SECOND YEAR PROGRESS REPORT

Harvard/MIT Air, Climate and Energy Center

Regional Air Pollution Mixtures: The Past and Future Impacts of Emission Controls and Climate Change on Air Quality and Health (EPA-G2014-STAR-J1)

Annual Center Progress Report

Period: December 1, 2016 – November 30, 2017
Date of Report: February 23, 2018

Introduction. This annual report presents an overview and summary of the progress and achievements during the second year of the Harvard University/Massachusetts Institute of Technology Air, Climate and Energy (ACE) Center: Regional Air Pollution Mixtures. The primary objective of the Center is to generate new scientific knowledge on past and future air quality in the US and the associated health impacts, which is of paramount importance to designing sound strategies and solutions. Our Center is investigating the sources, composition, trends, and effects of regional air pollutant mixtures across the US over a relatively long chronological period spanning past and future years (2000-2040), and examining the influence of climate, socioeconomic factors, policy decisions, and control strategies on air pollution, health, and economic outcomes. Our Center’s investigators are a multidisciplinary team with expertise in a wide range of fields, including Exposure Assessment, Air Pollution Modeling, Atmospheric Chemistry, Epidemiology, Climate Change Modeling, Meteorology, Biostatistics, Economics, Urban Planning, Social Behavior, Risk Assessment, Remote Sensing, and Public Policy.

Our ACE Center includes five Research Projects supported by the Air Pollution Core and Administrative Unit. Each Research Project addresses the scientific issues stated in one or more of the EPA ACE Center Research Questions (RQ), and together they address all four RQs using a synergistic and integrative approach. Achievement of our research goals will be made possible by sharing data and analytical methodologies among the research teams. This synergistic approach enables us to not only more efficiently utilize resources, but also foster collaborations among the Investigators that cross disciplinary boundaries, enhance productivity, and strengthen the overall Center. The researchers and investigators of this ACE Center are collaborating to: 1) investigate past and projected regional and sub-regional air pollutant mixtures and examine their characteristics and health effects and their relationship to emissions, climate, and modifiable factors; 2) project the potential impact of climate change on regional mixtures; 3) assess the effectiveness of past and future regulatory strategies, and; 4) project and quantify future changes in socioeconomic drivers of air pollution and its health related impact and characterize carbon policy measures with respect to their air co-benefits.

During Year 2 of the Center, each of the five Research Projects, the Air Pollution Core, and the Administrative Unit have continued to make progress on their objectives. In the remainder of this section of the report, we will discuss achievements of the Center. These will include a report of our data management activities and human subject activities, as well as Center milestones. In later sections of this report, each project is discussed separately. For each project, we summarize the objectives and present the progress and achievements, publications and presentations, changes to key personnel during the first year of the ACE Center.
During Year 2, we have continued our regular working group Research Meetings. These meetings feature presentations on their research by Center investigators, or by others whose research is relevant to Center objectives. These meetings also feature discussion of Center activities and serve as a forum for issues which may impact Center research. The Steering Committee (SC) has continued to work together to discuss project updates and data fusion. In addition, we have inaugurated research meetings focused on fostering interactions between researchers at Harvard and MIT, particularly among the post-doctoral fellows and doctoral students. These meetings are held quarterly.

During Year 2, we also held our first meeting of the Science Advisory Committee (SAC) for the Center, May 16-17, 2017 in Cambridge. At the SAC meeting, a summary of progress during Year 1 for the Center and the five Projects was presented by the Center Director and each Project PI. A poster session was also given, in which a total of 25 posters across all Projects and the Air Pollution Core were presented, the majority by young researchers including doctoral students and postdocs. The SAC Report provided feedback, comments, and suggestions, which were distributed to all Center Investigators; the Center Response addressed the questions, comments and suggestion of the SAC. Both the SAC Report and Center Response were provided to EPA.

June 1-2, 2017, our Center hosted the first EPA/ACE Centers annual meeting in Cambridge, MA. At the meeting, ACE Centers discussed their specific projects and presented highlights of their Y1 accomplishments, and EPA presented and discussed related activities, including lifecycle analysis, social sciences, and regional perspective. Sessions of the meeting were devoted to exploration of inter-Center collaborative avenues, including potential projects and opportunities for data/model sharing.

**Difficulties/Delays.** There have been no delays or difficulties to report in the Center or any of the Research Projects.

**Key Personnel Changes.** We have had no key personnel changes during Year 2 of the ACE Center to report.

**Expenditures.** Current expenditures are somewhat below the budget plan at the date of report. No significant changes in the size or scope of the Center or its projects or in the originally-negotiated total estimated costs are reported.

**Quality Assurance.** The Quality Management Plan (QMP) was accepted by the EPA on February 3, 2017. There have been no changes to the QMP since that date.

The Center held biweekly research meetings during the project year. The topics to be discussed at each meeting were distributed prior to the meetings via email. The purpose of these meetings is to review progress on the research projects, including discussion of any issues that may have an impact on data quality.

During the period December 2017 through January 2018, the Quality Assurance Manager (QAM) conducted a series of interviews (Audits) with project staff identified by each project PIs as having a research role on their project and individuals identified by the Air Pollution Core Director as having a role in environmental sample collection or analysis. The scope of the audits covered both
the program audit tasks and exposure system audit tasks listed in the QMP. The researchers did not report any significant changes to SOPs, data management plans, or code books. The audits consisted of reviewing the project specific audit forms generated during the previous year’s audit and updating any changes to the forms. In addition, three publications from the Center’s publication list, representative of the work being performed under the Center, were selected for Data Management Audits:


The purpose of the data management audit was to review if the archiving requirements for each project were being met and that data and programs used to prepare the manuscript were readily available. Researchers were asked to identify the source program for data, tables and figures in the publication and the location were the programs were archived. Responses to the audits were all in a timely fashion, and all documentation was consistent with each project's data management and archiving requirements.

For all Projects and the Air Pollution Core any items identified during the audits were addressed and no items were identified affecting data quality.

**Human Subjects.** Research Projects 1, 3, 4, 5 and the Air Pollution Core of this Center contain no human subjects research as defined by the Basic Federal Policy for the Protection of Human Research Subjects, also known as the Common Rule, 40 CFR Part 26 Subparts A, B and C. Research Project 2, however, uses indoor exposure data (geographic coordinates and indoor air pollutant concentrations) from (1) the Normative Aging Study (NAS; 2006-2011) and (2) the “Diabetes, Cardiac Disease, and Pollution Vulnerability” study (2006-2010). The HSPH Institutional Review Board (IRB) has reviewed the project plans for these projects and we have obtained non-human subjects determinations for use of data from our IRB, and submitted this documentation to the EPA in our original proposal for review and approval prior to beginning this work.

In addition, Research Project 2 uses indoor air pollution data (location and concentrations) from ongoing work by Harvard University and Department of Veterans Affairs (VA) researchers in the Chronic Obstructive Pulmonary Disease (COPD) Air Pollution Study. This data set was completed in July 2017 by the VA Researchers. At that time we submitted all work to be performed under Project 2 to the IRB for review. We obtained an exemption determination for the work on Project 2 from the HSPH IRB on August 3, 2017. All work of the center has been reviewed by the HSPH IRB and has been determined to be either Non-Human Subjects Research or Exempt from IRB review. No additional IRB review is required as long as the scope of the research remains the same and no other human subjects-related data is added to the research.
Research Conduct. All research effort in the Harvard/MIT ACE Center has been performed to the highest standard of conduct.

Future Activities. During Year 3, the Center will continue with activities discussed above. Specific Project activities for Year 3 are described in the Project Progress Reports included in later sections. Additional planned Center activities during Year 2 include our second annual SAC meeting, which is scheduled for May 30-31, 2018.

Center Publications. Our list of publications and manuscripts accepted during Year 2 of the Center was submitted separately on January 31, 2018. Publications, presentations, and submitted manuscripts for each Research Project are included within the Project annual reports in the following sections.
Period Covered by the Report: December 1, 2016 - November 30, 2017

Date of Report: February 23, 2018

EPA Agreement Number: Air, Climate and Energy (ACE) Centers (EPA-G2014-STAR-J1)

Title: Project 1, Regional Air Pollution: Mixtures Characterization, Emission Inventories, Pollutant Trends, and Climate Impacts

Investigators: Petros Koutrakis, Ph.D. (lead PI; petros@hsph.harvard.edu); Brent Coull, Ph.D.; Daniel J. Jacob, Ph.D.; Loretta J. Mickley, Ph.D.; Joel Schwartz, Ph.D.

Institution: Harvard School of Public Health, Boston, MA; Harvard School of Engineering and Applied Sciences, Cambridge MA

Research Category: Air, Health, Climate Change, Integrated Assessment of the Consequences of Climate Change, Air Quality and Air Toxics, Social Science, Airborne Particulate Matter Health Effects, Air Toxics, Health Effects

Project Period: Year 2

Objective(s) of the Research: The overall objective of Project 1 is to apply new approaches to characterize and analyze both historical and projected regional air pollution mixtures and emissions across the continental US. Project 1 characterizes temporal and spatial patterns of pollutant mixtures within and across regions. In addition, this project investigates factors influencing regional pollutant mixtures and predicts the impact of climate change on future air quality. Project 1 has four specific objectives.

Objective 1 is to compile comprehensive air pollution, weather, emissions, and GIS datasets for the entire continental US for the period 2000-2015. We will estimate gas and particle concentrations at a high spatial resolution by assimilating data from monitoring networks (compiled in collaboration with the Air Pollution Core), satellite platforms, air pollution models, and spatiotemporal statistical models. Objective 2 is to develop and make publically available a national \( \text{PM}_{2.5} \) emission inventory database of high spatial resolution (1 km) for 2000-2015. This will be achieved through the application of a novel methodology we developed that predicts point and area source emissions using aerosol optical thickness measured by satellite remote sensors. Objective 3 is to characterize spatial and temporal trends of pollutant mixtures. We will perform cluster analysis to group areas that exhibit distinct pollutant profiles or mixtures, referred to as “Air Pollution Regions,” then analyze their spatial patterns and temporal trends to investigate the impact of regulations, climate change, and modifiable factors on regional mixtures. Objective 4 is to forecast the impact of regional climate change on air quality for 2016-2040 using an ensemble of climate models. We will project the potential impact of climate change on regional pollutant mixtures and predict future regional air quality assuming no changes in anthropogenic emissions.

Progress Summary/Accomplishments (Outputs/Outcomes): In year 2, our efforts were focused on Objectives 1, 2 and 4.

Objective 1. We conducted a 3-year GEOS-Chem simulation for 2013-2015 to provide continuous information on ozone and PM concentrations for the purpose of epidemiological analyses. The simulation has 0.5°x0.625° resolution over North America and is driven by MERRA-2 assimilated meteorological data from the NASA Global Modeling and Assimilation Office (GMAO). We used
the most recent benchmarked version of GEOS-Chem (v11-2-c) including detailed ozone-aerosol chemistry. That version was previously evaluated in detail with observations of ozone, PM, and their precursors over the Southeast US during the SOAS and SEAC4RS campaigns in summer 2013. The evaluation tested a new model scheme for isoprene secondary organic aerosol (SOA) and was successful in simulating overall organic aerosol levels. Observations of NOx and its oxidation products showed that NOx emissions in the EPA National Emission Inventory (NEI) were biased high by up to 50% (depending on errors in soil NOx emissions), and implementation of this correction provided an improved simulation of ozone. All these developments were incorporated in the GEOS-Chem version used for our multi-year EPA ACE simulation. Long-term relative trends of NEI emissions were applied over the duration of the simulation period.

The simulation was completed in November 2017 and results were provided to Prof. Joel Schwartz’s group (Project 3) for epidemiological analyses. The GEOS-Chem output archive includes daily surface concentrations of maximum daily 8-hr average (MDA8) ozone, NO2, PM2.5 and its components, and aerosol optical depth (AOD). This was a major deliverable for the project.

**Objective 2.** Focus during year 2 has been on Particle Emission Inventory using Remote Sensing (PEIRS) to construct spatially- and temporally-resolved emission inventories for PM2.5. Our efforts on this objective during Year 2 have included extending our work to cover more of the USA and the complete targeted time span. Using a mass balance approach, we converted the daily PM2.5 fields to a net addition (reflecting contributions of primary emission, secondary formation, and losses due to deposition) within each 1x1 km grid cell. We have also devoted significant effort toward improving our models. Because the spatial scale of primary PM2.5 emissions is smaller than the scale of secondary formation, discrete wavelet decomposition can be used to separate the net addition fields into two parts: the high frequency fraction and low-frequency fraction. This allows us to better assess primary emissions versus secondary formation and regional emissions.

**Objective 4.** We wrapped up work begun on statistical models to investigate the meteorological drivers of inter-annual to multidecadal variability of air quality, including ozone and fine particulate matter (PM2.5), in the United States. We examined processes involving climate patterns at different spatial scales, including that of local weather (~100 km), synoptic circulation (~1,000 km) and large-scale climate patterns (~10,000 km).

In Shen et al. (2017a), we first constructed a statistical model for relating observed PM2.5 to regional meteorology across the United States from 1999 to 2013. We applied the model to an ensemble of global climate models under the RCP4.5 scenario, predicting an annual mean increase of 0.4-1.4 µg/m³ of PM2.5 in the eastern United States by the 2050s. This prediction assumes present-day anthropogenic sources of PM2.5. Mean summertime PM2.5 increases as much as 2-3 µg/m³ in the eastern United States due to faster oxidation rates and greater mass of organic aerosols from biogenic emissions.

In Shen and Mickley (2017a), we developed a seasonal prediction model for surface ozone in the East. The model predicts June–July–August (JJA) daily maximum 8-h average (MDA8) ozone concentrations using large-scale climate patterns during the previous spring. In Shen and Mickley (2017b), we examined the influence of El Nino on U.S. surface ozone from 1980 to 2016. We found that each standard deviation increase in the Niño 1+2 index is associated with an increase of 1–2 ppbv ozone in the Atlantic states and a decrease of 0.5–2 ppbv ozone in the south central states. These influences can be predicted 4 months in advance. Finally, in Shen et al. (2017b), we found that U.S. summertime air quality displays strong dependence on North Atlantic sea surface temperatures, resulting from large-scale ocean-atmosphere interactions. We further identified multidecadal variability in surface air quality driven by the Atlantic Multidecadal Oscillation.
Taken together, these studies (1) improve understanding of current trends in U.S. air quality, (2) provide a means to evaluate current chemistry dynamical models such as GEOS-Chem, and (3) allow projection of air quality trends into the future, given meteorological fields from global climate models.

**Publications/Presentations:** Project 1 has several publications during the last performance period. Publications listed in the Y1 annual report as “accepted” are marked with an asterisk (*) and included here to provide full citation information.

**Publications:**


**Presentations:**


**Future Activities:**

We have now begun to use the archive of GEOS-Chem output generated for Objective 1 to address major questions that arose from our SOAS and SEAC⁴RS work. These questions, which have
subsequently been supported by other analyses, are: why are NEI NO\textsubscript{x} emissions so overestimated, and can satellite (OMI) observations of tropospheric NO\textsubscript{2} columns provide useful constraints on the patterns of overestimates by sector, region, and season? Previous studies have shown consistency between 2005-present NEI NO\textsubscript{x} emission trends and OMI NO\textsubscript{2} trends; but that then implies the NEI overestimate of NO\textsubscript{x} emissions has persisted for over a decade. We plan to compare the detailed patterns observed by OMI to those seen in the GEOS-Chem simulation to identify specific source sectors (including soils) and regions that may be responsible for the overestimate, and the extent to which this overestimate persists across seasons. This work will constitute the final thesis project for Harvard PhD student Rachel Silvern.

During Year 3, we also expect to focus intensively on Objective 2, using PEIRS with wavelet decomposition to separate the high frequency/primary emissions from the net addition fields.

**Supplemental Keywords:** particles, pollutant mixtures, pollution trends, regional pollution, public policy, data fusion

**Relevant Web Sites:** The website for the Harvard/MIT ACE Center is:

[https://content.sph.harvard.edu/ace/](https://content.sph.harvard.edu/ace/)
Period Covered by the Report: December 1, 2016 - November 30, 2017

Date of Report: February 23, 2018

EPA Agreement Number: Air, Climate and Energy (ACE) Centers (EPA-G2014-STAR-J1)

Title: Air Pollutant Mixtures in Eastern Massachusetts: Spatial Multi-resolution Analysis of Trends, Effects of Modifiable Factors, Climate, and Particle-induced Mortality

Investigators: Brent Coull, Ph.D. (lead PI; bcoull@hsph.harvard.edu); Petros Koutrakis, Ph.D.; Joel Schwartz, Ph.D.

Institution: Harvard School of Public Health, Boston, MA

Research Category: Air, Climate Change, Global Climate Change, Air Quality and Air Toxics, Airborne Particulate Matter Health Effects

Project Period: Year 2

Objective(s) of the Research: The objective of Project 2 is to characterize historical air pollution in Eastern Massachusetts at a high spatial resolution and identify modifiable factors responsible for observed changes in PM$_{2.5}$ mass, emissions, elemental profiles, and ground air temperature. Project 2 investigates within-region variability of pollutant mixtures; examines the impact of modifiable factors on air quality; and evaluates the effectiveness of source control policies. Project 2 has four specific objectives.

Objective 1 is to use a novel, multi-resolution spatial analysis based on wavelet decomposition of high-resolution (1x1 km) remote sensing data on PM$_{2.5}$ mass and ground air temperature to identify daily regional, sub-regional (urban background) and locally generated variation in these fields. Objective 2 is to develop and apply spatiotemporal regression models to (a) quantify the impact of modifiable factors, including transportation, heating fuel use, energy, urban planning, PM$_{2.5}$ emissions, population statistics, and policy interventions, on (i) sub-regional and local variation in PM$_{2.5}$ mass and ground air temperature and (ii) high resolution local estimates of PM$_{2.5}$ emissions; (b) identify locations in which these impacts are greatest; and (c) identify lag times between implementation of a given control strategy and decreases in PM$_{2.5}$ emissions and mass. Objective 3 is to implement a novel multi-resolution correlation analysis to identify PM$_{2.5}$ elemental profiles that vary at regional, sub-regional, and local scales, and apply spatiotemporal regression models to these profiles to identify modifiable factors driving urban background and local variability in PM$_{2.5}$ composition. And Objective 4 is to use the spatial scale-specific (regional, sub-regional, and local) temporal variability in PM$_{2.5}$ mass and the PM$_{2.5}$ elemental profiles to identify source types (regional, urban background, or local) and the composition of their emissions driving pollution-induced mortality in Eastern Massachusetts. This project relies on existing remote-sensing satellite data, ambient monitoring data collected from numerous sampling campaigns (including the HSPH Boston Supersite daily samples collected since 1998 and samples from 600 locations), as well as new data collected from 2015-2018 in Eastern Massachusetts.

Progress Summary/Accomplishments (Outputs/Outcomes): This past year we finalized several manuscripts that ultimately appeared in print. Highlights include:

Black Carbon (BC), an indicator of particles generated from traffic sources, has been associated with a number of health effects however due to its high spatial variability its concentration is
difficult to estimate. We previously fit a model estimating BC concentrations in the greater Boston area; however this model was built using limited monitoring data and could not capture the complex spatio-temporal patterns of ambient BC. In order to improve our predictive ability, we obtained more data for a total of 24,301 measurements from 368 monitors over a 12 year period in Massachusetts, Rhode Island and New Hampshire. We also used Nu-Support Vector Regression (nu-SVR) - a machine learning technique which incorporates nonlinear terms and higher order interactions, with appropriate regularization of parameter estimates. We then used a generalized additive model to refit the residuals from the nu-SVR and added the residual predictions to our earlier estimates. Both spatial and temporal predictors were included in the model which allowed us to capture the change in spatial patterns of BC over time. The 10 fold cross validated (CV) $R^2$ of the model was good in both cold (10-fold CV $R^2 = 0.87$) and warm seasons (CV $R^2 = 0.79$). We have successfully built a model that can be used to estimate short and long-term exposures to BC in MA, RI and Southern NH. This work was published in *Environmental Research* (Abu Awad 2017).

Fine particulate matter (PM2.5) measured at a given location is a mix of pollution generated locally and pollution traveling long distances in the atmosphere. Therefore, the identification of spatial scales associated with health effects can inform on pollution sources responsible for these effects, resulting in more targeted regulatory policy. We proposed a two-dimensional wavelet decomposition that alleviates restrictive assumptions required for standard wavelet decompositions. Using this method we decomposed daily surfaces of PM2.5 to identify which scales of pollution are most associated with adverse health outcomes. A key feature of the approach is that it can remove the purely temporal component of variability in PM2.5 levels and calculate effect estimates derived solely from spatial contrasts. This eliminates the potential for unmeasured confounding of the exposure - outcome associations by temporal factors, such as season. We applied our method to a study of birth weights in Massachusetts, U.S.A from 2003-2008 and find that both local and urban sources of pollution are strongly negatively associated with birth weight. Results also suggest that failure to eliminate temporal confounding in previous analyses attenuated the overall effect estimate towards zero, with the effect estimate growing in magnitude once this source of variability is removed. This work was published in the *Annals of Applied Statistics* this past year (Antonelli et al. 2017).

Recent interest focuses on identifying critical windows of vulnerability associated with prenatal exposure to air pollution during pregnancy. An analysis based on a distributed lag model (DLM) can yield estimates of a critical window different from those from an analysis that regresses the outcome on each of the three trimester average exposures (TAEs), which is the standard approach typically used in the environmental health literature. Using a simulation study, we assessed bias in estimates of critical windows obtained using three regression approaches: 1) three separate models to estimate the association with each of the three TAEs; 2) a single model to jointly estimate the association between the outcome and all three TAEs; and 3) a DLM. We use weekly fine particulate matter (PM$_{2.5}$) exposure data for 238 births in a Boston-area birth cohort and a simulated outcome and time-varying exposure effect. Estimates using separate models for each TAE were biased and identified incorrect windows. This bias arose from seasonal trends in PM$_{2.5}$ that induce correlation between TAEs. Including all TAEs into one model reduced bias. DLM produced estimates that were unbiased and added flexibility to identify critical windows. Analysis of body mass index z-score and fat mass in the same cohort highlights inconsistent estimates from the three methods. This work was published in *American Journal of Epidemiology* this past year (Wilson et al. 2017a).
In related work, simultaneous estimation of windows of vulnerability and effect heterogeneity is typically accomplished by fitting a distributed lag model (DLM) stratified by subgroup. However, this does not allow for subgroups to have the same window of vulnerability but different effects within the window or to have different windows but the same within-window effect, which can make full characterization of effect heterogeneity difficult. We proposed a new approach that partitions the DLM into a constrained functional predictor that estimates windows of vulnerability and a scalar effect size that estimates the effect within the window. The proposed method allows for heterogeneity in only the window of vulnerability, only the effect within the window, or in both by allowing each component to be either shared or differ across groups. We used the proposed method to estimate windows of vulnerability in the association between prenatal exposures to fine particulate matter (PM$_{2.5}$) and each of birth weight and asthma incidence, and to estimate how these associations vary by sex and maternal obesity status, in a Boston-area prospective pre-birth cohort study (Wilson et al. 2017b). These methods have been implemented in several other analyses of prenatal, sex-specific critical windows of air pollution exposure on health outcomes in children (Brunst et al. 2017; Chiu et al. 2017; Lee 2017).

**Publications/Presentations:**


**Future Activities:** We will focus on two areas of research recommended by the Center’s Scientific Advisory Council last May. These are better characterization of uncertainty associated with exposure-prediction models based on remote-sensing satellite data, and the application of the methods applied in Objective 1 to mortality data, which is objective 4 in the proposal. More details are as follows:

In the last few years, several research teams have developed distinct spatio-temporal models (exposure models) to predict ambient air pollution exposures of study participants even in areas where air pollution monitors are sparse. Use of these exposure models has led to strong evidence of air pollution-related adverse health effects, and in some instances, evidence of heterogeneity of these health effects across sub-populations. A significant limitation, however, is that these health effect estimates and their statistical uncertainty are based on the very strong assumption that a single exposure model is correct. Over the coming year we will be evaluating the impact of this assumption and developing a way to characterize uncertainty associated with exposure prediction model choice. Our preliminary results show that even when distinct exposure models predict similarly well, the spatio-temporal pattern of their errors might differ significantly. These complex error structures can lead to misleading evidence of effect heterogeneity across sub-populations, which in turn could be wrongly attributed to other factors that vary in a similar manner either in space (e.g. rural vs. urban) or time (e.g. by season).

Ultimately, we will develop a novel ensemble model framework to integrate information across multiple existing air pollution prediction models across the United States and, for the first time, to
comprehensively quantify the uncertainty associated with air pollution exposure assessment. Subsequently, we propose to couple the ensemble predictions with a novel measurement error correction approach, to fully characterize the impact of air pollution on adverse health and the shape of the exposure-response curves. The proposed novel paradigm will greatly improve communication of exposure uncertainty in the health effect estimates both to policy makers and the public and can easily be extended for use at different locations and at a global scale, as well as for other environmental exposures.

During the coming year we will also complete work on Objective 4 of the project, which is to use the spatial scale-specific (regional, sub-regional, and local) temporal variability in PM$_{2.5}$ mass that we developed in Objective 1 (Antonelli et al. 2017) to identify source types (regional, urban background, or local) and the composition of their emissions driving pollution-induced mortality in Eastern Massachusetts.

**Supplemental Keywords:** climate change, regional pollution, multi-resolution spatial analysis, source emissions, local pollution control strategies, wavelet analysis

**Relevant Web Sites:** The Harvard/MIT ACE Center website, [https://content.sph.harvard.edu/ace/](https://content.sph.harvard.edu/ace/)
**Period Covered by the Report:** December 1, 2016- November 30, 2017

**Date of Report:** February 23, 2018

**EPA Agreement Number:** Air, Climate and Energy (ACE) Centers (EPA-G2014-STAR-J1)

**Title:** Causal Estimates of Effects of Regional and National Pollution Mixtures on Health: Providing Tools for Policy Makers

**Investigators:** Joel Schwartz, Ph.D. (lead PI; joel@hsph.harvard.edu); Brent Coull, Ph.D.; Petros Koutrakis, Ph.D.; Antonella Zanobetti, Ph.D.

**Institution:** Harvard School of Public Health, Boston, MA

**Research Category:** Air, Climate Change, Air Quality and Air Toxics, Airborne Particulate Matter Health Effects, Particulate Matter

**Project Period:** Year 2

**Objective(s) of the Research:** The objective of Project 3 is to estimate the causal impact of changes in pollution concentrations and mixtures (annual averages and daily patterns), how they vary by modifiable factors, the causal impacts of AQI triggers, how climate change that occurred in the last 20 years has increased mortality due to pollution, how temperature modifies the effects of pollution mixtures, and how these effects change for exposures less than the ambient standards for PM$_{2.5}$. Project 3 provides region-specific causal estimates of effects of pollution mixtures; provides causal estimates of the impact of modifiable factors; assesses the impact of climate change on mortality from air pollution using historic data, avoiding any dependence on the accuracy of climate models; and provides causal estimates of how changes in particular components of mixtures affect mortality, to guide region-specific policy decisions on air pollution. Project 3 has five specific objectives.

*Objective 1* is to identify and estimate the causal effects of air pollution and mixtures on human health. We will use methods of causal inference to a) identify the causal effects of regional annual air pollution concentration fluctuations and temperature fluctuations during the last 16 years on human health; b) identify the causal effects of regional air pollution trends during the last 16 years; c) identify the causal effects of pollution mixtures, sources, and emissions on health; d) identify differences in these effects by modifiable factors; e) conduct a national risk assessment on the causal impact of past pollution on mortality, including the regional differences in concentration-response; and f) investigate the causal impact of AQI thresholds for PM$_{2.5}$ and O$_3$ due to behavioral adaptation. *Objective 2* is to analyze relative acute toxicity of pollution mixtures. We will a) examine spatial (across regions) and temporal heterogeneity in the acute toxicity of pollution mixtures and emissions to understand which source types, atmospheric processes, and exposure factors influence the toxicity of regional mixtures and b) use causal mediation analysis to determine how much of the temperature effect on mortality is mediated by its effects of pollution concentrations, and how that varies regionally. This will allow us to obtain local- and region-specific estimates of future health effects and the benefits of changes in modifiable factors and adaptation. *Objective 3* is to estimate the excess deaths resulting from air pollutant concentration changes due to weather changes in the last 20 years. We will demonstrate the extent to which public health impacts of climate change through pollution have already occurred, by using causal estimates of C-R relationships. Regional health impacts will be assessed using region-specific mortality risks estimates from Objective 1. *Objective 4* is to estimate the causal health effects of low-level air pollution exposure. Specifically, we will examine whether the observed effects at
low pollutant levels are due to the synergistic effect of multiple pollutants (mixtures) present at low levels. And Objective 5 is to investigate air pollution-related health effects at high and low temperatures. We will examine this by region and determine whether populations, especially those that include sensitive individuals, adapt to abrupt temperature changes.

**Progress Summary/Accomplishments (Outputs/Outcomes):** The past year was a very successful year for our project. We published two major nationwide studies on the acute and chronic effects of exposure to PM$_{2.5}$ and ozone. In Di et al we analyzed the entire Medicare population and looked at the effect of annual average exposure to PM$_{2.5}$ and ozone on survival in a cohort study. This study included all Medicare beneficiaries for the first time, and importantly, people who live in small cities, towns, and rural areas as well as larger urban areas. We found a highly significant association between PM$_{2.5}$ and ozone and mortality rates. Importantly, we had 32 million people whose exposure was below the current ambient air quality standard of 12 μg/m$^3$, with 248 million person years of follow-up, and in analyses restricted to just those observations, we found a highly significant association with PM$_{2.5}$, demonstrating that the mortality effects continue to concentrations well below the current standard.

In the second paper, we performed a case-crossover analysis looking at ozone exposure and PM$_{2.5}$ exposure and the acute risk of mortality, again, in the entire Medicare population. Once more, both associations were highly significant, and continued well below current air quality standards with no evidence of a threshold.

In Abu Awad et al we extended a BC model we had previously fit in eastern MA to include all of Massachusetts as well as Rhode Island and Southern New Hampshire. We also improved the model by using machine learning, providing estimates with smaller mean squared error on left out monitors, and extended the time period up to 2012. Using this updated exposure we then published a paper (Kingsley 2017) using that exposure (and our PM$_{2.5}$ model) to look at the role of particles in preterm birth.

In Dai 2017 we used the measurements of particle metal components from our central site to examine the effects of different kinds of particles on DNA methylation in an agnostic analysis. We observed 20 Bonferroni significant (P-value < 9.4$\times$ 10$^{-9}$) CpGs for Fe, 8 for Ni, and 1 for V, demonstrating that metal particles influence DNA methylation.

In Wang 2017 we developed a novel method for assessing adaptation to heat waves using temperature and mortality data from 1961 onwards and applied that to the estimated temperatures under 4 different greenhouse gas emissions scenarios from 21 different global climate models. We found adaptation continued up until a maximum temperature. Hence in the Northern US, heat wave related mortality is expected to fall (taking into account continued adaptation) but in the Southeastern US, it is expected to increase as adaptation reaches its limits.

In Nwanaji-Enwerem et al, 2017, we made the important finding that DNA methylation age, a measure of biological aging, was increased by exposure to PM$_{2.5}$ (estimated at home addresses with one of our models) and that this effect was modified by genetic variations in genes related to miRNA processing. This effect varied by genotype from between 1 and 3 years of additional aging equivalent as PM$_{2.5}$ went from 8 to 12 μg/m$^3$, again showing effects below ambient standards.

In another paper by Nwanaji-Enwerem we examined PM$_{2.5}$ components from a chemical transport model as predictors of DNA methylation age. We found that sulfate and ammonium components were associated with increased methylation age, independent of PM$_{2.5}$ mass. This held when restricted to PM$_{2.5}$ concentrations below 12.
In Peng 2017 we showed that short term increases in air temperature resulted in increased lesions in mitochondrial DNA.

In Prada 2017 we reported that PM$_{2.5}$ from our models was associated with hospital admissions for osteoporosis-related fractures in the Medicare cohort. Simultaneously, in a more detailed cohort in the Boston area we followed a cohort of participants longitudinally and for that both black carbon (from the model above) and PM$_{2.5}$ were associated with lower serum parathyroid hormone, and black carbon was associated with higher bone mineral density loss over time in multiple sites.

In Schwartz 2017 we addressed the causality of the association between PM$_{2.5}$ and NO$_2$ with daily deaths, but implementing an instrumental variable analysis for local particles, as well as a negative exposure control. We found significant effects of both pollutants. The instrument was a combination of the height of the planetary boundary layer and wind speed, which influence local pollution concentrations, but are unlikely to otherwise influence daily deaths.

We followed up with another causal modeling approach (Wang 2017), this time for cohort analysis of long term exposure. Using the Medicare population of the Southeast as our cohort, we developed a doubly robust additive hazard model to study survival versus exposure. That is, like a tradition causal model, it give causal estimates if the model for exposure given the covariates is correct, however, it also gives causal estimates if the model for mortality given the covariates is correct. We found a highly significant, causal estimate for the effects of PM$_{2.5}$ on life expectancy. Separately, we fit a more traditional proportionate hazard model in the same population to examine effect modification. We found higher effects among males, non-whites, persons eligible for Medicaid as well as Medicare, and persons living in neighborhoods with lower SES.

In addition to these publications, we continued to update our pollution models in terms of improving modeling methods and extending the time frame. We expect to produce updated estimates to 2016 for BC, PM$_{2.5}$, and O$_3$ by early summer.

**Publications/Presentations:**


Wang, Y; Lee, M-h; Liu, P; Shi, L; Yu, Z; Awad, Y A; Zanobetti, A; Schwartz, J. D. Doubly Robust Additive Hazards Models to Estimate Effects of a Continuous Exposure on Survival. Epidemiology, 2017, 28:771-779.

Makar, M; Antonelli, J; Di, Q; Cutler, D; Schwartz, J; Dominici, F. Estimating the Causal Effect of Low Levels of Fine Particulate Matter on Hospitalization. Epidemiology, 2017 28:627-34.


**Future Activities:** Future plans include updating our BC, PM2.5, and ozone models, adding a NO2 model, and starting models for PM2.5 components nationwide. Simultaneously we plan to expand our outcomes to include the effects of long term exposure on hospital admissions, and to expand our causal modeling efforts for both acute and chronic exposures.

**Supplemental Keywords:** particles, particulate matter, pollutant mixtures, regional pollution, risk analysis, causal modeling

**Relevant Web Sites:** The Harvard/MIT ACE Center website, [https://content.sph.harvard.edu/ace/](https://content.sph.harvard.edu/ace/)
Period Covered by the Report: December 1, 2016-November 30, 2017

Date of Report: February 23, 2018

EPA Agreement Number: Air, Climate and Energy (ACE) Centers (EPA-G2014-STAR-J1)

Title: Project 4: A Causal Inference Framework to Support Policy Decisions by Evaluating the Effectiveness of Past Air Pollution Control Strategies for the Entire United States

Investigators: Corwin Zigler (PI), Francesca Dominici (co-PI), Joel Schwartz, Loretta Mickley, Steve Barrett

Institution: Harvard TH Chan School of Public Health/Harvard School of Engineering and Applied Sciences/Massachusetts Institute of Technology

Research Category: Air, Climate Change, Air Quality and Air Toxics, Airborne Particulate Matter Health Effects

Project Period: Year 2

Objective(s) of the Research: The overall objective of Project 4 is to develop a new methodological framework rooted in principles of causal inference to investigate the effectiveness of specific control strategies on impacting the largest power-generating units in the United States. In Project 4, we combine state-of-the-art atmospheric modeling, causal inference methods, and national data sets to conduct accountability research; that is, research that characterizes causal effects of well-defined regulatory actions at power plants on: 1) emissions; 2) air quality across distant locations in accordance with atmospheric fate, transport, and other factors; and 3) health outcomes. Project 4 has 4 specific objectives.

Objective 1 is to develop a national database on emissions control technologies employed at a large number of power-generating units in the US linked with: continuous emissions monitoring, ambient air quality monitoring, weather, population demographics, and Medicare hospitalization and mortality outcomes for the period 1995 to 2015. Objective 2 is to estimate and compare the causal effects of past control strategies implemented at the largest power-generating facilities on \( \text{SO}_2 \), \( \text{NO}_x \), \( \text{CO}_2 \), and \( \text{PM}_{2.5} \) emissions and population exposure to criteria pollutants (\( \text{PM}_{2.5} \) and \( \text{O}_3 \)) for the entire US for the period 2000 to 2015. This requires integrating new statistical methods for causal inference with atmospheric chemistry models of how changes in emissions impact ambient exposures across distant locations in accordance with atmospheric fate, transport, and other factors. Objective 3 is to estimate the causal effects of past control strategies implemented at the largest power-generating facilities on mortality and morbidity in the entire US both locally and nationally, and compare the differential health impact of different control strategies. And Objective 4 is to develop approaches for mediation analysis that will quantify the extent to which causal effects of regulatory actions on health outcomes can be attributable to changes in targeted modifiable factors (e.g., emissions, targeted pollutants), as opposed to being driven by co-benefits to other factors.

Progress Summary/Accomplishments (Outputs/Outcomes): The bulk of progress during year 2 for Project 4 has fallen into 5 categories:

1. Refining the national database on power plants, emissions, ambient pollution, population demographics, and health outcomes among Medicare beneficiaries. Efforts have included the acquisition of a curated database of census-like variables for as all the years available from ESRI Business Analyst (annual data, 2000-present). We also bought two reserved-
usage Dell servers and 30TB of secure storage, within the Harvard RCE high-performance computing cluster (https://rce-docs.hmdc.harvard.edu/).

2. **New reduced-form scalable strategies to produce source-receptor matrices linking individual power plants to US zip codes.** We have continued to refine our strategy based on the Hybrid Single Particle Lagrangian Integrated Trajectory model. In collaboration with the Carnegie Mellon/UW ACE Center, we have initiated a new strategy based on the recently-proposed Intervention Model for Air Pollution (InMAP) to generate similar source-receptor matrices. We have also explored purely statistical strategies based on scalable generalized additive models of daily emissions and pollution time series. Methods are being compared/validated against each other and against observations, and also being used for intervention analyses.

3. **Statistical methods development.** The various methods for obtaining source-receptor models have been used as inputs into novel statistical methods development, including development of new methodology for bipartite causal inference with interference. We have also continued to pursue methods for general causal inference, multiple mediating variables, spatial confounding adjustment, uncertainty in propensity score “design” stage, causal inference with interference, model averaging for confounder selection, measurement error, causal exposure-response estimation, and statistical network analysis, all in the context of evaluating power plant regulations.

4. **Environmental and Epidemiological Analyses.** The above methods development has been in service of analyses of air pollution interventions including analyses of the health benefits of PM nonattainment designations, evaluation of selective (non) catalytic reduction systems for reducing NOx emissions and ambient ozone, mediation analyses of how scrubbers on coal-fired power plants reduce ambient PM, evaluation of the health effects associated with air pollution derived specifically from coal-fired power plants, flexible estimation of causal exposure-response relationships at low levels of ambient pollution.

5. **Software.** Publications are paired with reproducible R scripts and/or packages, hosted on digital repositories. We deployed a customized version of the SplitR package (the R interface to HYSPLIT) on our new servers. We also developed and deployed the rinmap R package to generate source-receptor matrices using inMAP.

**Publications/Presentations:**

**Publications:**


Zigler CM, Choirat C, and Dominici F. Impact of National Ambient Air Quality Standards non-attainment designations on particulate pollution and health. Epidemiology in press.


**Submitted Manuscripts Currently Under Review:**


**Presentations:**


Georgia Papadogeorgou, Statistical challenges in air pollution research: from spatial confounding to interference. Invited seminar at the University of Minnesota, Department of Biostatistics, Minneapolis MN, Oct 25, 2017.


Kevin Cummiskey, A source-oriented approach to coal emissions health effects. Contributed Talk, Joint Statistical Meetings, Aug 2017

Kevin Cummiskey, A source-oriented approach to coal emissions health effects. Contributed Talk and Student Travel Award Winner, International Conference on Health Policy Statistics, Jan 2018

Kevin Cummiskey, A data-driven approach to source-receptor mapping of power plant emissions and exposed populations, Harvard Biostatistics Environmental Statistics Seminar, Dec 2017

**Future Activities:** We will continue to refine methodology for reduced-form models to produce spatially and temporally refined source-receptor mappings, including a dispersion implementation of HYSPLIT and investigation of other products produced by other ACE Centers (e.g., the EASIUR model from CMU). With a newly-hired postdoctoral fellow specializing in environmental engineering, we plan to formalize our comparison of different source-receptor mapping techniques for the specific purposes of evaluating interventions, as well as deploy these methods in a variety of health-outcome studies of interventions.

**Supplemental Keywords:** Accountability assessment, power-generating sector, intervention evaluation, source-receptor mapping

**Relevant Web Sites:** The Harvard/MIT ACE Center website, [https://content.sph.harvard.edu/ace/](https://content.sph.harvard.edu/ace/) In addition, the following websites provide tools and data to implement methods and analyses developed during this project:

- [https://github.com/czigler/arepa](https://github.com/czigler/arepa)
- [https://github.com/czigler/PM2.5-Nonattainment](https://github.com/czigler/PM2.5-Nonattainment)
- [https://dataverse.harvard.edu/dataverse/airqualregs](https://dataverse.harvard.edu/dataverse/airqualregs)
- [https://github.com/gpapadog/DAPSm](https://github.com/gpapadog/DAPSm)
- [https://github.com/gpapadog/DAPSm-Analysis](https://github.com/gpapadog/DAPSm-Analysis)
- [https://github.com/cchoirat/rinmap](https://github.com/cchoirat/rinmap)
- [https://github.com/lhenneman/SplitR](https://github.com/lhenneman/SplitR)
- [https://osf.io/s98qm/](https://osf.io/s98qm/)
**Period Covered by the Report:** December 1, 2016-November 30, 2017

**Date of Report:** February 23, 2018

**EPA Agreement Number:** Air, Climate and Energy (ACE) Centers (EPA-G2014-STAR-J1)

**Title:** Projecting and Quantifying Future Changes in Socioeconomic Drivers of Air Pollution and its Health-Related Impacts

**Investigators:** N.E. Selin, S.R.H. Barrett, S. Solomon, J. Reilley

**Institution:** Massachusetts Institute of Technology

**Research Category:** Air, Climate Change, Urban Air Toxics, Mercury, Air Quality and Air Toxics, Airborne Particulate Matter Health Effects, Air Toxics, Particulate Matter

**Project Period:** Year 2

**Objective(s) of the Research:** Project 5 investigates future changes in regional air pollution characteristics as a result of technological and societal changes. We will quantify the future implications of technologies and efficiency improvements in the energy and transportation sectors on regional differences in air pollution impacts. Selected case studies assess, *inter alia*, the environmental and health benefits of choices in state and regional carbon policy implementation relevant to recently proposed carbon dioxide emission reductions from the energy sector. We will examine the health-related benefits of reducing concentrations of ozone and particulate matter, as well as changing regional air pollution mixtures including air toxics.

**Progress Summary/Accomplishments (Outputs/Outcomes):** Work on Project 5 during the reporting year has been focused on three main goals. The first has been to continue the sector-, emitter-, and receiver-specific analysis of past and present emissions within the US. To that end, we have completed a study quantifying the fine particulate matter (PM$_{2.5}$) pollution exchange between the US states. Using the adjoint of the GEOS-Chem chemistry transport model, the state-level population exposure sensitivities to PM$_{2.5}$ precursor emissions in all other states in the continental US have been calculated, including “domestic” emissions (same-state). The EPA’s National Emissions Inventories from 2004 onwards have been leveraged to provide a high-quality input dataset, while the Sparse Matrix Operator Kernel (SMOKE) Modeling System has been applied to group individual sources into seven sectors: electric power generation, industry, commercial/residential, and four modes of transportation: road, marine, rail and aviation. By applying the estimated emissions distribution to the calculated sensitivity matrices, we have generated impact source-receptor matrices for every pair of states within the contiguous US. This allows each state to estimate not only the degree to which air quality degradation is a result of emissions from a specific sector, but also the degree to which those impacts are within the legislative control of that state. Furthermore, estimates were made for multiple years from 2004 to 2014, showing that while the overall impacts of anthropogenic emissions have been decreasing, the impact per unit emitted has been increasing. This speaks directly to objective 4, by quantifying the role that changes in the mix and magnitude of anthropogenic emissions has affected and will continue to affect local air quality. It also provides crucial information for the evaluation of the impact of related energy policy, by giving a numerical estimate of the co-benefits or tradeoffs which might be incurred if other policies penalize or incentivize specific states or industrial sectors.
A manuscript based on this work has been submitted for publication and is currently under peer review.

The second focus area has been on objectives 2 and 3, through extending and applying our economic analysis capabilities. We have made additional enhancements to the U.S. Regional Energy Policy Model (USREP) which will allow us to model energy policies and their impacts with greater precision. These enhancements include the incorporation into USREP of detailed data on the relative costs of energy technologies. On the basis of these changes, we have designed and implemented a new version of USREP (currently in the final stages of development) which models a larger number of sub-national regions and allows for the simulation of state-level energy policies. USREP’s comprehensive coverage of the U.S. economy allows us to account for the broad impacts that energy policies can have on the economy. We are using this model in combination with reduced-form air quality tools (including the source-receptor matrices described above) to simulate energy policies and estimate their effects with regards to both the economy and air quality. This modeling framework provides self-consistent estimates of the costs and benefits of U.S. energy policies. We are also exploring comparisons with other reduced-form tools including the Intervention Model for Air Pollution (InMAP), in collaboration with the Marshall group at the University of Washington.

We have continued an effort to develop a large ensemble of climate simulations relying on the MIT Integrated Global System Model, more specifically a version coupled with the NCAR Community Atmospheric Model, to provide the core basis to examine the impact of future climate change on air quality and health. This specifically serves project objectives 1, 3, and 5. This ensemble builds upon the work by Garcia-Menendez et al. (2015), which investigated the air quality and health benefits from avoided climate change under greenhouse gas mitigation using one of the largest ensemble of climate simulations to drive the global and three-dimensional atmospheric chemistry model CAM-Chem. Because the ensemble includes both multi-decadal and initial condition perturbations, leading to more than 1000 years of simulation air quality, Garcia-Menendez et al. (2017) further examined the impact of natural variability on the robustness of projections of climate change impacts on U.S. ozone pollution. However, this ensemble was designed as time slices, with three 30-year time periods (1981-2010, 2036-2065 and 2086-2115). Thus it does not allow a proper time of emergence analysis, which instead requires transient simulations, to determine how many years it will take for climate change, greenhouse mitigation policies or air pollution policies to be detected, given the large year-to-year variations in the climate and air quality systems. For this reason, we are running a new large ensemble of transient climate simulations to answer these questions. In addition, since the ongoing project will use the GEOS-Chem atmospheric chemistry model instead of the CAM-Chem model, we have modified the source code of the climate model to provide the appropriate input data for the project. The updated climate model is now production ready and the runs will start momentarily. A new, high-performance version of GEOS-Chem (GCHP) has now been adapted to accept these fields, allowing rapid and responsive air quality and composition simulations to be incorporated as a component in the modeling chain.

Publications/Presentations:

Publications:

**Presentations:**


Steven R.H. Barrett, Irene C. Dedoussi, “Pollution exchange between the US states (as part of Project 5)”. Presentation given at the Harvard/MIT ACE Center Science Advisory Committee (SAC) Meeting on May 16 & 17, 2017 (Boston, MA).

**Future Activities:** With the integrated methodology now complete, the first major objective of this reporting period is to complete a comprehensive set of climate simulations (Objectives 1 and 5). Once generated, these will be applied in the newly-configured global chemistry-transport model GCHP, to provide fully transient estimates of ozone and PM$_{2.5}$ impacts in the US over the next 100 years under different climate scenarios (Objectives 4 and 5). A central focus of year 3 will be to establish an analytical method through which the time of emergence of air quality impacts can be established. This will take advantage of our ensemble approach, with a separate evaluation for each climate scenario based on differences between climate change-driven change and local variability from 1980-2100. The new approach will provide a robust estimate of how and when we expect to be able to observe impacts of climate change on air quality, from state-level to regional and international analysis (Objective 5). On the economic analysis aspects of the project, we will continue work on USREP and expect to complete a study on the economic and air pollution effects of regional energy policies such as the Renewable Portfolio Standards (Objectives 2 and 3). We anticipate no significant additional timeline adjustments for the project.

**Supplemental Keywords:** PM$_{2.5}$, ozone, adjoint method, air quality, greenhouse gas, computable general equilibrium, mercury, polycyclic aromatic hydrocarbons, PAHs

**Relevant Web Sites:** Our web site [http://globalchange.mit.edu](http://globalchange.mit.edu) highlights recent work from across our research program, and links to the web sites of individual PIs. In addition, the Harvard/MIT ACE Center website, [https://content.sph.harvard.edu/ace/](https://content.sph.harvard.edu/ace/)