Faculty Journal Articles   Faculty Scholarship

2020

Enhancing models and measurements of traffic-related air pollutants for health studies using dispersion modeling and Bayesian data fusion

Stuart A. Batterman
University of Michigan - Ann Arbor

Veronica J. Berrocal
University of California, Irvine

Chad Milando

Owais Gilani
Bucknell University, og005@bucknell.edu

Saravanan Arunachalam

See next page for additional authors

Follow this and additional works at: https://digitalcommons.bucknell.edu/fac_journ

Part of the Biostatistics Commons, Environmental Health and Protection Commons, Environmental Indicators and Impact Assessment Commons, Environmental Monitoring Commons, Statistical Methodology Commons, and the Statistical Models Commons

Recommended Citation

Authors

Stuart A. Batterman, Veronica J. Berrocal, Chad Milando, Owais Gilani, Saravanan Arunachalam, and K. Max Zhang

Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion

Stuart Batterman, Veronica J. Berrocal, Chad Milando,
Owais Gilani, Saravanan Arunachalam, and K. Max Zhang

Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion

Stuart Batterman, Veronica J. Berrocal, Chad Milando,
Owais Gilani, Saravanan Arunachalam, and K. Max Zhang

with a Critique by the HEI Review Committee

Research Report 202

Health Effects Institute

Boston, Massachusetts

Trusted Science • Cleaner Air • Better Health

CONTENTS

ABOUT HEI

The Health Effects Institute is a nonprofit corporation chartered in 1980 as an independent research organization to provide high-quality, impartial, and relevant science on the effects of air pollution on health. To accomplish its mission, the institute

- Identifies the highest-priority areas for health effects research;

- Competitively funds and oversees research projects;

- Provides intensive independent review of HEI-supported studies and related research;

- Integrates HEI's research results with those of other institutions into broader evaluations; and

- Communicates the results of HEI's research and analyses to public and private decision makers.

HEI typically receives balanced funding from the U.S. Environmental Protection Agency and the worldwide motor vehicle industry. Frequently, other public and private organizations in the United States and around the world also support major projects or research programs. HEI has funded more than 340 research projects in North America, Europe, Asia, and Latin America, the results of which have informed decisions regarding carbon monoxide, air toxics, nitrogen oxides, diesel exhaust, ozone, particulate matter, and other pollutants. These results have appeared in more than 260 comprehensive reports published by HEI, as well as in more than 2,500 articles in the peer-reviewed literature.

HEI's independent Board of Directors consists of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization. The Research Committee solicits input from HEI sponsors and other stakeholders and works with scientific staff to develop a Five-Year Strategic Plan, select research projects for funding, and oversee their conduct. The Review Committee, which has no role in selecting or overseeing studies, works with staff to evaluate and interpret the results of funded studies and related research.

All project results and accompanying comments by the Review Committee are widely disseminated through HEI's website (*www.healtheffects.org*), printed reports, newsletters and other publications, annual conferences, and presentations to legislative bodies and public agencies.

ABOUT THIS REPORT

Research Report 202, *Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion,* presents a research project funded by the Health Effects Institute and conducted by Dr. Stuart Batterman of the University of Michigan, Ann Arbor, and his colleagues. The report contains three main sections.

The HEI Statement, prepared by staff at HEI, is a brief, nontechnical summary of the study and its findings; it also briefly describes the Review Committee's comments on the study.

The Investigators' Report, prepared by Batterman and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The Critique, prepared by members of the Review Committee with the assistance of HEI staff, places the study in a broader scientific context, points out its strengths and limitations, and discusses remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI's rigorous review process. When an HEI-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. This draft report is first examined by outside technical reviewers and a biostatistician. The report and the reviewers' comments are then evaluated by members of the Review Committee, an independent panel of distinguished scientists who have no involvement in selecting or overseeing HEI studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Critique reflects the information provided in the final version of the report.

PREFACE

HEI's Research Program to Improve Assessment of Exposure to Traffic-Related Air Pollution

INTRODUCTION

Traffic emissions are an important source of urban air pollution. Emissions from motor vehicles and ambient concentrations of most monitored traffic-related pollutants have decreased steadily over the last several decades in most high-income countries as a result of air quality regulations and improvements in vehicular emission control technologies, and this trend is likely to continue. However, these positive developments have not been able to fully compensate for the rapid growth of the motor vehicle fleet due to growth in population and economic activity and increased traffic congestion, as well as the presence of older or malfunctioning vehicles on the roads.

In 2010, HEI published Special Report 17, *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. The report "identified an exposure zone within a range of up to 300 to 500 m from a major road as the area most highly affected by traffic emissions (the range reflects the variable influence of background pollution concentrations, meteorologic conditions, and season)" and estimated that 30% to 45% of people living in large North American cities reside within this zone. Based on a review of health studies, the report concluded that exposure to traffic-related air pollution was causally linked to worsening asthma symptoms. It also found "suggestive evidence of a causal relationship with onset of childhood asthma, nonasthma respiratory symptoms, impaired lung function, total and cardiovascular mortality, and cardiovascular morbidity" (HEI 2010).

Special Report 17 also noted that exposure assessment of traffic-related air pollution is challenging because it is a complex mixture of pollutants in particulate and gaseous forms, many of which are also emitted by other sources. Traffic–related air pollution is also characterized by high spatial and temporal variability, with the highest concentrations occurring at or close to major roads. Therefore, it has been difficult to identify an appropriate exposure metric that uniquely indicates traffic-related air pollution, and to model the distribution of exposure at a sufficiently high degree of spatial and temporal resolution.

The most commonly used exposure metrics are measured or modeled concentrations of individual pollutants considered to be indicators of traffic-related air pollution (such as nitrogen dioxide or black carbon) and simple indicators of traffic (such as distance of the residence from busy roads or traffic density near the residence).

A range of models — such as dispersion, land-use regression, and hybrid models — has been developed to estimate exposure. Some attempts to account for outdoor air entering buildings and how people spend time outdoors versus indoors have been made to refine such estimates. Many improvements in these exposure models have occurred over time, especially with the advance of geographical information system approaches and the application of more sophisticated statistical methods. However, their usefulness still depends on the model assumptions and input data quality. Few studies have compared the performance of different models and evaluated exposure measurement error and possible bias in health estimations.

To start addressing these issues, HEI issued a Request for Applications in 2013. To inform the development of the RFA, the HEI Research Committee held a workshop in April 2012 with experts in the areas of atmospheric chemistry, pollutant measurements, exposure models, epidemiology, and health assessment in order to discuss and identify the highest priority research questions.

Preface

OBJECTIVES OF RFA 13-1

RFA 13-1, *Improving Assessment of Near-Road Exposure to Traffic Related Pollution*, aimed to solicit studies to improve exposure assessment for use in future work on the health effects of traffic-related air pollution. The RFA had three major objectives:

- Demonstrate novel surrogates of near-road traffic-related pollution, taking advantage of new sensors and/or existing monitoring data.

- Determine the most important variables that explain spatial and temporal variance of near-road traffic-related pollutant concentrations at the personal, residential, and/or community levels, and explain the implications of these for future monitoring, modeling, exposure, and health effects studies.

- Improve inputs for exposure models for traffic-related health studies; evaluate and compare the performance of alternative models to existing models and actual measurements to quantify exposure measurement error.

DESCRIPTION OF THE PROGRAM

Five studies were funded under RFA 13-1 to represent a variety of geographical locations and cover the various RFA objectives; they are summarized below. The study by Batterman and colleagues described in this report (Research Report 202) is the third to be published. In the meantime, HEI has funded additional studies on similar exposure assessment topics. All recent and ongoing exposure assessment studies are included in the Preface Table.

"The Hong Kong D3D Study: A Dynamic Three Dimensional Exposure Model for Hong Kong," Benjamin Barratt, King's College London, United Kingdom. Barratt and colleagues estimated exposure to traffic-related air pollution using a dynamic three-dimensional land-use regression model for Hong Kong, which has many high-rise buildings, resulting in street canyons. Different exposure models were developed with increasing complexity (e.g., incorporating infiltration indoors, vertical gradients, and time–activity patterns) and applied in an epidemiological study to evaluate the potential impact of exposure measurement error in mortality estimates (Research Report 194).

"Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion," Stuart Batterman, University of Michigan, Ann Arbor, Michigan. In the study presented in this report, Batterman and colleagues evaluated the ability to predict traffic–related air pollution using a variety of methods and models, including a line source air pollution dispersion model and sophisticated spatiotemporal Bayesian data fusion methods. The study made extensive use of data collected in the Near-road EXposures and effects of Urban air pollutants Study (NEXUS), a cohort study designed to examine the relationship between near-roadway pollutant exposures and respiratory outcomes in children with asthma who live close to major roadways in Detroit.

"Characterizing the Determinants of Vehicle Traffic Emissions Exposure: Measurement and Modeling of Land-Use, Traffic, Transformation, and Transport," Christopher Frey, North Carolina State University, Raleigh, North Carolina. Frey and colleagues investigated key factors that influence exposure to traffic-related air pollution: traffic and its composition; built environment including road characteristics and land use; and dispersion, transport, and transformation processes. They made extensive measurements of fine particulate matter, ultrafine particles, oxides of nitrogen, and semi-volatile organic compounds in various near-road locations in the Raleigh–Durham area. This study has been completed and, at the time of publication of this volume, was in review.

"Developing Multipollutant Exposure Indicators of Traffic Pollution: The Dorm Room Inhalation to Vehicle Emissions (DRIVE) Study," Jeremy Sarnat, Emory University, Atlanta, Georgia Sarnat and colleagues evaluated novel multipollutant traffic surrogates by collecting measurements in and around two student dormitories in Atlanta and explored the use of metabolomics to identify possible exposure-related metabolites. The DRIVE study made use of a unique emission-exposure setting in Atlanta, on the Georgia Institute of Technology campus, with one dorm immediately adjacent to the busiest and most congested highway artery in the city (with more than 300,000 vehicles per day) and another dorm located farther away (Research Report 196).

Preface

Preface Table. Summary of Recently Completed, Ongoing, and Projected Studies Funded by HEI to Improve Exposure Assessment for Health Studies

Principal Investigator	Title	Study Status
RFA 13-1, *Improving Assessment of Near-Road Exposure to Traffic Related Pollution*		
Benjamin Barratt, King's College London, United Kingdom	The Hong Kong D3D Study: A Dynamic Three Dimensional Exposure Model for Hong Kong	Research Report 194
Stuart Batterman, University of Michigan, Ann Arbor	Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion	Research Report 202*
Christopher Frey, North Carolina State University, Raleigh	Characterizing the Determinants of Vehicle Traffic Emissions Exposure: Measurement and Modeling of Land-Use, Traffic, Transformation, and Transport	In review
Jeremy Sarnat, Emory University, Atlanta	Developing Multipollutant Exposure Indicators of Traffic Pollution: The Dorm Room Inhalation to Vehicle Emissions (DRIVE) Study	Research Report 196
Edmund Seto, University of Washington, Seattle	Evaluation of Alternative Sensor-Based Exposure Assessment Methods	Unpublished report
RFA 17-1, *Assessing Adverse Health Effects of Exposure to Traffic-Related Air Pollution, Noise, and Their Interactions with Socioeconomic Status*		
Payam Dadvand and **Jordi Sunyer**, Barcelona Institute for Global Health (ISGlobal), Spain	Traffic-Related Air Pollution and Birth Weight: The Roles of Noise, Placental Function, Green Space, Physical Activity, and Socioeconomic Status (FRONTIER)	Ongoing
Ole Raaschou-Nielsen, Danish Cancer Society Research Center, Copenhagen, Denmark	Health Effects of Air Pollution Components, Noise and Socioeconomic Status ("HERMES")	Ongoing
Meredith Franklin, University of Southern California, Los Angeles	Intersections as Hot Spots: Assessing the Contribution of Localized Non-Tailpipe Emissions and Noise on the Association between Traffic and Children's Health	Ongoing
RFA 16-1, *Walter A. Rosenblith New Investigator Award*		
Joshua Apte, University of Texas, Austin	Scalable Multi-Pollution Exposure Assessment Using Routine Mobile Monitoring Platforms	Ongoing
RFA 19-1, *Applying Novel Approaches to Improve Long-Term Exposure Assessment of Outdoor Air Pollution for Health Studies*		
Scott Weichenthal, McGill University, Montreal, Canada	Comparing the Estimated Health Impacts of Long-Term Exposures to Traffic-Related Air Pollution Using Fixed-Site, Mobile, and Deep Learning Models	Projected start in 2020
Gerard Hoek, Utrecht University, The Netherlands	Comparison of Long-Term Air Pollution Exposure Assessment Based on Mobile Monitoring, Low-Cost Sensors, Dispersion Modelling and Routine Monitoring-Based Exposure Models	Projected start in 2020
Kees de Hoogh, Swiss Tropical and Public Health Institute, Basel, Switzerland	Accounting for Mobility in Air Pollution Exposure Estimates in Studies on Long-Term Health Effects	Projected start in 2020
Klea Katsouyanni, King's College London, United Kingdom	Investigating the Consequences of Measurement Error of Gradually More Sophisticated Long-Term Personal Exposure Models in Assessing Health Effects: The London Study (MELONS)	Projected start in 2020
Lianne Sheppard, University of Washington, Seattle	Optimizing Exposure Assessment for Inference about Air Pollution Effects with Application to the Aging Brain	Projected start in 2020

*Current study.

"Evaluation of Alternative Sensor-Based Exposure Assessment Methods," Edmund Seto, University of Washington, Seattle, Washington. Seto and colleagues performed an evaluation of novel, low-cost air pollution sensors to characterize traffic-related air pollution in the San Francisco Bay area. They deployed various sensors — including Shinyei particulate matter sensors and Alphasense electrochemical sensors — for an extended period of time. Sensors were colocated with reference monitors to evaluate sensor performance. This study resulted in an unpublished report, which can be obtained by contacting HEI at *pubs@healtheffects.org.*

FURTHER RESEARCH UNDERWAY

The studies funded under RFA 13-1 offer valuable lessons that can be integrated into new epidemiological research on the health effects of traffic–related air pollution. Thus, HEI issued RFA 17-1, *Assessing Adverse Health Effects of Exposure to Traffic-Related Air Pollution, Noise, and Their Interactions with Socioeconomic Status*, seeking studies to assess adverse health effects of short- and/or long-term exposure to traffic-related air pollution. The applicants were asked to consider spatially correlated factors that may either confound or modify the health effects of traffic-related air pollution, most notably, traffic noise, socioeconomic status, and factors related to the built environment, such as presence of green space. Three studies funded under RFA 17-1 are in progress as of the publication of this report (see Preface Table). In addition, HEI funded a related study under the Walter A. Rosenblith New Investigator Award to compare exposure estimates obtained from intensive air pollutant measurement campaigns with Google Street View cars with estimates from more conventional methods.

Subsequently, HEI issued RFA 19-1, *Applying Novel Approaches to Improve Long-Term Exposure Assessment of Outdoor Air Pollution for Health Studies* to address challenges in accurately assigning exposures of pollutants that vary highly in space and time to individuals, and to quantify the influence of exposure measurement error on estimated health risks. At the time of publication of this report, five studies have been selected for funding under RFA 19-1 and are expected to start in the spring of 2020. Three of the studies plan to combine measurements of air pollution from emerging sources — such as satellite data — and diverse exposure assessment approaches to improve exposure assignment in well-established cohorts. Two studies plan to test the added value of incrementally more complex statistical modeling approaches to improving exposure assessment and how this may affect uncertainty in health effect estimates in epidemiological studies.

In addition, since the release of HEI's critical review of the traffic literature in 2010, many additional studies about traffic-related air pollution have been published, and regulations and vehicular technology have advanced significantly. Therefore, HEI is undertaking a new review of the epidemiological literature on selected health effects of long-term exposure to traffic-related air pollution. Further information on these activities can be obtained at the HEI website, *www.healtheffects.org/air-pollution/traffic-related-air-pollution.*

REFERENCES

Barratt B, Lee M, Wong P, Tang R, Tsui TH, Cheng W, et al. 2018. A Dynamic Three-Dimensional Air Pollution Exposure Model for Hong Kong. Research Report 194. Boston, MA: Health Effects Institute.

Health Effects Institute. 2010. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. HEI Special Report 17. Boston, MA:Health Effects Institute.

Sarnat JA, Russell A, Liang D, Moutinho JL, Golan R, Weber RJ, et al. 2018. Developing Multipollutant Exposure Indicators of Traffic Pollution: The Dorm Room Inhalation to Vehicle Emissions (DRIVE) Study. Research Report 196. Boston, MA:Health Effects Institute.

Seto E, Austin E, Carvlin G, Shirai J, Hubbard A, Hammond K, et al. 2018. Evaluation of Alternative Sensor-based Exposure Assessment Methods. Unpublished report. Boston, MA: Health Effects Institute.

Dispersion and Bayesian Models of Traffic-Related Air Pollutants

INTRODUCTION

Traffic emissions are an important source of urban air pollution, and exposure to traffic-related air pollution has been associated with various adverse health effects. However, exposure assessment is challenging because traffic-related air pollution is a complex mixture of particles and gases that varies greatly by location and over time. This variability complicates the development of accurate models of traffic-related air pollution to assess exposure to air pollutants for epidemiological studies, in particular because of small-scale variations within cities. Dr. Stuart Batterman from the University of Michigan and his team aimed to improve estimates of traffic-related air pollution concentrations for use in health studies. They used a systematic approach to apply and test a dispersion model — RLINE — developed by the United States Environmental Protection Agency (U.S. EPA) and novel statistical approaches (called "Bayesian spatiotemporal data fusion models" by the investigators) that combine measurements with concentration estimates generated by RLINE. The long-term goal was to apply and improve existing models that could then be employed in other settings.

APPROACH

The study used data collected as part of NEXUS (Near-road EXposures and effects of Urban air pollution Study), a large study conducted in Detroit to evaluate health effects of air pollution in children with asthma living near major roads. All air pollution data were previously collected for the NEXUS study at central and near-road monitoring sites in 2011–2014 or by measuring concentrations at different distances from roads with a mobile monitoring platform during one week in December 2012 (see Statement Figure).

The investigators employed models of varying computational complexity — RLINE plus five different statistical methods — for particulate matter $\leq 2.5\ \mu m$ in aerodynamic diameter ($PM_{2.5}$), nitrogen oxides (NO_x), carbon monoxide (CO), and black carbon (BC). RLINE was designed to model concentrations of air pollutants by including factors such as traffic volume, meteorology, and other factors that influence how those pollutants spread after being emitted by motor vehicles. First, the investigators evaluated the RLINE model by predicting daily average ambient NO_x, CO, and $PM_{2.5}$ concentrations at five U.S. EPA monitoring sites in the Detroit area. Second, the investigators systematically applied and

What This Study Adds

- The investigators undertook to improve the estimation of air pollution exposure from traffic by applying and testing different statistical models of concentrations near major roads.

- Specifically, they evaluated whether inclusion of predictions from the RLINE model of traffic-related air pollution would improve sophisticated statistical models for potential use in exposure assessment.

- Each model provided different useful information, and inclusion of RLINE improved predictions of the increase in near-road concentrations of $PM_{2.5}$, but not of NO_x, relative to background levels.

- The application of the statistical models was an important contribution. However, the usefulness and generalizability of these models remain limited until they have been evaluated with long-term measurements.

Air Quality Station
Airport Meteorological Site
Mobile Monitoring Areas
State or Interstate Highway
Arterial Road
Local Roads
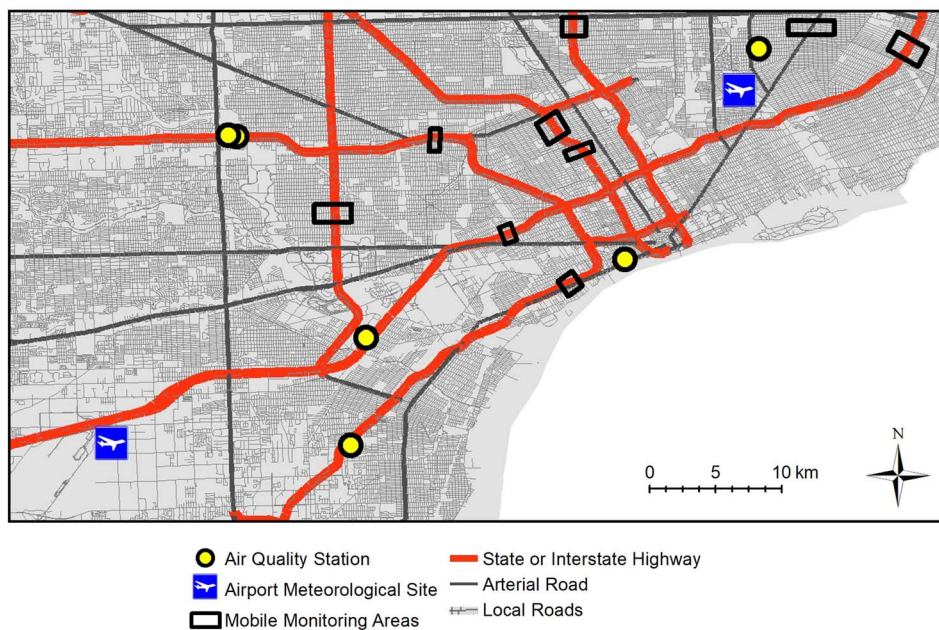
Statement Figure. Map of Detroit showing air quality monitoring stations, airport weather stations, and near-road mobile monitoring locations. (The map is based on Figure 1 of the Investigators' Report and Figure 6 in the Additional Materials, with background layers from Michigan GIS Open Data.)

evaluated the performance of a series of increasingly complex statistical models by including factors such as day of week, upwind versus downwind of the nearest major road, and traffic activity. They also developed a model that predicted both $PM_{2.5}$ and NO_x concentrations together instead of in separate models.

MAIN RESULTS AND INTERPRETATION

In its independent review of the study, the HEI Review Committee concluded that Batterman and colleagues had successfully evaluated the performance of the RLINE model, as well as the performance of universal kriging and sophisticated statistical models that combined RLINE output with measurements. The Committee agreed with the investigators that both RLINE and measurements contributed useful information to the concentration predictions from statistical models. The performance of the RLINE model depended on the pollutant as well as on spatial and temporal factors, such as distance from the nearest major road. In addition, statistical models with different sets of assumptions generally led to the same conclusions and provided complementary information on how the air pollutants were spatially distributed. Finally, adding RLINE to the statistical models or jointly modeling NO_x and $PM_{2.5}$ improved predictions only for $PM_{2.5}$ and not for NO_x.

The Committee thought the statistical models were state of the science and well executed and that the application of the statistical models was a novel and important contribution. They appreciated that the models were systematically compared using a number of performance statistics. On the other hand, the Committee thought that the report may have overstated the usefulness of the models for epidemiological studies for several reasons. First, the models appeared to have limited use over a broad geographic area. Second, the models performed better closer to roads than farther away, which might translate to biased health effect estimates because the exposure predictions would be more accurate for the most highly exposed people in an epidemiological cohort (with participants living at varying distances from major roads). In addition, the uncertainties in the predictions of air pollutant concentrations remained large, even for the most refined models.

There remains a need to further refine the models and distribute these new tools for wider use. In particular, these and similar models will need to be rigorously tested on large databases of measurements collected over long periods before they are used on a large scale in epidemiological studies.

Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion

Stuart Batterman[1], Veronica J. Berrocal[2], Chad Milando[3], Owais Gilani[4], Saravanan Arunachalam[5], and K. Max Zhang[6]

[1]Environmental Health Sciences, and Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan; [2]School of Public Health, University of Michigan, Ann Arbor, Michigan; [3]Department of Environmental Health, Boston University, Massachusetts; [4]Department of Mathematics, Bucknell University, Lewisburg, Pennsylvania; [5]Institute for the Environment at the University of North Carolina, Chapel Hill; [6]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York

ABSTRACT

INTRODUCTION

The adverse health effects associated with exposure to traffic-related air pollutants (TRAPs*) remain a key public health issue. Often, exposure assessments have not represented the small-scale variation and elevated concentrations found near major roads and in urban settings. This research explores approaches aimed at improving exposure estimates of TRAPs that can reduce exposure measurement error when used in health studies. We consider dispersion models designed specifically for the near-road environment, as well as spatiotemporal and data fusion models. These approaches are implemented and evaluated utilizing data collected in recent modeling, monitoring, and epidemiological studies conducted in Detroit, Michigan.

* A list of abbreviations and other terms appears at the end of this volume.

APPROACH

Dispersion models, which estimate near-road pollutant concentrations and individual exposures based on first principles — and in particular, high fidelity models — can provide great flexibility and theoretical strength. They can represent the spatial variability of TRAP concentrations at locations not measured by conventional and spatially sparse air quality monitoring networks. A number of enhancements to dispersion modeling and mobile on-road emissions inventories were considered, including the representation of link-based road networks and updated estimates of temporal allocation of traffic activity, emission factors, and meteorological inputs. The recently developed Research LINE-source model (RLINE), a Gaussian line-source dispersion model specifically designed for the near-road environment, was used in an operational evaluation that compared predicted concentrations of nitrogen oxides (NO_x), carbon monoxide (CO), and $PM_{2.5}$ (particulate matter ≤ 2.5 µm in aerodynamic diameter) with observed concentrations at air quality monitoring stations located near high-traffic roads. Spatiotemporal and data fusion models provided additional and complementary approaches for estimating TRAP exposures. We formulated both nonstationary universal kriging models that exploit the spatial correlation in the monitoring data, and data fusion models that leverage the information contained in both the monitoring data and the output of numerical models, specifically RLINE. These models were evaluated using observations of nitric oxide (NO), NO_x, black carbon (BC), and $PM_{2.5}$ monitored along transects crossing major roads in Detroit. We also examined model

assumptions, including the appropriateness of the covariance functions, errors in RLINE outputs, and the effects of jointly modeling two pollutants and using an updated emission inventory.

RESULTS

For CO and NO_x, dispersion model performance was best when monitoring sites were close to major roads, during downwind conditions, during weekdays, and during certain seasons. The ability to discern local and particularly the traffic-related portion of $PM_{2.5}$ was limited, a result of high background levels, the sparseness of the monitoring network, and large uncertainties for certain sources (e.g., area, fugitive) and some processes (e.g., formation of secondary aerosols). Sensitivity analyses of alternative meteorological inputs and updated emission factors showed some performance gain when using local (on-site) meteorological data and updated inventories. Overall, the operational evaluation suggested RLINE's usefulness for estimating spatially and temporally resolved exposure estimates. The application of the universal kriging models confirmed that wind speed and direction are important drivers of nonstationarity in pollutant concentrations, and that these models can predict exposure estimates that have lower prediction errors than do stationary model counterparts. The application of the Bayesian data fusion models suggested that the RLINE output had a spatially varying additive bias for NO_x and $PM_{2.5}$ and provided little additional information for NO_x, besides what is already contained in traffic and geographical information system (GIS) covariates, but had improved estimates of $PM_{2.5}$ concentrations. Results of the nonstationary Bayesian data fusion model that used RLINE output across a field spanning the measurement sites were similar to a regression-based Bayesian data fusion approach that used only RLINE output at the monitoring locations, with the latter being computationally less burdensome. Using the regression-based Bayesian data fusion model, we found that RLINE with the updated emission inventory provided results that were more useful for estimating NO_x concentration at unmonitored sites, but the updated emission inventory did not improve predictions of $PM_{2.5}$ concentrations. Joint modeling of NO_x and $PM_{2.5}$ was not useful, a result of differences in RLINE's utility in predicting $PM_{2.5}$ and NO_x — useful for the former, but not for the latter — and differences in the spatial dependence structures of the two pollutants. Overall, information provided by RLINE was shown to have the potential to improve spatiotemporal estimates of TRAP concentrations.

CONCLUSIONS

The study results should be interpreted and generalized cautiously given the limitations of the data used. Similar analyses in other settings are recommended for confirming and extending our findings. Still, the study highlights considerations that are relevant for exposure estimates used in health studies. The ability of a dispersion model to accurately reproduce and predict a pollutant depends on the pollutant as well as on spatial and temporal factors, such as the distance and direction from the road, time-of-day, and day-of-week. The nature and source of exposure measurement errors should be taken into consideration, particularly in health studies that take advantage of time–activity information that describes where and when individuals are exposed to pollution. Efforts to refine model inputs and improve model performance can be helpful; meteorological inputs may be the most critical. For both dispersion and spatiotemporal statistical models, sufficient and high-quality monitoring data are essential for developing and evaluating these models. Our analyses using Bayesian data fusion models confirm the presence of spatially varying errors in dispersion model outputs and allow quantification of both the magnitude and the spatial nature of these errors. This valuable information can be leveraged in health studies examining air pollution exposure as well as in studies informing regulatory responses.

INTRODUCTION

Starting in the 1990s, population-based observational epidemiological studies began to associate residential proximity to busy roads, and sometimes to truck traffic, with adverse respiratory health outcomes such as asthma exacerbations, lung function decrements, and hospital admissions (Brunekreef et al. 1997; English et al. 1999; Pönkä 1990; Wjst et al. 1993). Since these and other early investigations, literally hundreds of studies on TRAP have been conducted and published. In 2010, HEI's critical review of the literature concluded that exposure to TRAP was causally linked to worsening asthma, possibly sufficiently causally linked to incident childhood asthma, and suggestively linked to adult-onset asthma, deterioration of lung function, cardiovascular death, myocardial infarction, and atherosclerosis progression (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010). Subsequently, causal links have been established between exposure to TRAP and incident asthma (Anderson et al. 2013) and between diesel engine exhaust and lung cancer (International Agency for Research on Cancer [IARC] 2014). Links have been suggested for many other outcomes, for example,

adverse pregnancy outcomes (Stieb et al. 2016), childhood cancer (Heck et al. 2013), neurological and cognition effects (Woodward et al. 2015), accelerated aging-related declines in physical ability (Weuve et al. 2016), shorter sleep duration (Fang et al. 2015), and heart rate variability (Adar et al. 2007). The associations have been attributed to ambient exposure to high levels of TRAP near major roadways and have been corroborated by other epidemiological studies that have found weak or no association between ambient exposure to background concentrations of pollutants and adverse health outcomes (Hoek et al. 2002; Lewis et al. 2005; Venn et al. 2000). While the progression of emission standards has lowered exhaust emissions of many pollutants, growth in vehicle-kilometers-traveled (Office of Highway Policy Information 2014), increased urbanization of populations worldwide, and strengthened evidence regarding the health significance of air pollutants at even low concentrations suggest that TRAP exposure will remain a major public health concern.

Individuals living in urban areas, especially those living near major roads, may be exposed to high levels of TRAP. Elevated exposures are found in the near-road zone that extends to distances of roughly 100 to 500 meters from major urban roads and major highways, respectively; levels decline rapidly with distance from the road (Baldwin et al. 2015; Brauer et al. 2013; HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010; Karner et al. 2010; Shi et al. 1999; Zhu et al. 2002a,b). This zone frequently contains large and potentially susceptible populations. For example, an estimated 11.3 million individuals in the United States live within 150 meters of a major highway (Boehmer et al. 2013); 40 million live within 100 meters of a four-lane highway, railroad, or airport (U.S. Census Bureau 2007); and many schools and other facilities housing potentially vulnerable populations are near major roads (Wu and Batterman 2006). Residence location is an important determinant of TRAP exposure. Other factors include ambient concentrations of TRAP, the amount of time spent indoors and outdoors, building and vehicle-cabin air exchange and pollutant penetration rates, and breathing rates (Özkaynak et al. 2013). TRAP itself is a heterogeneous mixture, as summarized in Appendix 1 (see Additional Materials on the HEI website) of this report.

A critical element, and a major weakness in many epidemiological studies that have analyzed the health effects of TRAP exposure, is the assignment of pollutant exposure to study participants. Given the cost of air quality measurements, monitoring stations are spatially sparse and only a few pollutants are measured. Biological monitoring is rarely feasible or practical. Thus, exposure to study participants must be assigned using approaches that estimate pollutant concentrations at *unsampled* locations, typically residences or workplaces of participants. A wide variety of approaches has been used to derive concentration or exposure estimates (Appendix 2; see Additional Materials on the HEI website). Improved estimates of TRAP exposure can minimize exposure measurement and misclassification errors that can adversely affect results of epidemiology studies, thus increasing the accuracy of risk and disease burden projections in health impact studies, and better identify affected populations in environmental justice studies.

ROADMAP TO THIS REPORT

The Introduction provides background on the nature, causes, and implications of exposure measurement errors pertaining especially to TRAPs in epidemiological applications.

The Methods and Study Design and the Results sections sequentially describe the formulation, application, and evaluation of dispersion models and of spatiotemporal and Bayesian data fusion models for estimating exposure. The model applications build, in part, on work performed for the Near-road EXposures and effects of Urban air pollutants Study (NEXUS) exposure assessment and epidemiological studies conducted earlier in Detroit, Michigan. Dispersion modeling features the use of RLINE, a new line-source dispersion model developed by the U.S. Environmental Protection Agency (U.S. EPA) specifically for near-road applications (*https://cmascenter.org/r-line/*). We describe an operational evaluation of this model using a highly detailed link-based emission inventory and daily observations of NO_x, CO, and $PM_{2.5}$ concentrations measured at fixed sites in Detroit over the 2011 to 2014 period. These measurements represent the best dataset available for this application given the duration, completeness, and relevance of these data for TRAPs.

Additionally, sensitivity analyses are presented for critical model parameters, including different meteorological datasets that contrast site-specific versus airport data, and for receptor networks that contrast a population-weighted sample with a network for a vulnerable and susceptible population, the latter using the homes and school locations of children in the NEXUS. We then describe the formulation and evaluation of several spatiotemporal statistical models, including nonstationary universal kriging models, joint Bayesian data fusion models, and regression-based Bayesian data fusion models; these models are designed to test different approaches and assumptions. These applications focus on NO_x and $PM_{2.5}$ observations measured in a highway transect study in Detroit, along with observations from the fixed-site

monitors. These data provide the spatial resolution necessary for the spatiotemporal models. In addition, the data fusion models utilize RLINE dispersion model predictions and the same modeling framework just described. The performance evaluation of both dispersion and statistical models uses a variety of statistical measures and sensitivity analyses.

The Discussion and Conclusion sections highlight the strengths, limitations, lessons learned, and implications for epidemiological studies. The report is supplemented with 11 appendices (see Additional Materials available on the HEI website) that provide significant depth on RLINE and the spatiotemporal statistical models, data and methods used in the analysis, as well as general descriptions of TRAP and other models and approaches for exposure assessment.

EXPOSURE MEASUREMENT ERRORS

Exposure estimates should minimize *exposure measurement error*, defined as the difference between the measured or predicted exposure used in the analysis compared to the underlying true exposure. *Exposure misclassification* is the analogous term for a categorical exposure variable. Importantly, such errors can lead to incorrect inference in epidemiological studies; specifically, to biased and/or imprecisely estimated effect coefficients that can invalidate inferences regarding the effect of pollution on health (Carroll et al. 2006; Sheppard et al. 2012).

Exposure measurement errors can be classified either as *Berkson-like errors*, which originate when only part of the true exposure or aggregated exposure is measured, or *classical measurement errors*, which arise when the true exposure is measured with noise. While Berkson-like errors in the exposure measurements lead to unbiased but more uncertain and variable health effect estimates, classical measurement errors can cause the effect estimates to be biased, with standard errors that can be either larger or smaller than they would be in the case of no exposure measurement error. When exposure measurement error is due to the spatial misalignment between the monitoring data and the residential locations of the subjects, the resulting errors are a combination of Berkson-like errors (e.g., from predicting the true exposure surface with an invariably smoother one) and *classical-like measurement errors* (e.g., from noise in the observed concentrations that is not independent of exposure) (Szpiro and Paciorek 2013).

Additional types of errors are at play when dealing with TRAP exposure estimates. Specifically, these include:

- errors related to the *modifiable areal unit problem*, which results in biases when point-based measures

(e.g., concentrations measured or predicted at a site or model receptor) are aggregated into districts (e.g., census tracts or other arbitrary spatial boundary), which can lead to the *ecological fallacy* (Shafran-Nathan et al. 2017);

- *pure spatial location errors* associated with geocoding (e.g., identifying true locations of participants' homes and schools) (Zhang et al. 2016b);

- *location-based covariate measurement errors* (e.g., distance from highways) (Ganguly et al. 2015);

- differences between *ambient and personal exposures* (Kioumourtzoglou et al. 2014); and

- *exposure timing errors* (e.g., using current exposure conditions when evaluating chronic exposures or diseases that may develop over years to decades, or not considering the diurnal pattern of traffic) (Lipfert and Wyzga 2008).

Health studies likely involve the presence of multiple error types, with potentially important consequences regarding study outcomes as well as implications for the development of policies and regulations (Adam-Poupart et al. 2014; Dionisio et al. 2016; Jerrett et al. 2010; Shafran-Nathan et al. 2017; Sheppard et al. 2012; Szpiro et al. 2011).

EPIDEMIOLOGICAL STUDY DESIGNS AND APPLICATIONS

Accounting for exposure measurement error and deriving more accurate exposure metrics, are broadly beneficial for all epidemiological studies, but they are particularly important for epidemiological studies examining populations where TRAP exposures are expected to vary over time and space. This includes many types of cohort and longitudinal study designs. For example, NEXUS examined the relationship between near-roadway exposures to air pollutants and respiratory outcomes in children with asthma living near major roadways in Detroit, Michigan (Vette et al. 2013). Children in NEXUS were recruited into three groups based on residence location: children living within 175 meters of major roads, with this group divided into roads with low and high volumes of diesel trucks, and children living more than 500 meters from major roads. The NEXUS cohort is an example of susceptible and vulnerable populations, which are often the focus of epidemiological studies. Susceptibility generally refers to intrinsic factors that tend to intensify the biological response from exposure to a stressor, asthma in this case, while vulnerability refers to extrinsic factors that can increase exposures or reduce the ability to mitigate them (e.g., poverty or proximity to major roads) (O'Neill et al.

2012; Sacks et al. 2011). Health outcome measures in NEXUS were obtained on a seasonal basis over 14-day periods and included daily measures of pulmonary function, medication and health care use, diary reports of upper respiratory infection symptoms, fraction of exhaled NO, missed school days, sleep quality, and other aspects. In such studies, each subject's exposure is expected to vary temporally with diurnal, daily, weekly, and seasonal effects. These temporal variations are driven principally by variation in traffic patterns, temperature-dependent emission factors, meteorology, and time–activity factors. In addition, exposures will vary spatially, reflecting differences between locations where individuals spend most of their time, primarily home and school for children.

In NEXUS, a hybrid air quality modeling approach obtained quantitative, daily, and individual-level exposure estimates. The approach combined two dispersion models — AERMOD (*www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models*) (Cimorelli et al. 2004) and RLINE-source dispersion, which modeled local non-road (point and area) and road sources, respectively — local emission information, detailed road network information and traffic activity, local meteorological data, and a combination of the Community Multiscale Air Quality (CMAQ) model and space–time ordinary kriging models to estimate background concentrations of pollutants (contributions of pollutants from distant sources [Isakov et al. 2014]). While state-of-the-art, this exposure assessment framework utilized a number of deterministic and physically based (or process) models with results that may diverge from measured concentrations. As a result, NEXUS included a number of sensitivity analyses and performance evaluations to understand model performance (Heist et al. 2013; Milando and Batterman 2018a,b; Snyder et al. 2013a), an important step in applying model-derived exposure estimates in epidemiological studies. This report utilizes and extends the NEXUS modeling framework.

Considerations of spatial and temporal variability apply to other types of study designs. For example, in case–crossover studies examining short-term associations between exposure TRAPs and morbidity and/or mortality, the temporal variation of pollutant levels or traffic volumes may be aggregated across the study population (e.g., if a single monitoring site is used to represent pollutant levels). More typically, however, some approach is needed to account for spatial differences; for example, the distance between residence location and major roads may be used to select or group individuals in the study. However, such approaches may provide only qualitative exposure metrics, and the representation of spatial variation may be highly simplified.

ISSUES AND CHALLENGES IN DEVELOPING EXPOSURE METRICS FOR TRAPS

Several important issues and challenges regarding the development of exposure metrics for TRAPs are reviewed; these shaped the nature of modeling used in this report.

- *Desired spatial resolution* of exposure metrics ranges from point estimates at discrete locations (e.g., residence locations or dispersion model receptors) to large zonal aggregations or districts (e.g., ZIP code, census tract, and census block). Larger zonal units are poorly suited for TRAP investigations due to the spatial mismatch of concentration gradients and zone sizes; based on a modeling analysis, interpolations of TRAP exposure should be limited to not more than 40 meters near major roads and 100 meters at larger distances from major roads (Batterman et al. 2014b). This topic is further explored in Appendix 1.

- *Averaging times* should take into account the dynamics of biological processes governing the health outcomes, the diurnal and seasonal patterns of TRAP emissions, the time constants pertaining to pollutant entry and fate in exposure compartments (e.g., buildings and vehicle cabins), and the time–activity patterns of study participants governing exposure (e.g., locations and breathing rates at home, work, and while commuting). Averaging times must also consider the frequency of air quality measurements (typically hourly to daily) and the limited representativeness of short-term (e.g., hourly) wind fields in urban areas that drive dispersion-modeling-based predictions (Chang and Hanna 2004). Most studies use daily to annual averaging times.

- *Time–activity patterns* bring together the spatial and temporal factors affecting TRAP exposure. Few studies have utilized the microcompartmental exposure estimates needed to represent activity–travel patterns and dwell times spent at different locations (Chang et al. 2015; Dons et al. 2014; Gurram et al. 2015; Yu and Stuart 2016).

- *Variability and uncertainty* of emissions of TRAPs include large time-of-day, day-of-week, seasonal, and multiyear changes that affect emission rates, phase, chemical composition, and particle sizes of both exhaust and non-exhaust components of TRAP.

- *Limited measurements of ambient TRAPs* are due to the spatial sparseness of the monitoring network, the presence of high background levels for many pollutants

that make it difficult to distinguish TRAP from other sources of pollutants (especially for $PM_{2.5}$), and the lack of unique or cost-effective tracers of TRAP.

- *Complex micrometeorology* in near-road and urban environments is a result of influences of sound walls, road characteristics, nearby buildings, vehicle-induced turbulence, and other factors (Hanna and Chang 2012), which may not be reflected by typical airport observations.

- *Data gaps and complexity of physically based models* for predicting concentrations of TRAPs. While in theory, dispersion models can generate estimates of near-road concentrations at high spatial and temporal resolutions based on first principles, these models require extensive input data. The accuracy of their predictions, as well as the uncertainty in such predictions, particularly in urban settings, have not been well characterized (Claggett et al. 2009; Colvile et al. 2002; Hanna 2007; Jerrett et al. 2005; Rao 2005).

Such issues have led to a variety of approaches for estimating TRAP exposures. They include: the use of air quality monitoring data (ambient fixed site, ambient mobile, in-cabin, indoor, personal); biomonitoring measurements; surrogates such as residential proximity to high-traffic roads and traffic intensity measures; land use regression models; source-oriented dispersion (or simulation) models; spatiotemporal modeling; data fusion statistical models; and hybrid methods that combine several approaches. A number of these approaches have been reviewed and compared (Batterman et al. 2014a; Baxter et al. 2013; Dionisio et al. 2016; Hannam et al. 2013; HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010; Hoek et al. 2008; Huang and Batterman 2000; Jerrett et al. 2005; Lipfert and Wyzga 2008; Martenies et al. 2015; Patton et al. 2017; Wu et al. 2011). Further details are presented in Appendix 2. The methods differ with respect to their data demands, feasibility, cost, intended applications, strengths, and weaknesses.

As summarized above, concentration estimates of TRAPs used in health effect studies should incorporate sufficient spatial and temporal resolution to reflect conditions near major roads. This report evaluates the use of dispersion models, which *predict concentrations at unmonitored sites* based on physical mechanisms (e.g., emissions and dispersion), and spatiotemporal statistical models, which represent *observed dependencies in concentrations measured at different sites* as functions of spatial, meteorological, and other covariates. These approaches are considered both separately and together with the goal of improving exposure estimates for epidemiological studies. The combined approach, using Bayesian data fusion models, is potentially valuable given the spatial sparseness of ambient monitoring networks — which rarely represent locations of susceptible and vulnerable populations — and the data demands, uncertainty, and possible systematic biases in dispersion modeling.

SPECIFIC AIMS

This report investigates ways to improve estimates of TRAP concentrations for use in health effect studies, with specific attention to dispersion modeling and spatiotemporal statistical methods that can provide the spatial and temporal resolution needed to accurately determine near-road exposures. The specific aims are to:

1. explore potential enhancements for dispersion models, including alternate treatments of meteorological inputs, background levels, and traffic inputs;

2. assess the performance of dispersion models for predicting concentrations of TRAPs in a full-scale urban case study, including identification of critical inputs and uncertainties; and

3. apply spatiotemporal and Bayesian data fusion statistical techniques for combining dispersion model outputs and pollutant monitoring observations.

These aims are motivated by the need to improve estimates of exposure to TRAPs, particularly in the near-road zone extending up to 300 to 500 meters or possibly more from major roads, and to provide the spatial, temporal, and source resolution needed for studies examining asthma, cardiovascular, pregnancy, and other important health outcomes.

METHODS AND STUDY DESIGN

DISPERSION MODELING

This section presents the approach used to develop and evaluate dispersion model predictions of TRAPs in the Detroit application. We perform an *operational evaluation* relevant to health studies, investigating whether model estimates agree with observations in an overall sense. Routine observations of pollutant concentrations, emissions, meteorology, and other variables are utilized with the goal of characterizing prediction uncertainties and limitations of models for particular applications (Dennis et al. 2010). Daily average concentrations of NO_X, CO, and $PM_{2.5}$ measured at sites across Detroit for the 2011 to 2014 period are compared to dispersion model predictions. Performance is evaluated by pollutant, site, wind speed, meteorological

condition, averaging time, and other factors. The evaluation utilizes a state-of-the-science modeling system that includes an updated link-based roadway inventory, an updated point-source inventory; the MOVES2014b emission factor model (*www.epa.gov/moves/latest-version-motor-vehicle-emission-simulator-moves*) with monthly-adjusted fuel characteristics and ambient temperature in Detroit, and hourly-adjusted link-specific volumes. Local point and on-road mobile sources are modeled using AERMOD and RLINE, respectively. The modeling domain uses a 40 × 30 kilometer region, covering portions of Macomb, Oakland, and Wayne counties in Michigan. Hourly pollutant concentrations were modeled from 2011 (the first complete year of near-road monitoring in Detroit) to 2014 (the most recent year of point-source emission inventory data).

The evaluation compares observed and predicted concentrations using a 24-hour averaging period, an exposure metric frequently used in epidemiological and health impact studies; this averaging period is also supported by previous evaluations suggesting that meteorological variability makes comparisons at the hourly level "almost fruitless" (Chang and Hanna 2004). Analyses were conducted by pollutant, wind direction, monitoring site, season, and day-of-week. Wind directions were defined for wind speeds exceeding 1 m/s, and monitoring sites were considered to be *downwind* for directions within ±30° of perpendicular of the largest road near each site, and *parallel* for directions within ±15° of parallel (Venkatram et al. 2013). Daily average downwind or parallel concentrations were calculated for those hours of each (calendar) day that met these conditions if a minimum of 6 hours of valid model–observation pairs was available. Periods with fewer than five valid days were not considered. Sensitivity analyses were also performed for critical model parameters, for example, meteorological data. We also contrasted concentrations predicted for two receptor grids: one designed to represent a population-weighted sample and one representing a vulnerable and susceptible group using the homes and school locations of children in the NEXUS study (Vette et al. 2013), two-thirds of whom lived within 175 meters of major roads.

The evaluation emphasized four metrics following air quality model evaluation guidelines (Chang and Hanna 2004; Hanna and Chang 2012). The F2 statistic (percentage of modeled values within a factor of 2 of observed values) shows over- and underpredictions and provides a measure of overall model performance. The Spearman rank correlation coefficient (R_{SP}) assesses the similarity between ranked observations and predictions and may be particularly appropriate for epidemiological studies as it can indicate whether

exposures are correctly classified. The *fractional bias* (FB) — defined as $FB = \left(\overline{C_p} - \overline{C_o} \right) / \left[\left(\overline{C_p} + \overline{C_o} \right) / 2 \right]$, where $\overline{C_p}$ and $\overline{C_o}$ are mean predicted and observed concentrations, respectively — shows the tendency to over- or under-predict, in other words, the likelihood of false positives or false negatives. (Equal weight is given to under- and over-estimates.) The *geometric variance* (V_G), defined as $\exp \left[\overline{\left(\ln C_o - \ln C_p \right)^2} \right]$, indicates the irreducible (*systematic*) and reducible (*random*) errors. This metric can help identify conditions where performance potentially could be improved; in other words, the percentage of errors that are reducible (% Red) is the ratio between the natural logarithm of the reducible component of V_G and the total V_G (the product of the systematic and random components). Minimum performance criteria suggested for air quality models are F2 ≥ 50%, mean bias ≤ 30%, and V_G ≤ 1.6 (Chang and Hanna 2004). We also tabulate R^2 and mean standard error metrics.

Given the number of comparisons made, several rules were used to identify potentially meaningful differences and produce a summary measure. Each performance metric was compared to its best value (i.e., 1.00 for R_{SP} and V_G, 0.00 for FB and % Red). Results were recorded as whether the nominal model input improved model performance (•), gave results that were among those that improved results (~), did not conclusively improve model performance (' '), or diminished performance (○) (see Appendix 8, Tables 9 and 10; Additional Materials, available on the HEI website). A minimum of at least one set with R_{SP} = 0.1 was required for comparisons to be considered. Only potentially meaningful changes were distinguished, for example, changes in R_{SP} and other metrics had to exceed 0.05, a threshold selected to balance sensitivity and avoid false indications. Comparisons of 2010 (nominal) and 2015 emission factors, and comparisons of the U.S. default temporal allocation factor (TAF) (nominal) to the two alternative TAFs (Detroit-specific with commercial and noncommercial traffic separated and combined) used the same scheme.

Comparisons of the four sets of meteorological inputs were more complex. We checked whether on-site/KDET (Detroit City Airport) meteorology provided the best results (denoted as on-site/KDET highest?); whether KDET data provided better results than KDTW (Detroit Metro Airport) data when using National Weather Service (NWS) data alone or in conjunction with on-site data (KDET > KDTW?), and if on-site data generally improved results over NWS data alone (on-site > NWS?).

RLINE Dispersion Model

Concentrations from on-road mobile sources were predicted using RLINE version 1.2, a research-grade dispersion model developed by the U.S. EPA to support risk assessments and health studies related to near-road pollutants (*www.cmascenter.org/r-line/*). (RLINE and other dispersion models are described in Appendix 3 in Additional Materials on the HEI website.) Like its predecessors, RLINE is based upon a steady-state Gaussian formulation that simulates line-type emission sources (e.g., mobile sources on roadways) by numerically integrating point-source emissions along the line source. RLINE was designed to simulate concentrations at receptors (arbitrarily placed point locations) positioned very near the line source.

The current version (RLINE 1.2) was formulated for near-surface releases in flat terrain (simple terrain without surrounding complexities). It contains new formulations of vertical and lateral dispersion rates based on recent field and wind tunnel studies. The model also simulates low wind meander conditions, includes Monin-Obukhov similarity profiling of winds near the surface, and selects plume-weighted winds for transport and dispersion calculations (*www.cmascenter.org/r-line/*). The current version includes beta-option algorithms for simulating several complex near-source effects, for example, effects of noise and vegetative barriers and depressed roadways (these features have not been evaluated in the peer-reviewed literature). RLINE also provides an analytical approximation (an option to the default numerical integration), which can dramatically speed calculations, although the guidance notes that "this solution includes some simplifying assumptions that lead to slightly different results than the numerical solution, especially for receptors close to the source, or for sources and/or receptors significantly off the ground." RLINE requires hourly values of sensible heat flux, surface friction velocity, convective velocity, convective stable planetary boundary layer heights, Monin-Obukhov length, surface roughness, wind speed, and wind direction, and it utilizes the AERMET meteorological data preprocessor surface to process surface and upper air meteorological datasets for these purposes (U.S. EPA 2004). A simplified version of RLINE, called CLINE (community LINE-source model), is available as a web application (*https://cmascenter.org/c-tools/c-line.cfm*).

RLINE modeling used the numerical integration method, an iteration limit of 1,000 and an error limit of 0.001. Beta modules for roadside barriers and depressed roadways were not used. For large urban scale applications, RLINE is computationally intensive. For example, using the Detroit emissions inventory (described below)

with 9,701 links, hundreds to tens of thousands of model receptors (depending on the application), and long-term simulation (e.g., annual with 8,760 hours), could require run times of many days with a high-speed computer cluster. Moreover, RLINE runs a single hour at a time; multihour runs are normally accomplished by post-processing of RLINE output, which generates extremely large files. Thus, for some analyses, the model was modified to increase efficiency:

- a source-receptor cut-off distance was imposed (i.e., calculations were not performed for receptor-link distances exceeding 4 kilometers — these concentrations were very small);

- precomputed (lookup) tables were used for emission factors, following earlier work that binned emission factors by pollutant, vehicle type, speed, ambient temperature, hour-of-day, and month (Isakov et al. 2014);

- a more flexible input/output scheme was implemented; and

- data checks and other features were incorporated.

These changes made negligible differences in model predictions. The following sections summarize dispersion model inputs, which include the emission inventories, meteorological data, and receptor networks.

Emissions Inventory

In Detroit, we assembled an emissions inventory that contained mobile, point, and area sources. The 2011 National Emission Inventory data for Wayne County, Michigan, which includes Detroit, gives a high-level view of the emissions data. (See Appendix 4 in Additional Materials on the HEI website, especially Appendix Table 2 for details.) In the Detroit area, on-road mobile sources constitute 48% of NO_x emissions and 54% of CO emissions, but only 21% of $PM_{2.5}$ emissions. Area and non-road emissions of $PM_{2.5}$ substantially exceeded on-road mobile sources. Notably, the National Emission Inventory lacks both temporal and spatial information for these area, road, and non-road emissions. As described later, this can restrict the ability of dispersion models to portray small-scale variation in $PM_{2.5}$ concentrations and lead to model evaluation results (i.e., comparisons between dispersion model predictions and observations) that are not informative.

Mobile source emissions inventories suitable for exposure modeling in the near-road environment require spatially and temporally resolved estimates of on-road emissions. We used a *link-level* inventory that provides information for individual road segments or links, which was assembled using a *bottom-up* approach. This starts with the *road network configuration* (location, number of

lanes, depth above/below grade), adds *traffic activity* information (vehicle volume, speed, acceleration, and vehicle mix on each link), and then *emission factors*. Such inventories consolidate data from multiple sources, for example, GIS shape files representing roads, estimates of total vehicle-kilometers-traveled from metropolitan planning organizations, historical traffic measurements and estimates, traffic demand model estimates of vehicle volumes, and other data types. Comparable urban scale link-level mobile source inventories have been assembled for Detroit, Atlanta, Houston, Beijing, and Macau (Huo et al. 2009; Lindhjem et al. 2012; Snyder et al. 2014; Venugopal and Yang 2014; Zhang et al. 2016a).

In addition to on-road mobile sources, other emission sources in the Detroit area were modeled. Appendix 4 describes area and point-source inventories of CO, NO_X, and PM emissions for southeast Michigan (including Lenawee, Livingston, Macomb, Monroe, Oakland, Washtenaw, and Wayne counties) generated for the years 2011 to 2014.

Meteorological Data

Meteorological data for dispersion modeling should be representative of local conditions and thus is normally collected at or near the site of interest, most commonly, at the closest NWS sites. Meteorological surface observations also are collected at air quality monitoring sites, including near-road sites; however, such sites generally measure only a subset of parameters (e.g., wind direction and speed, temperature, humidity, and pressure). These variables are insufficient for running AERMET (Cimorelli et al. 2004) since surface friction velocity (Ustar or U*), convective velocity scale (W*), surface roughness length (Z_0), and other (hourly) parameters are missing. These parameters can be calculated using data collected at NWS stations (NWS 2016) and upper air data stations (National Oceanic and Atmospheric Administration [NOAA] 2016).

Meteorological data were obtained at five air quality system (AQS) sites (described in the next section), two local NWS stations located 33 kilometers apart (KDET and KDTW) (NWS 2016), and the Pontiac, Michigan radiosonde site (approximately 45 km north of Detroit) (NOAA 2016). The NWS datasets include the parameters needed by the AERMET preprocessor to develop the surface (SFC) files used by RLINE (Cimorelli et al. 2004), whereas the AQS sites collect only basic meteorological parameters, for example, surface wind speed and direction. We also obtained meteorological data from the weather research and forecasting (WRF) model, a mesoscale numerical weather prediction system designed to serve both atmospheric research and operational forecasting needs

(*www.wrf-model.org/*). WRF data are continuous in space and time, and can serve as a diagnostic reference providing spatial information (typically at 12-km intervals). The latest version (V3.0) of the Meteorological Model Interface tool (*www.epa.gov/ttn/scram/models/relat/mmif/MMIFv3.1 _Users_Manual.pdf*) was used to extract the meteorological fields from WRF for January and July 2010 for the grid cell containing the downtown Detroit files. Key meteorological variables were compared to AERMET-generated SFC files based on KDET airport data for the same periods. (This analysis was feasible given the availability of WRF data from other projects.)

Procedures recommended in the AERMET User's Guide (U.S. EPA 2004) and AERMET version 14134 were used to create quality-checked site meteorological data in calendar years 2010 to 2012 for screening and sensitivity analysis purposes, and in years 2011 to 2014 for detailed modeling. The NWS data at KDET was designated as the nominal input due to its central location and presumed representativeness (Isakov et al. 2014). Three sets of alternative meteorological inputs were developed: SFC files using NWS data at KDTW; AQS-site-specific meteorology supplemented with KDET data (on-site/KDET); and site-specific meteorology supplemented with KDTW data (on-site/KDTW). The correlation between wind direction measurements across the sites was evaluated using the circular correlation coefficient (Jammalamadaka and Sengupta 2001). Correlations of other meteorological variables used Pearson correlation coefficients. Hours missing any required parameter were excluded. The SFC files were mostly complete, for example, 6% to 15% of hours were missing across the five sites and four years. Impacts of the different meteorological datasets on RLINE predictions were evaluated in sensitivity analyses.

Monitoring Data and Background Concentrations

Ambient air quality monitoring data for CO, NO, NO_2, NO_X, and $PM_{2.5}$ collected at AQS sites in Wayne County were downloaded from the U.S. EPA database (U.S. EPA 2015). Data completeness, detection frequencies, quality assurance checks, maps of the five near-road AQS sites, and other details are given in Appendix 5 (see Additional Materials on the HEI website).

The model performance evaluation requires background concentrations, defined as contributions from both regional sources (outside the modeled area) and local but unmodeled area and mobile sources. The background sources are not explicitly modeled because they are distant, too numerous, or too difficult to simulate (Arunachalam et al. 2014), or because the data are incomplete. For NO_X and CO, background was estimated using a

conditional selection method that subtracted the geometric mean of monthly upwind modeled concentrations due to point and on-road sources from the observed geometric monthly mean concentration (Malby et al. 2013). Missing months were imputed by linear interpolation, and then leave-one-out nearest neighbor linear regressions were performed to obtain a smoothed sequence of monthly background estimates at each monitor. A different approach was used for $PM_{2.5}$, given the variation in daily levels. First, a daily $PM_{2.5}$ dataset was created that consolidated Detroit area daily measurements (9 sites) and 24-hour averages of hourly $PM_{2.5}$ measurements (3 sites using tapered element oscillating microbalances [TEOMs]), then a complete dataset was generated by imputation (50 iterations) using the mice package in R (van Buuren and Groothuis-Oudshoorn 2011). Using this complete dataset, the second lowest concentration at any monitor on each day was selected, and a leave-one-out nearest neighbor linear regression was performed on the imputed lowest values to obtain a smoothed time series of daily levels representing the $PM_{2.5}$ background. These background estimates reflect temporal changes and do not require additional model runs.

SPATIOTEMPORAL AND DATA FUSION MODELING

This section summarizes the development and evaluation of *nonstationary universal kriging models* and *nonstationary data fusion models* for TRAPs in localized near-road environments. We used measurements of NO, NO_x, BC, and $PM_{2.5}$ collected at transect areas surrounding major highways in Detroit, Michigan to investigate the performance of several types of models and applicability of model assumptions, including whether nonstationary covariance functions are appropriate modeling choices for TRAPs. In other words, we investigate whether the spatial correlation between the concentration of TRAPs at any two sites is just a function of their separation (e.g., their distance and the direction of one site with respect to the other), and not a function of where these two sites are actually located. If only the separation matters, the covariance function is stationary, otherwise the covariance is nonstationary. In addition, we develop Bayesian data fusion statistical models that address the following research questions:

1. Is incorporation of RLINE outputs useful for estimating ambient concentrations of TRAPs in near-road urban environments?

2. Does RLINE correctly capture the spatial dependence structure of TRAP concentrations in near-road environments?

3. Do updates in emission inventories translate into predictive improvements of the RLINE output?

4. Do estimations of TRAP concentrations improve if pollutants are modeled jointly versus independently?

In developing our proposed nonstationary covariance model that addresses possible differences in the spatial dependence structure between upwind and downwind pollutant concentrations, we followed and expanded the kernel mixing approach of Fuentes (2001) by incorporating covariate information in the covariance function. However, in contrast to Reich and colleagues (2011), instead of including the covariates only in the weighting kernels, we also included them in the covariance functions of the underlying spatial processes, similar in spirit to Schmidt and colleagues (2011). In particular, based on physical considerations, we incorporated wind speed and direction in the covariance function as key factors that influence the spatial distribution and variability of TRAPs. We evaluated the appropriateness of this assumption by comparing the predictive performance of nonstationary and stationary spatial statistical models. We first describe the data sources, followed by the statistical models.

Ambient Monitoring and Other Data

Ambient pollutant data were collected using a Mobile Air Pollution Lab (MAPL), a recreational vehicle equipped with a variety of air quality monitoring instruments, along nine transects that crossed major roadways in Detroit, Michigan, on seven consecutive days (December 14–20, 2012) during morning and afternoon rush-hour periods (Baldwin et al. 2015). No afternoon observations were taken on December 14 and December 20, 2012, and no morning observations were taken on December 16, 2012. Figure 1 shows the locations of the transect areas. In areas 1 through 8, sampling sites were located at a nominal distance of 50 meters (two sites), 150 meters (two sites), and 500 meters (one site) from both edges of the road. In area 9, six sampling sites were used at distances of 50, 150, and 500 meters from each edge of the road. Areas 1 through 8 were monitored during both morning (07:15–09:45) and afternoon (15:45–18:15) rush-hour periods, while area 9 was monitored once or twice each day, just after the morning rush hour, or before the afternoon rush hour. On any given day, up to three different areas were monitored. The MAPL visited each site and conducted measurements for 5 minutes before proceeding to the next site. The vehicle engine was turned off during measurements. Air was sampled at a 3.5 meter height. NO and NO_x concentrations were measured using a conventional federal reference monitor (Model 42i, Thermo, MA, USA); BC was measured using a two wavelength aethalometer (Model AE42, Magee Scientific, Berkeley, CA, USA); and particle number
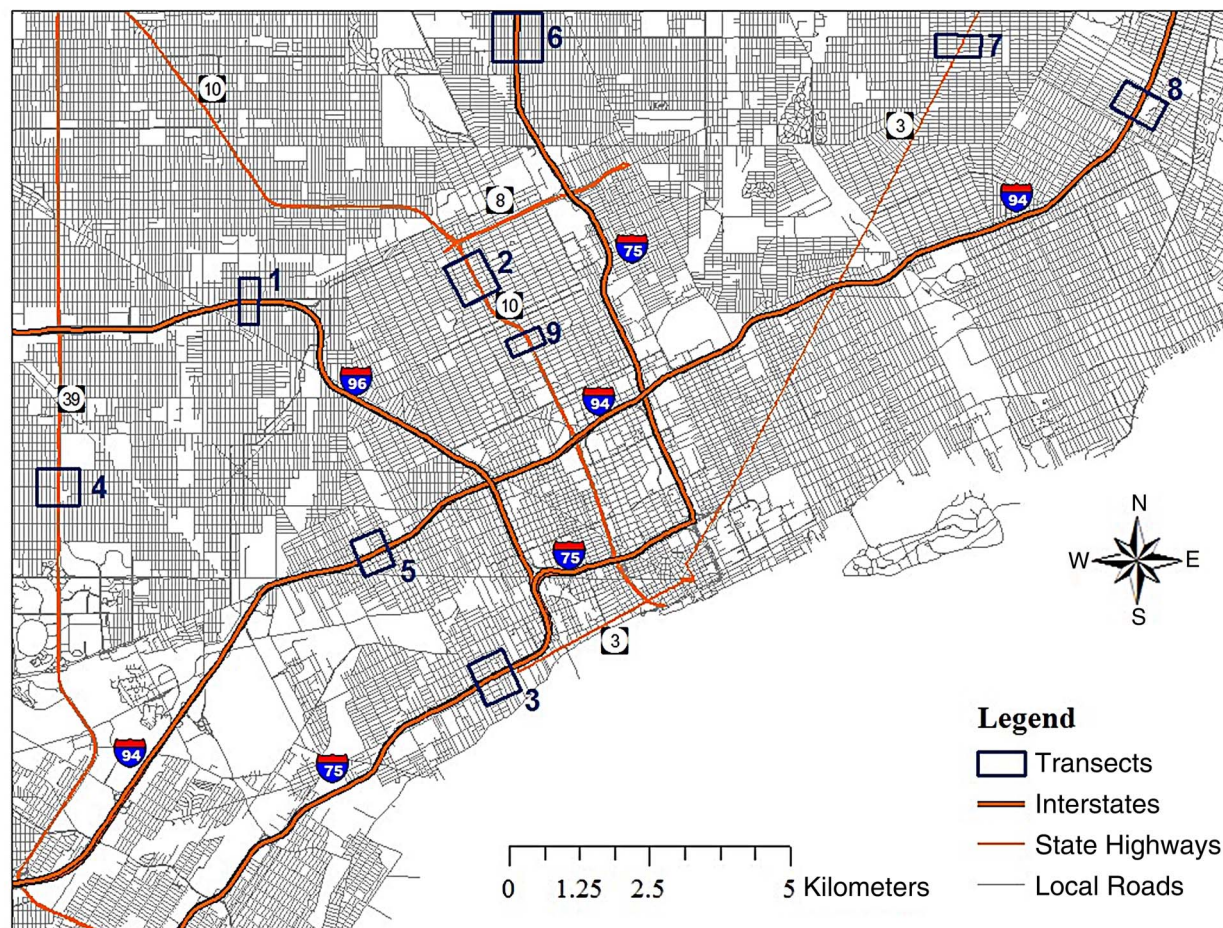
Figure 1. Map showing transect areas (boxes 1–9) and major roads in the study area. Two transects were used in each area except area 9 (single transect). (Reprinted from Baldwin et al. 2015 by permission of Elsevier.)

concentrations in multiple-size bins were measured with a GRIMM Model 1.109 Spectrometer (Grimm Aerosol Technik, Germany) and a Fast Mobility Particle Sizer Spectrometer (Model 3091, TSI Incorporated, Shoreview, MN, USA). Particle number concentrations were converted to mass concentrations (Grimm and Eatough 2009) assuming a particle density of 1.67 g/cm^3. PM$_{2.5}$ concentrations were calculated as the sum of mass concentrations for particles with a diameter less than or equal to 2.5 μm.

Across the nine areas, 5-minute average concentrations were collected for NO and NO$_x$ (a total of 286 measurements), BC (277 measurements), and PM$_{2.5}$ (235 measurements). As an example of the collected data, Figure 2

shows NO and BC concentrations at area 8 sites on the morning of December 20, 2012.

Besides the actual measured concentrations of NO$_x$ and PM$_{2.5}$ at monitoring sites, we also derived the NO$_x$ and PM$_{2.5}$ *near-road increment* (NRI) at the monitoring sites. These are defined as the difference between observed and background pollutant concentrations. For a transect area, we took as background concentration the lowest pollutant concentration observed in that area during the monitoring period. Using this definition, across the nine areas, a total of 254 5-minute average NRI concentrations were collected for NO$_x$, and 235 for PM$_{2.5}$. In creating the working dataset used for the Bayesian data fusion statistical models, we
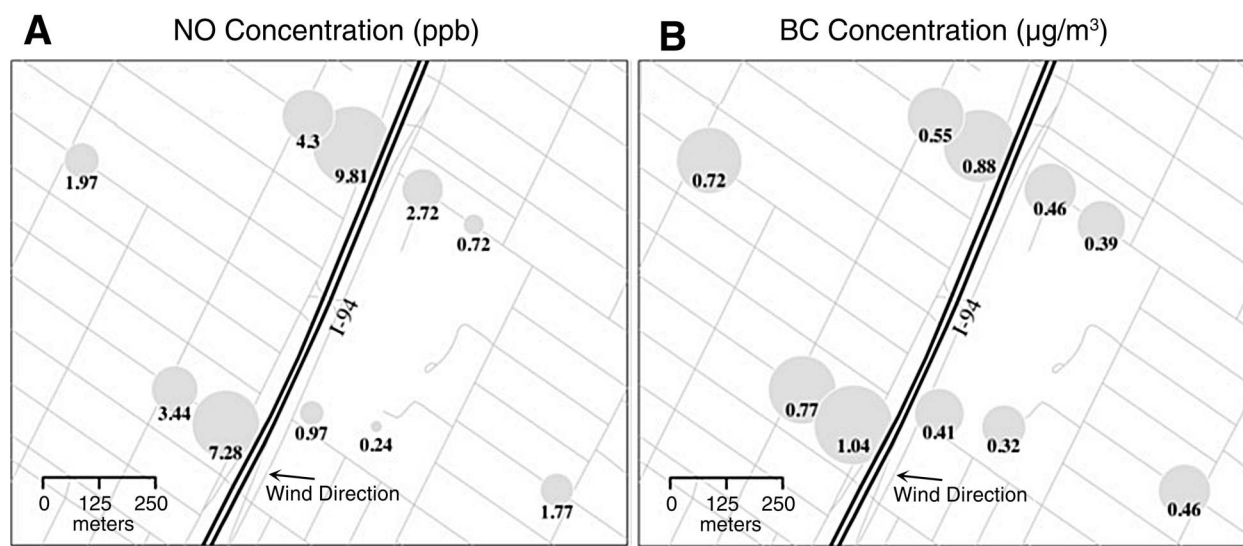
Figure 2. Maps of transect area 8 showing (A) NO concentrations (ppb) and (B) BC concentrations (µ*g*/m³). Measurements for each were taken at 10 sites on the morning of December 20, 2012. The area of the bubble is proportional to the concentration.

restricted ourselves to time periods (e.g., days and periods of the day — morning versus afternoon) for which RLINE output was available. In some time periods, the RLINE model could not be run due to unavailability of some inputs. The final number of monitoring NRI data used was 254 NO_x and 235 $PM_{2.5}$ values at the MAPL sites. Figure 3 shows the NRI for NO_x and $PM_{2.5}$ at sites within area 5 on the morning of December 18, 2012. These quantities will be used in Bayesian data fusion statistical models that combine monitoring observations with the output of the RLINE dispersion model, also referred to as the RLINE predictions (described below).

RLINE Predictions

Concentrations of NO_x and $PM_{2.5}$ in the near-road environment attributable to on-road vehicular traffic were modeled by the dispersion model RLINE (Snyder et al. 2013b), described earlier. The dispersion model used the spatially and temporally resolved link-based emissions inventory, the initial (2010) and updated (2012) emissions factors from the Motor Vehicle Emission Simulator (MOVES) that depend on vehicle class, speed, ambient temperature, and other factors, and the Detroit-specific hourly temporal allocation factors that separated commercial and noncommercial traffic patterns (Batterman et al. 2015).

Multiple runs of the RLINE model were used. Initially, RLINE model estimates of average hourly pollutant concentrations for the time periods corresponding to MAPL data collection were obtained at 96 regularly spaced point locations (receptors) within a 2-km square centered on the major road at each transect area. This modeling used the 2010 emission inventory. Given the 96 receptors, 9 areas, and 3 to 4 periods monitored at each area, a total of 3,038 1-hour average NRI concentrations were available for both NO_x and $PM_{2.5}$. Figure 3 provides an example of MAPL sites and RLINE receptors for NO_x and $PM_{2.5}$ for one sampling event. This first set of RLINE output is called *RLINE output set 1*. Because RLINE modeling reflects concentrations due only to local traffic emissions, modeling outputs are considered equivalent to the NRI. To determine whether the updated emission inventory yielded an improvement in the RLINE output, additional RLINE outputs were generated that estimated NO_x and $PM_{2.5}$ concentrations at receptors placed exactly at the MAPL sites. The set of RLINE runs using the original 2010 emission inventory is called *RLINE output set 2 – 2010*, and the set of otherwise identical RLINE runs using the updated inventory is called *RLINE output set 2 – 2012*.
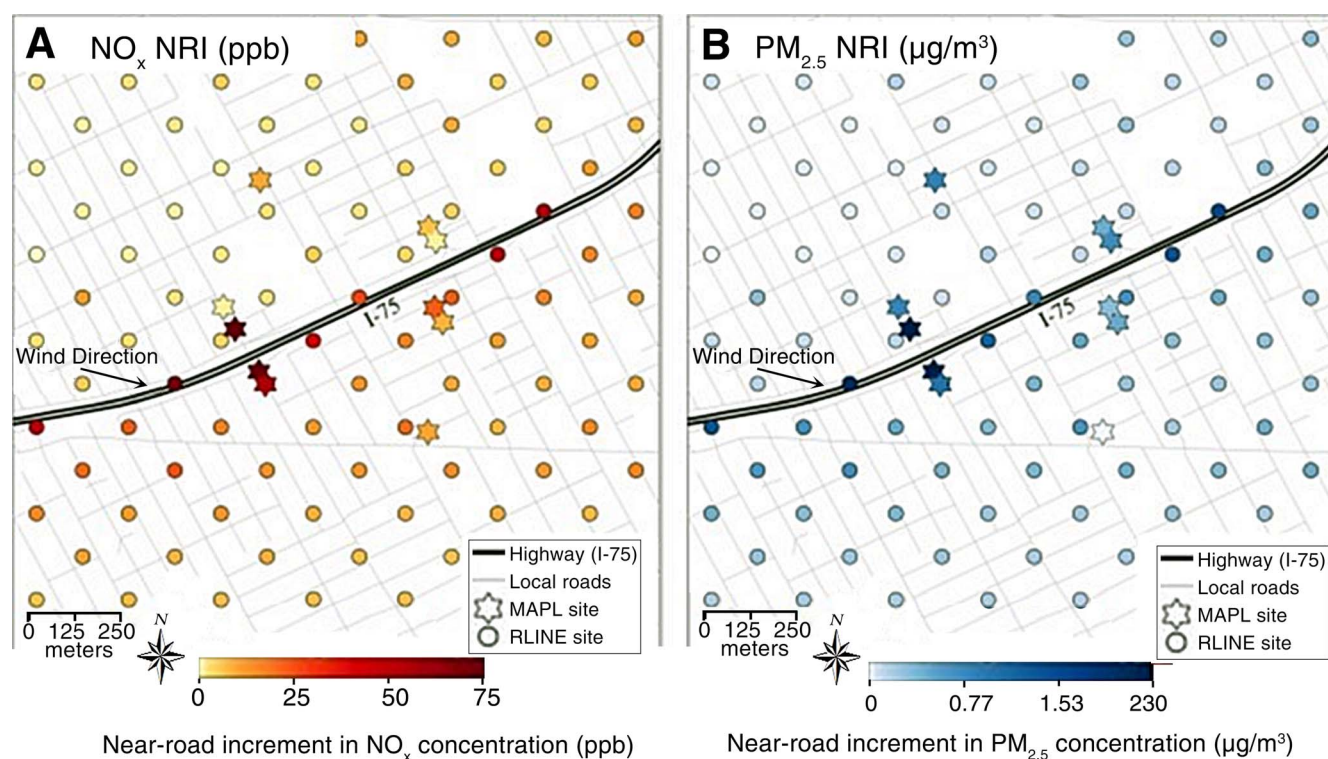
Figure 3. **Maps of transect area 5 showing NRI for (A) NO$_x$ and (B) PM$_{2.5}$.** Measurements for each were taken at 9 sites on the morning of December 18, 2012. The panels show the NRI concentrations determined at the MAPL sites and the RLINE predicted concentrations using a 150-meter receptor grid for the corresponding pollutant and period.

Exploratory investigation showed considerable variability in RLINE outputs for the two pollutants between the different areas. Reporting summary statistics for RLINE output set 1, the lowest mean concentrations of both pollutants occurred in area 7 for both NO$_x$ (8.0 ± 10.1 ppb) and PM$_{2.5}$ (0.23 ± 0.30 µg/m^3); the highest concentrations occurred in area 6 for both NO$_x$ (42.4 ± 52.1 ppb) and PM$_{2.5}$ (1.2 ± 1.5 µg/m^3). Figure 4A compares the observed NRI concentrations of NO$_x$ at the MAPL monitoring sites to RLINE outputs at the corresponding receptors using both sets of emissions; Figure 4B gives the comparable plot for PM$_{2.5}$. These plots suggest that RLINE tends to provide higher predictions of the NRI for both NO$_x$ and PM$_{2.5}$, with the RLINE PM$_{2.5}$ NRI being larger than the corresponding MAPL observations almost all the time.

Covariates

We obtained traffic and meteorological data for the study area. Annual average daily traffic (AADT) and commercial AADT (CAADT) volumes for major roads in Detroit were obtained from the Michigan Department of Transportation Traffic Monitoring Information System (2014), and adjusted to hourly volumes using Detroit-specific temporal allocation factors (Batterman et al. 2015) (see Appendices 4 and 7 in Additional Materials on the HEI website). Because emissions from commercial vehicles, especially heavy-duty diesel vehicles, can greatly exceed those from noncommercial traffic, which are mostly light-duty gasoline vehicles (Batterman et al. 2015; Watkins 2012), we defined an *adjusted traffic volume* to account for the relative contribution of these vehicles to air pollution. This covariate was calculated as NCAADT + c CAADT, where NCAADT is the noncommercial volume (estimated hourly as AADT — CAADT), and parameter c, sometimes referred to as the
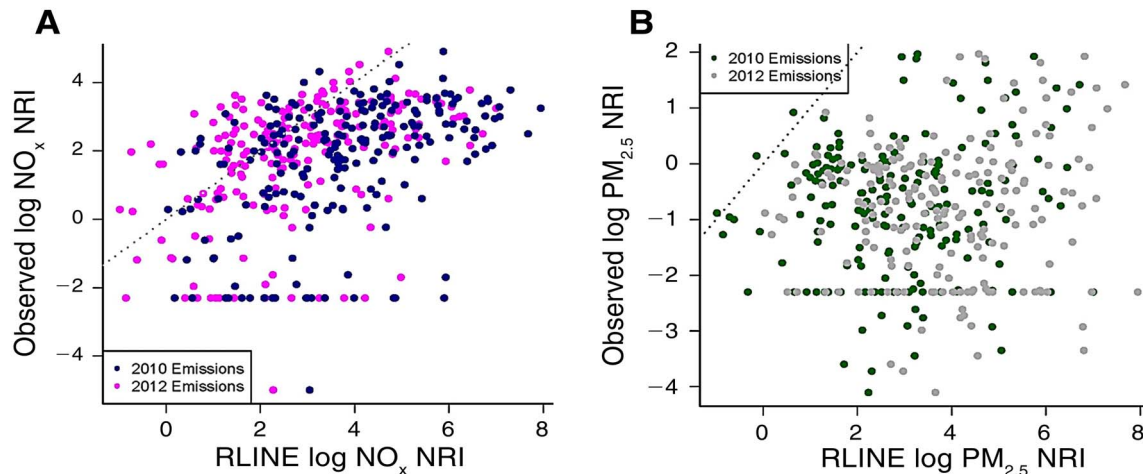
Figure 4. Scatterplots of NRI in monitored concentrations at MAPL sites versus RLINE model predictions at corresponding receptors for (A) NO$_x$ (ppb) and (B) PM$_{2.5}$ (µg/m^3). In each panel, the dotted line indicates 45 degrees.

passenger car equivalent, was set to 10 for NO and NO$_x$ and to 50 for BC and PM$_{2.5}$ following prior work in Detroit (Baldwin et al. 2015).

Wind speed and direction data measured at five airport meteorological sites were obtained as Quality Controlled Local Climatological Data (National Oceanic and Atmospheric Administration, *http://cdo.ncdc.noaa.gov/qclcd/QCLCD*). To obtain representative statistics, hourly wind direction and wind speed were averaged across sites. Hourly measurements of NO and NO$_x$ concentrations were obtained from the AQS site at the East 7 Mile site in Detroit (*http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html*). BC concentrations were not available for a representative site in the study area.

Model Formulation

Several spatiotemporal statistical models were developed for (1) universal kriging and spatial interpolation of observed concentrations of TRAP; and (2) data fusion of observations of NRIs of pollutant concentrations with dispersion model outputs. In each model, we assumed that the spatial dependence in TRAP concentration between any two sites depended on the actual locations of the two sites; in other words, we assumed that the spatial correlation in TRAP concentrations was nonstationary.

Each model uses the following notation: Let t represent the time period for the day and time of day (morning or afternoon) of pollutant monitoring. There were either $T = 10$ or $T = 11$ distinct time periods during the study period

depending on the pollutant considered. Due to the right-skewness in both observed concentrations and NRIs (Appendix 6 in Additional Materials on the HEI website), we decided to work on the log scale. Hence, $Y_t(\mathbf{s})$ denotes the natural log of an ambient pollutant concentration (NO, NO$_x$, or BC) or the NRI concentration (PM$_{2.5}$ and NO$_x$) at location $\mathbf{s}$ in a spatial domain $\mathcal{G}$ and time period $t = 1, …, T$. Let $X_t(\mathbf{s})$ indicate the natural log of the RLINE output, that is, the estimated NRI for a pollutant concentration. The general modeling framework adopted decomposes $Y_t(\mathbf{s})$ as the sum of three terms:

$$Y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \eta_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \quad \varepsilon_t(\mathbf{s}) \stackrel{iid}{\sim} N(0, \tau^2), \qquad (1)$$

where $\mu_t(\mathbf{s})$ accounts for the large-scale spatial trend in the pollutant log-concentration or in the log NRI at time period t, $\eta_t(\mathbf{s})$ accounts for the small-scale spatial variation in the pollutant concentration or NRI at site $\mathbf{s}$ at time period t not captured by $\mu_t(\mathbf{s})$, and $\varepsilon_t(\mathbf{s})$ is an independent error process, independent across time periods and space, with mean 0 and variance τ^2, often referred to in the geostatistical literature as the nugget effect (Banerjee et al. 2004; Cressie 1993).

Nonstationary Universal Kriging Models For the log of an ambient pollutant concentration, we formulated several single-pollutant nonstationary universal kriging models that accounted for day-of-week, upwind versus downwind, topographic features, meteorology, traffic activity, and fleet composition features identified as important predictors in an exploratory analysis (Appendix 6). Following

earlier work examining the near-road environment in Detroit (Baldwin et al. 2015), we modeled the large-scale spatial trend $\mu_t(\mathbf{s})$ at $\mathbf{s}$ as a function of various predictors with regression coefficients β: the background concentration (ambient air quality measured by AQS monitors), an indicator for whether a site $\mathbf{s}$ was downwind of the nearest road during time period t, an indicator for weekday, an indicator for morning, a scalar that quantifies the normalized amount of traffic on the nearest highway around $\mathbf{s}$ during time period t, the distance of site $\mathbf{s}$ from the edge of the nearest highway and some appropriate interactions:

$$\mu_t(\mathbf{s}) = \beta_0 + \beta_1 \operatorname{Down}_t(\mathbf{s}) + \beta_2 \log(\operatorname{AQS})_t(\mathbf{s}) \tag{2}$$

$$+ \beta_3 \operatorname{Weekday}_t + \beta_4 \operatorname{Morning}_t$$

$$+ \beta_5 \operatorname{Weekday}_t * \operatorname{Morning}_t$$

$$+ \beta_6 \left(\frac{\operatorname{Traffic}_t(\mathbf{s})}{\operatorname{Wind\ Speed}_t(\mathbf{s})} \right)$$

$$+ \beta_7 \left(\frac{\operatorname{Traffic}_t(\mathbf{s})}{\operatorname{Wind\ Speed}_t(\mathbf{s})} \right) * \operatorname{Down}_t(\mathbf{s})$$

$$+ \beta_8 \operatorname{Distance}(\mathbf{s})$$

$$+ \beta_9 \operatorname{Distance}(\mathbf{s}) * \operatorname{Down}_t(\mathbf{s}).$$

We further assume, and confirm through autocorrelation function plots, that the temporal variability in each pollutant log concentration $Y_t(\mathbf{s})$ is captured by the term $\mu_t(\mathbf{s})$, so that the term $\eta_t(\mathbf{s})$ in equation 1 accounts for purely spatial dependence and can be modeled as independent realizations over time periods of a Gaussian process with mean zero and a given covariance function. We hypothesize that the covariance function of said spatial process is nonstationary. Empirical evaluations provided with more detail in Gilani and colleagues (2015), show that this is indeed the case.

Given the potentially different spatial dependences in log concentrations between upwind and downwind sites, for each pollutant, we model the term $\eta_t(\mathbf{s})$, $t = 1, \ldots, T$ as independent realizations in time of a mixture of two mutually independent mean-zero Gaussian processes, $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$, which are independent over time and equipped with covariance functions $C_{\theta_1}(\cdot,\cdot)$ and $C_{\theta_2}(\cdot,\cdot)$, respectively. In other words, for each time period t and $\mathbf{s} \in \mathcal{G}$

$$\eta_t(\mathbf{s}) = w_{1,t}(\mathbf{s}) \cdot \eta_{1,t}(\mathbf{s}) + w_{2,t}(\mathbf{s}) \cdot \eta_{2,t}(\mathbf{s}). \tag{3}$$

In turn, the weights, $w_{1,t}(\mathbf{s})$ and $w_{2,t}(\mathbf{s})$, in equation 3 are spatially and temporally varying and sum to 1, which we achieve by setting

$$w_{i,t}(\mathbf{s}) = \frac{\tilde{w}_{i,t}(\mathbf{s})}{\tilde{w}_{1,t}(\mathbf{s}) + \tilde{w}_{2,t}(\mathbf{s})}, \ i = 1, 2. \tag{4}$$

Analogous to the kernel expression of the weights (Fuentes 2001), we specify the weights $\tilde{w}_{1,t}(\mathbf{s})$ and $\tilde{w}_{2,t}(\mathbf{s})$ in such a way that if a site $\mathbf{s}$ is downwind during time period t, the spatial process $\eta_{1,t}(\mathbf{s})$ in equation 3 receives a larger weight, and vice versa. A specification of $\tilde{w}_{i,t}(\mathbf{s})$ that allows this is:

$$\tilde{w}_{i,t}(\mathbf{s}) \tag{5}$$
$$= \begin{cases} \exp\left(-\psi \cdot \left(1 - \operatorname{Down}_t(\mathbf{s})\right) \cdot |\mathbf{s} - \mathbf{s}^*| \right), & i = 1 \\ \exp\left(-\psi \cdot \left(\operatorname{Down}_t(\mathbf{s})\right) \cdot |\mathbf{s} - \mathbf{s}^*| \right), & i = 2, \end{cases}$$

where $\mathbf{s}^*$ is defined as the projection of the site $\mathbf{s}$ onto the farther edge of the highway. Note that the parameter ψ in the definition of the unnormalized weights $\tilde{w}_{1,t}(\mathbf{s})$ and $\tilde{w}_{2,t}(\mathbf{s})$, and thus in the definition of the mixture weights $w_{1,t}(\mathbf{s})$ and $w_{2,t}(\mathbf{s})$, is the same across pollutants since the three traffic-related pollutants are recorded at the same monitoring sites and at the same time periods. Parameter ψ controls how quickly the unnormalized weights $\tilde{w}_{1,t}(\mathbf{s})$ and $\tilde{w}_{2,t}(\mathbf{s})$, decay to zero as the point $\mathbf{s}$ is farther away from the edge of the highway. At a distance of approximately 3ψ, the unnormalized weight $\tilde{w}_{2,t}(\mathbf{s})$ (or $\tilde{w}_{1,t}(\mathbf{s})$) will be equal to 0.05 if the point $\mathbf{s}$ is downwind (respectively, upwind) during time period t.

For the covariance functions, $C_{\theta_1}(\cdot,\cdot)$ and $C_{\theta_2}(\cdot,\cdot)$ of the serially independent copies $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$ across time periods t of the two mutually independent mean-zero Gaussian processes, we use different covariance models: a stationary one (e.g., exponential) and a nonstationary one. As we believe that for each pollutant and time period t, the covariance functions of the two underlying spatial processes $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$ are nonstationary and potentially influenced by wind speed, we follow Schmidt and colleagues (2011). Thus, we define a transformation $\tilde{\mathbf{f}}_t : \mathcal{G} \subset \mathbf{R}^2 \to \mathcal{D} \subset \mathbf{R}^3$ from the spatial domain $\mathcal{G}$ to a subset $\mathcal{D}$ of $\mathbf{R}^3$ that, for each time period t, maps a site $\mathbf{s} \in \mathcal{G}$ into $\tilde{\mathbf{f}}_t(s) = (\mathbf{s}_x, \mathbf{s}_y, h_t(\mathbf{s}))$ where $h_t(\mathbf{s})$ is the signed wind speed at $\mathbf{s}$ and time period t; in other words, wind speed has a positive sign if $\mathbf{s}$ is downwind at time t and with a negative sign otherwise. Finally,

for each pollutant and time period t, the covariance function of the underlying process $\eta_{i,t}(\mathbf{s})$, $i = 1, 2$ is modeled using an exponential covariance function and the Mahalanobis distance. Hence, for any pair of sites $\mathbf{s}, \mathbf{s}' \in \mathcal{G}$,

$$Cov\left(\eta_{i,t}\left(\mathbf{s}\right), \eta_{i,t}\left(\mathbf{s}'\right)\right) \qquad (6)$$

$$= \sigma_i^2 \exp\left\{-\sqrt{\left(\tilde{\mathbf{f}}_t\left(\mathbf{s}\right) - \tilde{\mathbf{f}}_t\left(\mathbf{s}'\right)\right)' \cdot \boldsymbol{\Phi}_i \cdot \left(\tilde{\mathbf{f}}_t\left(\mathbf{s}\right) - \tilde{\mathbf{f}}_t\left(\mathbf{s}'\right)\right)}\right\},$$

$i = 1, 2$.

Φ_i is a 3×3 diagonal matrix for each $i = 1, 2$,

$$\begin{pmatrix} \phi_i & 0 & 0 \\ 0 & \phi_i & 0 \\ 0 & 0 & \varphi_i \end{pmatrix},$$

with ϕ_i controlling the range or smoothness of the covariance function along the x-y direction and φ_i controlling it along the signed wind-speed direction.

This covariance function is a valid covariance function, and simply an application of the model proposed by Schmidt and colleagues (2011), which in turn extends an already well-known and well-accepted method to model nonstationary covariance functions, using the deformation method (Sampson and Guttorp 1992).

The model proposed for the spatial covariance function of the term $\eta_t(\mathbf{s})$ is a mixture of the two nonstationary covariance functions, and as such is a valid covariance function with nonstationarity driven by a covariate (e.g., signed

wind speed) with weights that vary spatially and temporally and that depend on a covariate as well (e.g., being upwind or downwind): for each pollutant, at time period t and sites $\mathbf{s}, \mathbf{s}' \in \mathcal{G}$,

$$Cov\left(\eta_t\left(\mathbf{s}\right), \eta_t\left(\mathbf{s}'\right)\right) = \sum_{i=1}^{2} w_{i,t}\left(\mathbf{s}\right) w_{i,t}\left(\mathbf{s}'\right) \cdot \sigma_i^2 \cdot \qquad (7)$$

$$\exp\left\{-\sqrt{\left(\tilde{\mathbf{f}}_t\left(\mathbf{s}\right) - \tilde{\mathbf{f}}_t\left(\mathbf{s}'\right)\right)' \cdot \boldsymbol{\Phi}_i \cdot \left(\tilde{\mathbf{f}}_t\left(\mathbf{s}\right) - \tilde{\mathbf{f}}_t\left(\mathbf{s}'\right)\right)}\right\}.$$

Note that the specification of the nonstationary covariance function of $\eta_t(\mathbf{s})$ in equation 7 contains simpler models as special cases. As examples, removing the influence of the covariate in the expression of the unnormalized mixture weights $\tilde{w}_{i,t}(\mathbf{s}), i = 1, 2$, leads to a model formulation reminiscent of Fuentes (2001); while setting one of the two mixture weights $w_{1,t}(\mathbf{s})$, $w_{2,t}(\mathbf{s})$ to zero for any $\mathbf{s}$ and time period t, but maintaining equation 6 as the model for the covariance functions of the underlying spatial processes, $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$, leads to the formulation of Schmidt and colleagues (2011).

In summary, we modeled the small-scale spatial variation $\eta_t(\mathbf{s})$ of the pollution concentration field as a weighted spatiotemporal mixture of two mutually independent spatiotemporal processes, loosely interpretable as concentrations upwind and downwind, with wind speed and wind direction influencing the mixture weights and the dependence structure of each latent process. Six models that differed in the approach used to model the spatial dependence were considered. Table 1 summarizes these models, the combination of weighting schemes, and their covariance functions.

Table 1. Description of the Universal Kriging Models[a, b]

Model Name	Weighting Scheme	Covariance Function
Model 1: Independence	None	Independence
Model 2: Stationary	None	Exponential covariance function
Model 3	Binary upwind-downwind	Exponential covariance function
Model 4	Binary upwind-downwind	Covariates in covariance function, as in Eq. 6
Model 5	As in Eq. 5	Exponential covariance function
Model 6	As in Eq. 5	Covariates in covariance function, as in Eq. 6

[a] For each model, the trend term $\mu_t(\mathbf{s})$ is modeled according to equation 2.

[b] For each model, we report the type of weighting scheme used in the mixture, and the covariance functions used for $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$. Models 1 and 2 do not express $\eta_t(\mathbf{s})$ as a mixture; thus the covariance function in the table refers to the covariance function of $\eta_t(\mathbf{s})$.

Joint-Modeling Bayesian Data Fusion Models Data fusion models are statistical models that aim to provide more accurate spatiotemporal estimates of air pollutant concentrations by combining different data sources — in this case, data observed at air quality monitors and the output of the dispersion model RLINE. Data fusion models for pollutants can be classified into two general categories: (1) joint modeling approaches in which both measured monitor data and RLINE model output are treated as stochastic realizations of an underlying unobserved true pollutant concentration field (Choi et al. 2009; Fuentes and Raftery 2005; McMillan et al. 2010; Sahu et al. 2010); and (2) regression-based approaches in which measured monitor data are the outcome variables and are regressed on the RLINE model output through a Bayesian hierarchical spatiotemporal model without additional covariates (Berrocal et al. 2010a,b, 2012; Crooks and Özkaynak 2014; Gilani et al. 2016; Reich et al. 2014; Rundel et al. 2015; Zidek et al. 2012). We considered both approaches. Specifically, using RLINE output set 1 (concentrations derived using the 2010 emission inventories and 96 regularly spaced receptors within a 2-km square centered on the major road), we developed a single-pollutant, nonstationary joint data fusion modeling approach for both the NO_x and the $PM_{2.5}$ NRIs. In *RLINE output set 2 – 2010* and *RLINE output set 2 – 2012* we assessed whether changes in the emission inventory improved the predictive power of RLINE output using a nonstationary, single-pollutant, regression-based framework.

Still in the context of nonstationary regression-based data fusion approaches, we also developed a multipollutant, nonstationary data fusion model that leverages the correlation between NO_x and $PM_{2.5}$ and jointly predicts them.

Single-Pollutant, Nonstationary Spatiotemporal, Joint Modeling Bayesian Data Fusion Models Let $\hat{Y}_t(\mathbf{s})$ now represent the *true*, underlying and unobserved natural log of the NRI concentration for a pollutant ($PM_{2.5}$ or NO_x) at location $\mathbf{s} \in \mathcal{S}$ ($\mathcal{S}$ with spatial domain) and time period $t = 1, ..., T$. In a similar fashion to equation 1, we decompose $\hat{Y}_t(\mathbf{s})$ as the sum of two terms

$$\hat{Y}_t(\mathbf{s}) = \hat{\mu}_t(\mathbf{s}) + \hat{\eta}_t(\mathbf{s}), \qquad (8)$$

where $\hat{\mu}_t(\mathbf{s})$ accounts for the large-scale spatial trend in the pollutant's log NRI concentration at location $\mathbf{s}$ and time period t, and $\hat{\eta}_t(\mathbf{s})$ accounts for the small-scale spatial structure. Since $\hat{Y}_t(\mathbf{s})$ is a latent, unobserved field, equation 8 does not include an error term with a nugget effect variance τ^2.

At a MAPL monitoring site $\mathbf{s}$ and time period t, the measured log NRI concentration, $Y_t(\mathbf{s})$ is an error-prone measurement of $\hat{Y}_t(\mathbf{s})$, in other words,

$$Y_t(\mathbf{s}) = \hat{Y}_t(\mathbf{s}) + e_t(\mathbf{s}), \qquad e_t(\mathbf{s}) \stackrel{iid}{\sim} N\left(0, \tau_e^2\right), \qquad (9)$$

where $e_t(\mathbf{s})$ represents the measurement error at site $\mathbf{s}$ and time period t and is independent of the true underlying process $\hat{Y}_t(\mathbf{s})$.

For the RLINE output, we assume that the log NRI concentration estimate at location $\mathbf{s}$ and time t, denoted by $X_t(\mathbf{s})$, displays both additive and multiplicative biases that may be spatially varying or constant. While previous studies have shown evidence of spatially varying additive bias, the multiplicative bias is generally modeled as constant in space and time. Therefore, the log RLINE NRI concentration is modeled as

$$X_t(\mathbf{s}) = a + a_t(\mathbf{s}) + b \cdot \hat{Y}_t(\mathbf{s}) + \delta_t(\mathbf{s}), \qquad (10)$$

where a is a constant (over space and time) additive bias, and $a_t(\mathbf{s})$ is a spatially correlated mean-zero additive bias, local deviation at $\mathbf{s}$ during period t from the overall additive bias a. This spatially varying additive bias in turn is modeled as a Gaussian process with mean 0 and with an exponential covariance function with parameters σ_a^2 (marginal variance) and ϕ_a (decay parameter). With this modeling choice, we assume that the additive bias of the RLINE output is similar at nearby locations but less so at distant locations, with a correlation that decays exponentially with distance. Finally, in equation 10, b is the constant (over space and time) multiplicative bias of the RLINE output, and $\delta_t(\mathbf{s})$ captures the random deviation of the RLINE output from the true underlying process $\hat{Y}_t(\mathbf{s})$ at location $\mathbf{s}$ and time t. We hypothesize that the deviation $\delta_t(\mathbf{s})$ is mean-zero, independent across space and time, and follows a Gaussian distribution, in other words, $\delta_t(\mathbf{s}) \stackrel{iid}{\sim} N\left(0, \tau_\delta^2\right)$.

Adopting an analogous model to equation 2 for the large-scale spatial trend $\hat{\mu}_t(\mathbf{s})$ of the true unobserved log NRI concentration $\hat{Y}_t(\mathbf{s})$

$$\hat{\mu}_t(\mathbf{s}) = \hat{\beta}_0 + \hat{\beta}_1 \, \text{Down}_t(\mathbf{s}) + \hat{\beta}_2 \, \text{Weekday}_t \qquad (11)$$
$$+ \hat{\beta}_3 \, \text{Morning}_t + \hat{\beta}4 \, \text{Weekday}_t * \text{Morning}_t$$
$$+ \hat{\beta}_5 \left(\frac{\text{Traffic}_t(\mathbf{s})}{\text{Wind Speed}_t(\mathbf{s})}\right) + \hat{\beta}_6 \left(\frac{\text{Traffic}_t(\mathbf{s})}{\text{Wind Speed}_t(\mathbf{s})}\right)$$
$$* \, \text{Down}_t(\mathbf{s}) + \hat{\beta}_7 \, \text{Distance}(\mathbf{s}) + \hat{\beta}_8 \, \text{Distance}(\mathbf{s})$$
$$* \, \text{Down}_t(\mathbf{s}),$$

where $\text{Down}_t(\mathbf{s})$ is an indicator for site $\mathbf{s}$ at time t being downwind (versus upwind) of the highway; Weekday_t is an indicator for time period t falling on a weekday (versus weekend); Morning is an indicator for morning versus afternoon time period t; $\dfrac{\text{Traffic}_t(\mathbf{s})}{\text{Wind Speed}_t(\mathbf{s})}$ is a scalar that quantifies the traffic volume on the nearest highway around $\mathbf{s}$ during time period t normalized by the wind speed at site $\mathbf{s}$ during time period t; and $\text{Distance}(\mathbf{s})$ records the distance of $\mathbf{s}$ from the nearest highway.

Additionally, for both NO_x and $PM_{2.5}$ we use the model formulation in equation 3 for the small-scale spatial structure $\hat{\eta}_t(\mathbf{s})$ of the true, unobserved log NRI field. In other words, for $t = 1, \ldots, T$, $\hat{\eta}_t(\mathbf{s})$ is assumed to be a weighted mixture of two mutually independent mean-zero Gaussian spatial processes $\hat{\eta}_{1,t}(\mathbf{s})$ and $\hat{\eta}_{2,t}(\mathbf{s})$. In turn, the latter are taken to be independent over time, and combined as in equation 3 with weights defined as in equation 5 to yield $\hat{\eta}_t(\mathbf{s})$. Finally, the nonstationary covariance functions $\hat{C}_{\theta_1}(\cdot,\cdot)$ and $\hat{C}_{\theta_2}(\cdot,\cdot)$ of $\hat{\eta}_{1,t}(\mathbf{s})$ and $\hat{\eta}_{2,t}(\mathbf{s})$, respectively, are taken as in equation 6.

We compare the predictive performance of six joint Bayesian data fusion models that differ in the approach used to model the bias of the log RLINE output (e.g., constant in space and time versus not) and for the type of spatial dependence structure hypothesized for the two latent processes, $\hat{\eta}_{1,t}(\mathbf{s})$ and $\hat{\eta}_{2,t}(\mathbf{s})$. Table 2 provides a description of the models considered. The simplest model, called model 1-JBDF, assumes that the error of the RLINE output in representing the true, unobserved field is constant in space, thus $a_t(\mathbf{s})$ is equal to 0. Model 2-JBDF postulates that even though the RLINE output has a spatially additive

error with mean 0, overall, the RLINE output does not have an additive bias, that is, $a \equiv 0$. Model 3-JBDF is the full model whether the small-scale spatial structure, $\hat{\eta}_t(\mathbf{s})$, of the unobserved true pollution field, $\hat{Y}_t(\mathbf{s})$, is stationary or not. For each of these three models, we contrast two cases: the first assumes that $\hat{\eta}_t(\mathbf{s})$ is a Gaussian process independent in time with mean 0 and with a stationary exponential covariance function; the second assumes $\hat{\eta}_t(\mathbf{s})$ is equipped with the nonstationary covariance function described in equation 7. These cases are distinguished by appending S or NS.

Single-Pollutant, Regression-Based Bayesian Data Fusion Models

The single-pollutant nonstationary regression-based Bayesian data fusion approach proposed to combine RLINE output with monitoring data can be interpreted as a compromise between the nonstationary universal kriging model and the Bayesian data fusion model. Specifically, letting $\hat{Y}_t(\mathbf{s})$ be the observed natural log NRI concentration at site $\mathbf{s}$ at time period t, we write $\hat{Y}_t(\mathbf{s})$ according to the general model (equation 1), for example, $Y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \eta_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$, $\varepsilon_t(\mathbf{s}) \sim N(0,\tau^2)$, with $\mu_t(\mathbf{s})$ being the large-scale spatial trend of the observed log NRI and $\eta_t(\mathbf{s})$ accounting for the small-scale residual spatial structure. The large-scale trend $\mu_t(\mathbf{s})$ is modeled as a linear function of the RLINE output:

$$\mu_t(\mathbf{s}) = \alpha_0 + \alpha_1 X_t(\mathbf{s}), \tag{12}$$

with α_0 representing the overall additive bias of the RLINE output and α_1 indicating the multiplicative bias. The small-scale spatial structure $\hat{Y}_t(\mathbf{s})$ in equation 1, $\eta_t(\mathbf{s})$, can now be interpreted both as the residual spatial structure of

Table 2. Summary of the Joint Bayesian Data Fusion Models for Log NRI of NO_x and $PM_{2.5}$

Model Name	Form of Additive Bias for RLINE	Covariance Structure of $\hat{\eta}_t(\mathbf{s})$
Model 1-JBDF-S	$\delta_t(\mathbf{s}) \equiv 0$	Stationary: exponential
Model 2-JBDF-S	$a_0 \equiv 0$	Stationary: exponential
Model 3-JBDF-S	Full	Stationary: exponential
Model 1-JBDF-NS	$\delta_t(\mathbf{s}) \equiv 0$	Nonstationary, as in Eq.7
Model 2-JBDF-NS	$a_0 \equiv 0$	Nonstationary, as in Eq.7
Model 3-JBDF-NS	Full	Nonstationary

the observed log NRI field after having accounted for the log RLINE output, or as the spatial additive error of RLINE. As before, the spatial additive error of RLINE, $\eta_t(\mathbf{s})$, is modeled as a spatially varying mixture of two nonstationary, independent in time, spatiotemporal processes as in equation 3 with weights $w_{1,t}(\mathbf{s})$ and $w_{2,t}(\mathbf{s})$ defined in equations 4 and 5. Moreover, both underlying spatiotemporal processes $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$ are taken to be mean-zero Gaussian processes provided with nonstationary covariance functions, as described earlier. Consequently, the resulting covariance function of the additive bias of RLINE is given by equation 7.

To evaluate whether RLINE adds any additional information to that contained in meteorological and traffic covariates when estimating the large-scale behavior of the log NRI field, we consider an additional regression-based Bayesian data fusion model. This regression-based Bayesian data fusion model maintains the same general formulation for the log NRI field used above, for example, $Y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \eta_t(\mathbf{s}) + \varepsilon_t(\mathbf{s})$, $\varepsilon_t(\mathbf{s}) \sim N(0,\tau^2)$, and it employs the same spatial dependence structure for the small-scale residual spatial structure $\eta_t(\mathbf{s})$, but models the large-scale trend $\mu_t(\mathbf{s})$ as:

$$\mu_t(\mathbf{s}) = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_t(\mathbf{s}) + \gamma_1 \, \mathrm{Down}_t(\mathbf{s}) + \gamma_2 \, \mathrm{Weekday}_t \quad (13)$$

$$+ \; \gamma_3 \, \mathrm{Morning}_t + \gamma_4 \, \mathrm{Weekday}_t * \mathrm{Morning}_t$$

$$+ \; \gamma_5 \left(\frac{\mathrm{Traffic}_t(\mathbf{s})}{\mathrm{Wind\ Speed}_t(\mathbf{s})} \right)$$

$$+ \; \gamma_6 \left(\frac{\mathrm{Traffic}_t(\mathbf{s})}{\mathrm{Wind\ Speed}_t(\mathbf{s})} \right)$$

$$* \; \mathrm{Down}_t(\mathbf{s})$$

$$+ \; \gamma_7 \, \mathrm{Distance}(\mathbf{s})$$

$$+ \; \gamma_8 \, \mathrm{Distance}(\mathbf{s}) * \mathrm{Down}_t(\mathbf{s}).$$

Even though the regression-based Bayesian data fusion approach uses the RLINE output only at locations where monitors are located, after substituting equation 12 (or equation 13, respectively) in equation 11, the log NRI concentration can be obtained at any prediction site. Thus, spatial maps of the NRI continuous surface can be generated if the RLINE output is available on a very fine grid.

Multiple-Pollutants, Regression-Based Bayesian Data Fusion Model The final modeling approach extends the single-pollutant regression-based Bayesian data fusion model to multiple pollutants using models that estimate the overall additive and multiplicative biases of the RLINE

output corresponding to NO_x and $PM_{2.5}$, as well as the spatial additive errors of the RLINE NRI output for the two pollutants. Specifically, the multiple-pollutant regression-based Bayesian data fusion model specifies a regression-based Bayesian data fusion model for each single pollutant. If $Y_t^{(k)}(\mathbf{s})$ and $X_t^{(k)}(\mathbf{s})$ are, respectively, the log observed NRI and the RLINE estimated log NRI for pollutant k ($1 = NO_x$, $2 = PM_{2.5}$) at location $\mathbf{s}$ and time period t, adopting the same model as equations 1 and 12 for each $Y_t^{(k)}(\mathbf{s})$ individually, yields:

$$Y_t^{(k)}(\mathbf{s}) = \mu_t^{(k)}(\mathbf{s}) + \eta_t^{(k)}(\mathbf{s}) + \epsilon_t^{(k)}(\mathbf{s}), \quad (14)$$

$$\epsilon_t^{(k)}(\mathbf{s}) \overset{iid}{\sim} N\left(0,\tau_k^2\right); k = 1,2 \;\; \text{and}$$

$$\mu_t^{(k)}(\mathbf{s}) = \alpha_0^{(k)} + \alpha_1^{(k)} + X_t^{(k)}(\mathbf{s}); \; k = 1,2,$$

where $\alpha_0^{(k)}$ and $\alpha_1^{(k)}$ are, respectively, the additive and multiplicative bias of the log RLINE output for pollutant k. These models will yield estimates of the overall additive $\alpha_0^{(k)}$ and multiplicative $\alpha_1^{(k)}$ bias of the RLINE output corresponding to NO_x and $PM_{2.5}$, respectively, while estimates of $\eta_t^{(1)}(\mathbf{s})$ and $\eta_t^{(2)}(\mathbf{s})$ will provide an idea of the magnitude of the spatial additive errors of the RLINE NRI output for NO_x and $PM_{2.5}$, respectively.

To explicitly account for nonstationarity in both pollutants and to exploit the correlation among the two pollutants, we express each $\eta_t^{(k)}(\mathbf{s})$, $k = 1, 2$, according to equation 3, that is, as a mixture with pollutant-specific weights, $w_{1,t}^{(k)}(\mathbf{s})$ and $w_{2,t}^{(k)}(\mathbf{s})$ of the same two underlying, latent processes, $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$, thus

$$\eta_t^{(k)}(\mathbf{s}) = w_{1,t}^{(k)}(\mathbf{s}) \cdot \eta_{1,t}(\mathbf{s}) + w_{2,t}^{(k)}(\mathbf{s}) \cdot \eta_{2,t}(\mathbf{s}). \quad (15)$$

In turn, while the pollutant-specific weights, $w_{1,t}^{(k)}(\mathbf{s})$ and $w_{2,t}^{(k)}(\mathbf{s})$, are modeled as in equations 4 and 5 with pollutant-specific decay parameter $\psi^{(k)}$, $k = 1, 2$, as in the other models, the latent underlying processes $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$ are modeled as mutually independent, independent in time and equipped with nonstationary covariance functions as in equation 6.

Fitting and Performance Evaluation

Details regarding fitting and performance evaluation of the spatiotemporal models are provided in Appendix 7 (see Additional Materials on the HEI website). Evaluation of the predictive performance of each model was made for

each pollutant and for each type of measurement (concentration or NRI) through comparisons between observations and predictions at out-of-sample sites. For example, for the nonstationary universal kriging models of ambient pollutant concentrations, of the 286 observed NO and NO$_x$ concentrations, we randomly selected 253 observations for model fitting and held out 33 observations for model validation. As our objective was to evaluate whether each model was able to capture both the spatial and temporal structure in the TRAPs and NRI concentration, for each model and pollutant, we held out 10%–15% of the observations, resulting in 1–3 observations per time period, which we randomly sampled from the complete dataset. Predictive performance was assessed by back transforming predictions to the original scale. Using the median of the posterior predictive distribution as the predicted value at each site, the predictive performance of each model was evaluated in terms of *mean absolute prediction error* (MAPE), average *length of the 90% prediction interval* (PI), and *empirical coverage* of the 90% PI. The latter is used to assess whether the uncertainty in the prediction is correctly quantified: if empirical coverage is below the nominal level,

assuming no bias in the predictions, the model is underestimating the variability/uncertainty in the predictions. Vice versa, empirical coverage of the PIs above the nominal level indicates that the model overestimates the variability/uncertainty. For the Bayesian data fusion models, along with the above-mentioned prediction metrics, we also reported the *root mean square error* (RMSE) and the *Pearson correlation* between the predicted concentrations and the held-out data.

RESULTS

DISPERSION MODELING

Operational Evaluation

Performance metrics are summarized in Table 3; scatterplots of observed versus modeled NO$_x$ and CO at the I-69 near-road site are shown in Figure 5. For NO$_x$, daily mean predictions (modeled background + modeled traffic contributions) were similar to observations (20–38 ppb and

Table 3. Performance Metrics for Dispersion Modeling of Daily Average NO$_x$ and CO Concentrations[a]

Pollutant / Site	Method	Days	Means (ppb)						F2	R$_{SP}$	FB		V$_G$		NMSE	R^2	R$_P$
			Obs	Back	Model	Ncom	Com	Point			FP	FN	Irr	Red			
NO$_x$																	
School	ICHEM	918	23	17	3	1.2	0	1	95	0.32	0.07	0.22	1.01	1.12	18	0.07	0.26
Near-road	ICHEM	334	37	16	21	18.6	2	1	92	0.58	0.17	0.17	1.01	1.18	30	0.37	0.61
Near-road	IGpCHEM	705	48	15	23	18.5	4	1	95	0.74	0.05	0.28	1.03	1.11	34	0.46	0.68
Urban	ICHEM	238	25	18	11	8.5	1	1	93	0.57	0.22	0.09	1.03	1.12	23	0.27	0.52
Urban	IGpCHEM	565	26	16	12	8.5	2	1	97	0.58	0.15	0.09	1.01	1.09	17	0.33	0.57
CO																	
Suburban	IGFC	40	673	671	27	19	3	5	100	0.21	0.11	0.07	1.00	1.04	153	0.01	0.11
Near-road	EC9830T	82	479	128	192	180	9	4	94	0.89	0.00	0.40	1.14	1.05	204	0.72	0.85
Near-road	INDiI	655	667	519	291	277	9	5	99	0.45	0.21	0.01	1.04	1.03	197	0.18	0.43
Urban	INDiI	284	639	545	126	115	5	6	99	0.17	0.12	0.07	1.00	1.05	166	0.02	0.15
Industrial	IGFC	63	585	535	115	100	10	5	100	0.00	0.14	0.03	1.01	1.03	132	0.00	−0.01

[a] Back = modeled background contribution; Com = modeled contribution from commercial traffic; F2 = % of model + background within a factor of 2 of observed; FB = fractional bias; FP = false positive component of fraction bias; FN = false negative component of fractional bias; ICHEM = instrumental chemiluminescence; IGpCHEM = instrumental gas-phase chemiluminescence; Irr = irreducible or systematic component of V$_G$; Model = modeled contribution from traffic (Ncom + Com); Ncom = modeled contribution from noncommercial traffic; NMSE = normalized mean square error; Obs = observed concentrations; Point = modeled contribution from point sources; R^2 = coefficient of determination; R$_P$ = Pearson correlation coefficient; R$_{SP}$ = Spearman rank correlation coefficient; Red = reducible or random component of V$_G$; RMSE = root mean square error; V$_G$ = geometric variance.
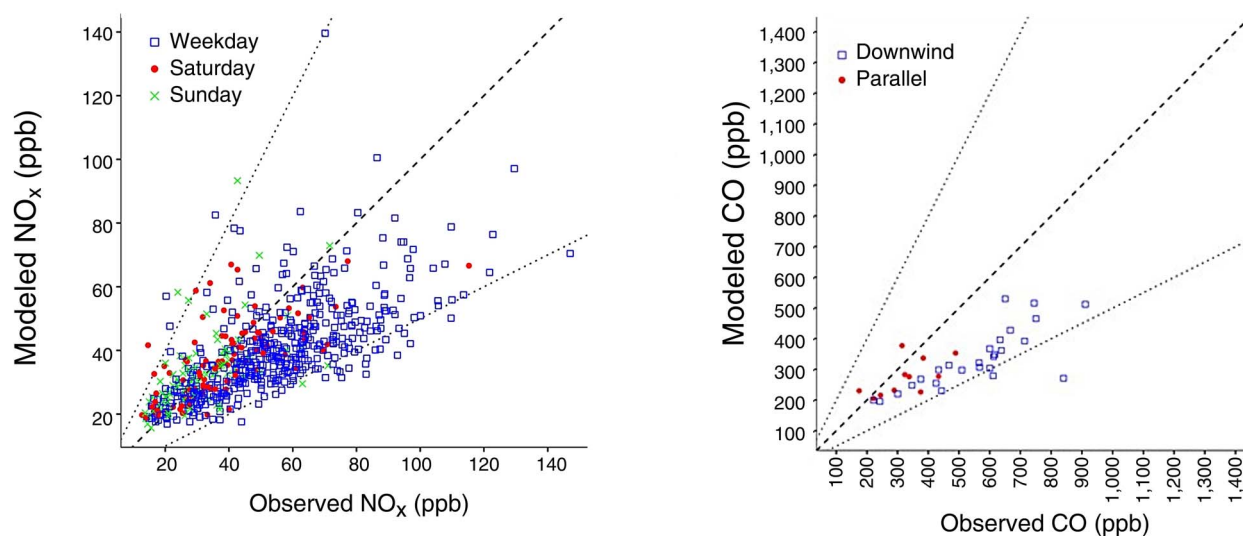
Figure 5. **Observed vs. modeled NO$_X$ and CO at the near-road site (using the IGpCHEM and EC9830T monitors, respectively).** Figures show 1:1 and factor of 2 lines.

23–48 ppb, respectively). Performance tended to decrease with distance from the roadway, for example, R$_{SP}$ ranged from 0.58 to 0.74 at the near-road site (10 meters from I-96), 0.57 to 0.58 at the urban site (100 meters from I-96), and 0.32 at the schools site (350 meters from MI-97). The near-road site using the instrumental gas-phase chemiluminescence (IGpCHEM) monitor had the highest R$_{SP}$, the lowest % reducible V$_G$ and the highest mean model-to-background ratio. However, the model of near-road NO$_X$ measured using IGpCHEM had the highest false negative FB, above that using the instrumental chemiluminescence (ICHEM) instrument, a result obtained mainly because the IGpCHEM measurements (average of 48 ppb) exceeded the ICHEM measurements (37 ppb), while predictions were similar (38 and 37 ppb, respectively). Performance for NO$_X$ at other sites varied. The schools site tended to underpredict daily averages; the near-road and urban sites were overpredicted; and reducible errors at all four sites exceeded systematic errors, suggesting improvements in model inputs or parameterization could improve model performance.

For CO, daily predictions (320 to 810 ppb) was in the range of observed levels (479 to 673 ppb). For NO$_X$, performance generally decreased with distance from the roadway (e.g., R$_{SP}$ was 0.45 to 0.89 at the near-road site, 0.17 at the urban site, and 0.21 at the suburban site). Despite its proximity to I-75 (150 meters), the industrial site had an R$_{SP}$ near zero, probably due to that monitor's high detection limit that falsely elevated the background estimates. (At the industrial site, for example, the estimated background averaged 92% of measurements.) Also as for

NO$_X$, the near-road site (with the EC9830T instrument) had the highest R$_{SP}$ for CO (Figure 5) and again, this case had the lowest ratio of reducible to overall V$_G$, the highest mean model-to-background ratio, but the highest negative FB. Patterns at the other sites were similar to those seen for NO$_X$. For CO, observations frequently fell below the detection limit for the less sensitive instruments (IGFC and INDiI), which yielded relatively high background estimates (averaging 519 to 671 ppb). Ideally, trace-level CO instrumentation would be used.

For PM$_{2.5}$, background averaged 8.8 μg/m^3, equivalent to 88% to 92% of observed levels (9.5 and 10 μg/m^3, respectively), and day-to-day variability was significant. Predicted contributions from point and on-road mobile sources at the monitoring sites were small (averaging from 0.1 to 0.8 μg/m^3), and including local sources did not increase model fit. This result can be attributed to the dominance of regional sources and the small signal remaining from local sources after considering background levels, the gaps and uncertainties of the PM$_{2.5}$ emission inventory, the absence of chemical transformations in RLINE, and the few near-road sites monitoring PM$_{2.5}$. Thus, RLINE performance evaluations for PM$_{2.5}$ were not attempted. Performance evaluations of the spatiotemporal models (described later) did include PM$_{2.5}$; however, the transect dataset used for this evaluation included monitoring sites much closer to major roads, and measurements were conducted at peak traffic periods — factors that tend to increase the NRI (as described in later sections of this report and in Appendix 6).

Sensitivity Analyses

The sensitivity of concentration predictions to meteorological, emission, and traffic allocation inputs is summarized below. (Detailed analyses are provided in the appendices.) The analysis uses the modeling system described earlier to estimate daily average concentrations of CO and NO_X, and it compares baseline (or nominal) and alternative inputs. We also examined effects of downwind versus parallel winds, day-of-week and seasonal effects, and the updated emissions inventory. These analyses include the use of four years of ambient monitoring data, exposures predicted for both general and vulnerable populations in Detroit, and the same performance metrics used in the models described earlier.

Downwind vs. Parallel Winds   We first compared model performance when receptors were downwind of the road, and when winds were parallel to the road. For NO_X, downwind conditions generally gave higher F2 and higher R_{SP} (0.30 to 0.64); other performance metrics were mixed (e.g., at the urban site during downwind periods, FB was slightly lower, V_G was unchanged, and the % Red was lower, mainly for the ICHEM monitor). Performance for CO was also generally better during downwind periods, albeit less conclusively than for NO_X. F2 exceeded 92% at all sites. The near-road and urban sites had higher R_{SP} (0.29 to 0.83) during downwind periods compared to parallel winds (−0.07 to 0.60). While limited by high detection limits, the CO dataset indicated better performance during downwind conditions. This is largely consistent with the more limited evaluation using hourly CO and NO_X at the near-road sites (Appendix 10, see Additional Materials on the HEI website) that suggested systematic model biases by wind direction (e.g., overprediction with downwind winds, and under-prediction during parallel winds). The better performance during downwind conditions may reflect the greater signal from local (on-road) emission sources.

Day of Week   For NO_X, performance on weekdays generally was better than on Saturdays and Sundays (e.g., weekdays gave higher F2 and higher R_{SP} in most cases, although weekdays tended to have more underpredictions). For CO, the evaluation was hampered by data limitations, but weekday performance again appeared better, although the metrics were inconsistent and the sample sizes was small. This may reflect the more regular traffic volume and fleet-mix patterns occurring on weekdays that are more accurately represented by temporal allocation factors (Batterman et al. 2015). In contrast, traffic patterns on weekends (especially Sundays) are more variable. Higher traffic volumes on weekdays also may increase traffic-related emissions

and concentrations. Underpredictions on weekdays might result from higher emissions, possibly due to a higher diesel fraction in the fleet mix, and possibly lower dispersion than assumed. These speculations might be examined using diagnostic (rather than operational) evaluations that examine rush-hour periods and traffic conditions (Zhang and Batterman 2010; Zhang et al. 2011).

Season   Performance trends by season were not strong, but performance appeared slightly better during winter. For example, R_{SP} at the near-road site was 0.79 to 0.84 in winter (depending on the instrument), and from 0.52 to 0.69 in other seasons (excluding spring with one instrument when R_{SP} was also 0.79). At the near-road site, F2 was highest in winter with the ICHEM instrument (but in spring with the IGpCHEM instrument), and the lowest relative reducible error was in winter. However, seasonal trends differed at other sites and data limitations restricted the reliability of the CO data. Potentially important seasonal changes in Detroit include shifts in prevailing wind directions, changes in the relative frequency of dispersion regimes (represented in RLINE as MO lengths), large temperature swings — which affects MOVES emission factors (Chan et al. 2013), changes in atmospheric composition (especially OH^-) that can alter pollutant transformation and fate, and changes in the level and composition of regional pollutants (particularly for $PM_{2.5}$). Only some of these processes are captured in dispersion models.

Emission Factors   Performance was slightly better using the updated emission factors, which changed emission factors for several vehicle classes. For example, overall emissions of NO_X and CO from light-duty gas vehicle and heavy-duty diesel vehicle classes increased by 48% and 30%, respectively (Appendix 8). The results suggest that NO_X emission estimates can be very sensitive to the estimated traffic activity (e.g., commercial traffic volume), especially during cold weather and congestion when speeds are lower and emissions are high relative to gasoline vehicles.

Temporal Allocation Factors   The three sets of TAFs yielded few differences in either NO_X and CO predictions that exceeded the significance thresholds (Appendix 9, see Additional Materials on the HEI website). Thus, the Detroit-specific TAFs that separated commercial and noncommercial traffic did not perform better than the simpler and default TAFs. This result was unanticipated, especially for NO_X, given the differences between commercial and noncommercial vehicles and the differences seen in the simplified analyses.

Meteorology Meteorological datasets obtained at NWS stations 18 kilometers or more apart caused large differences in daily concentration predictions on some days at both sets of receptors, which supports findings from comparisons at the monitoring sites (Appendix 10). Both NWS stations are at airports, and the surrounding terrain is flat and mostly urban, commercial, wooded, or agricultural. The differences in predicted concentrations seem likely to result mainly from changes in atmospheric stability that alters near-road concentration gradients, possibly due to very stable conditions — which can cause the highest concentrations (as seen in Appendix 3) (Snyder et al. 2013b). This suggests the possibility of significant exposure measurement error if the meteorological dataset is not representative. Errors may be higher for more vulnerable populations, as portrayed by the NEXUS receptors representing children who lived close to major roads (Appendix 11, see Additional Materials on the HEI website).

Due to siting and instrumentation limitations, few air quality monitoring sites (including near-road sites) measure all of the meteorological parameters required for dispersion modeling. For example, of the 79 near-road sites in the United States (2015; *www3.epa.gov/ttnamti1/near-road.html*), the geocoordinates of 7 sites were not available. For the remaining 72 sites, the distance to the nearest NWS station averaged 18.5 kilometers; 6 sites were within 5 kilometers and 28 were within 10 kilometers of an NWS station. This suggests that available meteorological inputs at many sites may not be representative of near-road settings. Blending local meteorological data with NWS (or other) datasets is workable and is incorporated in the AERMET processor. This procedure generally obtained the best performance in the Detroit application. Still, a full set of measurements at the local site of interest may be preferable for obtaining measurements that are most representative of near-road environments. This option, which could not be fully tested in Detroit, leads to a recommendation to collect a full set of local meteorological measurements for dispersion modeling when practicable, reinforcing long standing model guidance that recognizes the increased heat flux and surface roughness in urban areas and the general need for multiple monitoring sites in large urban areas (Giambini et al. 2012; U.S. EPA 2000). (No specific guidance was available during the study.) At larger roads in urban settings, such modeling involves dispersion transitioning from the road microenvironment to the adjacent suburban microenvironment. Differences between these microenvironments may be considerable. Many urban roads are large paved areas. For example, portions of the I-96 right-of-way in Detroit exceeds 150 meters in width as each traffic direction includes three local and three express lanes, a two-lane service road, multiple shoulders, and some vegetated buffers. The suburban road microenvironment has buildings and trees bordering a smaller area of flat and paved surfaces. Guidance defining the most representative meteorological data for traffic-related emissions in such settings, which differ from the general urban environment, would be helpful for improving near-road modeling. (See *Implications of the Findings* for additional dispersion modeling recommendations.)

Treatment of Low Concentrations Omitting low measured concentrations (e.g., below the detection limit) from the evaluation may have artificially increased correlations by limiting analyses to those observations when local-source impacts are seen. In another sensitivity analysis, values below the detection limit were set to half of the detection limit and all analyses were repeated. This dampened some trends (e.g., the wind direction analysis of NO_x, and in some cases R_{SP} and other metrics changed noticeably), although the general conclusions remain unchanged. Removing low values has the advantage of largely eliminating (meaningless) comparisons between modeled and measured background, which can be important if roadway impacts are small or if monitoring methods have low detection frequencies.

SPATIOTEMPORAL MODELING

Nonstationary Universal Kriging Models

The predictive performance of the nonstationary universal kriging models for the three pollutants is shown in Table 4. Model 6 provided the smallest MAPE among the six models and had at least 97% empirical coverage of the 90% PI for all pollutants. The independent and stationary models (models 1 and 2, respectively) had smaller average 90% PIs compared with the nonstationary models, but the PIs appeared to be overly conservative, as shown by the smaller empirical coverage across pollutants. The improved predictive performance of Models 4 and 6 for all three pollutants suggests that the signed wind speed is an important factor that explains the nonstationarity processes in the near-road environment.

Posterior summaries of regression coefficients and covariance parameters for Model 6 are shown in the top and bottom sections of Table 5, respectively. For all pollutants, decay parameters ϕ_1 and ϕ_2 of the underlying spatial processes $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$ relative to the *x-y* direction were very similar, giving effective ranges for NO of 55 and 45 meters for $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$, respectively, 35 and 28 meters for NO_x, and 44 and 32 meters for BC. In other words, 55 and 45 meters were distances at which the spatial

Table 4. Predictive Performance of the Nonstationary Universal Kriging Models Averaged Across Validation Sites and Time Periods[a]

Criteria	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
NO						
MAPE (ppb)	7.91	7.24	15.07	7.36	15.96	7.20
Avg. length of 90% PI (ppb)	57.4	65.0	132	69.96	3,146	72.45
Emp. coverage of 90% PI (%)	94	97	94	100	100	100
NO$_X$						
MAPE (ppb)	10.0	9.29	17.7	9.26	18.5	8.93
Avg. length of 90% PI (ppb)	65.7	69.9	200	75.5	5,794	77.2
Emp. coverage of 90% PI (%)	94	94	97	100	100	100
BC						
MAPE ($\mu g/m^3$)	0.54	0.54	0.69	0.43	0.65	0.42
Avg. length of 90% PI ($\mu g/m^3$)	2.96	3.15	7.00	3.09	203	3.18
Emp. coverage of 90% PI (%)	85	85	100	97	100	97

[a] Shown is the mean absolute prediction error (MAPE) and the average length and empirical coverage of 90% prediction intervals (PI).

[b] Models are defined in Table 2.

correlation in NO concentration at any two sites reduced to 0.05 upwind and downwind, respectively. Similarly, the correlation in NO$_X$ concentrations between any two sites upwind (respectively, downwind) is 0.05 or less if their distance is greater or equal than 35 meters (respectively, 28 meters). Analogous interpretations can be provided for the effective ranges for BC.

Hence, spatial correlation decayed quickly with increasing distance, with similar rates for the two underlying spatial processes (e.g., at downwind and upwind sites). On the other hand, decay parameters for the signed wind speed directions ϕ_1 and ϕ_2 were very different: the effective ranges for NO were 0.02 and 4.22 m/s for $\eta_{1,t}(\mathbf{s})$ and $\eta_{2,t}(\mathbf{s})$, respectively, 0.01 and 2.64 m/s for NO$_X$, and 0.02 and 2.99 m/s for BC. Considering the weighting scheme, these results suggest that downwind sites are practically uncorrelated in the signed wind speed direction, while correlation between upwind sites decays rather slowly with increasing differences in the signed wind speeds between sites. These results provide further evidence that the nonstationarity observed in the log concentrations is driven predominantly by the signed wind speed and not by the geographical distance, and that the wind speed influences the spatial correlation differently for upwind and downwind sites.

The coefficient estimated for the downwind indicator (β_1) was positive for all three pollutants. Controlling for other covariates in the model, downwind concentrations of NO averaged exp(0.98) = 2.7 ppb higher than upwind concentrations; NO$_X$ and BC concentrations were 1.8 ppb and 1.52 $\mu g/m^3$ higher, respectively. The coefficient for the interaction between downwind indicator and distance (β_9) was negative, indicating a significant difference in the decay rate of concentrations with distance from the highway between downwind and upwind sites. In particular, NO concentrations decreased exp(−1.85) = 0.16 times slower at downwind compared to upwind sites, 0.28 times slower for NO$_X$, and 0.41 times slower for BC. The estimated coefficients for morning (β_4) were positive for all three pollutants, and negative for the interactions between morning and weekday (β_5), suggesting that after controlling for the other covariates, morning concentrations did not vary greatly between weekdays and weekends, but evening concentrations did. In particular, on average, NO on weekday

Table 5. Posterior Medians and 95% Credible Intervals of Regression Coefficients and Covariance Parameters Estimated by Model 6 (Full Nonstationary Kriging Model) Fitted to NO, NO_x, and BC[a]

Coefficient / Parameter[b]		NO		NO_x		BC	
		Median	95% CI	Median	95% CI	Median	95% CI
β_0	Intercept	−0.66	(−1.24, −0.09)	0.60	(0.14, 1.08)	−0.23	(−0.49, 0.03)
β_1	DW	0.98	(0.60, 1.35)	0.58	(0.32, 0.83)	0.42	(0.13, 0.70)
β_2	log(AQS)	1.01	(0.81, 1.20)	0.82	(0.68, 0.96)	—	
β_3	Weekday	0.04	(−0.36, 0.42)	−0.19	(−0.45, 0.08)	−0.11	(−0.42, 0.19)
β_4	Morning	1.53	(1.02, 2.04)	0.75	(0.41, 1.10)	1.06	(0.66, 1.48)
β_5	Weekday*morning	−1.33	(−1.93, −0.72)	−0.65	(−1.08, −0.25)	−0.91	(−1.37, −0.42)
β_6	(Traffic/WS) $\times 10^{-5}$	0.79	(−1.79, 3.35)	0.87	(−0.84, 2.62)	1.60	(−0.26, 3.44)
β_7	(Traffic/WS)*DW $\times 10^{-5}$	−1.92	(−5.43, 1.56)	−0.76	(−3.13, 1.62)	−0.37	(−3.03, 2.32)
β_8	Distance	−0.69	(−1.52, 0.13)	−0.11	(−0.67, 0.46)	−0.19	(−0.87, 0.47)
β_9	Distance*DW	−1.85	(−3.23, −0.52)	−1.29	(−2.24, −0.4)	−0.88	(−1.95, 0.17)
τ^2	Nugget	0.34	(0.26, 0.46)	0.16	(0.12, 0.21)	0.20	(0.15, 0.27)
σ_1^2	Sill — DW	0.48	(0.25, 2.52)	0.21	(0.11, 1.25)	0.31	(0.16, 1.86)
ϕ_1	DW	1,700	(1,256, 2,248)	1,702	(1,262, 2,278)	1,701	(1,262, 2,278)
φ_1	DW, WS	12,420	(9,349, 16,133)	12,365	(9,225, 16,128)	12,413	(9,303, 16,268)
σ_2^2	Sill — UW	0.33	(0.24, 0.45)	0.16	(0.12, 0.22)	0.19	(0.14, 0.27)
ϕ_2	UW	1,732	(1,276, 2,275)	1,711	(1,283, 2,266)	1,735	(1,291, 2,266)
φ_2	UW, WS	0.20	(0.15, 0.26)	0.20	(0.15, 0.26)	0.20	(0.15, 0.27)

[a] AQS = Concentration measured by Air Quality System monitor; CI = credible interval; DW = downwind; UW = upwind; WS = wind speed. Regression coefficients are in the top part of the table (β_0–β_9); covariance parameters are in the bottom part of the table.

[b] Traffic was recorded in adjusted number of vehicles per hour, WS in m/s; and distance in km.

mornings was exp(0.2) = 1.22 ppb higher concentrations than on weekday evenings, and NO on weekend mornings was 4.62 ppb higher than on weekend evenings. Similarly, NO_x and BC concentrations on weekday mornings were respectively 1.11 ppb and 1.16 μg/m³ higher than on weekday afternoons; differences between weekend mornings and evenings were 2.12 ppb and 2.89 μg/m³, respectively. While the magnitude of these differences seems modest, they apply to observations from 50 to 500 meters from the road; downwind sites nearest roads are expected to have larger differences, as shown earlier. As expected, the estimated coefficients are positive both for the log of background concentration (β_2) for NO and NO_x, and for the ratio of traffic to wind speed (β_6) for all three pollutants.

The sensitivity of results for Models 4 and 6 to the parameter ψ used to calculate weights in equation 5 was evaluated by rerunning models for different values of ψ. This showed only negligible changes, suggesting that the choice of ψ did not have much influence.

Model 6 was used to predict concentrations of the three pollutants over the spatial domain. As an example, Figure 6 shows the mean and 90% PIs of NO concentrations around area 8, which straddles I-94 in Detroit, on December 20, 2012 on a 50-meter grid. Analogous plots for NO_x and BC concentrations are displayed in Figures 7 and 8, respectively. The prediction maps show clear concentration gradients with increasing distance from the highway, particularly in the downwind direction. Upwind concentrations are more homogenous, with slightly higher concentrations near the road than at distant sites.
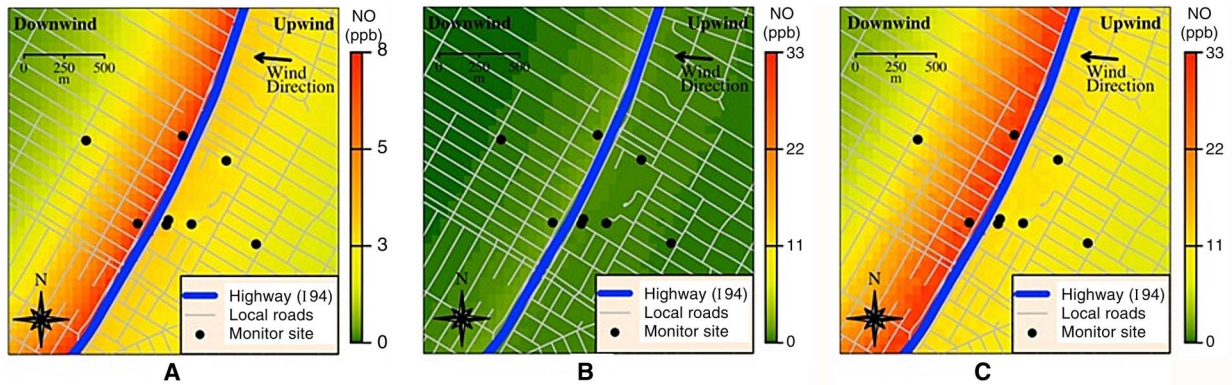
Figure 6. Mean and 90% PI of NO concentrations (ppb) yielded by model 6 on 50-meter grids. Measurements were taken around area 8 on December 20, 2012. **(A)** Predicted hourly NO concentrations; 90% PI **(B)** lower and **(C)** upper bounds.
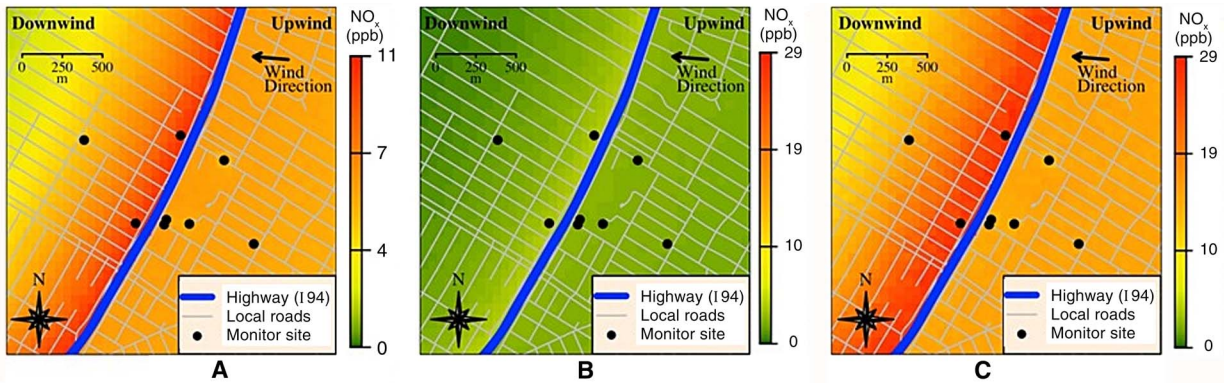

Figure 7. Mean and 90% PI of NO_x concentrations (ppb) yielded by model 6 on 50-meter grids. Measurements were taken around area 8 on December 20, 2012. **(A)** Predicted hourly NO_x concentrations; the 90% PI **(B)** lower and **(C)** upper bounds.
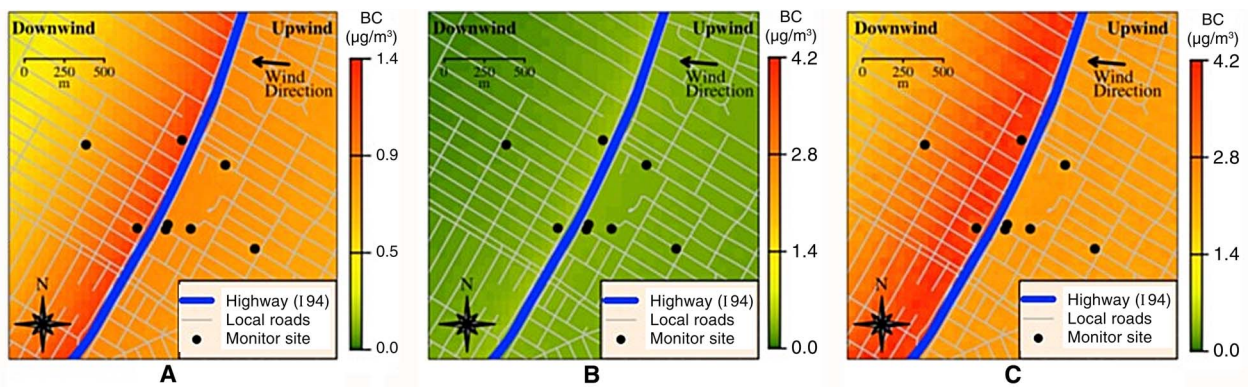

Figure 8. Mean and 90% PI of BC concentrations (µg/m^3) yielded by model 6 on 50-meter grids. Measurements were taken around area 8 on December 20, 2012. **(A)** Predicted hourly NO_x concentrations; the 90% PI **(B)** lower and **(C)** upper bounds.

Joint Bayesian Data Fusion Models

Performance statistics of the joint Bayesian data fusion models for NO_x and $PM_{2.5}$ NRI are shown in Table 6. No single model was clearly dominant with respect to all criteria for both pollutants, in other words, no single model had the lowest MAPE, the lowest RMSE, the highest correlation with the held-out data, and the shortest PI with empirical coverage close to the nominal level.

For NO_x, the stationary models provided better predictive performance than the nonstationary counterparts. Model 2-JBDF-S yielded the lowest MAPE and RMSE, and its 90% PIs had an empirical coverage that was the second closest to the nominal coverage, even though on average they were wider than the 90% PIs from Model 1-JBDF-S. (Estimated parameters for Model 2-JBDF-S are shown in Table 7.) This suggests that although the RLINE output displays a spatial error in representing the true, underlying NO_x NRI field, possibly the magnitude of the site-specific

errors is not very large and is uniform in space. This is in part confirmed by the estimate of the marginal variance σ_a^2 of the spatially varying error $a_t(\mathbf{s})$ of the log RLINE output, which was estimated to be quite small and equal to 0.60 (95% credible interval: 0.43 to 0.92). Although comparisons of estimates in Tables 5 and 7 are not meaningful since different variables are considered (observed concentrations in one case versus NRI in the other), and the observations and sites used for model fitting do not coincide, we infer that the additional information provided by the RLINE output reduces the uncertainty of the estimated model parameters. However, this may be achieved at the cost of the empirical coverage of the PIs.

For $PM_{2.5}$, the situation was different; the Bayesian data fusion models with nonstationary covariance functions provided strikingly better predictive performance than the stationary models. However, as for NO_x, no model was better in terms of all five criteria (MAPE, RMSE, correlation,

Table 6. Predictive Performance of the Bayesian Data Fusion Models Averaged Across Validation Sites and Time Periods[a,b]

Criteria	Model 1 JBDF-S	Model 2 JBDF-S	Model 3 JBDF-S	Model 1 JBDF-NS	Model 2 JBDF-NS	Model 3 JBDF-NS
NO_x						
MAPE (ppb)	6.59	6.29	7.00	6.95	6.69	8.13
RMSE (ppb)	8.75	8.65	9.34	9.42	8.92	11.99
Pearson correlation	0.55	0.44	0.38	0.49	0.42	0.28
Avg. length of 90% PI (ppb)	23.81	28.31	32.49	108.25	33.98	300.41
Emp. coverage of 90% PI (%)	84.0	77.0	77.0	84.0	74.0	74.0
$PM_{2.5}$						
MAPE ($\mu g/m^3$)	1.10	0.71	0.77	0.48	0.42	0.44
RMSE ($\mu g/m^3$)	1.63	0.90	0.97	0.67	0.60	0.63
Pearson correlation	0.02	0.44	0.41	0.14	0.36	0.34
Avg. length of 90% PI ($\mu g/m^3$)	4.72	7.10	7.66	3.30	4.87	4.22
Emp. coverage of 90% PI (%)	55.0	90.0	86.0	66.0	69.0	69.0

[a] Shown are the mean absolute prediction error (MAPE), root mean squared error (RMSE), Pearson correlation with held-out data, and the average length and empirical coverage of 90% prediction intervals (PI). JBDF = joint Bayesian data fusion; NS = nonstationary; S = stationary.

[b] Models are defined in Table 2.

Table 7. Posterior Medians and 95% Credible Intervals of Regression Coefficients and Covariance Parameters Estimated by the Best-Fitting Models, Model 2-JBDF-S and Model 2-JBDF-NS, Fitted to NO_x and $PM_{2.5}$, Respectively[a]

Coefficient / Parameter[b]		NO_x[c]		$PM_{2.5}$[d]	
		Median	(95% CI)	Median	(95% CI)
$\hat{\beta}_0$	Intercept	1.65	(1.26, 2.10)	−0.89	(−2.06, 0.51)
$\hat{\beta}_1$	DW	0.96	(0.83, 1.11)	1.37	(−0.65, 3.36)
$\hat{\beta}_2$	Weekday	0.57	(0.03, 1.15)	−0.25	(−1.57, 1.10)
$\hat{\beta}_3$	Morning	0.32	(−0.43, 1.07)	1.06	(−0.88, 3.05)
$\hat{\beta}_4$	Weekday*morning	−0.77	(−1.67, 0.06)	−0.63	(−2.79, 1.59)
$\hat{\beta}_5$	(Traffic/WS) $\times 10^{-5}$	5.78	(2.71, 8.63)	14.1	(4.79, 24.3)
$\hat{\beta}_6$	(Traffic/WS)*DW $\times 10^{-5}$	−4.54	(−6.04, −3.13)	−15.7	(−27.0, −5.22)
$\hat{\beta}_7$	Distance	−15.1	(−17.8, −12.6)	−12.9	(−19.2, −6.88)
$\hat{\beta}_8$	Distance*DW	−3.01	(−6.28, 0.15)	−1.49	(−8.44, 5.40)
a	Additive bias of RLINE output	—		—	
b	Multiplicative bias of RLINE output	1.25	(1.11, 1.35)	0.66	(0.59, 0.75)
τ_e^2	Nugget effect	0.42	(0.42, 0.43)	0.06	(0.06, 0.06)
τ_δ^2	Variance of the deviation between the RLINE output and the true NRI	0.07	(0.05, 0.09)	0.07	(0.06, 0.009)
σ_1^2	Sill — DW		—	6.34	(2.70, 13.00)
ϕ_1	DW		—	0.004	(0.001, 0.03)
φ_1	DW, WS		—	6.58	(6.18, 35.3)
σ_2^2	Sill — UW		—	8.90	(5.06, 14.3)
ϕ_2	UW		—	0.07	(0.02, 0.19)
φ_2	UW, WS		—	25.6	(64.5, 76.5)
$\sigma_{\hat{Y}}^2$	Marginal variance of the true, underlying NRI	0.49	(0.35, 0.69)		—
$\phi_{\hat{Y}}$	Range parameter of the true, underlying NRI	1.01	(0.65, 1.52)		—
σ_a^2	Marginal variance of the spatially varying additive error of RLINE	0.60	(0.43, 0.92)	1.01	(0.68, 1.69)
ϕ_a	Range parameter of the spatially varying additive error of RLINE	0.78	(0.44, 1.26)	0.15	(0.07, 0.26)

[a] CI = credible interval; DW = downwind; UW = upwind; WS = wind speed. Regression coefficients $\left(\hat{\beta}_0 - \hat{\beta}_8\right)$ are in the top part of the table, covariance parameters are in the bottom part of the table.

[b] Traffic was recorded in adjusted number of vehicles per hour, WS in m/s; and distance in km.

[c] NO_x: Model 2-JBDF-S.

[d] $PM_{2.5}$: Model 2-JBDF-NS.

average length, and empirical coverage of 90% PI). Weighting the MAPE and RMSE criterion the most, Model 2-JBDF-NS yielded the best predictions. As for NO_x, Model 2-JBDF-NS hypothesized a null overall additive bias, but localized additive error at each location. Again, as for NO_x and also for $PM_{2.5}$, the spatially varying additive bias of RLINE had a small marginal variance and a very short range, indicating that the spatial correlation decays very fast (Table 7).

Even though the best-performing joint Bayesian data fusion models are the models specified by equations 8 through 11 with $a \equiv 0$, that is, models that do not include a global additive calibration term for the RLINE output, both models include a global multiplicative calibration term b. Hence, according to these models RLINE should be rescaled to represent the true unobserved NO_x and $PM_{2.5}$ NRI fields. Given the observed data, the multiplicative calibration term b for NO_x and $PM_{2.5}$ are estimated to be 1.25 and 0.66, respectively (see Table 7), with 95% credible intervals that do not contain 1. This suggests that in light of the observed data, the RLINE output is not appropriately scaled to correctly represent the true NRI concentration for both pollutants.

In other words, the Bayesian data fusion models suggest that the RLINE output needs to be calibrated to be considered as a good representation of the true unobserved NRI field for both NO_x and $PM_{2.5}$, but the bias adjustment mostly consists of multiplicative correction (or rescaling) of the RLINE output for both pollutants, as the best performing model for both pollutants (in terms of MAPE) are models that did not include an overall additive bias. For both pollutants, RLINE displayed a spatial error that had a small to moderate marginal variance and a fast-decaying spatial correlation. Interestingly, while the universal kriging models for observed NO_x and $PM_{2.5}$ concentrations better matched the data when a nonstationary covariance function was used, data fusion models that exploited the information in the RLINE output to represent the true unobserved NO_x NRI field performed better when the covariance function was modeled to be stationary.

Regression-Based Bayesian Data Fusion Models

Single-Pollutant Models Performance statistics of the regression-based Bayesian data fusion models for NO_x (Table 8) were inferior to the nonstationary universal kriging model, which indicates that RLINE does not provide any additional useful information than what is contained in the GIS covariates used in the nonstationary kriging model. As reported in the previous section, when incorporating the RLINE output, a stationary covariance function described the residual small-scale spatial structure of the NRI field better, suggesting that the nonstationarity in the NRI field was partly accounted for by the RLINE output.

Results of joint Bayesian data fusion models and single-pollutant regression-based Bayesian data fusion models cannot be compared since different RLINE outputs were used; the former used *RLINE output set 1* (2010 emissions inventory and dense set of receptors), while the latter used *RLINE output set 2* (both 2010 and 2012 emissions inventory, receptors located at MAPL sites only). The many receptors used in the joint Bayesian data fusion model application likely explains the divergent widths of the 90% PIs of these two modeling approaches: 28.3 ppb (Table 6) for Model 2-JBDF-S versus 113.3 ppb (Table 8) for the stationary regression-based Bayesian data fusion model using the best-fitting model within their Bayesian data fusion model category. While the joint Bayesian data fusion modeling approach yielded smaller PIs, this imposed increased computational complexity and running time, in addition to potential identifiability issues.

The influence of the emission inventories used in the dispersion modeling on the predictive performance of the regression-based Bayesian data fusion models is displayed in Table 8, which compares the predictive performance of a model that uses the RLINE output derived using, respectively, the original (2010) and the updated (2012) emissions. For NO_x, the updated inventory led to a slight gain in predictive performance regardless of whether a stationary or nonstationary covariance function was used for the dependence structure of the spatial additive bias of the RLINE.

As noted for the joint Bayesian data fusion models, the RLINE output was helpful for predicting $PM_{2.5}$ NRI concentration (although not for NO_x), especially when the small-scale covariance structure of the observed NRI field was modeled using a nonstationary covariance function. In particular, this latter class of model yielded predictions that had a good correlation with the held-out data, while there was almost no correlation between the predictions obtained using the stationary Bayesian data fusion models. In addition, the empirical coverage of the 90% PI for $PM_{2.5}$ was very close to the nominal level. Finally, the updated emissions inventory made only minimal differences in terms of predictive performance for the $PM_{2.5}$ NRI, another difference from NO_x.

Parameters of the best-fitting regression-based Bayesian data fusion models for NO_x and $PM_{2.5}$ NRI are listed in Table 9. As already noted for the joint Bayesian data fusion model, the best-fitting model for NO_x had an overall additive bias equal to 0 and a local spatial additive bias with a

Table 8. Predictive Performance of Regression-Based Bayesian Data Fusion Models Averaged Across Validation Sites and Time Periods Using RLINE with 2010 vs. 2012 Emissions Inventories[a,b]

Criteria	Nonstationary Universal Kriging	Stationary Data Fusion (emissions)		Nonstationary Data Fusion (emissions)	
		2010	2012	2010	2012
NO_x					
MAPE (ppb)	6.60	7.25	6.97	7.10	7.07
RMSE (ppb)	9.01	10.22	9.65	9.92	9.90
Pearson correlation	0.71	0.64	0.67	0.66	0.64
Avg. length of 90% PI (ppb)	116.95	113.27	113.34	116.97	121.27
Emp. coverage of 90% PI (%)	66.67	69.70	69.70	69.70	69.70
$PM_{2.5}$					
MAPE ($\mu g/m^3$)	0.51	0.62	0.62	0.47	0.47
RMSE ($\mu g/m^3$)	0.82	1.05	1.05	0.86	0.87
Pearson correlation	0.56	0.01	0.001	0.75	0.73
Avg. length of 90% PI ($\mu g/m^3$)	3.08	3.41	3.35	2.60	2.55
Emp. coverage of 90% PI (%)	86.67	76.67	76.67	86.67	83.33

[a] The nonstationary universal kriging model is fitted to the two pollutants individually, the stationary data fusion model is defined using equation 12, but the two underlying spatial processes are assumed to have a stationary covariance function. The nonstationary data fusion model is defined using equation 12 with the covariance function defined using equations 4, 5, and 6 (see section *Single Pollutant Models in Regression-Based Bayesian Data Fusion Models*).

[b] Shown are the mean absolute prediction error (MAPE), root mean squared error (RMSE), Pearson correlation with held-out data, and the average length and empirical coverage of 90% prediction intervals (PI). RLINE = Research LINE-source model.

small marginal variance and a fast-decaying correlation. In contrast, the best regression-based Bayesian data fusion model for $PM_{2.5}$ had an overall additive bias that was estimated to be significantly different from 0, while the localized, spatially varying additive bias had a small marginal variance. Finally, for both pollutants, the best-fitting models contained a multiplicative RLINE bias term that was estimated to be less than 1.

As an example case, Figure 9 shows the predicted $PM_{2.5}$ NRI (and 90% PIs) for Area 8 on the morning of December 20, 2012, yielded by the nonstationary regression-based Bayesian data fusion model with the 2012 emission inventory. The figure displays also the observed NRI and the RLINE output, and again highlights that RLINE underestimated the NRI values of NO_x and $PM_{2.5}$.

The predictive performance of the regression-based Bayesian data fusion models that include, along with the RLINE NRI output, meteorological and traffic covariates, is summarized in Table 10. In agreement with the results presented in Table 9, the stationary data fusion model with the most recent emissions provided the best predictive performance for NO_x in terms of MAPE, RMSE, and correlation coefficient, while the nonstationary model yielded the best performance for $PM_{2.5}$ without much difference among the two sets of emissions.

Table 11 lists estimates of coefficients for the RLINE output, regression coefficients for the meteorological and traffic covariates, and covariance parameters for the best-fitting regression-based Bayesian data fusion models that included meteorological and traffic covariates for NO_x and

Table 9. Posterior Medians and 95% Credible Intervals of Additive and Multiplicative Bias of the RLINE Output Coefficients and Covariance Parameters as Estimated by the Best-Fitting Regression-Based Bayesian Data Fusion Model to NO_x and $PM_{2.5}$, Respectively[a]

Coefficient / Parameter		NO_x[b]		$PM_{2.5}$[b]	
		Median	95% CI	Median	95% CI
α_0	Additive bias	0.05	(−0.47, 0.57)	−0.85	(−1.02, −0.66)
α_1	Multiplicative bias	0.45	(0.32, 0.58)	0.12	(0.006, 0.23)
τ^2	Nugget	1.45	(1.06, 1.98)	0.40	(0.24, 0.65)
σ_1^2	Sill — DW		—	1.01	(0.67, 1.53)
ϕ_1	DW		—	39.25	(36.42, 41.05)
φ_1	DW, WS		—	36.38	(35.16, 37.71)
σ_2^2	Sill — UW		—	0.96	(0.64, 1.45)
ϕ_2	UW		—	39.33	(36.38, 40.79)
φ_2	UW, WS		—	0.21	(0.15, 0.27)
σ_Y^2	Still spat-vary add bias[c]	0.99	(0.67, 1.49)		—
ϕ_Y	Range covar spat-vary add bias[c]	39.34	(36.70, 40.66)		—

[a] CI = credible interval; DW = downwind; RLINE = Research LINE-source model; UW = upwind; WS = wind speed. RLINE output coefficients are in the top part of the table (α_0 & α_1), covariance parameters are in the bottom part of the table.

[b] For NO_x the best fitting model is the stationary regression-based Bayesian data fusion model; for $PM_{2.5}$ the best fitting model is the nonstationary regression-based Bayesian data fusion model.

[c] Sill spat-vary add bias = sill or marginal variance for the spatially varying additive bias; Range covar spat-vary add bias = range parameter for the covariance of the spatially varying additive bias.

$PM_{2.5}$, that is, the stationary model with 2012 emissions for NO_x and the nonstationary model with 2012 emissions for $PM_{2.5}$. After including RLINE in the data fusion model, meteorological and traffic covariates gave no additional information that was useful in explaining the variability in the NO_x NRI, with the exception of distance to the edge of the road, which is significantly and negatively associated with NO_x NRI (the farther from the edge of the road, the lower the NRI concentration for NO_x, as expected). However, the magnitude of the effect of distance from the edge of the road had little consequence on the NO_x NRI concentration: for each additional meter farther from the road, the NO_x NRI concentration is $\exp(−0.003) = 0.997$ times

smaller. Interestingly, estimates of the covariance parameters for the stationary Bayesian data fusion model for NO_x that includes meteorological and traffic covariates are very similar to those in Table 9, which did not include the additional covariate information.

For $PM_{2.5}$, the conclusions are quite different: after accounting for meteorological and traffic covariates, RLINE does not provide any additional information that is useful for estimating the $PM_{2.5}$ NRI concentration. In particular, as seen in Table 11, only the indicator for morning and the interaction of weekday and morning were significantly associated with $PM_{2.5}$ NRI. More specifically, the $PM_{2.5}$ NRI concentration is expected to be greater during
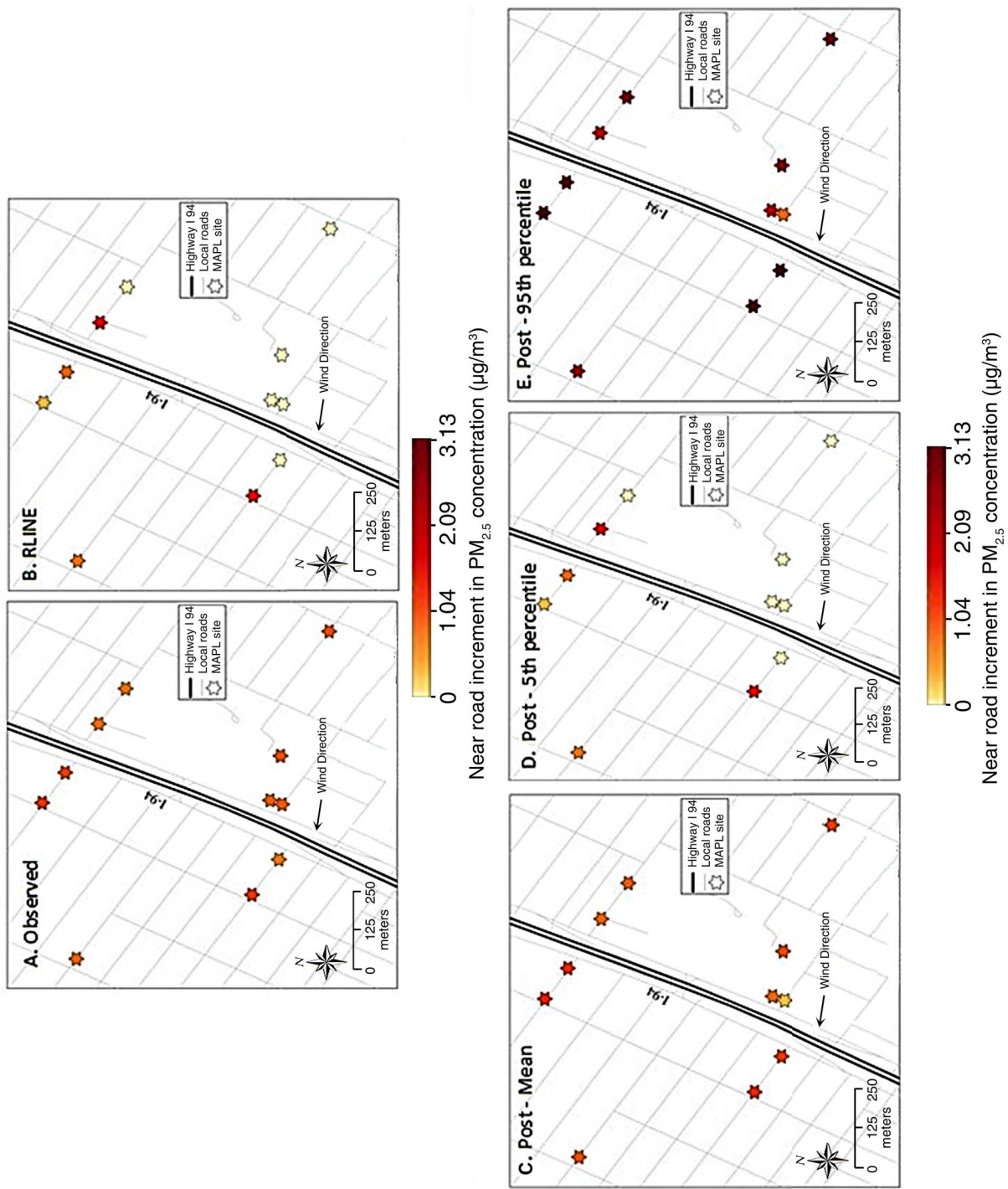
Figure 9. PM$_{2.5}$ NRI concentrations (μg/m³) measured at area 8 at 8:00 am on December 20, 2012 (A, B) and from the nonstationary regression-based Bayesian data fusion model with 2012 emissions (C–E). (A) Observed PM$_{2.5}$ NRI concentrations; **(B)** RLINE output; **(C)** predicted hourly PM$_{2.5}$ concentrations and 90% PI **(D)** lower and **(E)** upper bounds.

Table 10. Predictive Performance of the Two-Pollutant Regression-Based Bayesian Data Fusion Models, Including Meteorological and Traffic Covariates and Averaged Across Validation Sites and Time Periods, Using RLINE with 2010 vs. 2012 Emissions Inventories for Fitting Both NO_x and $PM_{2.5}$ Near-Road Increments [a,b]

Criteria	Stationary Data Fusion (emissions)		Nonstationary Data Fusion (emissions)	
	2010	2012	2010	2012
NO_x				
MAPE (ppb)	6.64	6.18	6.61	6.46
RMSE (ppb)	8.66	8.02	8.90	8.51
Pearson correlation	0.76	0.80	0.72	0.75
Avg. length of 90% PI (ppb)	112.21	114.01	112.76	116.57
Emp. coverage of 90% PI (%)	69.70%	69.70%	69.70%	69.70%
$PM_{2.5}$				
MAPE ($\mu g/m^3$)	0.71	0.71	0.52	0.51
RMSE ($\mu g/m^3$)	1.07	1.07	0.84	0.82
Pearson correlation	0.26	0.24	0.54	0.56
Avg. length of 90% PI ($\mu g/m^3$)	4.16	4.14	3.20	3.14
Emp. coverage of 90% PI (%)	73.33%	73.33%	86.67%	86.67%

[a] Shown are the mean absolute prediction error (MAPE), root mean squared error (RMSE), Pearson correlation with held-out data, and the average length and empirical coverage of 90% prediction intervals (PI). RLINE = Research LINE-source model.

[b] Meteorological and traffic covariates as indicated in equation 13.

the morning, while the NRI concentration is expected to be lower during weekday mornings than the weekday evenings. As observed with NO_x NRI, estimates of the spatial covariance parameters are very similar to those obtained for the regression-based Bayesian data fusion model that included only the RLINE output as predictor. Comparing the predictive performance of the nonstationary regression-Bayesian data fusion models for $PM_{2.5}$, when only RLINE is used as predictor, versus the case where other meteorological and traffic covariates also are included as predictors, we can see that adding additional covariates leads to a small deterioration of predictive performance with respect to MAPE and RMSE criteria. In addition, adding meteorological and traffic covariates leads to less precise $PM_{2.5}$ NRI estimates, as indicated by wider PIs but not higher empirical coverage of the Bayesian data fusion models with additional covariates. This suggests that for prediction purposes, a regression-based Bayesian data fusion model for $PM_{2.5}$ NRI concentration with only $PM_{2.5}$ RLINE output as predictor might

be preferable over a model that includes additional meteorological and traffic covariates.

Multiple-Pollutant Models   Given the large differences in the importance of the RLINE output in predicting the NO_x and $PM_{2.5}$ NRI, and the difference in the form of the spatial covariance structure of the RLINE error, we did not anticipate that jointly modeling these pollutants would be beneficial. Given results for the single-pollutant regression-based Bayesian data fusion models, we developed the multiple-pollutant regression-based Bayesian data fusion model using the updated emissions inventory. The predictive performance of various two-pollutant regression-based Bayesian data fusion models is presented in Table 12. Our expectations were confirmed by the results, which again showed that while the RLINE output was helpful for the $PM_{2.5}$ NRI, it was not for the NO_x NRI, and that the best performance for $PM_{2.5}$ was obtained using a nonstationary model.

Table 11. Posterior Medians and 95% Credible Intervals of the RLINE Output Coefficients, Regression Coefficients of Meteorological and Traffic Covariates, and the Covariance Parameters Estimated by the Best-Fitting Regression-Based Bayesian Data Fusion Models with Additional Covariates to NOx and $PM_{2.5}$[a]

	Parameter	NO_x[b] Median	95% CI	$PM_{2.5}$[c] Median	95% CI
$\tilde{\alpha}_0$		0.94	(0.22, 1.67)	−1.00	(−1.56, −0.44)
$\tilde{\alpha}_1$		0.38	(0.21, 0.55)	0.14	(−0.03, 0.31)
γ_1	DW	0.33	(−0.55, 1.18)	0.33	(−0.34, 1.00)
γ_2	Weekday	0.04	(−0.68, 0.81)	0.07	(−0.53, 0.67)
γ_3	Morning	0.02	(−0.81, 0.88)	1.75	(1.00, 2.46)
γ_4	Weekday*morning	0.01	(−1.06, 1.05)	−1.30	(−2.24, −0.32)
γ_5	(Traffic/WS) $\times 10^{-5}$	−1.87	(−12.1, 8.35)	−1.10	(−9.20, 7.00)
γ_6	(Traffic/WS)*DW $\times 10^{-5}$	4.91	(−8.50, 18.17)	−4.00	(−13.42, 6.05)
γ_7	Distance	−0.003	(−0.005, −0.001)	−0.0006	(−0.002, 0.0007)
γ_8	Distance*DW	0.0003	(−0.002, 0.003)	−0.0008	(−0.003, 0.001)
τ^2	Nugget	1.33	(0.96, 1.81)	0.33	(0.19, 0.53)
σ_1^2	Sill — DW	—		1.06	(0.70, 1.61)
ϕ_1	DW	—		39.43	(36.99, 39.98)
φ_1	DW, WS	—		34.55	(33.75, 37.10)
σ_2^2	Sill — UW	—		0.95	(0.66, 1.40)
ϕ_2	UW	—		39.45	(37.21, 39.98)
φ_2	UW, WS	—		0.21	(0.16, 0.27)
σ_Y^2	Marginal variance (or sill) of the spatially varying additive bias of RLINE	0.99	(0.67, 1.49)	—	
ϕ_Y	Range parameter of the spatially varying additive bias of RLINE	39.46	(37.30, 39.98)	—	

[a] CI = credible interval; DW = downwind; RLINE = Research LINE-source model; UW = upwind; WS = wind speed. RLINE Output calibration coefficients are in the top part of the table ($\tilde{\alpha}_0$ and $\tilde{\alpha}_1$), regression coefficients are in the middle part of the table, and covariance parameters are in the bottom part of the table.

[b] NO_x: the stationary model with 2012 emissions.

[c] $PM_{2.5}$: the nonstationary model with 2012 emissions.

Table 12. Predictive Performance of Various Two-Pollutant Regression-based Bayesian Data Fusion Models for Both NO_x and $PM_{2.5}$ Near-Road Increments[a,b]

Criteria	Nonstationary Universal Kriging	Stationary Data Fusion 2012 Emissions	Nonstationary Data Fusion 2012 Emissions
NO_x			
MAPE (ppb)	8.55	9.36	9.34
RMSE (ppb)	15.94	17.74	17.24
Pearson correlation	0.55	0.47	0.52
Avg. length of 90% PI (ppb)	192.91	145.30	147.38
Emp. coverage of 90% PI (%)	90.0	90.0	90.0
$PM_{2.5}$			
MAPE ($\mu g/m^3$)	0.78	0.82	0.77
RMSE ($\mu g/m^3$)	1.42	1.50	1.41
Pearson correlation	0.17	−0.11	0.23
Avg. length of 90% PI ($\mu g/m^3$)	4.23	3.90	3.76
Emp. coverage of 90% PI (%)	93.33	93.33	93.33

[a] The nonstationary universal kriging model is fitted to the two pollutants individually, the stationary data fusion model is defined using equations 14 and 15, but the two underlying spatial processes are assumed to have a stationary covariance function. The nonstationary data fusion model is defined using equations 14 and 15 and modeled as in equations 4, 5, and 6 (see section *Multiple-Pollutants, Regression-Based Bayesian Data Fusion Model*).

[b] Shown are the mean absolute prediction error (MAPE), root mean squared error (RMSE), Pearson correlation with held-out data, and the average length and empirical coverage of 90% prediction intervals (PI).

DISCUSSION AND CONCLUSIONS

The following summarizes and discusses the main findings pertaining to improving estimates of TRAP exposures, particularly in the near-road zone. Given the limitations of our data, we are cautious about interpreting and generalizing the results of each aim. Still, these findings have a number of implications for health studies and future research that lead to recommendations for improving dispersion modeling and practice, as described later in "Implications of the Findings."

DISPERSION MODELING

Synthesis and Comparison with the Literature

The operational evaluation characterized dispersion modeling performance for daily averages of NO_x and CO at multiple sites in Detroit over a four-year period. Overall, performance metrics for NO_x and CO met the criteria laid out in evaluation guidelines (Chang and Hanna 2004; Hanna and Chang 2012). The performance metrics often, but not always, provided consistent information, although some interpretations can be complex. For example, if R_{SP} is low, then comparisons of FB and V_G across sites might provide little information. Most downwind NO_x and CO predictions were within a factor of two of observations (F2 > 90%); correlation coefficients were moderate to high for NO_x (0.32 to 0.74) but were variable for CO (0 to 0.89), which was limited by instrument sensitivity. Agreement between observed and predicted concentrations improved when monitors were downwind of major roads, but the NO_x concentration was overpredicted and correlation decreased with low NO_x observations and parallel winds. While of significant interest, $PM_{2.5}$ predictions were not evaluated given the limited ability to detect $PM_{2.5}$ emitted

from local sources. This was due to the strength of background and regional sources of $PM_{2.5}$, as well as to the lack of spatially and temporally resolved emissions data for area and non-road mobile emissions, which are substantial, and the limited information regarding non-exhaust emissions (i.e., brake, tire, and road-wear emissions).

Dispersion models like RLINE are expected to perform best at sites that are close to roads without large obstructions, as nearby sources likely contribute a larger fraction of observed concentrations. Also, air flow around buildings and other features is not explicitly modeled by Gaussian plume models. (RLINE simulates near-source dispersion using a general surface roughness parameter and dispersion parameters.) For NO_x and CO, which are emitted primarily from traffic-related sources in urban areas, performance improved with proximity to major roads. The best performance in Detroit was attained at the Eliza Howell near-road site, located close to the busy I-96 freeway. Model performance was also better on weekdays as compared to weekends, which is consistent with higher traffic volume and more regular traffic activity patterns on weekdays that are more consistent with the assumed diurnal traffic trends.

Our findings are largely consistent with prior RLINE evaluations. For example, using downwind 3-hour averages of SF_6 tracer gas at near-road sites in Sacramento, California, F2 > 80%, geometric mean (M_G) =1.18 (Snyder et al. 2013b). Using this same dataset, another study obtained F2 > 78% (Heist et al. 2013). For downwind and hourly SF_6 gas data collected in rural Idaho, F2 = 73% (Heist et al. 2013), and F2 was 75% to 100% (Venkatram et al. 2013). For downwind hourly near-road NO data, F2 = 93% and M_G = 1.12 (Snyder et al. 2013b); for Detroit all-direction hourly NO_x at the schools site, the mean bias was 30% and F2 = 62% (Isakov et al. 2014); and for Detroit downwind near-road NO_x and CO, F2 = 100% (Chang et al. 2015a). We found positive FB at the near-road site, similar to previous work. Overprediction and increased scatter of NO_x during aggregated hours with low NO_x measurements have also been shown (Heist et al. 2013; Snyder et al. 2013b; Venkatram et al. 2013). Using near-road and downwind SF_6 measurements, FB = 0.05 and normalized mean standard error was 0.34 (Heist et al. 2013), comparable to that found in Detroit for NO_x and CO. In contrast to earlier work, we did not show significant overprediction reported for parallel winds (Snyder et al. 2013b) or downwind peaks (Venkatram et al. 2013), and our NMSE estimates were considerably smaller than values reported in a recent RLINE evaluation (Heist et al. 2013). These differences likely arose from our inclusion of background and point sources (also performed in one other study [Isakov et al. 2014]), use

of daily averages, and differences in the estimated background.

Operational evaluations should be distinguished from diagnostic, dynamic, and probabilistic types of model evaluation. Comparisons to previous RLINE evaluations, performed primarily for diagnostic purposes, are limited by several factors. First, we examined daily concentrations, which are relevant to many epidemiological applications, and we did not focus on performance as a function of meteorological conditions. Lower performance and overprediction have been reported during stable periods, in other words, periods with low wind speeds (Heist et al. 2013; Snyder et al. 2013b; Venkatram et al. 2013). Second, performance during upwind periods was not evaluated (observations during these periods were used to estimate background); prior studies show overprediction and increased scatter at upwind receptors (Heist et al. 2013; Snyder et al. 2013b). Third, our large-scale and multiyear urban application used data from a sparse (though typical) air quality monitoring network, so the ability to assess spatial performance was limited. In comparison, other studies have used tracer gases, a higher density of monitoring sites, and a small study domain (<1 km^2) containing few sources.

Evaluation of Dispersion Modeling

Dispersion model predictions have been compared to monitored observations using diagnostic and operational evaluations, as well as sensitivity analyses, in this report and elsewhere. Here we focused on operational evaluations using daily average concentrations measured over a four-year period at near-road monitoring sites across Detroit. This evaluation is important due to its long record, urban scale application, and presumed greater relevance for health studies.

Overall, predictions of NO_x and CO met performance criteria laid out in evaluation guidelines (Chang and Hanna 2004; Hanna and Chang 2012). For example, NO_x and CO predictions were mostly within a factor of two of observations (F2 > 90%), and correlation coefficients were moderately high for NO_x (R_{SP} = 0.32 to 0.74), but more variable for CO (0 to 0.89). However, CO comparisons were likely impaired by the monitor performance (many observations below the instrument's detection limits). As noted, performance improved when monitors were downwind of major roads, at sites closer to major roads, on weekdays, and during winter and spring seasons. The evaluation was not informative for $PM_{2.5}$ due to the scarcity of $PM_{2.5}$ monitors near major roads, the lack of spatially and temporally resolved emissions inventory from non-road sources (which appear significant), and the presence of high

regional or background levels of PM$_{2.5}$. Sensitivity analyses showed scant changes from using updated emission factors (2012 estimates from MOVES 2014 with link-match traffic activity, compared to MOVES 2011 with default traffic activity) or Detroit-specific temporal allocation factors for traffic activity (compared to national defaults), but showed larger differences and improved performance when using on-site (or local) meteorological data (compared to airport data). These findings generally were consistent across most sites and with the literature. Overall, they suggest the usefulness of dispersion modeling for estimating spatially and temporally resolved exposure estimates.

Several issues are highlighted regarding the use of dispersion models in health studies. First, modeling can be data intensive and computationally demanding, largely due to the size and complexity of link-based emission inventories, the temporal and spatial resolution needed, and the use of numerical approaches for dispersion calculations (e.g., the default option in RLINE). Even after revisions to facilitate input/output and other operations, modeling remained computationally intensive given the number of links (9,700) and receptors (often hundreds or thousands) of interest in the NEXUS application. Second, Gaussian plume models like RLINE cannot model calm conditions. This means that calculations are not performed if wind speeds fall below 0.5 or 1.0 m/s, thus, concentrations for these periods are unavailable or, if multihour averages are computed, not used in computing averages. Third, RLINE and most plume models have limited or no ability to address the evolution of aerosols, vapors, and gases that can produce organic aerosols, NO to NO$_2$ conversion, and other secondary pollutants (Pant and Harrison 2013).

Uncertainty and Limitations

Many factors affect comparisons between observed and predicted pollutant concentrations. Our results show the importance of selecting pollutants that are predominately traffic based and measured with sufficient sensitivity. The use of monitoring parameters more specific to TRAP, for example, BC, and possibly ultrafine PM, would be valuable. While detailed, the mobile source inventory used estimates of traffic volumes and time allocation factors derived from mostly larger roads; the MOVES emission factors for the greater Detroit area may not have fully reflected local traffic volume, vehicle mix, and emissions. Point sources were aggregated to the facility level and used average emission rates. Temporal variability was not modeled. Background estimates only partly accounted for regional sources and may not have fully represented short-term fluctuations and gradients. (Some studies have used

complex regional chemical models to estimate background [Arunachalam et al. 2014].) The classification of downwind and parallel periods refers to only the nearest major road. Handling secondary pollutants, including secondary organic aerosols, is a gap that may increase in importance as primary emissions continue to be reduced due to more stringent emission controls. Finally, the fewer observations available on weekends may have influenced results.

Overall, results highlight the sensitivity of evaluation results to monitor placement, instrument sensitivity, and the ability to observe contributions from local sources. Results for NO$_x$ appear most meaningful given the NO$_x$ instrumentation's greater sensitivity and ability to detect traffic-related emissions. In contrast, the CO evaluation was limited by low detection frequencies at some sites, a product of high detection limits that reduced the number of valid observations (especially important when analyses were stratified by wind direction, day of week, and season).

Finally, to confirm and extend our results, other operational performance evaluations and sensitivity analysis should be conducted across a range of urban settings.

Enhancements to Dispersion Modeling

We identified a number of opportunities to enhance urban scale dispersion models and model inputs. Here we focus on the emissions inventory and the meteorological inputs to dispersion models.

The link-based on-road mobile source emissions inventories used in RLINE and other dispersion models are based on estimates of emission factors, traffic activity (e.g., vehicle volume, speed, and mix), and the representation of the road network. These inventories can have large uncertainties for many reasons, including the spatial and temporal variability of many of the parameters and governing processes and the limited nature and number of measurements supporting traffic activity and emission factor estimates. We examined differences between the inventory originally used in NEXUS modeling and an updated inventory that corrected some of the link geometry, updated AADT and vehicle mix information, incorporated site-specific temporal allocation factors, and used more recent MOVES emission factors. These are intensive and nontrivial revisions, in part due to the size of urban networks (9,700 links to represent the larger roads in Detroit alone), the diversity of the datasets, and the complex and many choices needed to match available data to requirements of MOVES and RLINE inputs. While some of the revisions improved dispersion modeling performance, often they did not appear to make substantial differences as shown by sensitivity analyses (for NO$_x$, CO, and PM$_{2.5}$) and operational evaluations (for NO$_x$ and CO), especially

when daily average concentrations were examined. Effects of the revisions may be masked or compensated by other factors that affect pollutant levels in real-world urban-scale settings.

For PM$_{2.5}$, as noted earlier, we could not clearly distinguish the traffic-related component from other sources of PM$_{2.5}$ using routine observations collected at the AQS monitoring sites in Detroit. This component could be distinguished, however, in the evaluation of the Bayesian fusion models, which utilized short-term PM$_{2.5}$ measurements collected on transects across major Detroit roads during rush hours. In this case, RLINE provided useful, though biased, information regarding PM$_{2.5}$ concentrations. While an important driver of health effects, the PM$_{2.5}$ emission estimates appear to have especially large uncertainties. This includes sizable uncertainties from urban non-road and area sources. While beyond the scope of this report, contributions from these sources contribute to urban PM levels and need appropriate quantification to evaluate model performance in real-world settings. Possibly, uncertainties of PM$_{2.5}$ emissions in urban settings may not be greatly reducible without improved approaches and data for characterizing these sources, as well as exhaust and non-exhaust emissions of on-road sources.

The analysis of meteorological inputs to RLINE demonstrated the sensitivity of RLINE predictions to the choice of meteorological inputs. In a limited diagnostic evaluation, for example, downwind concentrations were overpredicted and concentrations with winds parallel to the major nearby road were underpredicted, as were concentrations at low wind speeds (Appendix 10). Similar results were seen in the operational evaluation that may be more pertinent to health study investigations. Typically, the best performance was found using on-site (or local) meteorological inputs, that is, surface data collected at or near the road being modeled, as compared to airport data, which is the dataset most commonly employed in dispersion modeling. Meteorological inputs determine the stability conditions that govern dispersion, and very large differences in concentrations can result from convective and neutral conditions as compared to stable and very stable conditions (Appendix 3). It is unsurprising but worth reiterating that the selection of meteorological inputs is critical. Unfortunately, few near-road monitoring sites (or other air quality monitoring sites) have the full suite of instrumentation required to generate input files for dispersion models.

APPLICATION OF SPATIOTEMPORAL MODELS

Several sets of spatiotemporal models were developed and fitted to concentrations and near-road concentration increments of traffic-related pollutants measured in transects around major roads. The first set, called *nonstationary universal kriging models*, leveraged information in geographical and traffic covariates to capture the nonstationarity in concentrations observed upwind and downwind of the road. This nonstationarity was hypothesized to result in concentration differences due to effects of the spatial configuration of emissions (e.g., being upwind or downwind) and meteorological factors (e.g., wind speed and direction). To formalize such hypotheses, the concentration field was modeled as a weighted spatiotemporal mixture of two mutually independent spatiotemporal processes, loosely interpretable as concentrations upwind and downwind, with wind speed and wind direction influencing the mixture weights and the dependence structure of each latent process. This model specification is flexible and admits simpler stationary and nonstationary models as special cases. The application of this model to NO, NO$_x$, and BC confirmed that wind speed and direction are important drivers of the observed nonstationarity in pollutant concentrations. In particular, the analysis indicated that concentrations at downwind sites disperse quickly even at low wind speeds, and that concentrations at upwind sites decay rather slowly and with increasing wind speed. Accounting for such spatial dependencies improved predictions of TRAP concentration at unsampled locations compared to other stationary and nonstationary models, for example, resulting in a lower MAPE and appropriate quantification of uncertainty in the prediction.

While the general specification of the universal kriging models can be applied to large spatial domains, our goal was to model concentrations in the near-road environment. This motivated our specification of concentrations as a mixture of two underlying spatial processes, one receiving larger weight at downwind sites and the other at upwind sites. This particular specification is applicable only within a small neighborhood around an urban highway. Modeling on a larger scale is likely to encounter the presence of multiple roads, in other words, the same site can be downwind of one road and upwind of another. In such cases, modeling would require appropriate modifications.

We next assessed whether outputs from the RLINE dispersion model could improve predictions of NRI of TRAP concentrations. Two sets of spatiotemporal Bayesian data fusion models were developed and implemented: joint models that provide a stochastic formulation of both the observed NRI and the RLINE output, and regression-based models that use the RLINE output as the only covariate in spatiotemporal models for the observed NRI. This analysis used essentially the same RLINE modeling considered in

the operational evaluation. Building upon findings of the nonstationary universal kriging modeling, we postulated that the spatial dependence structure of the NRI field is also nonstationary and used the same covariance function with wind speed and direction as drivers of nonstationarity. Potential biases in the RLINE output were explored, including whether the additive bias of the RLINE output should be modeled as constant or spatially varying. Interestingly, even though the two Bayesian data modeling approaches used different data, the same conclusion was reached: the RLINE model displays a spatially varying additive bias for both NO_x and $PM_{2.5}$. In addition, RLINE does not provide much information for NO_x beyond what is already contained in traffic and GIS covariates, but for $PM_{2.5}$ the RLINE output is useful. While the RLINE output accounts for some of the nonstationarity in the NO_x NRI field to the point that the residual spatial correlation could be modeled as stationary, that is not the case for $PM_{2.5}$.

Given the similarity in results of the two Bayesian data fusion modeling approaches, in choosing between them we opted for the regression-based approach, simply because it is computationally less challenging. Using this approach, we investigated whether the updated (2012) emission inventory yielded a better RLINE output than the original (2010) inventory. While the updated inventory was more useful for NO_x, there was not much difference between the two for $PM_{2.5}$. This implies that with improvements in the emissions information (and possibly with other enhancements in dispersion modeling), RLINE could become more useful for deriving estimates of near-road NO_x concentrations. Somewhat similarly, the dispersion modeling evaluation showed only modest changes in sensitivity analyses examining the emission inventory.

Lastly, we investigated whether modeling NO_x and $PM_{2.5}$ jointly led to predictive benefits. We found that leveraging information on each pollutant's NRI was not useful for predicting the other, as expected given that RLINE was helpful in predicting concentrations of $PM_{2.5}$ but not NO_x, and given the noted differences in the spatial dependence structures of the NRI fields for these pollutants.

Evaluation of Spatiotemporal Modeling

The spatiotemporal and Bayesian data fusion models fitted to ambient concentrations and near-road concentration increments of TRAPs showed the ability to improve predictions at unsampled locations by: (1) accounting for characteristics of the spatiotemporal processes in the observed monitoring data, and (2) exploiting information in the RLINE output or in a small set of covariates plausibly related to emissions and dispersion of pollutants from on-road sources (e.g., wind speed, wind direction,

and traffic volume). The specification of these models can be quite flexible and can incorporate a range of spatiotemporal structures (e.g., stationarity assumptions and covariates). With appropriate formulation, these models can also serve important diagnostic functions, indicating parameters that may drive the variation in concentrations (e.g., nonstationarity such as differences found between upwind and downwind conditions), and also importantly, showing the existence and characteristics of errors in dispersion model predictions overall and in model components. For example, we found spatially varying additive bias in RLINE predictions of NO_x and $PM_{2.5}$, but generally negligible effects for using an updated emissions inventory. (The latter also was shown in the operational evaluation and sensitivity analysis).

We did not find benefits in jointly modeling NO_x and $PM_{2.5}$. Our results are based on short-term measurements on transects across a variety of major roads in Detroit collected during a limited number of wintertime morning and afternoon rush-hour periods. Additional evaluations using other datasets are necessary for confirmation and generalization of this result.

Even though the class of spatiotemporal models presented in this report has been developed to accommodate characteristics of the Detroit transect dataset, we believe that the models could be adapted and utilized outside of the Detroit study area. Adaptations to our model formulation will be dictated by the study design used to collect the data. For example, data might not be collected in different, separate areas of a metropolitan region, or at different times of day, as was the case in our transect dataset. The characteristics of our dataset led us to formulate a model with temporal independence across time periods and spatial independence across areas, which might not be appropriate for a different dataset. Regardless of considerations of potential temporal and spatial independence, the formulations of the models shown as equations 2 through 12 are very general. They could be applied to datasets outside of the Detroit area as long as information on traffic volume (ideally, commercial and noncommercial), wind speed, and wind direction are available over time and space.

Our dataset of observed TRAP concentrations was of moderate size, which led to rather fast computation with sophisticated and complex statistical spatiotemporal models. In modeling applications utilizing large datasets, it might be advisable to either incorporate some of the newly proposed methods for large spatial datasets, such as the nearest neighbor Gaussian process modeling approach (Datta et al. 2016) or the multiresolution approximation method (Katzfuss 2017). Alternatively, one could replace estimation via a Markov chain Monte Carlo algorithm with

the nested approximation approach of integrated nested Laplace approximations (Rue et al. 2009).

Overall, we show that RLINE can provide useful information when combined with observations that can be exploited to derive improved spatiotemporal estimates of TRAP concentrations for use in health studies. The Detroit application suggests that this applies to PM$_{2.5}$ and that more work is needed to improve NO$_x$ predictions (e.g., potentially addressing the RLINE model itself, the emissions inventory, and possibly other factors). Empirical evaluation will suggest whether this is true for other TRAPs not considered here. In addition, further effort is warranted to ensure that RLINE correctly captures the non-stationary behavior of near-road PM$_{2.5}$ concentrations seen in the Detroit data.

A key advantage of the spatiotemporal models is their ability to provide appropriate quantification of uncertainty in the prediction, information that is highly relevant to improving estimates of TRAP exposures in epidemiological studies given the nature of exposure measurement errors.

Uncertainties

Our results are based on short-term measurements on transects across a variety of major roads in Detroit collected during wintertime morning and afternoon rush-hour periods. Transect locations were selected to represent predominantly residential areas across the city, and the selected roads differ significantly with respect to traffic volume and fleet mix. In these ways, the collected data are representative and relevant to health studies using cohorts based on residence location. Areas with large industrial emissions sources, which are common in the Detroit area, were avoided. Still, PM$_{2.5}$, NO$_x$, and other pollutants in TRAP are emitted from numerous sources in addition to the roads studied, including point, non-road mobile, area, and regional sources. Also, the traffic-related component of PM$_{2.5}$ is small, as discussed elsewhere in this report (e.g., Appendix 4). The use of the NRI and multiple time periods monitored for each transect clearly showed the roadway influence and expected trends, for example, higher concentration gradients during mornings. In addition, the RLINE modeling benefited from its detailed application to Detroit and the many refinements described previously.

The implementation and evaluation of spatiotemporal models using additional datasets would be useful for extending our findings. In particular, applications using long-term (e.g., seasonal to annual average) observations might be particularly revealing, especially since dispersion modeling performance at short averaging periods often is not very good. Moreover, spatiotemporal models require a sufficient number of sampling locations to estimate parameters. Given the correlation scale found, some pairs of monitoring sites should be quite close, for example, within several hundred meters, as well as near major roads. Unfortunately, few datasets provide the sufficient spatial coverage needed to develop and assess long-term urban-scale spatiotemporal models, and current monitoring networks are not designed to capture the small-scale gradients of TRAP pollutants. There may be applications using satellite data or possibly low-cost monitoring sensors if sufficient spatial resolution, accuracy, sensitivity, and selectivity for TRAPs is attainable.

As suggested above, results of our analyses should be interpreted and generalized cautiously given the limited spatiotemporal nature of the observational data. Future statistical and scientific efforts directed toward predicting TRAP concentrations at unsampled locations should be cognizant of the importance and necessity of the extensive data collection and monitoring efforts needed to obtain valid and robust results.

IMPLICATIONS OF THE FINDINGS

IMPLICATIONS FOR EPIDEMIOLOGICAL STUDIES

Dispersion modeling to develop spatially and temporally resolved exposure estimates of TRAP for epidemiological studies has potentially significant advantages over other approaches. However, it is important to account for model performance and exposure measurement errors (or exposure misclassification in the case of categorical exposure variables). These errors may vary spatially or temporally, and they may differentially affect groups of study participants, with the potential of affecting health study outcomes.

The operational evaluation and sensitivity analyses suggested that dispersion model performance is best at near-road sites (e.g., within 10 to 100 meters of the road) and that uncertainty increases with distance from roadways. RLINE represented much of the day-to-day variation observed in daily average concentrations, suggesting that dispersion modeling can provide near-road (and potentially on-road) exposure predictions with good fidelity. This is important since many people live or work near roads where TRAP concentrations are highest (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010). While these results may be driven by the ability to discern contributions from local emission sources, dispersion model performance is likely to degrade with distance in urban settings for several reasons (Jerrett et al. 2005), for example, shifts in wind fields, the presence of unknown or

unmodeled sources (including other local roads), and atmospheric transformation and other unmodeled processes. Thus, at farther distances, daily fluctuations in concentrations may be less accurately estimated. This increases the likelihood of errors from dispersion model-based estimates if the study population is exposed over a range of distances from major roads. Such studies might benefit from weighting exposure estimates by their uncertainties. In contrast, study designs using only participants exposed near roads will have the advantages of higher concentrations of TRAP and potentially lower and more comparable exposure measurement error.

A second concern is the effect of wind direction relative to the orientation of (major) roads and locations of study participants. Dispersion models perform best at downwind receptors, in other words, when winds are approximately perpendicular to the road's orientation. Correlation between the prevailing wind direction(s), road alignment(s), and study participant locations might lead to differential errors. For example, in Detroit, prevailing winds come from the west and southwest. Thus, models will perform best for roads with north-south and northwest-southeast alignments with study participants on the downwind side; conversely, performance will be worse for roads that are aligned with (or parallel to) the prevailing wind directions or with participants in upwind locations. These errors were investigated in Detroit by identifying the nearest major road (within 150 meters; AADT > 10,000) for a random sample of residences (n = 4,000). Most roads are aligned on a north-south or east-west axis, and thus the direction from a residence to the nearest major road is mainly north and south. Based on prevailing winds and the largest roads, individuals living downwind are east of north-south roads (e.g., M-10, M-39, I-75), individuals living upwind are on the west side of the same roads, while individuals living south or north of east-west roads (e.g., I-96, I-94) will often experience parallel winds. Even if all individuals in a study lived at similar distances and/or had similar TRAP exposure, there is an increased likelihood of exposure measurement error for upwind and parallel groups. In general, population patterns and the importance of directional effects can depend on many factors, for example: clustering of residences, schools, workplaces (Cable 2013; Fessenden and Roberts 2011), geographic boundaries (mountains, coastlines), economic factors (real estate), and administrative factors (municipal boundaries). Some concerns might be addressed by selecting appropriate areas, or by using weights to account for prediction uncertainty.

Other implications for epidemiological studies arise from the day-of-week variation in model performance and the reliability of the time–activity information needed to assign exposures. Consider a statistical model associating health outcomes with the prior day's exposure (e.g., outcomes on Sundays and Mondays require exposure estimates for Saturdays and Sundays). Many models use 3- to 5-day lags. With a 3-day lag, for example, the Sunday through Wednesday outcomes require weekend exposure data. Given lower performance of the dispersion model and greater uncertainty (as well as variability) of weekend time–activity information, exposure measurement errors may increase from Saturday through Wednesday. Thus, a study incorporating 3-day exposure lags might have the effect of emphasizing health data for Thursdays, Fridays, and possibly Saturdays when exposure uncertainty is smaller. Again, weights accounting for the greater uncertainty of weekend exposure estimates (and possibly sensitivity analyses) might help control for these effects. A related concern is RLINE's tendency to underpredict on weekdays, which might: (1) bias concentration–outcome relationships if the (estimated) exposure variability is compressed; (2) increase uncertainty, since health models typically include both weekday and weekend periods; and (3) falsely attribute variation to day-of-week or weekday/weekend covariates, if used. Such effects are hypothetical. Calibrating the dispersion model (i.e., mobile source inventory, TAFs) and the exposure assumptions might help to resolve this issue.

Seasonal variation in dispersion model performance, while less consistent than the day-of-week effects, raises other concerns in epidemiological applications. This variation can be coupled to seasonal time–activity information that affects exposure; for example, the summer school holiday period for children can increase uncertainty since the home-school-home pattern is absent or less consistent and because of increased time spent outdoors. In addition, summer traffic patterns can have greater variability, a result of vacation and holiday travel and decreased commuting.

Model estimates are sensitive to input data, and our applications highlighted the need for representative meteorological data to predict near-road exposures.

RECOMMENDATIONS

1. Quantify uncertainty in the exposure estimates. This is a key advantage of and motivation for spatiotemporal statistical models, which have the ability to provide appropriate quantification of uncertainty in the prediction. This information should be utilized in health studies given the potentially deleterious effects of exposure measurement errors.

2. Conduct operational performance evaluations across a range of urban settings, thus developing an ensemble of evaluations that can provide robust and

representative results. Potentially utilize the online CLINE application at each of the near-road sites for this purpose.

3. Develop and evaluate spatiotemporal and Bayesian data fusion models using additional datasets, specifically those longer duration records with multiple daily, seasonal, or annual averages.

4. Utilize instruments at air quality monitoring sites for TRAPs that have appropriately low detection limits (i.e., trace-level capability) to ensure high detection frequencies. This applies to CO and other pollutants.

5. Harmonize traffic data types (e.g., vehicle classifications) collected by federal, state, and local authorities with MOVES (emission factor model) vehicle categories.

6. Equip air quality monitoring sites, and especially the near-road sites, with sufficient meteorological instrumentation to generate the AERMET meteorological preprocesser files necessary to run dispersion models.

7. Utilize on-site or local meteorological inputs in dispersion modeling.

8. Develop guidance that defines appropriate and representative meteorological data for dispersion modeling of the complex near-road environment.

9. Undertake studies to better characterize the non-exhaust component of $PM_{2.5}$ emissions from on-road sources, and more generally, improve the spatial and temporal resolution of urban emission inventories.

10. Consider locations and time–activity factors of participants in health studies.

11. Improve the computational performance of dispersion and spatiotemporal models to increase the feasibility of application. For dispersion modeling, analytical approaches, distance and concentration cut-offs, and other changes might be made to speed algorithms and data handling. For spatiotemporal models, we found that regression-based Bayesian data fusion approaches were not computationally burdensome and provided results that were generally comparable to joint models.

12. Consider the development and support of a standardized model package for spatiotemporal models, thus minimizing or avoiding tailored formulation and custom programming.

13. Develop databases appropriate for developing and evaluating models for TRAP. For dispersion modeling, consolidating data collected in the near-road ambient monitoring network and extending this with necessary meteorology, emission inventory, and meteorological information (recommendations 4–7) should be considered. For spatiotemporal models, more

spatial coverage in the near-road environment is required, and potentially a hybrid approach using conventional fixed-site, mobile, and/or transportable monitoring approaches could be used to estimate the 24-hour concentrations most relevant to TRAP exposures.

ACKNOWLEDGMENTS

REFERENCES

Adam-Poupart A, Brand A, Fournier M, Jerrett M, Smargiassi A. 2014. Spatiotemporal modeling of ozone levels in Quebec (Canada): A comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy-LUR approaches. Environ Health Perspect 122: 970–976.

Adar SD, Gold DR, Coull BA, Schwartz J, Stone PH, Suh H. 2007. Focused exposures to airborne traffic particles and heart rate variability in the elderly. Epidemiology 18:95–103.

Anderson HR, Favarato G, Atkinson RW. 2013. Long-term exposure to air pollution and the incidence of asthma: Meta-analysis of cohort studies. Air Qual Atmos Health 6:47–56.

Arunachalam S, Valencia A, Akita Y, Serre ML, Omary M, Garcia V, et al. 2014. A method for estimating urban background concentrations in support of hybrid air pollution modeling for environmental health studies. Int J Environ Res Public Health 11:10518–10536.

Baldwin N, Gilani O, Raja S, Batterman S, Ganguly R, Hopke P, et al. 2015. Factors affecting pollutant concentrations in the near-road environment. Atmos Environ 115: 223–235.

Banerjee S, Gelfand A, Knight JR, Sirmans C. 2004. Spatial modeling of house prices using normalized distance-weighted sums of stationary processes. J Bus Econ Stat 22: 206–213.

Batterman S, Burke J, Isakov V, Lewis T, Mukherjee B, Robins T. 2014a. A comparison of exposure metrics for traffic-related air pollutants: Application to epidemiology studies in Detroit, Michigan. Int J Environ Res Public Health 11:9553–9577.

Batterman S, Chambliss S, Isakov V. 2014b. Spatial resolution requirements for traffic-related air pollutant exposure evaluations. Atmos Environ 94:518–528.

Batterman S, Cook R, Justin T. 2015. Temporal variation of traffic on highways and the development of accurate temporal allocation factors for air pollution analyses. Atmos Environ 107:351–363.

Baxter LK, Dionisio KL, Burke J, Sarnat SE, Sarnat JA, Hodas N, et al. 2013. Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. J Expo Sci Environ Epidemiol 23:654–659.

Berrocal VJ, Gelfand AE, Holland DM. 2010a. A bivariate space-time downscaler under space and time misalignment. Ann Appl Stat 4:1942–1975.

Berrocal VJ, Gelfand AE, Holland DM. 2010b. A spatiotemporal downscaler for outputs from numerical models. J Agric Biol Environ Stat 15:176–197.

Berrocal VJ, Gelfand AE, Holland DM. 2012. Space-time data fusion under error in computer model output: An application to modeling air quality. Biometrics 48:837–848.

Boehmer TK, Foster SL, Henry JR, Woghiren-Akinnifesi EL, Yip FY, Centers for Disease Control and Prevention. 2013. Residential proximity to major highways — United States, 2010. MMWR Suppl 62:46–50.

Brauer M, Reynolds C, Hystad P. 2013. Traffic-related air pollution and health in Canada. Can Med Assoc J 185: 1557–1558.

Brunekreef B, Janssen NA, de Hartog J, Harssema H, Knape M, van Vliet P. 1997. Air pollution from truck traffic and lung function in children living near motorways. Epidemiology 8:298–303. Available: *www.jstor.org/stable/3702257 ?origin=JSTOR-pdf&seq=1#page_scan_tab_contents*.

Cable D. 2013. 2010 Racial dot map. Weldon Cooper Center for Public Service, University of Virginia [Online 11/28/2016] Available:*https://demographics.cooper center.org/racial-dot-map*.

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. Measurement error in nonlinear models: A modern perspective. Boca Raton, FL:CRC Press.

Chan TW, Meloche E, Kubsh J, Brezny R, Rosenblatt D, Rideout G. 2013. Impact of ambient temperature on gaseous

and particle emissions from a direct injection gasoline vehicle and its implications on particle filtration. SAE Int J Fuels Lubricants 6:350–371.

Chang JC, Hanna SR. 2004. Air quality model performance evaluation. Meteorol Atmos Phy 87:167–196.

Chang SY, Vizuete W, Breen M, Isakov V, Arunachalam S. 2015a. Comparison of highly resolved model-based exposure metrics for traffic-related air pollutants to support environmental health studies. Int J Environ Res Public Health 12:15605–15625.

Chang SY, Vizuete W, Valencia A, Naess B, Isakov V, Palma T, et al. 2015b. A modeling framework for characterizing near-road air pollutant concentration at community scales. Sci Total Environ 538:905–921.

Choi J, Fuentes M, Reich BJ. 2009. Spatial-temporal association between fine particulate matter and daily mortality. Comp Stat Data Anal 53:2989–3000.

Cimorelli AJ, Perry SG, Venkatram A, Weil JC, Paine RJ, Wilson RB, et al. 2004. AERMOD: A dispersion model for industrial source applications. Part I: General model formulation and boundary layer characterization. J Appl Meteorol 44:682–693.

Claggett M, Niemeier D, Eisinger D, Bai S, Chen H. 2009. Predicting near-road $PM_{2.5}$ concentrations. Transport Res Rec 2123:26–37.

Colvile RN, Woodfield NK, Carruthers DJ, Fisher BEA, Rickard A, Neville S, et al. 2002. Uncertainty in dispersion modelling and urban air quality mapping. Environ Sci Policy 5:207–220.

Cressie NAC. 1993. Statistics for Spatial Data, 2nd edition. Hoboken, New Jersey:Wiley.

Crooks JL, Özkaynak H. 2014. Simultaneous statistical bias correction of multiple $PM_{2.5}$ species from a regional photochemical grid model. Atmos Environ 95:126–141.

Datta A, Banerjee S, Finley AO, Gelfand AE. 2016. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. J Amer Stat Assoc 111:800–812.

Dennis R, Fox T, Fuentes M, Gilliland A, Hanna S, Hogrefe C, et al. 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. Environ Fluid Mech 10:471–489.

Dionisio KL, Baxter LK, Burke J, Özkaynak H. 2016. The importance of the exposure metric in air pollution epidemiology studies: When does it matter, and why? Air Qual Atmos Health 9:495–502.

Dons E, Van Poppel M, Kochan B, Wets G, Panis LI. 2014. Implementation and validation of a modeling framework to assess personal exposure to black carbon. Environ Int 62:64–71.

English P, Neutra R, Scalf R, Sullivan M, Waller L, Zhu L. 1999. Examining associations between childhood asthma and traffic flow using a geographic information system. Environ Health Perspect 107:761–767.

Fang SC, Schwartz J, Yang M, Yaggi HK, Bliwise DL, Araujo AB. 2015. Traffic-related air pollution and sleep in the Boston Area Community Health Survey. J Expo Sci Environ Epidemiol 25:451–456.

Fessenden F, Roberts S. 2011. Then as now — New York's shifting ethnic mosaic. NYT. Available: *https://archive.nytimes.com/query.nytimes.com/gst/full page-9B05E4D7123EF930A15752C0A9679D8B63.html* [accessed 15 April 2019].

Fuentes M. 2001. A high frequency kriging approach for nonstationary environmental processes. Environmetrics 12:469–483.

Fuentes M, Raftery AE. 2005. Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. Biometrics 66:36–45.

Ganguly R, Batterman S, Isakov V, Snyder M, Breen M, Brakefield-Caldwell W. 2015. Effect of geocoding errors on traffic-related air pollutant exposure and concentration estimates. J Expo Sci Environ Epidemiol 25:490–498.

Giambini P, Salizzoni P, Soulhac L, Corti A. 2012. Influence of Meteorological Input Parameters on Urban Dispersion Modelling for Traffic Scenario Analysis. In: Air Pollution Modeling and its Application (Steyn DG, Trini Castelli S, eds). Dordrecht, Netherlands:Springer, 453–457.

Gilani O, Berrocal VJ, Batterman SA. 2015. Non-stationary spatio-temporal modeling of traffic-related pollutants in near-road environments. Spatial Spatio-temporal Epidemiol 18:24–37.

Gilani O, McKay LA, Gregoire TG, Guan Y, Leaderer BP, Holford TR. 2016. Spatiotemporal calibration and resolution refinement of output from deterministic models. Stat Med 35:2422–2440.

Grimm H, Eatough DJ. 2009. Aerosol measurement: The use of optical light scattering for the determination of particulate size distribution, and particulate mass, including

the semi-volatile fraction. J Air Waste Manage Assoc 59:101–107.

Gurram S, Stuart AL, Pinjari AR. 2015. Impacts of travel activity and urbanicity on exposures to ambient oxides of nitrogen and on exposure disparities. Air Qual Atmos Health 8:97–114.

Hanna SR. 2007. Chapter 4.0. A review of uncertainty and sensitivity analyses of atmospheric transport and dispersion models. In: Developments in Environmental Science, Vol. 6: Air Pollution Modeling and Its Application XVIII, (Borrego C, Renner E, eds). Atlanta, GA:Elsevier, 331–351.

Hanna S, Chang J. 2012. Acceptance criteria for urban dispersion model evaluation. Meteorol Atmos Phy 116:133–146.

Hannam K, McNamee R, De Vocht F, Baker P, Sibley C, Agius R. 2013. A comparison of population air pollution exposure estimation techniques with personal exposure estimates in a pregnant cohort. Environ Sci Process Impacts 15:1562–1572.

Heck JE, Wu J, Lombardi C, Qiu JH, Meyers TJ, Wilhelm M, et al. 2013. Childhood cancer and traffic-related air pollution exposure in pregnancy and early life. Environ Health Perspect 121:1385–1391.

HEI Panel on the Health Effects of Traffic-Related Air Pollution. 2010. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Special Report 17. Boston, MA:Health Effects Institute.

Heist D, Isakov V, Perry S, Snyder M, Venkatram A, Hood C, et al. 2013. Estimating near-road pollutant dispersion: A model inter-comparison. Transport Res Part D: Transport Environ 25:93–105.

Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, et al. 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 42:7561–7578.

Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA. 2002. Association between mortality and indicators of traffic-related air pollution in the Netherlands: A cohort study. Lancet 360:1203–1209.

Huang Y-L, Batterman S. 2000. Residence location as a measure of environmental exposure: A review of air pollution epidemiology studies. J Expos Anal Environ Epidemiol 10:66–85.

Huo H, Zhang Q, He K, Wang Q, Yao Z, Streets DG. 2009. High-resolution vehicular emission inventory using a link-based method: A case study of light-duty vehicles in Beijing. Environ Sci Tech 43:2394–2399.

International Agency for Research on Cancer (IARC). 2014. Diesel and gasoline engine exhausts and some nitroarenes. IARC monographs on the evaluation of carcinogenic risks to humans. IARC Monogr Eval Carcinog Risks Humans 105:9–699.

Isakov V, Arunachalam S, Batterman S, Bereznicki S, Burke J, Dionisio K, et al. 2014. Air quality modeling in support of the near-road exposures and effects of urban air pollutants study (NEXUS). Int J Environ Res Public Health 11:8777–8793.

Jammalamadaka SR, Sengupta A. 2001. Topics in Circular Statistics. Series on Multivariate Analysis vol 5. Singapore:World Scientific.

Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, et al. 2005. A review and evaluation of intraurban air pollution exposure models. J Expo Anal Environ Epidemiol 15:185–204.

Jerrett M, Gale S, Kontgis C. 2010. Spatial modeling in environmental and public health research. Int J Environ Res Public Health 7:1302–1329.

Karner AA, Eisinger DS, Niemeier DA. 2010. Near-roadway air quality: Synthesizing the findings from real-world data. Environ Sci Technol 44:5334–5344.

Katzfuss M. 2017. A multi-resolution approximation for massive spatial datasets. J Am Stat Assoc 112:201–214.

Kioumourtzoglou MA, Spiegelman D, Szpiro AA, Sheppard L, Kaufman JD, Yanosky JD, et al. 2014. Exposure measurement error in $PM_{2.5}$ health effects studies: A pooled analysis of eight personal exposure validation studies. Environ Health 13:2.

Lewis S, Antoniak M, Venn A, Davies L, Goodwin A, Salfield N, et al. 2005. Secondhand smoke, dietary fruit intake, road traffic exposures, and the prevalence of asthma: A cross-sectional study in young children. Am J Epidemiol 161:406–411.

Lindhjem CE, Pollack AK, DenBleyker A, Shaw SL. 2012. Effects of improved spatial and temporal modeling of on-road vehicle emissions. J Air Waste Manag Assoc 62:471–484.

Lipfert FW, Wyzga RE. 2008. On exposure and response relationships for health effects associated with exposure to vehicular traffic. J Expo Sci Environ Epidem 18:588–599.

Malby AR, Whyatt JD, Timmis RJ. 2013. Conditional extraction of air-pollutant source signals from air-quality monitoring. Atmos Environ 74:112–122.

Martenies SE, Wilkins D, Batterman SA. 2015. Health impact metrics for air pollution management strategies. Environ Int 85:84–95.

McMillan NJ, Holland DM, Morara M, Feng J. 2010. Combining numerical model output and particulate data using Bayesian space-time modeling. Environmetrics 21:48–65.

Milando CW, Batterman SA. 2018a. Operational evaluation of the RLINE dispersion model for studies of traffic-related air pollutants. Atmos Environ 182:213–224.

Milando CW, Batterman SA. 2018b. Sensitivity analysis of the near-road dispersion model RLINE — an evaluation at Detroit, Michigan. Atmos Environ 181:135–144.

National Oceanic and Atmopsheric Administration (NOAA). 2016. Earth system research laboratory (ESRL) radiosonde database. Available: *https://ruc.noaa .gov/raobs/.*

National Weather Service. 2016. Integrated surface hourly data (ISHD) directory.

Office of Highway Policy Information. 2014. Highway statistics 2013. Washington, DC:U.S. Department of Transportation.

O'Neill MS, Breton CV, Devlin RB, Utell MJ. 2012. Air pollution and health: Emerging information on susceptible populations. Air Qual Atmos Health 5:189–201.

Özkaynak H, Baxter LK, Dionisio KL, Burke J. 2013. Air pollution exposure prediction approaches used in air pollution epidemiology studies. J Expo Sci Environ Epidemiol 23:566–572.

Pant P, Harrison RM. 2013. Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review. Atmos Environ 77:78–97.

Patton AP, Milando C, Durant JL, Kumar P. 2017. Assessing the suitability of multiple dispersion and land use regression models for urban traffic-related ultrafine particles. Environ Sci Tech 51:384–392.

Pönkä A. 1990. Absenteeism and respiratory disease among children and adults in Helsinki in relation to low-level air pollution and temperature. Environ Res 52:34–46.

Rao KS. 2005. Uncertainty analysis in atmospheric dispersion modeling. Pure Appl Geophys 162:1893–1917.

Reich BJ, Chang HH, Foley KM. 2014. A spectral method for spatial downscaling. Biometrics 70:932–942.

Reich BJ, Eidsvik J, Guindani M, Nail AJ, Schmidt AM. 2011. A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration. Ann Appl Stat 5:2425–2447.

Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol 71:319–392.

Rundel CW, Schliep EM, Gelfand AE, Holland DM. 2015. A data fusion approach for spatial analysis of speciated $PM_{2.5}$ across time. Environmetrics 26:515–525.

Sacks JD, Stanek LW, Luben TJ, Johns DO, Buckley BJ, Brown JS, et al. 2011. Particulate matter-induced health effects: Who is susceptible? Environ Health Perspect 119:446–454.

Sahu SK, Gelfand AE, Holland DM. 2010. Fusing point and areal level space-time data with application to wet deposition. J R Stat Soc Ser C Appl Stat 59:77–103.

Sampson PD, Guttorp P. 1992. Nonparametric estimation of nonstationary spatial covariance structure. J Am Stat Assoc 87:108–119.

Schmidt AM, Guttorp P, O'Hagan A. 2011. Considering covariates in the covariance structure of spatial processes. Environmetrics 22:487–500.

Shafran-Nathan R, Levy I, Levin N, Broday DM. 2017. Ecological bias in environmental health studies: The problem of aggregation of multiple data sources. Air Qual Atmos Health 10:411–420.

Sheppard L, Burnett RT, Szpiro AA, Kim S-Y, Jerrett M, Pope CA 3rd, et al. 2012. Confounding and exposure measurement error in air pollution epidemiology. Air Qual Atmos Health 5:203–216.

Shi JP, Khan A, Harrison RM. 1999. Measurements of ultra-fine particle concentration and size distribution in the urban atmosphere. Sci Total Environ 235:51–64.

Snyder M, Arunachalam S, Isakov V, Heist D, Batterman S, Talgo K, et al. 2013a. Sensitivity analysis of dispersion model results in the NEXUS health study due to uncertainties in traffic-related emissions inputs. In: Proceedings of the Air & Waste Management Association Annual

Conference & Exhibition, 2013. June 25–28, 2013. Chicago, IL: A&WMA.

Snyder M, Arunachalam S, Isakov V, Talgo K, Naess B, Valencia A, et al. 2014. Creating locally-resolved mobile-source emissions inputs for air quality modeling in support of an exposure study in Detroit, Michigan, USA. Int J Environ Res Public Health 11:12739–12766.

Snyder MG, Venkatram A, Heist DK, Perry SG, Petersen WB, Isakov V. 2013b. RLINE: A line source dispersion model for near-surface releases. Atmos Environ 77:748–756.

Stieb DM, Chen L, Hystad P, Beckerman BS, Jerrett M, Tjepkema M, et al. 2016. A national study of the association between traffic-related air pollution and adverse pregnancy outcomes in Canada, 1999–2008. Environ Res 148:513–526.

Szpiro AA, Paciorek CJ. 2013. Measurement error in two-stage analyses, with application to air pollution epidemiology. Environmetrics 24:501–517.

Szpiro AA, Paciorek CJ, Sheppard L. 2011. Does more accurate exposure prediction necessarily improve health effect estimates? Epidemiology 22:680–685.

U.S. Census Bureau. 2007. American Housing Survey for the United States, 2007. Available: *www.census.gov/library/publications/2008/demo/h150-07.html* [Online Feb. 2, 2015].

U.S. Environmental Protection Agency. 2000. Meteorological Monitoring Guidance for Regulatory Modeling Applications. EPA-454/r-99-005. Research Triangle Park, NC:Office of Air Quality Planning and Standards.

U.S. Environmental Protection Agency. 2004. User's guide for the AERMOD meteorological preprocessor (AERMET). EPA-454/B-03-002. Research Triangle Park, NC:Office of Air Quality Planning and Standards.

U.S. Environmental Protection Agency. 2015. Air quality system (AQS) data mart. Available: *https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html.*

van Buuren S, Groothuis-Oudshoorn K. 2011. Mice: Multivariate imputation by chained equations in R. J Stat Softw 45:1–67.

Venkatram A, Snyder MG, Heist DK, Perry SG, Petersen WB, Isakov V. 2013. Reformulation of plume spread for near-surface dispersion. Atmos Environ 77:846–855.

Venn A, Lewis S, Cooper M, Hubbard R, Hill I, Boddy R, et al. 2000. Local road traffic activity and the prevalence,

severity, and persistence of wheeze in school children: Combined cross sectional and longitudinal study. Occup Environ Med 57:152–158.

Venugopal M, Yang H. 2014. Development and application of link-level vehicle emission inventory for the metropolitan area. 14th COTA International Conference of Transportation Professionals. pp. 2769–2781. Available: *https://doi.org/10.1061/9780784413623.265*.

Vette A, Burke J, Norris G, Landis M, Batterman S, Breen M, et al. 2013. The near-road exposures and effects of urban air pollutants study (NEXUS): Study design and methods. Sci Total Environ 448:38–47.

Watkins N. 2012. Near-road NO_2 Monitoring Technical Assistance Document. Research Triangle Park, NC:U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Ambient Air Monitoring Group.

Weuve J, Kaufman JD, Szpiro AA, Curl C, Puett RC, Beck, T, et al. 2016. Exposure to traffic-related air pollution in relation to progression in physical disability among older adults. Environ Health Perspect 124:1000–1008.

Wjst M, Reitmeir P, Dold S, Wulff A, Nicolai T, von Loeffelholz-Colberg EF, et al. 1993. Road traffic and adverse effects on respiratory health in children. BMJ 307:596–600.

Woodward N, Finch CE, Morgan TE. 2015. Traffic-related air pollution and brain development. AIMS Environ Sci 2:353–373.

Wu J, Wilhelm M, Chung J, Ritz B. 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. Environ Res 111:685–692.

Wu Y-C, Batterman SA. 2006. Proximity of schools in Detroit, Michigan to automobile and truck traffic. J Expo Sci Environ Epidemiol 16:457–470.

Yu H, Stuart AL. 2016. Exposure and inequality for select urban air pollutants in the Tampa Bay area. Sci Total Environ 551–552:474–483.

Zhang K, Batterman S. 2010. Near-road air pollutant concentrations of CO and $PM_{2.5}$: A comparison of MOBILE6.2/CALINE4 and generalized additive models. Atmos Environ 44:1740–1748.

Zhang K, Batterman S, Dion F. 2011. Vehicle emissions in congestion: Comparison of work zone, rush hour and free-flow conditions. Atmos Environ 45:1929–1939.

Zhang S, Wu Y, Huang R, Wang J, Yan H, Zheng Y, et al. 2016a. High-resolution simulation of link-level vehicle emissions and concentrations for air pollutants in a traffic-populated eastern Asian city. Atmos Chem Phy 16:9965–9981.

Zhang ZJ, Manjourides J, Cohen T, Hu Y, Jiang QW. 2016b. Spatial measurement errors in the field of spatial epidemiology. Int J Health Geog 15:21.

Zhu Y, Hinds WC, Kim S, Shen S, Sioutas C. 2002a. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. Atmos Environ 36:4323–4335.

Zhu Y, Hinds WC, Kim S, Sioutas C. 2002b. Concentration and size distribution of ultrafine particles near a major highway. J Air Waste Manag Assoc 52:1032–1042.

Zidek JV, Le ND, Liu Z. 2012. Combining data and simulated data for space-time fields: Application to ozone. Environ Ecol Stat 19:37–56.

MATERIALS AVAILABLE ON THE HEI WEBSITE

Appendices 1 through 11 for this report are available as Additional Materials on the HEI website at *www.healthef fects.org/publications.*

Appendix 1. Characteristics of Traffic-Related Air Pollution Composition

Appendix 2. Exposure Metrics and Methods for TRAP

Appendix 3. Near-Road Dispersion Models

Appendix 4. Mobile, Point, and Area Emission Inventories

Appendix 5. Air Quality Monitoring Data in Detroit

Appendix 6. Exploratory Analysis of Near-Road Increments

Appendix 7. Fitting and Evaluation of Spatiotemporal Models

Appendix 8. Emission Factors

Appendix 9. Temporal Allocation Factors

Appendix 10. Meteorology

Appendix 11. Receptor Sets

ABOUT THE AUTHORS

Stuart Batterman is a professor of environmental health sciences and of civil and environmental engineering at the University of Michigan. He received his Ph.D. in civil and environmental engineering from the Massachusetts Institute of Technology. His research and teaching interests address environmental impact assessment and exposure science, especially as they pertain to air quality, health risk and impact assessment, and environmental management. He has over three decades of experience in air quality, exposure assessment, modeling, measurements, data interpretation, laboratory and field measurements, and related analyses. As the principal investigator on this project, he provided overall supervision and direction, assisted in data interpretation and manuscript preparation, and was responsible for quality control.

Veronica Berrocal was an associate professor of biostatistics at the University of Michigan until Summer 2019, when she joined the Department of Statistics at University of California, Irvine as associate professor. She received her Ph.D. in statistics from the University of Washington, working on developing spatial statistical models for probabilistic weather forecasting. Her research focuses on developing and applying statistical models for data collected over space and time, with a particular focus on developing methods to infer upon environmental and social determinants of health (e.g., air pollution, weather, built environment, poverty); characterizing environmental exposure; and estimating the impact of socio-economic context and environmental risk factors on health. She was responsible for the design and implementation of the spatiotemporal models in this report.

Chad Milando is currently a postdoctoral associate in environmental health at Boston University. He received his Ph.D. in environmental health sciences from the University of Michigan. His interests include exposure science with the goal of developing policies for local and regional air pollution interventions that improve health and mitigate existing health disparities. His role in this report included responsibilities for RLINE air quality dispersion modeling and the model performance evaluation.

Owais Gilani is an assistant professor of mathematics at Bucknell University. He received his Ph.D. in biostatistics at Yale University. His research interests span spatial and spatiotemporal statistics, environmental epidemiology, longitudinal data analysis, and hierarchical Bayesian modeling. He implemented and evaluated many of the spatiotemporal models presented in this report.

Saravanan Arunachalam is a research professor at the Institute for the Environment at the University of North Carolina at Chapel Hill. He received his Ph.D. in chemical engineering from Rutgers University. His research focuses on air quality modeling, atmospheric chemistry, near-road emissions, and other air pollution topics addressing the

development of models and analytical methods for air pollution and health risk. In this report, he helped to implement and evaluate dispersion model algorithms in RLINE and chemical transformation for on-road emissions, estimated background pollutant levels, and performed sensitivity analyses.

K. Max Zhang is an associate professor in the Sibley School of Mechanical and Aerospace Engineering at Cornell University. He received his Ph.D. in mechanical engineering from the University of California at Davis. His research interests include aerosols, air quality, climate change, near-road air pollution, plume characterization and air quality modeling, and energy systems. In this project, he utilized the Comprehensive Turbulent Aerosol dynamics and Gas chemistry model (CTAG) to examine on-road and near-road pollutant dispersion and transformation.

OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Gilani O, Berrocal VJ, Batterman SA. 2019. Nonstationary spatiotemporal Bayesian data fusion for pollutants in the near-road environment. Environmetrics 30:e2581. Available: *https://doi.org/10.1002/env.2581*.

Milando CW, Batterman SA. 2018a. Operational evaluation of the RLINE dispersion model for studies of traffic-related air pollutants. Atmos Environ 182:213–224. Available: *https://doi.org/10.1016/j.atmosenv.2018.03.030*.

Milando CW, Batterman SA. 2018b. Sensitivity analysis of the near-road dispersion model RLINE — An evaluation at Detroit, Michigan. Atmos Environ 181:135–144. Available: *https://doi.org/10.1016/j.atmosenv.2018.03.009*.

Yang B, Zhang KM, Xu WD, Zhang S, Batterman S, Baldauf RW, et al. 2018. On-road chemical transformation as an important mechanism of NO_2 formation. Environ Sci Technol 52(8):4574–4582; doi:10.1021/acs.est.7b05648.

Gilani O, Berrocal VJ, Batterman S. 2016. Non-stationary spatio-temporal modeling of traffic-related pollutants in near-road environments. Spat Spatiotemporal Epidemiol 18:24–37; doi:10.1016/j.sste.2016.03.003.

Milando C, Huang L, Batterman S. 2016. Trends in $PM_{2.5}$ emissions, concentrations and apportionments in Detroit and Chicago. Atmos Environ 129:197–209. Available: *https://doi.org/10.1016/j.atmosenv.2016.01.012*.

Milando CW, Martenies SE, Batterman SA. 2016. Assessing concentrations and health impacts of air quality management strategies: Framework for Rapid Emissions Scenario and Health impact ESTimation (FRESH-EST). Environ Int 94:473–481; doi:10.1016/j.envint.2016.06.005.

Batterman S. 2015. Temporal and spatial variation in allocating annual traffic activity across an urban region and implications for air quality assessment. Transportation Research, Part D: Transport and the Environment 41:401–415. Available: *https://doi.org/10.1016/j.trd.2015.10.009*.

Batterman S, Cook R, Justin T. 2015. Temporal variation of traffic on highways and the development of accurate temporal allocation factors for air pollution analyses. Atmos Enviro 107:351–363. Available: *http://dx.doi.org/10.1016/j.atmosenv.2015.02.047*.

Baldwin N, Gilani O, Raja S, Batterman S, Ganguly R, Hopke P, et al. 2015. Factors affecting pollutant concentrations in the near-road environment. Atmos Environ 115:223–235. Available: *http://dx.doi.org/10.1016/j.atmosenv.2015.05.024*.

Research Report 202, *Enhancing Models and Measurements of Traffic-Related Air Pollutants for Health Studies Using Dispersion Modeling and Bayesian Data Fusion*, S. Batterman et al.

INTRODUCTION

Traffic emissions are an important source of urban air pollution, and exposure to traffic-related air pollution has been associated with various adverse health effects. In 2010, HEI published Special Report 17, *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. That report, developed by the HEI Panel on the Health Effects of Traffic-Related Air Pollution, summarized and synthesized research related to the health effects from exposure to traffic emissions. The Panel concluded that exposure to traffic-related air pollution was causally linked to worsening asthma symptoms. It also found suggestive evidence of a causal relationship with onset of childhood asthma, nonasthma respiratory symptoms, impaired lung function, total and cardiovascular mortality, and cardiovascular morbidity (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010). Many additional studies have been published since the earlier review, and therefore HEI is currently conducting a new systematic review of the epidemiological literature on the health effects of long-term exposure to traffic-related air pollution.

Because traffic-related air pollution exposure is of public health interest, it is important to understand where and at what level people are exposed to air pollution from traffic emissions. However, exposure assessment is challenging because traffic-related air pollution is a complex mixture of many particulate and gaseous pollutants and is characterized by high spatial and temporal variability (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010). The highest levels of traffic-related air pollution occur within a few hundred meters of major roads, with the extent of the impact zone depending on the pollutant, geographical and land-use characteristics, and meteorological conditions (Karner et al. 2010; Zhou and Levy 2007).

Because it is difficult and labor intensive to measure concentrations of traffic-related pollutants on a large scale, scientists generally rely on modeling to assign exposures to air pollution in health studies. However, developing accurate models of traffic-related air pollution for use in exposure assessment for epidemiological studies has been challenging, in part because of the variation in air pollution levels on small spatial scales within cities. Approaches to assess exposure to traffic-related air pollution have included incorporating measurements made at various distances from busy roads using fixed sites or mobile platforms as well as employing various models such as land-use regression and dispersion models. In some cases, infiltration and time–activity patterns have been included for more accurate estimates of personal exposure to air pollution from traffic and other outdoor sources. Each of these exposure estimation approaches has limitations that have been discussed before (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010).

In 2013, following the recommendation of the HEI Traffic Review Panel to improve exposure assessment of traffic-related air pollution for use in health studies, HEI issued Request for Applications (RFA*) 13-1, *Improving Assessment of Near-Road Exposure to Traffic Related Pollution*. Since then, HEI has funded five studies under RFA 13-1 and nine other studies related to exposure assessment or health effects of traffic-related air pollution under other RFAs (see Preface). In response to RFA 13-1, Dr. Stuart Batterman and colleagues from the University of Michigan, the University of North Carolina, and the Cornell Institute proposed a 2.5-year study, "Integrating enhanced models and measurements of traffic related air pollutants for epidemiological and risk studies using Bayesian melding." They aimed to improve estimates of traffic-related air pollution concentrations for use in health studies, with specific attention to air pollution dispersion models and statistical approaches that combine measurements with dispersion model outputs. The HEI Research Committee recommended Dr. Batterman's application for funding because the study would apply and improve existing models that could be employed in other settings. The Bayesian approach in particular was considered novel.

---

* A list of abbreviations and other terms appears at the end of this volume.

This Critique provides the HEI Review Committee's evaluation of the study. It is intended to aid the sponsors of HEI and the public by highlighting both the strengths and limitations of the study and by placing the Investigators' Report in scientific and regulatory context.

SUMMARY OF THE STUDY

SPECIFIC AIMS AND APPROACH

This report investigates ways to improve estimates of traffic-related air pollution for potential use in exposure assessment applications for health effect studies, with specific attention to dispersion modeling and data fusion methods. The specific aims were to:

1. explore potential enhancements for dispersion models, including alternative treatments of meteorological inputs, background levels, and traffic inputs;

2. assess the performance of dispersion models for predicting concentrations of traffic-related air pollutants in a full-scale urban case study, including identification of critical inputs and uncertainties; and

3. apply spatiotemporal and Bayesian data fusion statistical techniques for combining dispersion model outputs and pollutant monitoring observations.

Because the dispersion model predicted concentrations at point locations as compared to Bayesian melding techniques — which require model output representing an average over a grid cell or other spatial unit — the investigators opted to use data fusion techniques rather than Bayesian melding as they had originally planned.

The study aimed to obtain improved exposure estimates of traffic-related pollutants utilizing data collected as part of NEXUS (Near-road EXposure and effects of Urban air pollution Study), a large cohort study conducted in Detroit focusing on the health effects of children with asthma living near major roads (Vette et al. 2013). The NEXUS study was a cooperative agreement between the United States Environmental Protection Agency (U.S. EPA) and the University of Michigan. In the current study, the investigators applied an existing U.S. EPA dispersion model designed for modeling impacts of traffic emissions on air pollution (called the Research LINE-source model [RLINE]) to predict short- and long-term exposure from number and types of motor vehicles, meteorology, and other factors affecting how air pollution is spread after being emitted by motor vehicles. They used or developed models of different levels of computational complexity — RLINE, universal kriging, and four types of joint Bayesian

data fusion — for particulate matter ≤ 2.5 μm in aerodynamic diameter ($PM_{2.5}$), nitrogen oxides (NO_x), carbon monoxide (CO), and black carbon (BC) (see Critique Table). Then, they compared the models by evaluating their performance relative to preexisting measurements at stationary central and near-road sites and along mobile monitoring transects from busy roads.

METHODS

Measurements

All air pollution data used in this study were previously collected in the Detroit area as part of the NEXUS study (Critique Figure 1). For evaluation of the RLINE dispersion model (Aim 1 and Aim 2), they used 24-hour average measurements of NO_x, CO, and $PM_{2.5}$ from five U.S. EPA monitors in Detroit from 2011–2014. For development and evaluation of the statistical models (Aim 3), the investigators made use of a previous 1-week mobile monitoring campaign in 2012 where they collected short-term (5-minute) measurements while parked at 3–6 locations along nine transects on one or both sides of major roads in Detroit.

RLINE Dispersion Model

To address Aims 1 and 2, air pollution concentrations were modeled using RLINE, a Gaussian line-source dispersion model specifically designed for the near-road environment (Snyder et al. 2013). Contributions of motor vehicles to ambient NO_x, CO, and $PM_{2.5}$ concentrations were modeled using RLINE, with background concentrations estimated based on measurements expected to be unaffected by nearby major roads. The investigators compared the modeled 24-hour concentrations with those measured at each U.S. EPA monitoring site and at two sets of receptors (i.e., locations where concentrations were predicted). The first set of receptors was selected from all the residences in Detroit; the second set of receptors was selected from the home and school addresses of children with asthma participating in the NEXUS study, in order to represent a vulnerable population. The investigators reported various statistical measures for model performance, such as the percentage of modeled values within a factor of two of observed values, Spearman rank correlation coefficient, fractional bias, and geometric variance. Several additional analyses were conducted testing the sensitivity of the model to wind direction, season, day of week, emission factors, and traffic inputs.

Spatiotemporal Models

For Aim 3, the investigators systematically applied and evaluated the performance of a series of increasingly complex

Critique Table. Summary of Model Types Evaluated by Batterman and Colleagues in the Investigators' Report

Model Class	Measurements to Develop and Evaluate Models	RLINE Model Receptors[a]	Modeled Concentrations	Pollutants[b]	Performance Evaluation[c]	Model Parameters Available
Dispersion model	Fixed[d]	Sets 1 and 2	Ambient	NO_x, CO, $PM_{2.5}$	IR Table 3	N/A, although Additional Materials[f] include emissions (Table 2) and road links (Figure 3)
Universal kriging	Mobile[e]	Set 1	Ambient	NO, NO_x, BC	IR Table 4	Model 6 (IR Table 5)
JBDF	Mobile[e]	Set 1	Near-road increment	NO_x, $PM_{2.5}$	IR Table 6	Stationary (NO_x) and non-stationary ($PM_{2.5}$) forms of Model 2 (IR Table 7)
Single-pollutant, regression-based JBDF	Mobile[e]	Set 1	Near-road increment	NO_x, $PM_{2.5}$	IR Table 8	Stationary (NO_x) and nonstationary ($PM_{2.5}$) models using 2012 emissions (IR Table 9)
Single-pollutant, regression-based JBDF with meteorology and traffic	Mobile[e]	Set 1	Near-road increment	NO_x, $PM_{2.5}$	IR Table 10	Stationary (NO_x) and nonstationary ($PM_{2.5}$) data fusion using 2012 emissions (IR Table 11)
Two-pollutant regression-based JBDF	Mobile[e]	Set 1	Near-road increment	NO_x, $PM_{2.5}$	IR Table 12	Not provided

Abbreviations: CO = carbon monoxide; BC = black carbon; JBDF = joint Bayesian data fusion model combining measurements and RLINE model output; IR = Investigators' Report; NO = nitrogen oxide; NO_x = oxides of nitrogen; $PM_{2.5}$ = mass of particulate matter ≤ 2.5 µg/m^3

[a] Receptor Set 1 was a dense network of 96 receptors in nine 2-km Detroit-area boxes (see Critique Figure 1). Receptor Set 2 included the location of NEXUS study homes and residences as a sensitivity analysis for a vulnerable population.

[b] Pollutants were modeled as natural logarithms of concentrations, except in the dispersion model, where they were modeled as actual concentrations. When incorporated into spatiotemporal models, RLINE output was also log-transformed.

[c] Performance of the dispersion model was evaluated based on daily average concentrations. All other models were assessed based on comparison of model predictions back-transformed to the original scale with measurements averaged across validation sites and time periods.

[d] Fixed-site measurements were reported for daily averages at five U.S. EPA air quality monitoring sites in the years 2011–2014. See Additional Materials, Appendix 5 (available on the HEI website) for information on U.S. EPA air quality monitoring data completeness.

[e] Mobile measurements were made for 5 minutes at a time at sites 50 m, 150 m, and 500 m from the edge of major roads in nine Detroit areas. Measurements were made on one side of the major road in eight areas and both sides of the road in one area. Monitoring was conducted in 2.5-hour morning and afternoon shifts on all days from December 14–20, except for December 14 (morning only) and December 16 (afternoon only), for a total of 30 hours of measurements. See Critique Figure 1 for the mobile monitoring transect locations.

[f] Additional Materials are available on the HEI website.

Bayesian spatiotemporal models (Critique Figure 2). First, they developed universal kriging models of ambient concentrations for three traffic-related air pollutants (nitric oxide [NO], NO_x, and BC), accounting for covariates such as day of week, upwind versus downwind of nearest road, topographical features, meteorology, traffic activity, and fleet composition features. Next, they combined mobile monitoring data with RLINE model outputs in Bayesian data fusion models of the increase in near-road concentrations of two traffic-related air pollutants (NO_x and $PM_{2.5}$) above ambient concentrations. They used two approaches for fusing: combining RLINE output with the mobile monitoring data as a multivariate realization of the unmeasured pollutant (joint Bayesian data fusion), and using RLINE
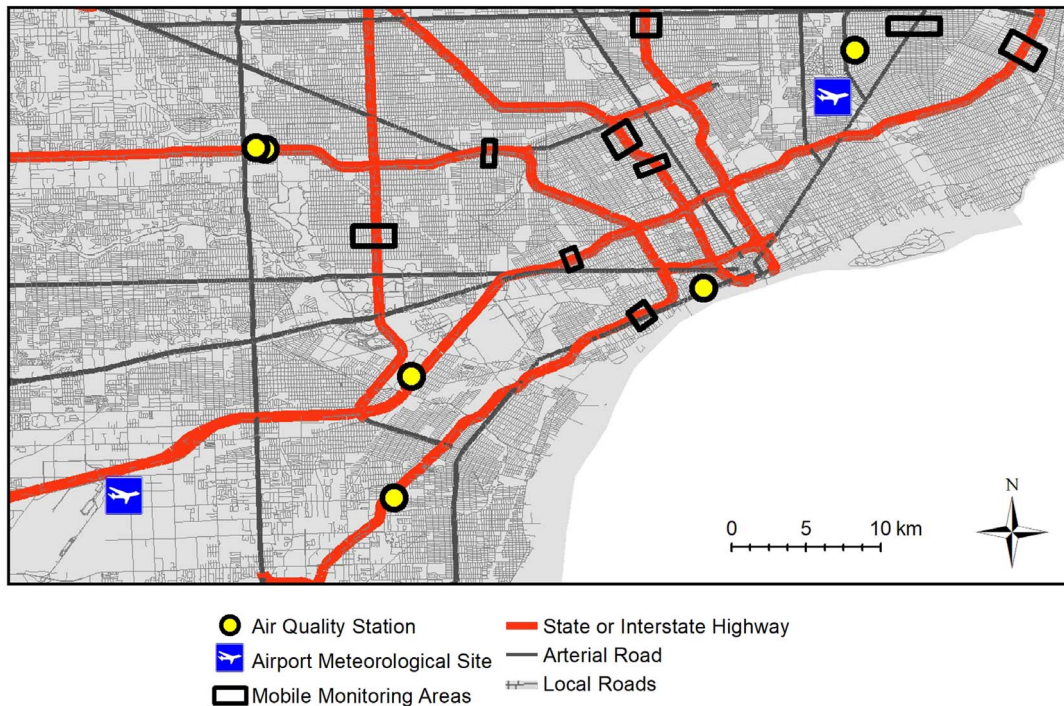
Critique Figure 1. Map of Detroit showing air quality monitoring stations, airport weather stations, and near-road mobile monitoring locations. (The map is based on Figure 1 of the Investigators' Report and Figure 6 in the Additional Materials, with background layers from Michigan GIS Open Data.)
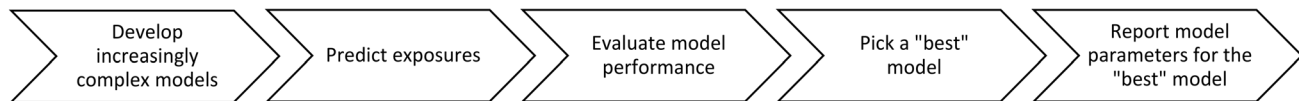
Critique Figure 2. Steps followed to compare spatiotemporal Bayesian models of each class considered.

output instead of the covariates in the prediction model for the mobile monitoring data (regression-based Bayesian data fusion). Both types of data fusion models were tested with different sets of assumptions of how the variables are related to each other; the difference in the sets of assumptions was whether the relationships among variables in the model stay the same (stationary) or change (nonstationary) over time. For the regression-based data fusion approach, they also developed a two-pollutant model that incorporated the near-road increments of both $PM_{2.5}$ and NO_x.

Samples of 10%–15% of the full dataset were excluded from model development and used for model evaluation. Model performance was first reported in terms of mean absolute prediction error, average length of the 90%

prediction interval, and empirical coverage of the 90% prediction interval. Additional performance statistics of root mean square error and Pearson correlation were later added. The investigators also examined several model assumptions, including several different covariance functions (i.e., the relationships among variables in the models), the different methods of including RLINE output in the statistical methods, and whether jointly modeling two pollutants would improve the statistical models. They also tested the influence of the emissions inventory used in RLINE by comparing results using an updated emissions inventory with detailed information from 2012, as opposed to the standard emissions inventory from 2010.

RESULTS

RLINE Dispersion Model

The RLINE model generally performed better closer to the major road; when the receptors were downwind rather than parallel relative to the nearest major road; on weekdays as opposed to weekends; and for winter rather than for other seasons. It also performed better for emission inventories using location-specific emission factors and temporal allocation factors versus national defaults from 2010. Sensitivity analyses showed that the results were more sensitive to meteorology than to other inputs evaluated.

As interpreted by the investigators, comparison of errors suggested that improvements in model inputs or parameterization could improve model performance for NO_x, but not for CO at the near-road sites. The models could not be evaluated when concentrations were below the detection limit of the instrumentation, which happened frequently for CO. Performance evaluation of the $PM_{2.5}$ modeled concentrations was not attempted because the predicted contributions of $PM_{2.5}$ from local sources were small relative to contributions from more regional sources.

Spatiotemporal Models

Universal Kriging. The universal kriging models generally predicted high and low concentrations and gradients: concentrations were higher immediately downwind rather than upwind of the nearest major road and stayed elevated at distances farther downwind. Key covariates in the models included background concentrations, traffic volume, location upwind or downwind of the nearest major road, wind speed, and distance from the nearest major road. The stationary model, with wind speed and wind direction relative to the nearest major road, performed similarly to the nonstationary model without those variables. The best-performing universal kriging model was the most computationally complex one, perhaps limiting its practical utility.

Single-Pollutant Joint Bayesian Data Fusion Models.
The investigators reported that no single model performed best for all evaluation performance criteria. Among the data fusion models, stationary models containing RLINE predictions as inputs performed better than nonstationary models for NO_x, whereas the opposite was true for $PM_{2.5}$ (nonstationary covariance functions performed better than stationary models).

For NO_x, the data fusion approach was inferior to the universal kriging model, suggesting that RLINE did not add useful information to the model inputs. However, for

$PM_{2.5}$, the RLINE output did improve the data fusion models. The investigators commented that even though the two Bayesian data modeling approaches (universal kriging and data fusion) used different measurements as inputs, the same conclusion was reached: the RLINE model displayed biases that were additive but had values that differed across locations for both NO_x and $PM_{2.5}$.

Additional traffic and meteorology variables did not strongly affect the results from single-pollutant regression-based models including RLINE output. For NO_x, only distance from the edge of the nearest major road gave additional information on NO_x near-road increments. The $PM_{2.5}$ models including these variables performed worse than the simpler models without them.

Two-Pollutant Regression-Based Joint Bayesian Data Fusion Models. The investigators produced two-pollutant models of near-road increments of NO_x and $PM_{2.5}$. However, they did not expect joint modeling to be particularly useful because (1) RLINE improved the single-pollutant model for $PM_{2.5}$ near-road increments but did not improve the model for the NO_x near-road increments, and (2) single-pollutant models with different sets of covariance assumptions performed best for near-road increments of NO_x (stationary structure) and $PM_{2.5}$ (nonstationary structure). As for the single-pollutant Bayesian models, the best-performing two-pollutant model (with $PM_{2.5}$ and NO_x) for $PM_{2.5}$ used a nonstationary covariance structure. Modeling NO_x jointly with $PM_{2.5}$ did not improve the NO_x models. Relative to single-pollutant models, the joint Bayesian data fusion models had smaller prediction intervals, increased computational complexity and runtime, and might not have been able to find a single best-fitting model. Overall, the investigators preferred the single-pollutant models over the two-pollutant data fusion models because they were less computationally challenging.

Summary

The investigators concluded that both dispersion and spatiotemporal statistical models contribute useful information to air pollution modeling for exposure assessment. They stated that the ability of a dispersion model to accurately predict near-road concentrations of a pollutant depends on the pollutant as well as spatial and temporal factors, such as the distance and direction of the receptor from the nearest major road relative to wind direction, and the temporal trends in traffic volume (i.e., time of day and day of week). They reported that their analysis using Bayesian data fusion models confirmed the presence of spatially varying errors in dispersion model outputs and allowed for quantification of both the magnitude and the

spatial nature of these errors. To improve the models for wider use in epidemiology, they recommended generating additional air pollution and meteorology datasets for more rigorous model testing. They also thought that improvements in the computational efficiency and accessibility of the models would be needed for wider adoption.

HEI REVIEW COMMITTEE'S EVALUATION

Batterman and colleagues conducted a study to evaluate a dispersion model and sophisticated spatiotemporal statistical models. In its independent review of the study, the HEI Review Committee thought that the report was well written, that the discussion, conclusions, and recommendations were generally appropriate, and generally agreed with the investigators' interpretation of the results.

EPIDEMIOLOGICAL APPLICATION OF SPATIOTEMPORAL AND BAYESIAN MODELS

The Review Committee thought the study was thorough and interesting, and that the application of the spatiotemporal and Bayesian models was sophisticated, well executed, and an important contribution. They were pleased that the investigators evaluated the added value of RLINE estimates in the regression-based Bayesian data fusion models, based on an earlier suggestion by the Review Committee. The investigators observed that after adding RLINE to the regression-based Bayesian models, none of the other covariates was associated with the observed near-road increment. The Committee thought this was worth noting, because RLINE appears to add value that can't be as readily captured by information that is easier to collect.

While the overall goal of the study was to improve estimates of traffic-related air pollution for use in health studies, the Committee concluded that the report may have overstated the usefulness of the models for epidemiological studies, which was one of the main goals of this RFA. For example, the models presented in this report appeared to have limited use over a broad geographic area, and it was not clear to the Committee how these narrowly applied models would scale up. In particular, the lack of long-term exposure data limited the possible insights to be gained from the application of the spatiotemporal and fusion models over longer time periods, as would be needed for most epidemiological applications. This was disappointing because the project's goal had originally been to improve long-term exposure estimates of traffic-related air pollution. Specifically, the application in NEXUS was limited to short-term exposures and did not make optimal use of the longitudinal design of the study.

The Committee agreed with the investigators that the generalizability of the analyses is limited because the measurements used for the statistical model development were from only one winter month. Although appropriate to evaluate traffic-related air pollution, the focus on residential areas in Detroit primarily affected by traffic sources limits generalizability to evaluating air pollution from other sources.

MODEL PERFORMANCE

A focus of the current study was evaluating the performance of a large number of air pollution models (Critique Table). The Committee was pleased that the authors included several measures typically used for evaluation of exposure-assessment models in epidemiological studies, such as the root mean square error and the Pearson correlation coefficient. They thought that some of the other statistics were less informative (e.g., the fraction of predictions within a factor of two of the measurements) and not typically used for evaluation of exposure models in epidemiological studies, while recognizing that these measures have been commonly used in evaluating dispersion models.

The Committee noted that the models may not have had sufficiently high predictive performance to provide reliable estimates of exposure. They thought that there were large uncertainties in the model predictions of air pollutant concentrations — for example, root mean square errors were as large as the mean for NO_x, and there was low variability in $PM_{2.5}$. These uncertainties in concentrations would result in large uncertainties in exposure assignment and even larger uncertainties in the associations of health outcomes with exposures. At the same time, differential measurement error in the RLINE predictions because of better prediction near roads than farther away may translate to biased health effect estimates when applied to an epidemiological cohort because the exposure predictions would be more accurate for the most highly exposed people (with participants living at varying distances from major roads).

At the same time, to address the question of whether inclusion of RLINE output improves predictive performance of the regression-based fusion models, the investigators emphasized statistical significance of coefficient estimates rather than predictive model performance statistics, which the Committee considered to be more useful. For example, to compare the predictive performance of the regression-based Bayesian models, the Committee observed that the root mean square error was about 1 ppb lower for the NO_x model with additional covariates but around the same for $PM_{2.5}$, though not completely consistent across stationary and nonstationary models and emission years. They

would have appreciated elaboration on why adding traffic and meteorology to regression-based Bayesian data fusion models improved predictive performance for NO_x but degraded predictive performance for $PM_{2.5}$.

While comparisons within model classes were presented, the Committee thought it would have been useful if the model performance evaluation had also been structured to compare model performance of the same pollutants at the same model receptors across the various model classes. For example, the Bayesian fusion models were analyzed on two different sets of model receptors, and the dispersion, universal kriging, and Bayesian data fusion models predicted three different sets of ambient concentrations or near-road increments in pollutant concentrations.

EXPOSURE TO A MULTIPOLLUTANT MIXTURE

Given increasing interest in assessing the health effects of traffic-related air pollution as a multipollutant mixture, the Review Committee appreciated that the investigators explored joint modeling of $PM_{2.5}$ and NO_x. The investigators reported that the joint models introduced computational complexity, while not improving the predictions of either pollutant, because the spatial structure differed between pollutants. The Committee thought that it may have been a missed opportunity for a more in-depth exploration of the relationships among different traffic-related air pollutants including an explanation of the two-pollutant model and its implications. However, they thought this was a reasonable interpretation of the joint $PM_{2.5}$ and NO_x model, and that it was not necessary to go beyond these two pollutants in this application.

SUMMARY AND CONCLUSIONS

Batterman and colleagues conducted a study to evaluate a dispersion model and sophisticated spatiotemporal statistical models to estimate pollutant exposures. They evaluated the performance of a U.S. EPA dispersion model of traffic-related air pollution (RLINE), as well as the performance of universal kriging, and sophisticated statistical models that combined RLINE and pollutant measurements. The Committee agreed with the investigators that both dispersion and statistical models contributed useful information to the air pollutant concentration predictions. The performance of the RLINE model depended on the pollutant as well as on spatial and temporal factors, such as distance from the major road. In addition, statistical

models with different sets of assumptions generally led to the same conclusions and provided complementary information on how the air pollutants were spatially distributed. Finally, adding RLINE to the statistical models or jointly modeling NO_x and $PM_{2.5}$ improved predictions only for $PM_{2.5}$ and not for NO_x.

The Committee thought the spatiotemporal and Bayesian models were state of the science and well executed, and that the application of the models was, in particular, novel and an important contribution. They appreciated the systematic quantification of both the magnitude and the spatial nature of model uncertainty, although they thought the information was not complete, given that the true underlying distribution of air pollutant concentrations could not be measured. On the other hand, they thought that the report may have overstated the usefulness of the models for epidemiological studies, because the models appeared to have limited use over a broad geographic area. Also, the models performed better closer to roads than farther away, which might translate to biased health effect estimates because the exposure predictions would be more accurate for the most highly exposed people in an epidemiological cohort (with participants living at varying distances from major roads). In addition, the uncertainties in the predictions of air pollutant concentrations remained large, even for the most refined models.

There remains a need to further refine the models and distribute these new tools for wider use. In particular, these and similar models will need to be rigorously tested on large databases of measurements collected over long periods before they are used on a large scale in epidemiological studies.

REFERENCES

HEI Panel on the Health Effects of Traffic-Related Air Pollution. 2010. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. HEI Special Report 17. Health Effects Institute:Boston, MA.

Karner AA, Eisinger DS, Niemeier DA. 2010. Near-roadway air quality: Synthesizing the findings from real-world data. Env Sci Technol 44:5334–5344; doi:10.1021/es100008x.

Snyder MG, Venkatram A, Heist DK, Perry SG, Petersen WB, Isakov V. 2013. RLINE: A line source dispersion model for near-surface releases. Atmos Environ 77:748–756; doi:10.1016/j.atmosenv.2013.05.074.

Vette A, Burke J, Norris G, Landis M, Batterman S, Breen M, et al. 2013. The Near-road Exposures and Effects of Urban Air Pollutants Study (NEXUS): Study design and methods. Sci Total Environ 448:38–47; doi:10.1016/j.scitotenv.2012.10.072.

Zhou Y, Levy JI. 2007. Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis. BMC Public Health 7:89; doi:10.1186/1471-2458-7-89.

Number	Title	Principal Investigator	Date
Research Reports			
196	Developing Multipollutant Exposure Indicators of Traffic Pollution: The Dorm Room Inhalation to Vehicle Emissions (DRIVE) Study	J.A. Sarnat	2018
194	A Dynamic Three-Dimensional Air Pollution Exposure Model for Hong Kong	B. Barratt	2018
183	Development of Statistical Methods for Multipollutant Research *Part 3.* Modeling of Multipollutant Profiles and Spatially Varying Health Effects with Applications to Indicators of Adverse Birth Outcomes	J. Molitor	2016
179	Development and Application of an Aerosol Screening Model for Size-Resolved Urban Aerosols	C.O. Stanier	2014
175	New Statistical Approaches to Semiparametric Regression with Application to Air Pollution Research	J.M. Robins	2013
152	Evaluating Heterogeneity in Indoor and Outdoor Air Pollution Using Land-Use Regression and Constrained Factor Analysis	J.I. Levy	2010
140	Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality	D. Krewski	2009
139	Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study	B. Brunekreef	2009
131	Characterization of Particulate and Gas Exposures of Sensitive Subpopulations Living in Baltimore and Boston	P. Koutrakis	2005
130	Relationships of Indoor, Outdoor, and Personal Air (RIOPA) *Part I.* Collection Methods and Descriptive Analyses *Part II.* Analyses of Concentrations of Particulate Matter Species	C.P. Weisel B.J. Turpin	2005 2007
127	Personal, Indoor, and Outdoor Exposures to $PM_{2.5}$ and Its Components for Groups of Cardiovascular Patients in Amsterdam and Helsinki	B. Brunekreef	2005
Special Reports			
17	Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects	HEI Panel on the Health Effects of Traffic-Related Air Pollution	2010

Copies of these reports can be obtained from HEI; PDFs are available for free downloading at www.healtheffects.org/publications.

ABBREVIATIONS AND OTHER TERMS

% Red   percentage of errors that are reducible

AADT   annual average daily traffic

AERMET   a meteorological data preprocessor for AERMOD

AERMOD   a dispersion model

AQS   air quality system

BC   black carbon

CAADT   commercial annual average daily traffic

CI   credible interval

CO   carbon monoxide

CMAQ   Community Multiscale Qir Quality model

CLINE   Community LINE-source model

F2   percentage of modeled values within a factor of 2 of observed values

FB   fractional bias

GIS   geographical information system

IARC   International Agency for Research on Cancer

ICHEM   instrumental chemiluminescence

IGpCHEM   instrumental gas-phase chemiluminescence

KDET   Detroit City Airport

KDTW   Detroit Metro Airport

MAPE   mean absolute prediction error

MAPL   Mobile Air Pollution Lab

MOVES   Motor Vehicle Emission Simulator

NEXUS   Near-road EXposures and effects of Urban air pollutants Study

NO   nitric oxide

NOAA   National Oceanic and Atmospheric Administration

NO_x nitrogen oxides

NRI   near-road increment

NWS   National Weather Service

PI   prediction interval

$PM_{2.5}$ particulate matter ≤2.5 μm in aerodynamic diameter

RFA   request for applications

RLINE   Research LINE-source model

RMSE   root mean square error

R_{SP} Spearman rank correlation coefficient

SFC   format for surface meteorological data

TAF   temporal allocation factor

TEOMS   tapered element oscillating microbalances

TRAP   traffic-related air pollutant

U.S. EPA   U.S. Environmental Protection Agency

V_G geometric variance

WRF   Weather Research and Forecasting model

HEI BOARD, COMMITTEES, and STAFF

Board of Directors

Richard F. Celeste, Chair *President Emeritus, Colorado College*

Enriqueta Bond *President Emerita, Burroughs Wellcome Fund*

Jo Ivey Boufford *President, International Society for Urban Health*

Homer Boushey *Emeritus Professor of Medicine, University of California, San Francisco*

Michael T. Clegg *Professor of Biological Sciences, University of California, Irvine*

Jared L. Cohon *President Emeritus and Professor, Civil and Environmental Engineering and Engineering and Public Policy, Carnegie Mellon University*

Stephen Corman *President, Corman Enterprises*

Martha J. Crawford *Dean, Jack Welch College of Business and Technology, Sacred Heart University*

Michael J. Klag *Dean Emeritus and Second Century Distinguished Professor, Johns Hopkins Bloomberg School of Public Health*

Alan I. Leshner *CEO Emeritus, American Association for the Advancement of Science*

Henry Schacht *Managing Director, Warburg Pincus; Former Chairman and Chief Executive Officer, Lucent Technologies*

Research Committee

David A. Savitz, Chair *Professor of Epidemiology, School of Public Health, and Professor of Obstetrics and Gynecology, Alpert Medical School, Brown University*

Jeffrey R. Brook *Senior Research Scientist, Air Quality Research Division, Environment Canada, and Assistant Professor, University of Toronto, Canada*

Francesca Dominici *Professor of Biostatistics and Senior Associate Dean for Research, Harvard T.H. Chan School of Public Health*

David E. Foster *Phil and Jean Myers Professor Emeritus, Department of Mechanical Engineering, Engine Research Center, University of Wisconsin, Madison*

Amy H. Herring *Sara & Charles Ayres Professor of Statistical Science and Global Health, Duke University, Durham, North Carolina*

Barbara Hoffmann *Professor of Environmental Epidemiology, Institute of Occupational, Social, and Environmental Medicine, University of Düsseldorf, Germany*

Allen L. Robinson *Raymond J. Lane Distinguished Professor and Head, Department of Mechanical Engineering, and Professor, Department of Engineering and Public Policy, Carnegie Mellon University*

Ivan Rusyn *Professor, Department of Veterinary Integrative Biosciences, Texas A&M University*

Review Committee

James A. Merchant, Chair *Professor and Founding Dean Emeritus, College of Public Health, University of Iowa*

Kiros Berhane *Professor of Biostatistics and Chair, Department of Biostatistics, Mailman School of Public Health, Columbia University*

Michael Jerrett   *Professor and Chair, Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles*

Frank Kelly *Professor of Environmental Health and Director of the Environmental Research Group, King's College London*

Jana B. Milford *Professor, Department of Mechanical Engineering and Environmental Engineering Program, University of Colorado, Boulder*

Jennifer L. Peel *Professor of Epidemiology, Colorado School of Public Health and Department of Environmental and Radiological Health Sciences, Colorado State University*

Roger D. Peng *Professor of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Officers and Staff

Daniel S. Greenbaum *President*

Robert M. O'Keefe *Vice President*

Rashid Shaikh *Director of Science*

Jacqueline C. Rutledge *Director of Finance and Administration*

Emily Alden *Corporate Secretary*

Lee Ann Adelsheim *Research Assistant*

Hanna Boogaard *Consulting Principal Scientist*

Sofia Chang-DePuy *Digital Communications Manager*

Aaron J. Cohen *Consulting Principal Scientist*

Robert M. Davidson *Staff Accountant*

Philip J. DeMarco *Compliance Manager*

Hope Green *Editorial Project Manager*

Joanna Keel *Research Assistant*

Lissa McBurney *Science Administrative Assistant*

Janet I. McGovern *Executive Assistant*

Pallavi Pant *Staff Scientist*

Allison P. Patton *Staff Scientist*

Hilary Selby Polk *Managing Editor*

Anna S. Rosofsky *Staff Scientist*

Robert A. Shavers *Operations Manager*

Annemoon M.M. van Erp *Managing Scientist*

Eleanne van Vliet *Staff Scientist*

Donna J. Vorhees *Director of Energy Research*

Katherine Walker *Principal Scientist*