEM_PJM, "/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/July2006_CEM_PJMonly.txt", row.names=F)

# match NEI and PJM ORIS/BLR pairs and make 2 data frames: 1. NEI PJM data and 2.
NEI data without PJM units
> NEI_PJM <- subset(NEI, NEI$pair%in%PJM$pair)
> NEI_PJM_not <- subset(NEI, !(NEI$pair%in%PJM$pair))
```

```
# put NEI ptipm file in correct SMOKE input format and write to csv form
> NEI_PJM$ann_value <- as.numeric(NEI_PJM$ann_value)
> NEI_PJM$ann_pct_red <- as.numeric(NEI_PJM$ann_pct_red)
> NEI_PJM$stkhgt<- as.numeric(NEI_PJM$stkhgt)
> NEI_PJM$stkdiam <- as.numeric(NEI_PJM$stkdiam)
> NEI_PJM$stktemp <- as.numeric(NEI_PJM$stktemp)
> NEI_PJM$stkflow <- as.numeric(NEI_PJM$stkflow)
> NEI_PJM$stkvel <- as.numeric(NEI_PJM$stkvel)
> NEI_PJM$longitude <- as.numeric(NEI_PJM$longitude)
> NEI_PJM$latitude <- as.numeric(NEI_PJM$latitude)
> NEI_PJM$design_capacity <- as.numeric(NEI_PJM$design_capacity)
> NEI_PJM$fac_source_type <- as.numeric(NEI_PJM$fac_source_type)
> NEI_PJM$unit_type_code <- as.numeric(NEI_PJM$unit_type_code)
> NEI_PJM$current_cost <- as.numeric(NEI_PJM$current_cost)
> NEI_PJM$cumulative_cost <- as.numeric(NEI_PJM$cumulative_cost)
> NEI_PJM$projection_factor <- as.numeric(NEI_PJM$projection_factor)
> NEI_PJM$calc_method <- as.numeric(NEI_PJM$calc_method)
> NEI_PJM$data_set_id <- as.numeric(NEI_PJM$data_set_id)
> NEI_PJM$calc_year <- as.numeric(NEI_PJM$calc_year)
> NEI_PJM$date_updated <- as.numeric(NEI_PJM$date_updated)
> NEI_PJM$fug_height <- as.numeric(NEI_PJM$fug_height)
> NEI_PJM$fug_width_ydim <- as.numeric(NEI_PJM$fug_width_ydim)
> NEI_PJM$fug_length_xdim <- as.numeric(NEI_PJM$fug_length_xdim)
```

```
> NEI_PJM$fug_angle <- as.numeric(NEI_PJM$fug_angle)
> NEI_PJM$annual_avg_hours_per_year <- as.numeric(NEI_PJM$annual_avg_hours_
      per_year)
> NEI_PJM$jan_value <- as.numeric(NEI_PJM$jan_value)
> NEI_PJM$feb_value <- as.numeric(NEI_PJM$feb_value)
> NEI_PJM$mar_value <- as.numeric(NEI_PJM$mar_value)
> NEI_PJM$apr_value <- as.numeric(NEI_PJM$apr_value)
> NEI_PJM$may_value <- as.numeric(NEI_PJM$may_value)
> NEI_PJM$jun_value <- as.numeric(NEI_PJM$jun_value)
> NEI_PJM$jul_value <- as.numeric(NEI_PJM$jul_value)
> NEI_PJM$aug_value <- as.numeric(NEI_PJM$aug_value)
> NEI_PJM$sep_value <- as.numeric(NEI_PJM$sep_value)
> NEI_PJM$oct_value <- as.numeric(NEI_PJM$oct_value)
> NEI_PJM$nov_value <- as.numeric(NEI_PJM$nov_value)
> NEI_PJM$dec_value <- as.numeric(NEI_PJM$dec_value)
> NEI_PJM$jan_pctred <- as.numeric(NEI_PJM$jan_pctred)
> NEI_PJM$feb_pctred <- as.numeric(NEI_PJM$feb_pctred)
> NEI_PJM$mar_pctred <- as.numeric(NEI_PJM$mar_pctred)
> NEI_PJM$apr_pctred <- as.numeric(NEI_PJM$apr_pctred)
> NEI_PJM$may_pctred <- as.numeric(NEI_PJM$may_pctred)
> NEI_PJM$jun_pctred <- as.numeric(NEI_PJM$jun_pctred)
> NEI_PJM$jul_pctred <- as.numeric(NEI_PJM$jul_pctred)
> NEI_PJM$aug_pctred <- as.numeric(NEI_PJM$aug_pctred)
```

```
> NEI_PJM$sep_pctred <- as.numeric(NEI_PJM$sep_pctred)
> NEI_PJM$oct_pctred <- as.numeric(NEI_PJM$oct_pctred)
> NEI_PJM$nov_pctred <- as.numeric(NEI_PJM$nov_pctred)
> NEI_PJM$dec_pctred <- as.numeric(NEI_PJM$dec_pctred)
> NEI_PJM <- NEI_PJM[,1:77]
>  write.csv(NEI_PJM, "/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/2008PTIPM_PJMonly.csv", row.names=F)

# put NEI ptnonipm files (6 of them) in correct SMOKE input format and write to csv
form. (NOTE: Only the first of 6 is pasted here because the rest is repetition and many
pages long)
> NONIPM1_PJM <- subset(NONIPM1, NONIPM1$pair%in%PJM$pair)
> NONIPM1_PJM_not <- subset(NONIPM1, !(NONIPM1$pair%in%PJM$pair))

> NONIPM1_PJM$ann_value <- as.numeric(NONIPM1_PJM$ann_value)
> NONIPM1_PJM$ann_pct_red <- as.numeric(NONIPM1_PJM$ann_pct_red)
> NONIPM1_PJM$stkhgt<- as.numeric(NONIPM1_PJM$stkhgt)
> NONIPM1_PJM$stkdiam <- as.numeric(NONIPM1_PJM$stkdiam)
> NONIPM1_PJM$stktemp <- as.numeric(NONIPM1_PJM$stktemp)
> NONIPM1_PJM$stkflow <- as.numeric(NONIPM1_PJM$stkflow)
> NONIPM1_PJM$stkvel <- as.numeric(NONIPM1_PJM$stkvel)
> NONIPM1_PJM$longitude <- as.numeric(NONIPM1_PJM$longitude)
> NONIPM1_PJM$latitude <- as.numeric(NONIPM1_PJM$latitude)
```

```
> NONIPM1_PJM$design_capacity <- as.numeric(NONIPM1_PJM$design_capacity)
> NONIPM1_PJM$fac_source_type <- as.numeric(NONIPM1_PJM$fac_source_type)
> NONIPM1_PJM$unit_type_code <- as.numeric(NONIPM1_PJM$unit_type_code)
> NONIPM1_PJM$current_cost <- as.numeric(NONIPM1_PJM$current_cost)
> NONIPM1_PJM$cumulative_cost <- as.numeric(NONIPM1_PJM$cumulative_cost)
> NONIPM1_PJM$projection_factor <- as.numeric(NONIPM1_PJM$projection_factor)
> NONIPM1_PJM$calc_method <- as.numeric(NONIPM1_PJM$calc_method)
> NONIPM1_PJM$data_set_id <- as.numeric(NONIPM1_PJM$data_set_id)
> NONIPM1_PJM$calc_year <- as.numeric(NONIPM1_PJM$calc_year)
> NONIPM1_PJM$date_updated <- as.numeric(NONIPM1_PJM$date_updated)
> NONIPM1_PJM$fug_height <- as.numeric(NONIPM1_PJM$fug_height)
> NONIPM1_PJM$fug_width_ydim <- as.numeric(NONIPM1_PJM$fug_width_ydim)
> NONIPM1_PJM$fug_length_xdim <- as.numeric(NONIPM1_PJM$fug_length_xdim)
> NONIPM1_PJM$fug_angle <- as.numeric(NONIPM1_PJM$fug_angle)
> NONIPM1_PJM$annual_avg_hours_per_year <-as.numeric(NONIPM1_PJM$annual_
      avg_hours_per_year)
> NONIPM1_PJM$jan_value <- as.numeric(NONIPM1_PJM$jan_value)
> NONIPM1_PJM$feb_value <- as.numeric(NONIPM1_PJM$feb_value)
> NONIPM1_PJM$mar_value <- as.numeric(NONIPM1_PJM$mar_value)
> NONIPM1_PJM$apr_value <- as.numeric(NONIPM1_PJM$apr_value)
> NONIPM1_PJM$may_value <- as.numeric(NONIPM1_PJM$may_value)
> NONIPM1_PJM$jun_value <- as.numeric(NONIPM1_PJM$jun_value)
> NONIPM1_PJM$jul_value <- as.numeric(NONIPM1_PJM$jul_value)
```

```
> NONIPM1_PJM$aug_value <- as.numeric(NONIPM1_PJM$aug_value)
> NONIPM1_PJM$sep_value <- as.numeric(NONIPM1_PJM$sep_value)
> NONIPM1_PJM$oct_value <- as.numeric(NONIPM1_PJM$oct_value)
> NONIPM1_PJM$nov_value <- as.numeric(NONIPM1_PJM$nov_value)
> NONIPM1_PJM$dec_value <- as.numeric(NONIPM1_PJM$dec_value)
> NONIPM1_PJM$jan_pctred <- as.numeric(NONIPM1_PJM$jan_pctred)
> NONIPM1_PJM$feb_pctred <- as.numeric(NONIPM1_PJM$feb_pctred)
> NONIPM1_PJM$mar_pctred <- as.numeric(NONIPM1_PJM$mar_pctred)
> NONIPM1_PJM$apr_pctred <- as.numeric(NONIPM1_PJM$apr_pctred)
> NONIPM1_PJM$may_pctred <- as.numeric(NONIPM1_PJM$may_pctred)
> NONIPM1_PJM$jun_pctred <- as.numeric(NONIPM1_PJM$jun_pctred)
> NONIPM1_PJM$jul_pctred <- as.numeric(NONIPM1_PJM$jul_pctred)
> NONIPM1_PJM$aug_pctred <- as.numeric(NONIPM1_PJM$aug_pctred)
> NONIPM1_PJM$sep_pctred <- as.numeric(NONIPM1_PJM$sep_pctred)
> NONIPM1_PJM$oct_pctred <- as.numeric(NONIPM1_PJM$oct_pctred)
> NONIPM1_PJM$nov_pctred <- as.numeric(NONIPM1_PJM$nov_pctred)
> NONIPM1_PJM$dec_pctred <- as.numeric(NONIPM1_PJM$dec_pctred)
> NONIPM1_PJM <- NONIPM1_PJM[,1:77]
> write.csv(NONIPM1_PJM, "/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/2008NONIPM1_PJMonly.csv", row.names=F)
```

```
##############################################################################
# Figure 3-1a of Caroline Farkas Dissertation (2016)
# Location and Fuel Type of PJM Peaking Unit Map
##############################################################################
> library(maps)
#read in peaking units (544 lines)
> col_names2 <- c("State","Plant_name","ORIS","BLR","Year","Lat","Lon","pair")
> col_classes2 <- c(rep("character",4), rep("numeric",3),"character")
> peakers<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/2008_NEI_Peakers_Removed_Lat_Long.csv", sep=",", skip=1, header=F,
      col.names=c(col_names2), colClasses=c(col_classes2), na.strings=c("NA","-
      999"))
#fuel types
> col_names3 <- c("State", "FacilityName", "ORISID", "BLRID", "Year", "UnitType",
      "FuelType_P", "FuelType_S", "blank")
> col_classes3 <- c(rep("character",9))
> fuels<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
      data/PJMRegion2006FuelTypes.csv",sep=",", skip=1, header=F,
      col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-
      999"))
#make ORIS/BLR ID unique pair as one column so it is easier to match
>fuels$pair <- paste(fuels$ORISID,fuels$BLRID,sep=" ")
```

```
#add fueltype to peakers dataframe
> peakers$fuel <- fuels$FuelType_P[match(peakers$pair,fuels$pair)]

#coal
> peakcoal <- subset(peakers,fuel==”Coal”)
#gas
> peakgas <- subset(peakers,fuel=="Pipeline Natural Gas" | fuel=="Other Gas" |
        fuel=="Natural Gas" | fuel=="Other Gas, Pipeline Natural Gas")
#oil
> peakoil <- subset(peakers, fuel=="Other Oil" | fuel=="Diesel Oil" | fuel=="Residual
        Oil")

# make map
> map(database="state",regions=c("Virginia","New Jersey","Pennsylvania","West
        Virginia", "Delaware", "Maryland", "DC", "Ohio", "Indiana", "Maine", "New
        Hampshire", "Vermont", "Connecticut" ,"Rhode Island", "Massachusetts",
        "Michigan", "Kentucky", "New York", "Illinois", "Wisconsin"), mar=c(0,3,0,0))
> points(peakcoal$Lon,peakcoal$Lat,cex=1.5,pch=17,col="black")
> points(peakgas$Lon,peakgas$Lat,cex=1.5,pch=20,col="green3")
> points(peakoil$Lon,peakoil$Lat,cex=1.5,pch=0,col="red")
> legend(-91.5,45.8, c("Coal", "Oil", "Gas"), pch=c(17,0,20), col=c("black", "red",
        "green3"), cex=1.5)
> title(main="Location and Fuel Type of PJM Peaking Unit EGUs")
```

```
##############################################################################
# Figure 3-1b of Caroline Farkas Dissertation (2016)
# Location and Fuel Type of All PJM EGUs Map
##############################################################################
#Read in CEM_PJMonly data
> col_names2 <- c("ORISID", "BLRID", "Date", "Hour", "NOXmass", "SO2mass",
        "NOXrate", "optime", "gload", "sload", "heatinput", "heatinput_measure",
        "so2measure", "noxmmeasure", "noxrmeasure", "unitflow")
> col_classes2 <- c(rep("character",16))
> PJMonly <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/July2006_CEM_PJMonly.txt",sep=",", skip=1, header=F,
        col.names=c(col_names2), colClasses=c(col_classes2), na.strings=c("NA","-999",
        ""))
> PJMonly$pair <- paste(PJMonly$ORISID,PJMonly$BLRID,sep=" ")


#read in CAMD data for lat/lon and primary fuel source
> col_names9 <- c("State", "FacilityName", "ORISPL", "UnitID", "AssociatedStacks",
        "Month", "Year", "Programs", "SO2", "AvgNOxRate", "NOx", "CO2",
        "HeatInput", "FuelType", "Lat", "Lon","other")
> col_classes9 <- c(rep("character",16))
```

```
> CAMD <-read.csv(file="/Users/Carna/Downloads/EPADownload/emission_12-11-
        2015.csv", sep=",", skip=1, header=F, col.names=c(col_names9),
        colClasses=c(col_classes9), na.strings=c("NA","-999", ""))

#make ORIS/BLR pairs in one column to match with other files
> CAMD$pair <- paste(CAMD$ORISPL,CAMD$UnitID, sep=" ")

# merge and match data
> PJMonly2 <- merge(PJMonly,CAMD, by.x="pair",by.y="pair")
# remove duplicates
> PJMonly3 <- PJMonly2[!duplicated(PJMonly2$pair),]

#separate into fuel types for mapping
> PJM_coal <- subset(PJMonly3,FuelType=="Coal" | FuelType=="Coal Refuse")
> PJM_oil <- subset(PJMonly3, FuelType=="Diesel Oil" | FuelType=="Other Oil" |
        FuelType=="Residual Oil" | FuelType=="Petroleum Coke" | FuelType=="Other
        Oil, Tire Derived Fuel")
> PJM_gas <- subset(PJMonly3, FuelType=="Pipeline Natural Gas" | FuelType=="Other
        Gas" | FuelType=="Process Gas" | FuelType=="Natural Gas" |
        FuelType=="Natural Gas, Pipeline Natural Gas" | FuelType=="Other Gas,
        Pipeline Natural Gas")
```

```
> map(database="state",regions=c("Virginia","New Jersey","Pennsylvania","West
      Virginia", "Delaware", "Maryland", "DC", "Ohio", "Indiana", "Maine", "New
      Hampshire", "Vermont", "Connecticut", "Rhode Island", "Massachusetts",
      "Michigan", "Kentucky", "New York", "Illinois", "Wisconsin"), mar=c(0,3,0,0))
> points(PJM_coal$Lon,PJM_coal$Lat,cex=1.5,pch=17,col="black")
> points(PJM_gas$Lon,PJM_gas$Lat,cex=1.5,pch=20,col="green3")
> points(PJM_oil$Lon,PJM_oil$Lat,cex=1.5,pch=0,col="red")
> legend(-91.5,45.8, c("Coal", "Oil", "Gas"), pch=c(17,0,20), col=c("black", "red",
      "green3"), cex=1.25)
> title(main="Location and Fuel Type of PJM EGUs")
```

```
##############################################################################
# Figure 3-1c of Caroline Farkas Dissertation (2016)
# Location and Fuel Type of Other RTO EGUs Map
##############################################################################
#Read in CEM_PJMonly data
> col_names2 <- c("ORISID", "BLRID", "Date", "Hour", "NOXmass", "SO2mass",
        "NOXrate", "optime", "gload", "sload", "heatinput", "heatinput_measure",
        "so2measure", "noxmmeasure", "noxrmeasure", "unitflow")
> col_classes2 <- c(rep("character",16))
> noPJM <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/July2006_CEM_noPJM.txt", sep=",", skip=1, header=F,
        col.names=c(col_names2), colClasses=c(col_classes2), na.strings=c("NA","-999",
        ""))
#make ORIS/BLR pairs in one column for matching
> noPJM$pair <- paste(noPJM$ORISID,noPJM$BLRID,sep=" ")

#merge and match data with CAMD
> noPJM2 <- merge(noPJM,CAMD, by.x="pair",by.y="pair")
#remove duplicates and NC and TN
> noPJM3 <- noPJM2[!duplicated(noPJM2$pair),]
> noPJM3 <- subset(noPJM3, noPJM3$State!="NC")
> noPJM3<- subset(noPJM3, noPJM3$State!="TN")
```

```
#separate fuel types
> nonPJM_coal <- subset(noPJM3,FuelType=="Coal" | FuelType=="Other Solid Fuel" |
        FuelType=="Coal, Wood")
> nonPJM_oil <- subset(noPJM3, FuelType=="Diesel Oil" | FuelType=="Other Oil" |
        FuelType=="Residual Oil" | FuelType=="Petroleum Coke" |
        FuelType=="Residual Oil, Pipeline Natural Gas" | FuelType=="Diesel Oil, Other
        Oil")
> nonPJM_gas <- subset(noPJM3, FuelType=="Pipeline Natural Gas" |
        FuelType=="Other Gas" | FuelType=="Natural Gas")

#make map
> map(database="state",regions=c("Virginia", "New Jersey", "Pennsylvania", "West
        Virginia", "Delaware", "Maryland", "DC", "Ohio", "Indiana", "Maine", "New
        Hampshire", "Vermont", "Connecticut", "Rhode Island", "Massachusetts",
        "Michigan", "Kentucky", "New York", "Illinois", "Wisconsin"), mar=c(0,3,0,0))
> points(nonPJM_coal$Lon,nonPJM_coal$Lat,cex=1.5,pch=17,col="black")
> points(nonPJM_gas$Lon,nonPJM_gas$Lat,cex=1.5,pch=20,col="green3")
> points(nonPJM_oil$Lon,nonPJM_oil$Lat,cex=1.5,pch=0,col="red")
> legend(-91.5,39.8, c("Coal", "Oil", "Gas"), pch=c(17,0,20), col=c("black", "red",
        "green3"), cex=1.25)
> title(main="Location and Fuel Type of Other RTO EGUs")
```

```
##############################################################################
# Figure 3-1d of Caroline Farkas Dissertation (2016)
# Total Population by County Map
##############################################################################
> library(acs)
> library(sqldf)
> library(ggplot2)
> library(maps)
> library(choroplethr)
> library(choroplethrMaps)

#sign into census API (request key here: http://api.census.gov/data/key_signup.html)
> api.key.install(key="key_goes_here")

# load the boundary data for all counties in domain
> county.df=map_data("county",region=c("Maine", "Vermont", "New Hampshire",
        "Rhode Island", "Massachusetts", "Connecticut", "New York", "New Jersey",
        "Delaware", "Pennsylvania", "Virginia", "West Virginia", "Ohio", "Kentucky",
        "Maryland", "Indiana", "Illinois", "Michigan", "Wisconsin", "DC"))

# rename fields for later merge
> names(county.df)[5:6]=c("state","county")
#load boundary data for all states in domain
```

```
> state.df=map_data("state",region=c("Maine", "Vermont", "New Hampshire", "Rhode
        Island", "Massachusetts", "Connecticut", "New York", "New Jersey", "Delaware",
        "Pennsylvania", "Virginia", "West Virginia", "Ohio", "Kentucky", "Maryland",
        "Indiana", "Illinois", "Michigan", "Wisconsin", "DC"))

#fetch data, first get table number
acs.lookup(endyear = 2009, span = 5, dataset = "acs", table.name = "population",
case.sensitive=F )

> census_pop <- acs.fetch(geography=counties, endyear=2010, span=5,
        table.number="B01003", col.names="pretty")

#make into data frame for easier handling
> census_pop.dat <- data.frame(county=geography(census_pop)[[1]],
        pop=as.numeric(estimate(census_pop)))

#clean up county names and find the states
> census_pop.dat$state=tolower(gsub("^.*County, ", "", census_pop.dat$county))
> census_pop.dat$county=tolower(gsub(" County,.*", "", census_pop.dat$county))

#fix spelling differences that make counties not match (blank data)
> a <- which(census_pop.dat$county=="st. lawrence")
> census_pop.dat$county[a] <- "st lawrence"
```

```
> b <- which(census_pop.dat$county=="st. croix")
> census_pop.dat$county[b] <- "st croix"
> c <- which(census_pop.dat$county=="st. clair") #illinois & michigan
> census_pop.dat$county[c] <- "st clair"
> d <- which(census_pop.dat$county=="st. joseph") #indiana & michigan
> census_pop.dat$county[d] <- "st joseph"
> e <- which(census_pop.dat$county=="laporte")
> census_pop.dat$county[e] <- "la porte"
> f <- which(census_pop.dat$county=="dekalb") #indiana & illinois
> census_pop.dat$county[f] <- "de kalb"
> g <- which(census_pop.dat$county=="lasalle")
> census_pop.dat$county[g] <- "la salle"
> h <- which(census_pop.dat$county=="dupage")
> census_pop.dat$county[h] <- "du page"
> i <- which(census_pop.dat$county=="prince george's")
> census_pop.dat$county[i] <- "prince georges"
> j <- which(census_pop.dat$county=="queen anne's")
> census_pop.dat$county[j] <- "queen annes"
> k <- which(census_pop.dat$county=="st. mary's")
> census_pop.dat$county[k] <- "st marys"

#merge county boundaries with the new data frame
> choropleth=merge(county.df, census_pop.dat, by=c("state","county"), all.x=TRUE)
```

```
> choropleth=choropleth[order(choropleth$order), ]

#make cuts in the data for the map scale
> choropleth$pop_cuts=cut(choropleth$pop, breaks=c(0, 5000, 10000, 50000, 100000,
        250000, 500000, 750000, 1000000, 2500000, 5172848), include.lowest=T)

#map
> ggplot(choropleth, aes(long, lat, group = group)) + geom_polygon(aes(fill = pop_cuts),
        colour = "black", size = 0.1) + geom_polygon(data = state.df, colour = "black",
        fill = NA) + scale_fill_manual(values = c("#FFFFFF", "#F2F0F6", "#C4BCD3",
        "#A99DBE", "#9282AC", "#7F6B9D", "#6D5690", "#5F4186", "#522A7E",
        "#4A007E"), na.value="pink") + theme(axis.line=element_blank(),
        axis.text.x=element_blank(), axis.text.y=element_blank(),
        axis.ticks=element_blank(), axis.title.x=element_blank(),
        axis.title.y=element_blank(), panel.background=element_blank(),
        panel.border=element_blank(), panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(), plot.background=element_blank())+
        ggtitle("Total Population by County") + theme(plot.title = element_text(size=20,
        face="bold", vjust=0.5))
```

```
##############################################################################
# Figure 3-2 of Caroline Farkas Dissertation (2016)
# PJM peaking unit heat input compared to PJM average summer temperature and U.S.
annual Gross Domestic Product from 2004-2014
##############################################################################
#read in data
> col_names <- c("Year", "MMBTU", "AvgTemp", "oilPrice", "gasPrice", "coalPrice",
        "Ap_Sept_AvgTemp", "JJA_AvgTemp", "JJA_MaxTemp", "GDP_growthRate")
> col_classes <- c("numeric", "numeric","numeric",rep("character",3), "numeric",
        "numeric", "numeric", "numeric")
> peak <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/PJM_peaking_heat_inputs.csv", sep=",", skip=1, header=F,
        col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA","-999"))

#plot
> par(mar=c(4,4.5,2,8.25))
> plot(peak$Year,peak$MMBTU/1000000, type="l", col="burlywood4", ylab=NA,
        xlab=NA, las=1, lwd=5, lty=1,cex.axis=1.25)
> mtext(side=1,expression(bold("Year")),line=2.25,cex=1.25)
> mtext(side=2, expression(bold("Heat Input (Millions of MMBtu)")), line=2.75,
        col="burlywood4", cex=1.25)
> par(new=T)
```

```
> plot(peak$Year,peak$JJA_AvgTemp, type="l", axes=F, xlab=NA, ylab=NA,
        col="hotpink", lwd=5, lty=2)
> axis(side=4, las=1,cex.axis=1.25)
> mtext(side=4,line=3.5, expression(paste(bold("Temperature ("), bold(degree),
        bold("F")))), col="hotpink",cex=1.25)
> par(new=T)
> plot(peak$Year, peak$GDP_growthRate, type="l", axes=F, xlab=NA, ylab=NA,
        col="darkolivegreen", lwd=5, lty=4)
> axis(side=4,las=1,line=5,cex.axis=1.25)
> mtext(side=4,line=7.35,expression(bold("GDP Growth Rate (annual %)")),
        col="darkolivegreen",cex=1.25)
> legend("bottomright", c("Average Heat Input - PJM Peaking Units","PJM Average
        Summer Temperature", "Annual U.S. GDP Growth Rate"), cex=1, lty=c(1,2,4),
        lwd=c(3,3, 3), col=c("burlywood4","hotpink", "darkolivegreen"))
```

```
##############################################################################
# Figure A-1 of Caroline Farkas Dissertation (2016)
# 4 panel average monthly stagnation days by season for Delaware, Maryland, New
Jersey, Pennsylvania, and Virginia
##############################################################################
> library(sp)
> library(maps)
> library(maptools)
> gpclibPermit()
> library(plyr)
> library(lubridate)

#reading in ~500 files using for loop
> filenames <- list.files(path = "/Users/Carna/Dropbox/Rutgers/Research/Stagnation_
        days/data", pattern = NULL, all.files = FALSE, full.names = TRUE, recursive =
        FALSE, ignore.case = FALSE)

> stag=NULL
> for (i in 1:length(filenames)){
        stag2<-read.table(filenames[i], header=T, quote="\"", skip=0,
                col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA",
                "", "CC"))
        stag2$Date <- paste(substr(filenames[i],60,65))
```

```
        stag=rbind(stag,stag2)
        }
> write.csv(stag,"/Users/Carna/Dropbox/Rutgers/Research/Stagnation_
        days/all_years.csv",row.names=F)

#creating function to return state from lat and lon
> latlong2state <- function(stag) {
        # Prepare SpatialPolygons object with one SpatialPolygon per state (plus DC,
        minus HI & AK)
        > states <- map('state', fill=TRUE, col="transparent", plot=FALSE)
        > IDs <- sapply(strsplit(states$names, ":"), function(x) x[1])
        > states_sp <- map2SpatialPolygons(states, IDs=IDs,
                        proj4string=CRS("+proj=longlat +datum=wgs84"))

        # Convert pointsDF to a SpatialPoints object
        > stagSP <- SpatialPoints(stag, proj4string=CRS("+proj=longlat
                +datum=wgs84"))
        # Use 'over' to get _indices_ of the Polygons object containing each point
        > indices <- over(stagSP, states_sp)
        # Return the state names of the Polygons object containing each point
        > stateNames <- sapply(states_sp@polygons, function(x) x@ID)
                stateNames[indices]
        }
```

```
# Test the function using points in Wisconsin and Oregon.
> testPoints <- data.frame(x = c(-90, -120), y = c(44, 44))
> testPoints <- data.frame(x = c(-74.44, -80.08), y = c(40.48, 36.11))

> latlong2state(testPoints)

#make lat and lon subset
> stag_latlon <- subset(stag, select=c(Lon,Lat), row.names=F)
> stag_states<- latlong2state(stag_latlon)

#put into data frame
> stag$state <- stag_states
> stag_5states <- subset(stag, stag$state == "new jersey" | stag$state == "maryland"|
        stag$state == "delaware" | stag$state == "virginia" | stag$state == "pennsylvania")

# subset into seasons, average by month, per year, make a four panel plot
> stag_5states$Month <- paste(substr(stag_5states$Date,5,6))

#plot
> par(mfrow=c(2,2),oma = c(3,4,1,0) + 0.1,mar = c(0,0,1,1) + 0.1)
#december-january-february
> djf <- subset(stag_5states, stag_5states$Month == "12" | stag_5states$Month == "01" |
        stag_5states$Month == "02")
```

```
> djf_avg <- aggregate(Stagnation_days~Date, djf, mean)
> djf_avg$Date2 <- ymd(as.character(paste(djf_avg$Date,"01",sep="")))
> plot(djf_avg$Date2,djf_avg$Stagnation_days, col="blue", ylim=c(0,15), xaxt="n",
      labels=FALSE)
> axis(1,labels=FALSE)
> abline(lm(djf_avg$Stagnation_days~djf_avg$Date2), col="red", lty=2, lwd=2)
> mtext(13, 3000000,"December-January-February")

#march-april-may
> mam <- subset(stag_5states, stag_5states$Month == "03" | stag_5states$Month == "04"
      | stag_5states$Month == "05")
> mam_avg <- aggregate(Stagnation_days~Date, mam, mean)
> mam_avg$Date2 <- ymd(as.character(paste(mam_avg$Date,"01",sep="")))
> plot(mam_avg$Date2,mam_avg$Stagnation_days, col="blue", ylim=c(0,15),
      labels=FALSE)
> abline(lm(mam_avg$Stagnation_days~mam_avg$Date2), col="red", lty=2, lwd=2)

#june-july-august
> jja <- subset(stag_5states, stag_5states$Month == "06" | stag_5states$Month == "07" |
      stag_5states$Month == "08")
> jja_avg <- aggregate(Stagnation_days~Date, jja, mean)
> jja_avg$Date2 <- ymd(as.character(paste(jja_avg$Date,"01",sep="")))
```

```
> plot(jja_avg$Date2,jja_avg$Stagnation_days, col="blue", ylim=c(0,15),
        labels=FALSE)
> abline(lm(jja_avg$Stagnation_days~jja_avg$Date2), col="red", lty=2, lwd=2)


#september-october-november
> son <- subset(stag_5states, stag_5states$Month == "09" | stag_5states$Month == "10" |
        stag_5states$Month == "11")
> son_avg <- aggregate(Stagnation_days~Date, son, mean)
> son_avg$Date2 <- ymd(as.character(paste(son_avg$Date,"01",sep="")))
> plot(son_avg$Date2,son_avg$Stagnation_days, col="blue", ylim=c(0,15), yaxt="n")
> axis(2,labels=FALSE)
> abline(lm(son_avg$Stagnation_days~son_avg$Date2), col="red", lty=2, lwd=2)
```

```
##############################################################################
# Figure B-2 of Caroline Farkas Dissertation (2016)
# Plot of daily fraction of PJM grossload (generation) that is Peaking Units
##############################################################################
> library(lubridate)
> library(plyr)

#read in onlyPJM CEM data
> col_names3 <- c("ORISID", "BLRID", "Date", "Hour", "NOXmass", "SO2mass",
        "NOXrate", "optime", "gload", "sload", "heatinput", "heatinput_measure",
        "so2measure", "noxmmeasure", "noxrmeasure", "unitflow")
> col_classes3 <- c("numeric","character","character", rep("numeric",13))
> CEM <- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/July2006_CEM_PJMonly.txt", sep=",", skip=1, header=F,
        col.names=c(col_names3), colClasses=c(col_classes3), na.strings=c("NA","-999",
        "-9"))
> CEM$Date <- ymd(CEM$Date)
> CEM$Hour <- paste(CEM$Hour,":00", sep="")
> CEM$DateTime <- paste(CEM$Date,CEM$Hour)
> CEM$DateTime <- ymd_hm(CEM$DateTime)

#pair ORIS and Boiler Ids with space in the middle to make a joint unique identifier
> CEM$pair <- paste(CEM$ORISID,CEM$BLRID,sep=" ")
```

```
#make hourly sum of gload for CEM (all PJM)
> CEM_hSum <- ddply(CEM, .(DateTime), summarise, hSum_gload = sum(gload,
        na.rm=TRUE))

#read in 544 peaking units (chosen based on EPA peaking unit requirements)
#these already have the ORIS and Boiler IDs paired
> col_names2 <- c("State","Plant_name","ORIS","BLR","Year","Lat","Lon","pair")
> col_classes2 <- c(rep("character",4), rep("numeric",3),"character")
> peakers<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/2008_NEI_Peakers_Removed_Lat_Long.csv", sep=",", skip=1, header=F,
        col.names=c(col_names2), colClasses=c(col_classes2), na.strings=c("NA","-
        999"))

#subset peaking unit data using peakers list
> peaking_CEM <- CEM[CEM$pair%in%peakers$pair,]

#make hourly averages of gload for Peaking Units (PJM)
> peaking_CEM_hSum <- ddply(peaking_CEM, .(DateTime), summarise,
        peak_hSum_gload = sum(gload, na.rm=TRUE))

#merge datasets
> hSum <- merge(CEM_hSum, peaking_CEM_hSum, by. x="DateTime",
        by.y="DateTime")
```

```
#calculate fraction of total gross load that is peaking units
> hSum$peak_frac <- hSum$peak_hSum_gload / hSum$hSum_gload


#plot
> par(mar=c(4,5,3.5,1))
> plot(hSum$DateTime,hSum$peak_frac, type= "l",col="blue",
        xlab=expression(bold("Days")), ylab=NA, las=1, main="Fraction of Hourly PJM
        Gross Load that is from Peaking Units", cex.lab=1.25, cex.main=1.25,
        cex.axis=1.25)
> mtext(side=2,expression(bold("Fraction of Total Gross Load")),cex=1.25,line=3.5)
```

```
##############################################################################
# Figure B-3a of Caroline Farkas Dissertation (2016)
# U.S. peaking unit heat input compared to U.S. average summer temperature and U.S.
annual Gross Domestic Product from 2004-2014
##############################################################################
# read in peaking unit data
> col_names <- c("Year", "MMBTU", "AvgTemp", "oilPrice", "gasPrice",
        "coalPrice","Ap_Sept_AvgTemp","JJA_AvgTemp","JJA_MaxTemp","GDP_gro
        wthRate")
> col_classes <- c("numeric", "numeric","numeric",rep("character",3), "numeric",
        "numeric", "numeric", "numeric")
> peak <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/PJM_peaking_heat_inputs.csv", sep=",", skip=1, header=F,
        col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA","-999"))
#US JJA average temp (from NCDC)
> peak$US_JJA_temp <- c(70.38, 72.57, 73.52, 73.07, 71.99, 71.02, 73.08, 73.65, 73.70,
        72.53, 71.71)

#add US_MMBTU to data frame (from Data Analysis for Chapter 3 to determine U.S.
peaking units)
> peak$US_MMBTU <- c(293076865, 635089063, 487538037, 608736607, 436649216,
        405063587, 626581211, 562045024, 699362633, 474566612, 522125170)
```

```
#plot U.S. JJA avg. temp vs MMBTU WITH GDP growth rate
> par(mar=c(4,4.5,2,8.25))
> plot(peak$Year,peak$US_MMBTU/1000000, type="l",col="burlywood4", ylab=NA,
      xlab=NA, las=1, lwd=5, lty=1,cex.axis=1.25)
> mtext(side=1,expression(bold("Year")), line=2.25,cex=1.25)
> mtext(side=2, expression(bold("Heat Input (Millions of MMBtu)")), line=2.75,
      col="burlywood4",cex=1.25)
> par(new=T)
> plot(peak$Year,peak$US_JJA_temp, type="l", axes=F, xlab=NA, ylab=NA,
      col="hotpink", lwd=5, lty=2)
> axis(side=4, las=1,cex.axis=1.25)
> mtext(side=4,line=3.5, expression(paste(bold("Temperature ("), bold(degree),
      bold("F")))), col="hotpink",cex=1.25)
> par(new=T)
> plot(peak$Year, peak$GDP_growthRate, type="l", axes=F, xlab=NA, ylab=NA,
      col="darkolivegreen", lwd=5, lty=4)
> axis(side=4,las=1,line=5,cex.axis=1.25)
> mtext(side=4,line=7.35,expression(bold("GDP Growth Rate (annual %)")),
      col="darkolivegreen",cex=1.25)
> legend("bottomright", c("Average Heat Input - U.S. Peaking Units","U.S. Average
      Summer Temperature", "Annual U.S. GDP Growth Rate"), cex=1, lty=c(1,2,4),
      lwd=c(3,3, 3), col=c("burlywood4","hotpink", "darkolivegreen"))
```

```
##############################################################################
# Figure B-3b of Caroline Farkas Dissertation (2016)
# PJM peaking unit annual heat input compared to U.S. annual average oil, gas, and coal price
##############################################################################
# read in peaking unit data
> col_names <- c("Year", "MMBTU", "AvgTemp", "oilPrice", "gasPrice",
        "coalPrice","Ap_Sept_AvgTemp","JJA_AvgTemp","JJA_MaxTemp","GDP_gro
        wthRate")
> col_classes <- c("numeric", "numeric","numeric",rep("character",3), "numeric",
        "numeric", "numeric", "numeric")
> peak <-read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/PJM_peaking_heat_inputs.csv", sep=",", skip=1, header=F,
        col.names=c(col_names), colClasses=c(col_classes), na.strings=c("NA","-999"))

# make fuel prices numeric values
> peak$oilPrice <- as.numeric(sub('\\$',",as.character(peak$oilPrice)))
> peak$gasPrice <- as.numeric(sub('\\$',",as.character(peak$gasPrice)))
> peak$coalPrice <- as.numeric(sub('\\$',",as.character(peak$coalPrice)))

#plot Fuel price vs. MMBTU
> par(mar=c(4,6,2,5))
```

```
> plot(peak$Year, peak$MMBTU/1000000, type="l",col="blue", ylab=NA,
        xlab=expression(bold("Year")), cex.lab=1.25, las=1, lwd=5, cex.axis=1.25)
> mtext(side=2, expression(bold("Heat Input (Millions of MMBtu)")), line=3.1,
        cex=1.25, col="blue")
> par(new=T)
> plot(peak$Year,peak$oilPrice, type="l", axes=F, xlab=NA, ylab=NA, col="red",
        lwd=5, lty=2)
> axis(side=4, las=1, cex.axis=1.25)
> mtext(side=4,line=3,expression(bold("Price per unit ($)")),cex=1.25)
> par(new=T)
> plot(peak$Year,peak$gasPrice, type="l", axes=F, xlab=NA, ylab=NA, col="green3",
        lwd=5, lty=3)
> par(new=T)
> plot(peak$Year,peak$coalPrice, type="l", axes=F, xlab=NA, ylab=NA, col="black",
        lwd=5, lty=4)
> legend(2012.05,42.1, c(sprintf('%s\n%s',"Avg. Peaking Unit", "Heat
        Input"),sprintf("Avg. U.S. Oil Price"),sprintf("Avg. U.S. Gas Price"),
        sprintf("Avg. U.S. Coal Price")), cex=1, lty=c(1,2,3,4), lwd=c(1.5,1.5, 1.5, 1.5),
        col=c("blue","red","green3","black"))
```

```
##############################################################################
# Figure B-10 of Caroline Farkas Dissertation (2016)
# Annual median household income adjusted by state cost of living index
##############################################################################
> income2005_2010.dat <- data.frame(county=geography(census_income)[[1]],
        income=as.numeric(estimate(census_income)))

#clean up county names and find the states (tolower is lowercase)
> income2005_2010.dat$state=tolower(gsub("^.*County, ", "",
        income2005_2010.dat$county))
> income2005_2010.dat$county=tolower(gsub(" County,.*", "",
        income2005_2010.dat$county))

#read in 2015 cost of living index per state (from MERIC)
> col_names <- c("state", "rank", "index", "grocery", "housing", "utilities",
        "transportation", "health", "misc.")
> col_classes <- c("character",rep("numeric",8))
> COLi<- read.csv(file="/Users/Carna/Dropbox/Rutgers/Research/Manuscript2_
        data/2015_costofliving.csv",sep=",", skip=1, header=F, col.names=c(col_names),
        colClasses=c(col_classes), na.strings=c("NA", "", "-999"))

#change state names to lower case
> COLi$state <- tolower(COLi$state)
```

```
#merge 2006 income and 2003 cost of living together by state
> income_COLi <- merge(income2005_2010.dat,COLi,by.x=”state”,by.y=”state”)

#normalize incomes by cost of living by multiplying median income by cost of living
index divided by 100 (standard cost of living)
> income_COLi$normalized <- (income_COLi$income / (income_COLi$index / 100))

#map new normalized values
# load the boundary data for all counties
> county.df=map_data("county",region=c("Maine","Vermont","New Hampshire","Rhode
        Island", "Massachusetts","Connecticut","New York","New Jersey", "Delaware",
"Pennsylvania", "Virginia","West Virginia", "Ohio", "Kentucky", "Maryland", "Indiana",
        “Illinois”, “Michigan”, “Wisconsin”, “DC”))

# rename fields for later merge
> names(county.df)[5:6]=c("state","county")
> state.df=map_data("state",region=c("Maine","Vermont","New Hampshire","Rhode
Island", "Massachusetts", "Connecticut", "New York", "New Jersey", "Delaware",
        "Pennsylvania", "Virginia","West Virginia", "Ohio", "Kentucky", "Maryland",
        "Indiana", "Illinois", "Michigan","Wisconsin", "DC"))

#fix spelling differences that make counties not match (blank data)
> a <- which(income_COLi$county=="st. lawrence")
```

```
> income_COLi$county[a] <- "st lawrence"
> b <- which(income_COLi$county=="st. croix")
> income_COLi$county[b] <- "st croix"
> c <- which(income_COLi$county=="st. clair") #illinois & michigan
> income_COLi$county[c] <- "st clair"
> d <- which(income_COLi$county=="st. joseph") #indiana & michigan
> income_COLi$county[d] <- "st joseph"
> e <- which(income_COLi$county=="laporte")
> income_COLi$county[e] <- "la porte"
> f <- which(income_COLi$county=="dekalb") #indiana & illinois
> income_COLi$county[f] <- "de kalb"
> g <- which(income_COLi$county=="lasalle")
> income_COLi$county[g] <- "la salle"
> h <- which(income_COLi$county=="dupage")
> income_COLi$county[h] <- "du page"
> i <- which(income_COLi$county=="prince george's")
> income_COLi$county[i] <- "prince georges"
> j <- which(income_COLi$county=="queen anne's")
> income_COLi$county[j] <- "queen annes"
> k <- which(income_COLi$county=="st. mary's")
> income_COLi$county[k] <- "st marys"
```

```
#merge county boundaries with the new data frame
> choropleth2=merge(county.df, income_COLi, by=c("state","county"),all.x=TRUE)
> choropleth2=choropleth2[order(choropleth2$order), ]

#make cuts in the data for the map scale
> choropleth2$normalized_cuts=cut(choropleth2$normalized, breaks=c(20000, 30000,
        40000, 50000, 60000, 70000, 80000, 90000, 100000, 110000, 120000, 130000),
        include.lowest=T)

#map
> ggplot(choropleth2, aes(long, lat, group = group)) + geom_polygon(aes(fill =
        normalized_cuts), colour = "black", size = 0.1) + geom_polygon(data = state.df,
        colour = "black", fill = NA) + scale_fill_manual(values = c("#FFFFFF",
        "#EEF0FC", "#C6CBE7", "#ADB4DA", "#98A1CF", "#8590C6", "#7480BD",
        "#6372B5", "#5265AF", "#4058A9", "#2B4BA6", "#023FA5"), na.value="pink")
        + theme(axis.line=element_blank(), axis.text.x=element_blank(),
        axis.text.y=element_blank(), axis.ticks=element_blank(),
        axis.title.x=element_blank(), axis.title.y=element_blank(),
        panel.background=element_blank(), panel.border=element_blank(),
        panel.grid.major=element_blank(), panel.grid.minor=element_blank(),
        plot.background=element_blank()) + ggtitle("Annual Median Household
        Income\n Adjusted by State Cost of Living Index") + theme(plot.title =
        element_text(size=20, face="bold", vjust=0.5))
```