FINAL REPORT

California Air Resources Board

CONTRACT NO. 07-310

IN-VEHICLE AIR POLLUTION EXPOSURE MEASUREMENT AND MODELING

Submitted by

Ralph J. Delfino, MD, PhD, and Jun Wu, PhD
Co-Principal Investigators

Department of Epidemiology, School of Medicine, University of California, Irvine, 92697-7550

Prepared for the California Air Resources Board and
the California Environmental Protection Agency.

Co investigators:

Scott Fruin, PhD, Keck School of Medicine, Environmental Health Division, University of Southern California

Constantinos Sioutas, ScD, Department of Civil & Environmental Engineering, University of Southern California.

Lianfa Li, PhD, Program in Public Health, University of California, Irvine

Rufus Edwards, PhD, Department of Epidemiology, School of Medicine, University of California, Irvine

Beate Ritz, MD, PhD, Department of Epidemiology, UCLA School of Public Health

Norbert Staimer, PhD, Department of Epidemiology, School of Medicine, University of California, Irvine

June 8, 2012

CONTENTS

FINAL REPORT ..... *i*

CONTENTS ..... *ii*

Disclaimer ..... *vii*

Acknowledgements ..... *vii*

LIST OF FIGURES ..... *viii*

LIST OF TABLES ..... *x*

ABSTRACT ..... *xiii*

EXECUTIVE SUMMARY ..... *xiv*

BODY OF REPORT 16

1. Chapter One: Introduction 16

1.1 Background 16

1.2. Scope and Purpose of the Project 19

1.3. Tasks 20

Overview 20

2. Chapter Two: A Predictive Model for Vehicle Air Exchange Rates based on a Large, Representative Sample 29

2.0 Introduction 29

2.1 Materials and Methods 31

2.1.1 Vehicle selection 31

2.1.2 Instruments. 32

2.1.3 Air Exchange Rate Determinations. 32

2.1.4 Mathematical Equation and Assumptions. 32

2.1.5 Determination of Source Strength 33

2.1.6 Determination of Equilibrium Concentration 33

2.1.7 Speed 34

2.1.8 Data Analysis 34

2.2. Results and Discussion 35

2.2.1 Vehicles Tested 35

2.2.2 Equilibrium Values and AERs Calculated 35
2.2.3 GEE Model Results. 36
2.3 Summary and Conclusions 39
References 41
3. Chapter Three: Factors that Determine Ultrafine Particle Exposure in Vehicles 43
3.0 Introduction 43
3.1 Methods 44
3.1.1 Vehicle Selection and Conditions Tested 44
3.1.2 Particle Concentration Measurements 45
3.1.3 Air Exchange Rate Measurements 46
3.2 Results and Discussion 48
3.2.1 Effect of Air Exchange Rate on I/O Ratios 48
3.2.2 Effect of Vehicle Speed and Age on AER and I/O Ratios 49
3.2.3 Effect of Particle Size on I/O Ratios 51
3.2.4 Effect of Ventilation Fan Setting on I/O Ratios 51
3.2.5 Effect of Cabin Air Filter and Loading on I/O Ratios 52
3.3 Implications for In-Vehicle Particle Models 55
3.4 implications for Exposure assessment 56
3.5 Summary and Conclusions 58
References 59
4. Chapter Four: Freeway Emission Rates and Vehicle Emission Factors of Air Pollutants in Los Angeles 61
4.0 Introduction 61
4.1 Methods 63
4.1.1 Mobile Measurement Platform (MMP) and continuous measurement instruments 63
4.1.2 Sampling Routes 64
4.1.3 Mathematical calculations and equations 65
4.1.3.1 Emission Factor (EF) 65
4.1.3.2 Traffic Characterization 68

4.1.3.3 Freeway emission rate calculations 69

4.2 Results and Discussion 69

4.2.1 Pollutant Concentrations 69

4.2.2 LDV and HDV emission factors 69

4.2.3 Fraction contribution of HDV to total emissions 72

4.2.4 Freeway Pollutant Emission Rates 73

4.2.4.1 Annual average emission rates 73

4.2.4.2 Diurnal variation in freeway emission rates 75

4.2.4.3 Freeway-to-freeway variability in emission rates 76

4.3 Summary and Conclusions 77

References 78

5. Chapter Five, Part I. Linking In-Vehicle Ultrafine Particle Exposures to On-Road Concentrations 81

5.0 Introduction 81

5.1 Methods 82

5.1.1 Vehicle selection and ventilation conditions tested 82

5.1.2 Speed and routes driven 84

5.1.3 Particle concentration measurement, I/O and AER determination 84

5.1.4 Predictive models 85

5.2 Results and Discussion 86

5.2.1 In-vehicle-to-roadway concentration ratios 86

5.2.2 Predictive model for ln(AER) at RC and OA setting 87

5.2.3 Predictive model for logit(I/O) under RC and OA setting 89

5.2.4. Fleet-wide distributions of AER and I/O 93

5.2.5 Expected in-cabin concentrations for given roadway concentrations 96

5.3. Summary and Conclusions 97

References 98

Chapter Five, Part II. Develop and validate the on-road exposure models for particle-bounded PAH, PNC, $PM_{2.5}$, NO_x, and BC (based on Task 4: Develop and validate in-vehicle exposure models for BC, UFP number, $PM_{2.5}$, particle-bounded PAH, and NO_x.) 100

5.4. *Introduction* 100
5.5. *Materials* 101
5.5.1 Mobile Measurement Platform and Concentrations Measured 101
5.5.2 Road and Traffic Classification 102
5.5.3 Meteorological Parameters 102
5.5.4 Independent and Dependent Variables 103
5.6. *Methods* 105
5.6.1 Exploratory Data Analysis 105
5.6.2 Selection of Predictor variables 106
5.6.3 General Linear and Non-Linear Models with Inclusion of Factor Variables 107
5.6.3.1 Basic model: linear regression with factor variables 107
5.6.3.2 Non-linear model: generalized additive model with factor variables 107
5.6.4 Time series model with temporal autocorrelation and factor variables 109
5.6.5 Model validation 110
5.6.5.1 Holdout validation as an independent test and validation 110
5.6.5.2 3x3 cross-validation 111
5.6.5.3 Measurement criteria 111
5.7. *Results and Discussion* 112
5.7.1 Dependent variable concentrations 112
5.7.2 Transformation and correlation analysis 114
5.7.3 Grouping Comparison 119
5.7.3.1 Roadway types 120
5.7.3.2 Time of day 121
5.7.3.3. Atmospheric Stability 124
5.7.4 Regression models for prediction 126
5.7.4.1 PAH modeling 126
5.7.4.2 PNC modeling 128
5.7.4.3 $PM_{2.5}$ modeling 131
5.7.4.4 NO_X modeling 133

5.7.4.5 BC modeling 135
5.7.5 Time series analysis 138
5.7.6 Discussion 140
5.7.6.1 Correlation analysis and scatter plots 140
5.7.6.2 Influence of roadway types 141
5.7.6.3 Influence of time of day 141
5.7.6.4 Influence of traffic variables 141
5.7.6.5 Influence of meteorological factors 142
5.7.6.6 Linear vs. non-linear models 142
5.7.6.7 Validation of predictive models 143
5.7.6.8 Consideration of temporal autocorrelation 143
5.8 *Summary and Conclusions* 144
5.9 *References* 145
6. *Chapter Six. Task 5: Validate the in-vehicle exposure model for PAH against measurements from representative subjects.* 146
6.1. *Materials and Methods* 146
6.2. *Results and Discussion* 149
6.3 *Summary and Conclusions* 152
6.4 References 154
7. Chapter 7. Study Limitations 154
8. Chapter 8. Overall Summary and Conclusions 156
References 160
9. Chapter 9. *Recommendations* 162
10. LIST OF PUBLICATIONS PRODUCED 163

DISCLAIMER

The statements and conclusions in this report are those of the University and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not construed as actual or implied endorsement of such products.

ACKNOWLEDGEMENTS

We thank Neelakshi Hudda, graduate student research assistant at the Department of Civil & Environmental Engineering, University of Southern California, for her diligent and superb work on this project. As she was lead or co-author of several papers resulting from this research, with her permission, some of the chapters in this report are partly adapted from her Ph.D. thesis. We also thank James Liacos, Winnie Kam, Evangelia Kostenidou, Sandrah P. Eckel at the Department of Civil & Environmental Engineering, University of Southern California, and Luke D. Knibbs at Queensland University of Technology, Brisbane, Australia. We thank Thomas Tjoa, Department of Epidemiology, UCI, for his work in constructing datasets and help in programming the data analysis.

This Report was submitted in fulfillment of California Air Resources Board contract no. 07-310 by the University of California, Irvine under the sponsorship of the California Air Resources Board. Work was completed as of June 28, 2012.

LIST OF FIGURES

Figure 2.1: Typical Time-series plot for runs conducted at Cemetery along with the initial build up and freeway run. Average speed during Freeway run was 89 ± 10 km hr-1 for stable portion highlighted in black). The second black highlight corresponds to stable values during 51.1 ± 9.4 km hr-1 and 31.3 ± 5.5 km hr-1 speed runs.

Figure 2.2: AER results for all 59 vehicles tested.

Figure 2.3: Model-predicted AER increase with age and speed for median age study vehicle.

Figure 2.4: Model predictions versus actual measurements, and the normality of the residuals. Each data point represents a measured AER used to populate the predictive model.

Figure 2.5:. Comparison of model predictions and results from Knibbs et al., 2009.

Figure 3.1: I/O ratio dependence on AER for 25-400 nm particles under re-circulation (RC) and outside air (OA) ventilation setting.

Figure 3.2: Agreement between measured I/O during variable speed driving and regression-predicted I/O.

Figure 3.3: Size range specific I/O ratios at three speeds and two ventilation conditions tested. The dashed lines join values from the same vehicle.

Figure 3.4: Comparison of I/O ratios at different speeds and fan settings.

Figure 3.5: I/O ratios by filter condition or absence under OA conditions in a 2010 Toyota Prius.

Figure 3.6: Effect of filter loading on particle removal for filters tested in Honda Civic vehicle under OA conditions.

Figure 3.7: Effect of presence of filter in RC ventilation mode. C0 is the concentration at the beginning of experiment, i.e., at time t = 0 and is equal to the ambient concentration.

Figure 3.8: Impact of change in particle size distribution on number concentration weighed I/O ratios.

Figure 3.9: Progression of particle loss within vehicles at recirculation and outside air intake condition. The subscripts in the legend indicate the experiment time during which the scan was made. The measured AER at 35 miles h-1, medium fan, recirculation mode was 7.3 h-1 and at 0 miles h-1, medium fan outside air intake mode was 93 h-1.

Figure 4.1: Freeway segments where measurements were conducted (generated using Google Maps).

Figure 4.2: Contribution of HDV to total emissions.

Figure 4.3: Annual average hourly freeway emission rates.

Figure 4.4: Vehicle miles travelled, truck vehicle miles travelled and fraction of total miles traveled by truck on four Los Angeles freeways in LA County during 12/1/2010 – 30/11/2011.

Figure 4.5: Diurnal profiles for freeway emission rates.

Figure 4.6: Diurnal vehicle activity trend on two Los Angeles freeways, representative of general trend on all freeways.

Figure 5.1: Distribution of Dependent Variables.

Figure 5.2: Predicted values for lnAER plotted against the two most significant variables under RC and OA ventilation modes.

Figure 5.3: Predicted values for I/O under RC and OA ventilation mode versus two most important model variables for each mode. Bottom subsets show actual measurements versus surface of median model predictions.

Figure 5.4: Distribution of passenger cars in various volume-based size classes.

Figure 5.5: Age distribution of US passenger car fleet and comparison with study fleet.

Figure 5.6: Typical distribution of driving speed on urban freeways and arterial roads during congested or peak and not congested or off-peak traffic conditions.

Figure 5.7: Distribution for AER and I/O for a fleet similar to U.S. passenger car fleet in terms of manufacturer's market share, vehicle volume and age.

Figure 5.8: Expected in-cabin concentration for U.S. vehicle fleet travelling on Los Angeles arterial roads and freeway.

Figure 5.9: Routes of on-road pollutant measurements from Task 4.

Figure 5.10: Box plots for four concentrations, PAH, PNC, PM2.5, NOX and BC.

Figure 5.11: Histograms for raw air pollutant concentrations without transformation.

Figure 5.12: Normal histograms for the transformed values of air pollutant concentrations.

Figure 5.13: Scatter plots of several covariates with the log dependent variable of PAH.

Figure 5.14: Scatter plots of several covariates with the square root dependent variable of PNC.

Figure 5.15: Scatter plots of several covariates with the log dependent variable of $PM_{2.5.}$

Figure 5.16: Scatter plots of several covariates with the log dependent variable of $NO_{X.}$

Figure 5.17: Scatter plots of several covariates with the log dependent variable of BC.

Figure 5.18: Box plots of pollutant concentrations across roadway types.

Figure 5.19: Box plots of pollutant concentrations by time of day.

Figure 5.20: Box plots of pollutant concentrations by stability groups.

Figure 5.21: Autocorrelation and partial-autocorrelation autocorrelogram for the residuals from the ordinary least squares (OLS) regression of concentrations.

Figure 6.1: Box plot for 1-minute average PAH concentrations (N=8785 in 25 subjects).

Figure 6.2: Box plot for series average PAH concentrations (N=36 weekly series in 25 subjects).

LIST OF TABLES

Table 2.1: AER model coefficients, 95% confidence intervals, and P values.

Table 3.1: List of vehicles tested.

Table 3.2: AER rates (h-1) at recirculation setting for the vehicles tested.

Table 3.3: AER rates (h-1) at outside air intake setting for the vehicles tested.

Table 3.4: I/O ratios for three filter scenarios.

Table 4.1: Instruments used in this study.

Table 4.2: Sampling days, hours and meteorological conditions.

Table 4.4: Comparison of Emission factors from current study to previous studies.

Table 4.3: Pollutant concentrations on freeways.

Table 4.5: P-values for non-parametric ANOVA analysis of freeway-to-freeway differences in hourly emission rates (p-value < 0.05 for freeways having different distribution of hourly emission rates).

Table 5.1: List of Vehicles tested in the study.

Table 5.2: AER under RC Model Coefficients, Confidence Intervals, and P Values.

Table 5.3: AER under OA Model Coefficients, Confidence Intervals, and P Values.

Table 5.4: I/O GEE Model Coefficients, confidence intervals and p-values.

Table 5.5: Manufacturer share of the vehicles in operation in U.S.

Table 5.6: Summary statistics for the one-minute average on-road air pollutants.

Table 5.7: Correlation of predictor variables with dependent air pollutant variables PAH, PNC, $PM_{2.5}$ and NO_X.

Table 5.8: Correlations of predictor variables with BC concentration measurements.

Table 5.9: Grouping statistics by roadway types.

Table 5.10: Grouping statistics by time of day.

Table 5.11: Grouping statistics by modeled atmospheric stability.

Table 5.12: Prediction performance for grouping PAH by roadway types.

Table 5.13: Prediction performance for grouping PAH by time of day.

Table 5.14: Prediction performance for grouping PAH by stability.

Table 5.15: Coefficients regressed and variance explained for the prediction of PAH.

Table 5.16: Independent 1/3 holdout and 3 x3 cross validation of predictive models for PAH.

Table 5.17: Prediction performance for grouping particle number concentrations by roadway type.

Table 5.18: Prediction performance for grouping particle number concentrations by time of day.

Table 5.19: Prediction performance for grouping particle number by atmospheric stability.

Table 5.20: Coefficients regressed and variance explained for the prediction of particle number.

Table 5.21: Independent 1/3 hold-out and 3x3 cross validation of predictive models for particle number.

Table 5.22: Prediction performance for grouping $PM_{2.5}$ by roadway type.

Table 5.23: Prediction performance for grouping $PM_{2.5}$ by time of day.

Table 5.24: Prediction performance for grouping $PM_{2.5}$ by stability class.

Table 5.25: Coefficients regressed and variance explained for the prediction of $PM_{2.5}$.

Table 5.26: Independent 1/3 holdout and 3x3 cross validation of predictive models for $PM_{2.5}$.

Table 5.27: Prediction performance for grouping NOx by roadway type.

Table 5.28: Prediction performance for grouping NOx by time of day.

Table 5.29: Prediction performance for grouping NOx by stability.

Table 5.30: Coefficients regressed and variance explained for the prediction of NO_x.

Table 5.31: Independent holdout and 3x3 cross validation of predictive models for NO_x.

Table 5.32: Prediction performance for grouping BC by roadway type.

Table 5.33: Prediction performance for grouping BC by time of day.

Table 5.34: Prediction performance for grouping BC by stability.

Table 5.35: Coefficients regressed and variance explained for predictive models of BC.

Table 5.36: Independent holdout and 3x3 cross validation of predictive models for BC.

Table 5.37: Temporal autocorrelation among different daily lags.

Table 5.38: Evaluation of the time series models constructed.

Table 5.39: Shrinkage on 3x3 cross validation of predictive time series models for the air pollutants.

Table 6.1: Subject Vehicles.

Table 6.2: Multivariate regression models for the prediction of particulate PAH: continuously measured or estimated predictors.

Table 6.3: Multivariate regression models for the prediction of particulate PAH: continuously measured or estimated predictors plus time-invariant subject-reported vehicle characteristics.

ABSTRACT

On-road concentrations of traffic-related pollutants are typically much higher than concentrations measured at ambient monitoring stations. This results in in-vehicle microenvironments contributing disproportionately to the total exposure with exposures frequently being as high as on-road concentrations. However, under conditions of low air exchange rate, pollutants with significant in-vehicle losses, such as particles, can have in-vehicle concentrations that are significantly lower than those outside the vehicle. We tested a large sample of vehicles selected to be representative of the California fleet for air exchange rate (AER) at various speeds and found that AER is a predictable function of vehicle age or mileage, speed, and ventilation setting choice (outside air, recirculation, or open windows). We demonstrated that AER is the dominant factor in determining the inside-to-outside ratio for pollutants like ultrafine particles. Models were developed that explain over 79% of the variability in AER and ultrafine particle indoor/outdoor ratios across the California fleet and across the expected range of normal driving conditions. To better determine on-road concentrations, we also conducted extensive on-road measurements using a mobile platform hybrid vehicle with real-time instrumentation. Models were developed and validated to estimate on-road traffic-related pollutant concentrations (variance explained was 37% to 73% depending on the air pollutant and modeling method). Models developed in this study can be combined with subject information about their vehicle, ventilation choices, and commute route to estimate in-vehicle exposures in future studies.

EXECUTIVE SUMMARY

Background: In-vehicle exposures to vehicle-related pollutants can be up to a magnitude higher than ambient levels for traffic-related pollutants such as ultrafine particles (UFP) and black carbon. Such exposures have been estimated to contribute as much as half of the total daily exposure to ultrafine particles, for example, by nonsmoking Los Angeles urbanites for open window driving conditions. However, under some conditions of low air exchange rate (e.g., low speed, newer vehicles, and recirculating air setting) in-vehicle particle losses are significant and in-vehicle concentrations can be significantly reduced. To assess differences in in-vehicle exposure in a systematic way, we measured in-cabin concentrations of key air pollutants in the Los Angeles air basin and modeled the factors determining their variability. We then applied the results of this work to develop models for use in estimating in-transit exposures of subjects in epidemiological studies.

Methods: We conducted the following five tasks:
Task 1. a) Examine the primary differences between vehicles for in-cabin pollutant concentrations by vehicle type and age during realistic driving conditions in southern California, and conduct a comprehensive evaluation of air exchange rate (AER); and
b) (from Phase I of the proposal revisions, page 6). Test a large, representative sample of vehicle AERs at various fixed speeds and ventilation conditions.
Task 2). Examine the impact of important influential factors that contribute to in-cabin pollutant concentrations. Factors included roadway type (freeway, major arterial, and minor surface streets), total traffic counts diesel truck counts, mixing height, temperature, relative humidity, AC use, season (summer, winter), day of week, time of day (morning rush hours, noon, afternoon rush hours, night).
Task 3). Estimate emission factors of pollutants based on roadway and urban background site measurements and CO_2-based dilution adjustments.
Task 4). Develop and validate in-vehicle exposure models for selected pollutants measured in this study using data collected from Tasks 1-3.
Task 5). Validate the in-vehicle exposure model from Task 4 for PAH against measurements in representative subjects under realistic driving conditions.

Results: Task 1: We developed a simplified yet accurate method for determining AER using the occupants' own production of CO_2. By measuring initial CO_2 build-up rates and equilibrium values of CO_2 at fixed speeds, AER was calculated for 59 vehicles representative of California's fleet. Multivariate models captured 70% of the variability in observed AER using only age, mileage, manufacturer and speed. AER increases strongly with increasing vehicle age and mileage, speed, and is very high (up to a magnitude higher) if windows are open or outside air ventilation settings are chosen. High AER (75-150 h^{-1}) results in in-vehicle concentrations equaling on-road concentrations. Low AER (< 35 h^{-1}) tends to significantly reduce particle mass and number concentrations.

Task 2: We focused on ultrafine particle (UFP) number concentrations, the particle pollutant with the highest and most widely-varying loss rates. Six vehicles were tested at different driving speeds, fan settings, cabin filter loadings, and ventilation conditions (outside air or recirculation). During outside air conditions, the inside-to-outside ratios averaged 0.67 ± 0.10 (SD). I/O ratios under outside air intake ventilation mode did not vary with vehicle speed but decreased at the higher ventilation flow rates of higher fan settings. During recirculation

conditions, AER was much lower and resulted in lower I/O ratios. Ratios averaged 0.17 ± 0.13 and were highly positively correlated with AER. Under both ventilation condition types, particle removal was primarily due to losses unrelated to filtration. Filter condition, or even the presence of a filter, played a minor role in particle fraction removed.

Extensive on-road measurements were made on two arterial and three freeway routes. Measurements of real-time black carbon, UFP, $PM_{2.5}$, NO, NO_2, CO, CO_2, and particle-bound PAHs were made, with GPS and video to capture time, location, and surrounding traffic conditions. Analyses below combined these data into freeway and arterial roadway concentration models.

Task 3: Using data from Task Two, fuel-based emission factors (EF) were calculated based on simultaneous pollutant and CO_2 measurements. EFs for light-duty vehicles (LDV) were generally in agreement with the most recent studies but lower for heavy-duty vehicles (HDV), and significantly lower only for oxides of nitrogen (NO_x). Annually on I-710, a major truck route, the 6.5% fraction of total vehicle miles travelled (VMT) associated with HDV, was estimated to contribute 69% to total NO_x emissions.

Task 4: We developed models for predicting in-cabin UFP concentrations if roadway concentrations are known, taking into account vehicle characteristics, ventilation settings, driving conditions and air exchange rates (AER). Particle concentrations and AER were measured in 43 and 73 vehicles, respectively, under various ventilation settings and driving speeds. AER was the most significant determinant of UFP indoor/outdoor ratios, and most strongly influenced by ventilation setting (recirculation or outside air intake). Additional inclusion of ventilation fan speed, vehicle age or mileage, and driving speed explained greater than 79 % of the variability in measured UFP indoor/outdoor ratios.

We also developed and validated predictive models for on-road concentrations of PAH, UFP, $PM_{2.5}$, NO_X and BC that can be combined with information from above tasks to evaluate exposure to in-vehicle pollutants among study subjects. The on-road measured data were compiled with traffic variables, meteorological factors and time of day to develop regression models and non-linear generalized additive. We found that time of day was significant, accounting for 5.2%-30.3% of variance explained. Traffic variables, roadway type, and number of lanes were significant for traffic-derived pollutants but not $PM_{2.5}$. Final prediction models showed the variance explained ranged from 37% to 73% depending on the pollutant and modeling method (linear or nonlinear).

Task 5: Using personal in-vehicle PAH exposure for 25 subjects participating in another study (NIH, NIEHS R21 ES016379, Wu) we examined the predictive ability of model variables also tested in other tasks. Although many predictors from Task 4 were significant and in the direction anticipated, the overall predictive power of models was lower compared to the models from the technician-administered testing for Tasks 1-4.

Conclusions: We conclude that models developed in this study will enable us to directly study the relationship between in-vehicle air pollutant exposures and the health of the people of California. The findings of this study will have direct application to health effect studies or epidemiological studies, to the CARB' Vulnerable Populations Research Program, and eventually to evaluations of air quality standards for PM and gas pollutants.

BODY OF REPORT

1. CHAPTER ONE: INTRODUCTION

1.1 Background

Exposure to traffic related pollutants has been associated with detrimental health outcomes like asthma (Brauer et al., 2007; McConnell et al., 2010) (1-2), and cardiovascular outcomes (Delfino et al., 2005) (3), coronary artery atherosclerosis (Hoffmann et al. 2007) (4), and an increase in mortality (Hoek et al., 2002) (5). Numerous studies (e.g., Leung and Harrison 1999; Westerdahl et al., 2005, Zhu et al. 2007) (6-8) have shown that pollutant concentrations on or in the vicinity of roadways are frequently almost one order of magnitude higher than ambient levels.

In-vehicle exposures to vehicle-related pollutants are frequently high, due to a vehicle's proximity to relatively undiluted emissions from other vehicles and the typically rapid air exchange rate (AER) inside vehicles (Fletcher and Saunders 1994; Ott et al. 2007; Rodes et al. 1998) (9-11), which drives pollutants to and from the cabin. Often, In-vehicle pollutant concentrations are approximately a magnitude higher than ambient levels for ultrafine particles (UFP) and volatile organic compounds (VOCs) (Leung and Harrison 1999; Westerdahl et al. 2005; Zhu et al. 2007; Chan et al. 1991; Duffy and Nelson 1997) (6-8, 12,13). This has important implications for exposure assessment. For example, the less than 10% of daily time that is estimated to be spent in vehicular transit microenvironments (Klepeis et al., 2001) (14) has been predicted by Fruin et al. (2008) (15) to contribute 35-50% of total UFP and 30-55% of black carbon (BC) exposure for nonsmoking urbanites in Los Angeles under open window conditions and 17% of UFP by Wallace and Ott (2011) (16) for more suburban locations.

On an average 95 min per day spent in the in-vehicle microenvironment among Californians Furthermore, the Southern California Association of Governments (SCAG) predicts that commuting times will double by 2020 due to population growth in the LA area (SCAG 1997) (17), adding urgency to research evaluating the impact of increased vehicle-related exposures on people’s health. But despite the demonstrated contribution of transit/vehicular microenvironment to the total exposure, it remains largely uncharacterized to date.

Compared to other microenvironments, vehicles typically have rapid air exchange rates and more complex structure whereby a multitude of factors including traffic mix and density, type and age of the vehicle, roadway type, vehicle speed, ventilation setting, and weather conditions combine interactively to determine the in-vehicle pollution levels. In effect, these factors can be divided into two categories; one of those that affect the I/O ratios (mostly physical characteristics of the car and drivers ventilation preferences) and the other of those factors that influence the roadway concentrations (like traffic and meteorological parameters). Therefore, to accurately assess the in-vehicle exposure, not only is it crucial to know the AERs and I/O ratios for pollutants but also the roadway concentrations. Also, large variations in exposure incurred inside

vehicles are expected to occur not only due to differences in roadway environments but also because inside-to-outside ratios (i.e., in-vehicle to roadway concentration ratios) (I/O) vary from vehicle to vehicle due to differences in ventilation conditions and other vehicle characteristics that affect air exchange rate (AER), which is defined as the number of times per hour vehicle cabin air is replaced by roadway/outside air. In general, I/O ratios in vehicles can range from nearly zero to nearly one. Recent studies (Knibbs et al., 2010) (18) have shown that I/O is strongly dependent on AER.

A few studies also characterized in-cabin air exchange rates (AER) under different driving conditions (Fletcher and Saunders 1994; Ott et al. 2007; Rodes et al. 1998) (9-11). All the studies showed a wide range of AERs during commuting. For example, the Ott et al. study (2007) (10) found that the in-cabin AER ranged from 1.6 h^{-1} to 71 h^{-1}, depending on vehicle speed, window position, ventilation system, and air conditioner setting. For closed windows and passive ventilation, the AER was linearly related to the vehicle speed over a range from 15 to 72 mph. The lowest AERs (<6.6 h^{-1}) occurred when windows were closed and the ventilation system was off. Opening a single window by 7.6 cm increased the AER by 8–16 times.

Pui et al. 2008 (19) and Qi et al. (20) 2008 have experimentally demonstrated that a dramatic reduction can be achieved in UFP concentration in-cabin with use of recirculation setting and consequent filtration, but did not establish a link with AER. Only two major studies have been identified that have observed the behavior of UFP inside the vehicle cabin when moving in real on road conditions. Recently, Knibbs et al. (2010) (18) studied the UFP in five cars and reported a high correlation between inside-to-outside UFP concentration ratios and AER (r^2 = 0.81), with somewhat higher losses with the recirculation ventilation setting. They report ratios in the range 0.08-0.47 when recirculation setting was on with low fan and 0.17-0.68 with fan off. Another major study by Zhu et al. 2007 (8) also report that maximum particle losses (~85% reduction in in-cabin concentrations) were observed at recirculation settings. This study tested three vehicles on Los Angeles freeways. They also suggest that an hour of commute in Los Angeles can be responsible for as much as 50% of daily UFP exposure. Both of these studies performed measurements under real driving conditions (multiple speed and ventilation conditions) and found that ventilation preference (windows open, outside air intake or in-cabin air recirculation) and ventilation fan setting strongly influences AER and the resulting I/O ratio. However, these studies are limited in nature and the current state of literature does not sufficiently address the epidemiological needs to assess exposure at a large scale. No studies have been identified that systematically assess exposure to gaseous pollutants during in-vehicle transit.

As mentioned previously, accurate prediction of in-vehicle exposures requires not only an estimate for I/O but also knowledge of on-road pollutant concentration.
Recently, a few studies have tried to characterize the roadway concentration of pollutants. Westerdahl et al. (2005) (7) conducted a mobile platform study in Spring, 2003 using a 1998 electric Toyota RAV4 SUV. The study showed relatively high correlation of UFP with nitrogen oxides (NO_x), black carbon (BC), and PM-bound polycyclic aromatic hydrocarbons (PM-PAH). Fruin et al. (2008) (15) conducted in

depth data analysis for this mobile platform study. They showed that on freeways, concentrations of UFP, BC, NO_x, and PM-PAH are generated primarily by diesel-powered vehicles, despite the relatively low fraction (6%) of diesel-powered vehicles on Los Angeles freeways. However, UFP concentrations on arterial roads appeared to be driven primarily by proximity to gasoline-powered vehicles undergoing hard accelerations. The Fruin et al. (2008) (15) results were promising since it demonstrated that in-vehicle exposures can be estimated using statistical models that incorporate traffic activity, meteorological conditions, and other relevant parameters. However, most of their measurements were conducted from 9 AM to 3 PM, which deliberately avoided traffic congestion times. Results from the southern California Regional Travel Survey showed that approximately 35% of trips occur from 9 AM to 3 PM, while 41% of the trips occur during rush hour from 6 AM to 9 PM and from 3 PM to 7 PM (SCAG 2003) (17). Considering much higher pollutant concentrations during congestion and the significant fraction of commuting time people spent during morning and afternoon traffic congestion hours, it is important to characterize and model in-cabin exposure during traffic congestion conditions. Moreover, the study provided insight into different patterns of concentrations on freeways and arterials; however, their results were not generalized for arterials streets although approximately half of the vehicle miles traveled are on major arterials in the region (Houston et al. 2004) (21).

Nonetheless, the interest in assessing transit time exposures has been growing. A recent review by Knibbs et al. 2011 (22) identified 47 UFP exposure studies performed across 6 transport modes: automobile, bicycle, bus, ferry, rail and walking. They concluded the following

> "*While the mean concentrations were indicative of general trends, we found that the determinants of exposure (meteorology, traffic parameters, route, fuel type, exhaust treatment technologies, cabin ventilation, filtration, deposition, UFP penetration) exhibited marked variability and mode-dependence, such that it is not necessarily appropriate to rank modes in order of exposure without detailed consideration of these factors*."

In summary, at the time of undertaking this study, the following limitations existed in the literature.

1. AER measurement existed for only 16 vehicles in real driving conditions, which were not systematically tested. Only Fletchers and Sunders et al. 1994 (9) had made an attempt to quantify the AER. As a result of a) small sample size b) differing methodologies in different studies, and c) missing information on variables that determine AER in many studies, the results on AER could not be conclusively tied to determinant factors. Furthermore, they could not be extrapolated with confidence to produce estimates at a fleet-wide level as is desired in an epidemiological study or for population risk assessment.
2. I/O measurements existed in even fewer vehicles. Only two studies measured them under realistic condition (Zhu et al., 2007, Knibbs et al., 2010) (8, 18). Of them only one study Knibbs et al., 2010 (18), measured I/O rations in such a manner that they

could be related to a quantifiable parameter like AER or vehicle speed. Nonetheless, no systematic attempt had been made towards understanding the following: a) what factors drive I/O under real driving conditions, and b) the order of influence of these factors that could help epidemiologists design a questionnaire to gather such data for large population studies.

The gap in knowledge prior to the present study prevented any generalization based on the above previous studies to predict in-transit exposure of individual subjects in epidemiological studies. In addition, none of the previous studies were directly intended for linkage to any health outcome research.

1.2. Scope and Purpose of the Project

The main purpose of the study was to collect in-vehicle air pollution data in Southern California, develop and validate in-vehicle exposure models, and apply the model results to help estimate in-vehicle exposure for subjects in future epidemiologic studies. The results of exposure modeling in this study may be used to develop similar models elsewhere in California or the US, but they would likely have to be validated by other investigators for the specific region. We were especially interested in developing these modes for use in future cohort studies. The modeling approach was designed for such use as discussed below and was intended to be useable with data collected in epidemiologic studies having detailed time-activity data and information about a subject's vehicle including easily-obtainable information like make, mileage and year.

To this end, we measured and modeled in-cabin concentrations of key air pollutant that are expected to serve as markers of exposure to complex mixtures of primary combustion aerosols and gases (e.g., BC, NO_x and ultrafine particle numbers [UFP]). We aimed to produce data to fully characterize the variability in a range of different pollutant concentrations in vehicles, including validating the in-vehicle exposure models using separate measurements of particle-bound PAH.

This project was intended to enhance our ability to estimate personal exposure to vehicle-related air pollutants. This could then be used in future studies to evaluate hypotheses regarding the role of air pollution exposure from the in-vehicle environment on the development and exacerbation of chronic diseases, including asthma and cardiovascular disease. The results and products of the proposed study are anticipated to be crucial in obtaining funding to study the health impacts of in-vehicle exposures. There are few published studies to our knowledge that have systematically examined in-vehicle exposure and the health effects of such exposure. The exception being studies of acute cardiorespiratory effects using quasi-experimental in-vehicle exposures (Adar et al. 2007a; 2007b; Riediker et al. 2005) (23-25). However, epidemiologic studies have explored the associations between traffic generated pollution and risk of developing asthma (Brauer et al. 2007; McConnell et al. 2010), (e.g., 1,2) adverse birth outcomes (Wu et al. 2009, Gehring et al. 2011, Brauer et al. 2008) (e.g., 26-28), and evidence of coronary artery disease (Hoffmann et al. 2009) (4). The present study will provide

measurement data and develop models to estimate chronic exposure-response relationships. This research is envisioned to augment exposure assessments for the work, home and neighborhood environments.

We built models that are intended to enhance our ability to incorporate estimated exposure from time in vehicles into health effects models. The in-vehicle environment has been largely ignored in previous epidemiological studies. The availability of data generated from this study will present a unique opportunity in future studies. For instance, it is not known whether in-vehicle exposure to air pollution during pregnancy adversely affects birth outcomes and promotes the occurrence of atopic sensitization and childhood respiratory diseases, including asthma. It is conceivable that health impacts from in-vehicle exposures will be as important, or more so, than exposures linked to the outdoor home environment, especially in the region of study. The exposure modeling provides results that will allow quantitative estimates of in-vehicle exposure to key pollutants given known driving conditions and other parameters. It will guide epidemiological studies focused on commuters' health outcomes, and help inform policy decision makers concerning motor vehicle emissions control.

The present research is among the first to systematically examine in-transit exposure and the conditions that drive major changes in exposure, and to develop models that can be used to estimate in-transit exposure for subjects in epidemiological studies. Methods could be adapted to regions where driving conditions and meteorology differ from southern California. The real-time data on gases and particulate air pollutants within vehicles will also provide information needed to support emission regulations for vehicles and effective pollution control strategies.
Three major strengths of this study are:
1) Use of representative vehicle types, roadway types, traffic fleets, driving conditions, seasons, and time of day;
2) Combining in-cabin measurements with real-time route information (through a GPS device), roadway information, and available traffic count data; and
3) Testing of identified predictors of exposure using subjects under normal commuting conditions.

Models developed in this study will enable us to directly study the relationship between in-vehicle air pollutant exposures and the health of Californians. The findings of this study will have direct application to CARB's Vulnerable Populations Research Program and to evaluations of air quality standards for PM and gas pollutants. Results are expected to advance understanding of the potential for adverse effects of vehicle-related air pollutants.

1.3. Tasks

Overview
We conducted an in-vehicle exposure monitoring and modeling study. The target study region for this proposal included the counties of the South Coast Air Basin that are anticipated to be of interest for epidemiologic research, namely, Los Angeles, San

Bernardino, Riverside and Orange Counties. Effort was made to collect representative measurements on not only freeways but also arterial roads – over varying traffic conditions, time of day, day of week and seasons. This enhanced variability in characteristics of particles, and enhanced the external validity of findings to populations at risk.

The following tasks were completed:

1. **Examine differences between vehicles for in-cabin pollutant concentrations by vehicle type and age during realistic driving conditions in southern California.**

   1a. Field Measurements. We measured AERs in over 60 vehicles at 3-4 speeds per vehicle (Phase I of III), in addition to stationary measurements to establish baseline AER. Two lower speeds (20 and 35 MPH) helped estimate AERs during typical arterial driving conditions and two higher speeds (55 and 65 MPH) helped estimate AERs during freeway driving. In addition to AERs, measurements were made for $PM_{2.5}$ and total particle number concentration. Furthermore, vehicles were selected to match the distribution in California fleet for age, mileage, and vehicle class and manufacturer.

   1b. Data Analyses. First, we developed a novel methodology to derive AER measurements. Second, we examined the influence of vehicle type, age, mileage, manufacturer and driving speed on AER, in addition to the most crucial determinant of AER ventilation choice (outside air intake or recirculation of cabin air). Third, we developed a model to estimate AER.

2. **Examine the impact of important influential factors that contribute to in-cabin pollutant concentrations.**

   2a. Field Measurements.

   This task was conducted in two additional phases. Phase II explored the factors that determine I/O ratios. We sought to examine the factors that influence I/O ratio and factors that influence roadway concentrations separately. This approach allowed the development of a systematic understanding in each phase and allowed us to conduct additional roadway sampling to successfully capture data under varying conditions (ranging from seasons to time of day).

   In Phase II, we measured a number of pollutant concentrations using a hybrid-electric vehicle on five selected routes that covered the southern California region of interest. Measurements were conducted on weekday/weekends, different times of day, and in both warm and cool conditions. Air pollutants included Aethalometer BC, total particle counts (CPC), particle-bound PAHs (PAS), NO-NO_x, CO, and CO_2. In addition to measuring $PM_{2.5}$ mass using a DustTrak we also stored particle filter samples for future analysis of chemical species as a function of particle size (PCIS) after measuring gravimetric mass.

Phase III explored the factors that determine I/O ratios. In Phase III, six representative cars were chose from the fleet previously tested in Task 1 and tested at different driving speeds, fan settings, cabin filter loadings, and ventilation conditions (outside air or recirculation).

2b. Data Analyses.

We examined the impact of roadway types, traffic characteristics, temporal factors, and meteorology (including seasonal effects) on roadway pollutant concentrations. Further the influence of speed, ventilation fan setting, filter loading and particle size was quantified for UFP I/O.Estimate emission factors of PM pollutant concentrations based on roadway and urban background site measurements and CO_2-based dilution adjustments.

Measurements for gas and particulate phase pollutants were performed using a mobile platform during the summer of 2011 on various Los Angeles freeways. Fuel-based emission factors (EF) were calculated for light-duty vehicles (LDV) and heavy-duty vehicles (HDV). The fractional contribution of HDV to total NO_x was calculated for different freeways including those with larger proportions of HDV. We also compared morning and afternoon rush hours, and midday traffic for speeds, truck fraction, VMT and per mile emissions.

3. Develop and validate in-vehicle exposure models for BC, UFP number, $PM_{2.5}$, particle-bounded PAH, and NO_x.

The models incorporated data from Tasks 1-2 on time of day, season, car types, driving conditions, roadway types, traffic characteristics, and meteorological conditions and were developed based on a training dataset (70% of randomly-selected measurements) that was validated against the remaining 30% random validation sample. K-fold cross-validation was also used to validate the models for each of the selected pollutants.

4. Validate the in-vehicle exposure model for particle-bound PAH against measurements in representative subjects.

We used data from a pregnancy cohort of 92 women who completed a time-activity questionnaire at baseline and carried a GPS devise to track movements over one-week for three different pregnancy periods (<20 weeks, 20-30 weeks, and >30 weeks of gestation) (NIH, NIEHS R21 ES016379, Wu). Twenty-five of these subjects also carried portable personal exposure monitors for particle-bound PAH (EcoChem PAS) for one-week (including weekdays and weekends) during their commutes. However, BC data collected in 9 of those subjects were insufficient for modeling due to instrument problems. These data are from working subjects in real world driving conditions. They were used as a first test the predictive ability of variables identified from the models developed in Task 4.

In-vehicle Testing Procedures for Tasks 1-3

Vehicle testing was conducted in three phases as follows:

Phase I (Task 1) tested a large, representative sample of vehicles for air exchange rate (AER). This was performed by measuring the decay rate of CO_2 at various fixed speeds and ventilation conditions. In addition, a series of alternating closed (with recirculation) and open window tests were conducted to test each vehicle's air movement systems for losses of particle number or particle mass.

Phase II comprehensively measured on-road concentrations on various road types across the LA Basin for multiple pollutants at different times of day and in different traffic conditions.

Phase III involved simultaneously measuring inside and roadway (outside) concentrations for various pollutants under different ventilation conditions to measure attenuation factors (AF), the loss rates for each pollutant.

On-road concentrations drive in-vehicle concentrations. We can assume that in-vehicle concentrations are a predictable function of on-road concentrations with losses reflected by some pollutant-specific AF such that:

$C_{in\text{-}cabin} = C_{on\text{-}road} * (1 - AF)$

where $C_{on\text{-}road}$ = f (traffic and truck volumes, meteorology, road type, lane, speed, etc.);

AF = f (AER, pollutant, cabin surface-to-volume ratio, fan setting) (see Phase III); and

AER = f (speed, vehicle type, age / mileage) if windows closed and ventilation is set to recirculate (see Phase I), otherwise,

AER = f (speed) if windows open or ventilation is set to outside air with fan on.

The latter situation tends to produce much higher AERs.

Of all the measurements proposed, the on-road concentrations are the most widely-ranging and rapidly-changing measurement we needed to make, being a function of constantly-changing traffic mix, traffic conditions and meteorology, which all vary greatly. By measuring AERs and AFs under more controlled conditions in separate tests, we were able to determine each with greater accuracy. Furthermore, by measuring on-road concentrations directly without the modifying effects of different AERs and AFs, we gain simplicity and reduce measurement variability, which was intended to make the effects of on-road variables more distinct and easier to model.

Phase I. Testing of Air Exchange Rates

We measured air exchange rates (AERs) in vehicles using a hand-held QTrak (and/or LI-COR 820) to measure CO_2 decay rates while driving at near-constant speeds (e.g., 20, 40 and 60 mph, or similarly-spaced intervals, depending on available routes and speed limits). Windows were closed and the ventilation conditions set to:

-- fan off and recirculation off

-- fan on low and recirculation off (outside air)

-- fan on medium or next higher level and recirculation off (outside air)
-- fan on low and recirculation on (with air conditioning on)

The LI-COR 820 CO_2 monitor was used because it has a faster response since it is pump driven and has a higher upper range than the QTrak, although the QTrak was certainly adequate for AER tests. The LI-COR was needed for on-road measurements (Phase II) where CO_2 is used to calculate dilution rates or emission factors. In those tests, CO_2 concentrations frequently fluctuate rapidly. During and after the AER tests, several battery-operated instruments (DustTrak, Aethalometer, and CPC) were run to provide supplementary pollution concentration measurements, since these instruments could be included with no additional fixturing required and little additional labor.

Routes were chosen for low traffic levels and the ability to drive continuously at a given speed with no stops for the duration needed. Duration needed is determined by the lowest AER expected. The lowest AER for a moving vehicle reported in the literature is 1.6 hr^{-1} at 20 mph, as reported by Ott et al. (2007) (10). This AER would require about 26 minutes to halve a given CO_2 concentration, and require about 9 miles of driving. Minimum distances to reduce CO_2 a given amount will decrease as speeds increase due to the non-linear increase in AER with speed. Most vehicles will have much higher AERs than this example and require much shorter driving distances.

The test began at the start of the selected route with two researchers building up in-cabin CO_2 levels inside the test vehicle via respiration with windows closed and the car motionless. Because the rate of CO_2 build-up rate will reflect the source CO_2 term and the air exchange rate while stationary, the QTrak also recorded during this build-up time. During build-up and all decay tests, cabin air was kept well mixed by a battery-operated fan or vehicle fan set to recirculate. The target CO_2 level for build-up was 4000 ppm.

When 4000 ppm CO_2 was reached, the car was driven at a fixed speed, ideally within ±2 mph, according to the judgment of the driver, traffic conditions, and safety (to be later verified by on-board GPS). The passenger seat observer recorded the time, to the second, for each 100 ppm decrease in CO_2 as back-up to the QTrak memory. The test was complete when the CO_2 concentrations reach 1000 ppm or begin to flatten out, whichever comes first. If constant speed was significantly interrupted, the test was repeated. If the vehicle AER appeared too low to complete the test on the selected route, the next higher speed was attempted.

When AER tests were complete, a series of alternating open and closed window tests (with air set to recirculate) were made at constant speed to test the effect of each vehicle's air handling unit on particle losses. Losses were determined by comparing inside and outside PM mass, black carbon, and particle number from the battery-operated instruments. Each condition was held for two minutes or until conditions reach steady state, whichever was longest, and a minimum of five alternating pairs of measurements were collected for each vehicle.

Cars were chosen in an attempt to get representative data for the California fleet. Each vehicle tested for AER had its mileage, age, internal and external condition recorded, its internal dimensions measured, and the ventilation system options and operation were carefully noted, especially as to what the default ventilation settings are and if the system is semi-automated, what the most common settings end up being.

Phase II. Measurement of On-Road Concentrations

Depending on instrument availability, black carbon, particle number, $PM_{2.5}$, particle-bound PAHs, NO_x, CO and CO_2 were continuously measured. Measurements took place in a hybrid vehicle outfitted with instruments, batteries and inverter, along with GPS and video. Hybrid vehicles have the advantage of no emissions while stopped, which is a situation where a vehicle's own exhaust can sometimes get sampled. Measurements were made in morning rush hour, noontime non-rush hour, afternoon rush hour, and nighttime non-rush hour with realistic driving.

Phase III Measurement of Pollutant Loss Rates (Attenuation Factors)

Pollutants with significant surface reaction or deposition loss rates will have potentially important losses at low AERs, and these losses will increase as AER is reduced. The losses will likely be highest for ultrafine particles (UFP), semi volatile species and may be potentially significant at sufficiently low AERs for black carbon, $PM_{2.5}$, NO_2 and CO. Although CO is non-reactive, significant CO removal rates can occur due to uptake from passengers at low AERs. Under conditions of low AER, measurable particle uptake from passengers can also occur, but we could not distinguish between occupant-driven particle losses and those due to surfaces. Thus, we assumed that under most circumstances, loss rates were not significantly different between one, two, and three occupants and that in the case where particle losses due to occupants is significant, our measurements reflected particle losses with two occupants present. We also assumed that any non-reactive, non-depositing pollutant will have 0% attenuation.

We characterized the AF for each pollutant as a function of three variables: 1) AER, 2) cabin volume to cabin surface ratio, and 3) fan setting (at low AER and recirculating air). We also included low fan settings of outside air since this is a frequent default setting in many cars. (The case of newer cars with particle filtration systems is addressed at the end of this section.)

AER is a dominant factor because it drives the renewal rate of the pollutant being removed. We can assume 0% AF for all pollutants when AER is high enough, such as with open windows at moderate speeds or higher, so these loss tests should emphasize closed windows conditions with ventilation set to recirculate. After our extensive AER testing described below we knew identified vehicles with low AERs when outside air was being pulled in by the ventilation system (the common default setting noted above). We also included both recirculation and outside air fan settings in our tests. Because AER is a non-linear function of speed for closed window conditions, relatively constant speeds were important for these tests, as described in Phase I.

The ratio of vehicle surface area to cabin volume may affect AF by increasing or decreasing the relative fraction of pollutant available to interact with surfaces, but we expect this effect to be not as pronounced as the effect of AER. Vehicle interior surface area is difficult to measure, but can be approximated by assuming surface area from vehicle to vehicle is proportional to the seat area plus the area equivalent to the rectangular inner cabin dimensions. Likewise, the cabin volume can be approximated by the rectangular volume of the inner cabin dimensions. A distribution of the ratios based on these dimensions was collected from the vehicles used in the AER tests earlier in the study.

For removal processes that are diffusion rate limited, fan setting may also affect losses by enhancing mixing at higher fan speeds (and reducing boundary layer depletion next to surfaces) and also by inducing turbulence in the air movement system, which tends to increase deposition rates.

To include all of these variables, we used a measurement matrix of 6 surface-to-volume ratios (using three vehicles) x 10 combinations of AER and fan settings (low and high settings for recirculating air and a low setting for outside air, all with windows closed). AERs can be based on our measured AER quartiles (25, 50, and 75th percentiles). Surface-to-volume ratios were chosen to cover low, medium, and high ratios.

For newer cars less than five years old that may have particle filtration systems, we first tested for the presence of filtration by observing the difference in UFP concentration when the ventilation setting is set to outside air and the fan is on medium or high, while alternating between open and closed windows. If incoming air is being filtered, closed windows will cause sharp drops in UFP levels. For cars with filtration systems, we established the filtration efficiency for UFP and $PM_{2.5}$ at low vehicle speeds (i.e., 20 mph) by multiple iterations of the above closed versus open window tests, alternating every 60 seconds on roadways with low traffic and reasonably stable UFP concentrations. When the filter efficiency is established, we assumed this is the dominant loss mechanism for particles and the AF were 1.0 minus the filter efficiency. Tests were then conducted as described above. We assumed that few if any vehicles have working activated carbon filtration systems for removal of gaseous pollutants.

Lastly, if the open/closed window tests in Phase I indicated that significant particle losses occur in certain vehicle types (or certain air movement system types) when the ventilation system is set to recirculation, one or more of each of these vehicle types (or air movement system) were included in Phase III tests, excluding the modifications of surface-to-volume ratios.

In the each of the following Chapters, which are divided by Tasks, we give an introductory overview, describe the materials and methods, present the results with discussion, and end with a summary and conclusions section.

References

1. Brauer M, Hoek G, Smit HA, de Jongste JC, Gerritsen J, Postma DS, Kerkhof M, Brunekreef B. Air pollution and development of asthma, allergy and infections in a birth cohort. Eur Respir J 2007;29:879-88.
2. McConnell R, Islam T, Shankardass K, Jerrett M, Lurmann F, Gilliland F, Gauderman J, Avol E, Künzli N, Yao L, Peters J, Berhane K. Childhood incident asthma and traffic-related air pollution at home and school. Environ Health Perspect 2010;118:1021-6.
3. Delfino, R.J.; Malik, S. and Sioutas, C. Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health. *Environ. Health Perspectives*. 2005, 113, 934-946.
4. Hoffmann B, Moebus S, Möhlenkamp S, Stang A, Lehmann N, Dragano N, Schmermund A, Memmesheimer M, Mann K, Erbel R, Jöckel KH; Heinz Nixdorf Recall Study Investigative Group Residential exposure to traffic is associated with coronary atherosclerosis. Circulation 2007, 116, 489-496.
5. Hoek, G.; Brunekreef, B.; Goldbohm, S.; Fischer, P. and van den Brandt P.A. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. The Lancet. 2002, 360, 1203-1209.
6. Leung P.L. and Harrison R.M. Evaluation of personal exposure to monoaromatic hydrocarbons, Occup. Environ. Med. 1998, 55, 249–257.
7. Westerdahl, D.; Fruin, S.; Sax, T.; Fine, P.M. and Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmos. Environ*. 2005, 39, 3597–3610.
8. Zhu, Y., Eiguren-Fernandez, A., Hinds, W.C., Miguel, A.H.. In-Cabin Commuter Exposure to Ultrafine Particles on Los Angeles Freeways. *Environ Sci Technol* 2007, 41(7): 2138-2145.
9. Fletcher, B., Saunders, C.J.. Air Change Rates in Stationary and Moving Motor-Vehicles. *Journal of Hazardous Materials* 1994, 38(2): 243-256.
10. Ott, W., Klepeis, N., Switzer, P.. Air change rates of motor vehicles and in-vehicle pollutant concentrations from secondhand smoke. *J Expo Anal Environ Epidemiol* 2007, 1-14.
11. Rodes, C., Sheldon, L., Whitaker, D., Clayton, A., Fitzgerald, K., Flanagan, J., et al. 1998. Measuring Concentrations of Selected Air Pollutants Inside California Vehicles. Final Report. Contract No. 95-339. Sacramento, CA: California Air Resources Research Division Board.
12. Chan, C.C., Ozkaynak, H., Spengler, J.D., Sheldon, L.. Driver Exposure to Volatile Organic-Compounds, Co, Ozone, and No2 under Different Driving Conditions. *Environmental Science & Technology* 1991, 25(5): 964-972.
13. Duffy, B.L., Nelson, P.F., 1997. Exposure to emissions of 1,3-butadiene and benzene in the cabins of moving motor vehicles and buses in Sydney, Australia. *Atmospheric Environment* 31(23): 3877-3885.
14. Klepeis, N. E., Nelson, W. C., Ott , W. R., Robinson, J. P., Tsang, A. M., Switzer, P., et al., 2001. The national human activity pattern survey (NHAPS): A resource for assessing exposure to environmental pollutants. J Expos Anal Environ Epidemiol. 11(3), 231-252.

15. Fruin S., Westerdahl D., Sax T., Sioutas C., and Fine P.M., 2008. Measurements and predictors of on-road ultrafine particle concentrations and associated pollutants in Los Angeles. Atmos Environ 42, 207–219.
16. Wallace L., and Ott W. Personal Exposure to Ultrafine Particles., 2011. J Expo Sci Environ Epidemiol 21, 20–30.
17. SCAG. 2003. Year 2000 Post-Census Regional Travel Survey. Final Report of Survey Results. Los Angeles, CA: Southern California Association of Governments.
18. Knibbs, L.D., de Dear, R.J. and Morawska, L., 2010. Effect of cabin ventilation rate on ultrafine particle exposure inside automobiles. Environ. Sci. Technol. 44, 3546-3551.
19. Pui, D. Y. H.; Qi, C.; Stanley, N.; Oberdorster, G. and Maynard, A. Recirculating air filtration significantly reduces exposure to airborne nanoparticles. *Environ. Health Perspect.* 2008, 116, 863-866.
20. Qi, C.; Stanley, N.; Pui, D. Y. H. and Kuehn, T. H. Laboratory and on-road evaluations of cabin air filters using number and surface area concentration monitors. *Environ. Sci. Technol.* 2008, 42, 4128-4132.
21. Houston, D., Wu, J., Ong, P., Winer, A., 2004. Structural disparities of urban traffic in Southern California: Implications for vehicle-related air pollution exposure in minority and high-poverty neighborhoods. *Journal of Urban Affairs* 26(5): 565-592.
22. Knibbs LD, Cole-Hunter T, Morawska L (2011) A review of commuter exposure to ultrafine particles and its health effects. Atmos Environ 45: 2611-2622.
23. Adar, S.D., Adamkiewicz, G., Gold, D.R., Schwartz, J., Coull, B.A., Suh, H., 2007a. Ambient and microenvironmental particles and exhaled nitric oxide before and after a group bus trip. *Environ Health Perspect* 115(4): 507-512.
24. Adar, S.D., Gold, D.R., Coull, B.A., Schwartz, J., Stone, P.H., Suh, H., 2007b. Focused exposures to airborne traffic particles and heart rate variability in the elderly. *Epidemiology* 18(1): 95-103.
25. Riediker M., W.E. Cascio, T.R. Griggs, M.C. Herbst, P.A. Bromberg, L. Neas, R.W. Williams, R.B. Devlin Particulate matter exposure in cars is associated with cardiovascular effects in healthy young men American Journal of Respiratory and Critical Care Medicine, 169 (2004), pp. 934–940
26. Wu J., C. Ren, R.J. Delfino, J. Chung, M. Wilhelm, B. Ritz, Association between local traffic-generated air pollution and preeclampsia and preterm delivery in the south coast air basin of California, Environ. Health Perspect., 117 (2009), pp. 1773–1779
27. Gehring U, van Eijsden M, Dijkema MB, van der Wal MF, Fischer P, Brunekreef B. Traffic-related air pollution and pregnancy outcomes in the Dutch ABCD birth cohort study. *Occup Environ Med.* 2011;68:36–43. doi: 10.1136/oem.2009.053132.
28. Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. A cohort study of traffic-related air pollution impacts on birth outcomes. *Environ Health Perspect.* 2008;116:680–686. doi: 10.1289/ehp.10952.

2. CHAPTER TWO: A PREDICTIVE MODEL FOR VEHICLE AIR EXCHANGE RATES BASED ON A LARGE, REPRESENTATIVE SAMPLE

(based on Task 1. Examine the primary differences between vehicles for in-cabin pollutant concentrations by vehicle type and age during realistic driving conditions in southern California, and add a comprehensive evaluation of air exchange rates [AER])

2.0 INTRODUCTION

The in-vehicle microenvironment is an important route of exposure to traffic-related pollutants. In-vehicle exposures are high due to vehicles' frequent proximity to relatively undiluted emissions from other vehicles, particularly in urban areas; the typically rapid air exchange rate (AER) inside vehicles (1-6); and the average 80 min per day spent by people in the U.S. in the in-vehicle microenvironment (7). Jenkins et al. (8) reported that Californians spend 7% of their time (100 minutes) in enclosed transit.

On-road and in-vehicle concentrations of traffic-related pollutants are typically an order of magnitude higher than urban ambient concentrations (9-11). The pollution concentrations inside a vehicle generally match the roadway concentrations when there is sufficiently high air turnover. This occurs whenever windows are open, whenever outside air is drawn into the vehicle through the ventilation system, or when a vehicle is sufficiently leaky. However, under conditions of sufficiently low air exchange rate, i.e., only a few air changes per hour, there can be significant reductions in particle mass and particle number due to losses to vehicle internal surfaces (12, 13). Conditions of low air exchange usually only occur for newer cars, for which door seals and insulation are tightest, and/or at low speeds where air flow dynamics are not producing large differences in pressure around the vehicle. If the air exchange rate (AER) of a vehicle is known, the particle losses can be estimated (12); however, AERs are usually not known, and are highly variable even for the same vehicle, as they vary widely with speed (1, 4, 6). For example, Knibbs et al. (1) found AERs to vary from 1 to 33 air changes per hour (hr^{-1}) across six cars at a speed of 60 km hr^{-1}.

Few studies have characterized AERs. The largest study to date has been Knibbs et al. (1) who measured AER using SF_6 as a tracer gas for six vehicles spanning an age range of 18 years at various speeds and under different ventilation settings. At speeds of 60 km hr^{-1} and 110 km hr^{-1}, they found AER to range from 1 to 33 hr^{-1} (mean 11.2) and 2.6 to 47 hr^{-1} (mean 18), respectively. They also tested cars at zero speed and reported AERs within the range 0.1-3.3 hr^{-1} with five cars having AERs <1/hr. Ott et al. (4) reported AERs in the range of 1.6-8.2 hr^{-1} for vehicles up to 10 years old tested using CO as a tracer gas. They also provide an excellent review of previous studies on the subject. Batterman et al. (14) report an AER of 92 hr^{-1} for a vehicle (Subaru Legacy 2000) traveling at 105 km hr^{-1} when ventilation was set to intake fresh air. Kvisgaard and Pejtersen et al. (15) also report AERs for a traveling vehicle for fresh air intake ventilation setting at various speeds.

AERs observed during conditions that bring fresh air into the cabin (either via ventilation system set to fresh air supply or by opening windows) can be a magnitude or higher compared to those observed at internal air recirculation settings. Knibbs et al. (1) conducted experiments for six cars and showed that even at lowest fan settings, AERs were typically over 100 hr^{-1}, even at zero speed, thus making the determination of AER unnecessary from an exposure standpoint, as in-vehicle concentration will equal on-road concentration at such high levels of air turnover. Ott et al. (4) found similar results when opening the windows by 3 inches increased AERs 8-16 times.

When windows are closed and recirculation is used, AER tends to be minimized and in-vehicle particle concentrations are also minimized due to particle losses. Knibbs et al. (16) tested the same five cars used in previous AER measurements of 2009 and found high correlation between inside-to-outside UFP concentration ratios and AER (r^2 = 0.81), with somewhat higher losses with the recirculation fan on. They report ratios in the range 0.08-0.47 when recirculation setting was on with low fan and 0.17-0.68 with fan off. Zhu et al. (11) also report that maximum particle losses (~85% reduction in in-cabin concentrations) were observed at recirculation settings. Pui et al. (17) and Qi et al. (18) have experimentally demonstrated that a dramatic reduction can be achieved in UFP concentration in-cabin with use of recirculation setting and consequent filtration.

Beside the work by Ott et al. (3, 4), Knibbs et al. (1), Rhodes et al. (6) and Fletcher and Saunders (19), (a total of 16 cars tested), others have tested AERs in stationary vehicles, but not during on-road conditions, where most of the travel time exposure occurs.

The purpose of this task was to test a sufficiently large number of cars to develop robust predictive models of AER that allow estimating vehicle AER as a simple function of readily-available information, such as vehicle age, mileage, manufacturer, and average speed. One important application of these models is epidemiological studies of particulate matter (PM), especially for coarse PM (CPM, $PM_{2.5\text{-}10}$, 2.5 µm< D_p< 10 µm) or ultrafine particulate matter (UFP, D_p < 0.1µm). CPM and UFP show sharp near-road gradients and high on-road concentrations. For these pollutants, excluding commute and/or travel time in exposure assessment introduces large exposure estimate errors. However, excluding CPM or UFP in-vehicle loss rates in in-vehicle exposure assessment would also produce significant exposure estimate errors for drivers of newer cars and drivers with significant time at slow speeds. Nevertheless, there is a particularly important need to better characterize exposure to ultrafine particles, since few such epidemiological investigations have been attempted. Fruin et al. (20) calculated that 33-45 % of UFP exposure occurs while driving based on typical micro-environmental concentrations and time spent in each.

In this task, we measured AERs at three speeds for each of 59 California vehicles, chosen to represent the California fleet with regard to age, vehicle type, and manufacturer. These results more than triple the number of vehicle AERs reported in the literature and provide for the first time a sample of vehicles that is large enough to

be considered reasonably representative of the current fleet of California vehicles and/or the U.S.

This task also demonstrated that using CO_2 to calculate vehicle AER is a relatively straightforward and accurate alternative to the use of tracer gases, which require more specialized measurement instruments. The ease of this method was one reason for the large number of vehicles tested. Since vehicle AER varies more than an order of magnitude between vehicles, a large sample number is necessary to fully characterize vehicle AERs.

2.1 MATERIALS AND METHODS

2.1.1 Vehicle selection

Vehicles were selected to approximate the distribution of the California fleet in terms of vehicle size type (e.g., subcompact, compact, midsize, etc.), mileage, and age. Vehicle size data were based on the dataset of the 2002 report by the California Department of Motor Vehicles to the California Air Resources Board in support of their mobile source Emission Factors model (EMFAC) database), the latest available at the time of initial study design (21). Data on fleet mileage and age were based on 2009 data. Target numbers of test vehicles for each size category were calculated based on the frequency of these size categories multiplied by the fraction of the fleet that was five years old or newer (30%), 6 to 14 years (53%), and 15 years or older (17%) (California New Car Dealers Association (CNDCA), 2010) (22). Within these categories, an attempt was also made to select vehicles from the manufacturers having the largest sales in California (e.g., Toyota Corolla, Honda Civic, etc.) but there were no specific requirements by manufacturer.

80% of the cars tested were obtained through California Air Resources Board (CARB) vehicle testing programs. The CARB selects cars for its dynamometer emissions testing program through randomly selecting cars from California Department of Motor Vehicle records. Each car tested was selected for AER testing if it fulfilled any of the size and age categories described above. Thus, the cars tested for our AER testing were randomly selected within a size and age category. However, there is some bias in actual participation rates of the program, with fewer very new cars obtained than in the California fleet. To remedy this under-representation of very new cars, we rented additional cars of model year 2007 and newer from an hourly car rental business. Lastly, certain size categories of older cars were relatively rare, so to obtain older cars of certain size, word-of-mouth recruiting was conducted among USC graduate students. This provided five cars of average age 1998 and one new 2010 model car. The three groups of cars, new rentals, CARB-selected, and USC student-owned were analyzed both as separate groups and collectively, as a test to ensure that no particular group gave AER results that indicate significantly different AER behavior, as described in the results section.

2.1.2 Instruments.

CO_2 was measured both inside and outside the vehicle simultaneously using two or more TSI Q-Traks, Model 7565 (TSI Inc., MN, USA) and one or more LI-COR Li-820 units (LI-COR Biosciences, NE, USA). Both units use a non-dispersive infrared (NDIR) detection technique, but the LI-COR unit is pump driven, thus allowing a faster response time than the Q-Trak unit, e.g., several seconds versus 20 seconds. The LI-COR's optical bench requires 10 minutes to warm-up to specified temperature but a longer warm-up of approximately 1.5 hours is required to bring the performance of the unit to within 1 to 2% of reading. All instruments used for a given vehicle test were run simultaneously and ambient concentrations before and after a run were checked for consistency. An on-board GPS device (Garmin GPSMAP 76CSC) recorded the location and speed of the car at 1-second intervals. All instruments were synced to within 1 sec of the time recorded by GPS.

2.1.3 Air Exchange Rate Determinations.

Carbon dioxide was chosen as an AER indicator for its low toxicity, ease of measurement, and its ready availability when using car occupants as the source. At a fixed vehicle speed (and hence fixed AER), in-vehicle CO_2 concentrations change until an equilibrium concentration is reached whereby the source of CO_2 from vehicle occupants is balanced by the losses of CO_2 due to exchange of low-CO_2 outside air with high-CO_2 inside air. This difference is typically hundreds or thousands of parts per million (ppm) of CO_2, so it is easy to measure with high relative accuracy. Well-mixed conditions were created by mixing the in-vehicle air with a fan during these measurements. The well-mixed assumption was verified for each test by checking agreement with Q-Trak and Li-820 instruments located in different locations within the car.

2.1.4 Mathematical Equation and Assumptions.

AER increases with increasing vehicle speed due to pressure differences and/or turbulence around the vehicle. However, for a given vehicle speed (strictly-speaking, the vehicle air speed), the AER is nearly constant and the CO_2 concentrations inside the car will reach an equilibrium value when the CO_2 source rate is balanced by the replacement of high, in-vehicle CO_2 with lower outside CO_2 concentrations, according to the mass balance Equation 2.1:

Equation 2.1

$$\left(\frac{dC_{\text{in}}}{dT}\right)V = S + (C_{\text{amb}} - C_{\text{in}})(AER_{\text{s}})V \quad or \quad \left(\frac{dC\text{in}}{dT}\right) = S/V + (C_{\text{amb}} - C_{\text{in}})(AER_{\text{s}})$$

where, S/V is the vehicle-volume-specific source strength in ppm per hour, C_{amb} and C_{in} the outdoor and in-vehicle CO_2 concentrations (ppm), respectively, and AER_s is the speed- and vehicle-specific air exchange rate (hr^{-1}).

If we assume a small air exchange rate when the car is stationary, and we keep the interior air well mixed, the vehicle-specific source term can be determined by the initial

build-up rate of CO_2 when inside and outside CO_2 concentrations are similar, i.e., the ((C_{amb} - C_{in}) * AER) term in Equation 2.1 is much smaller than the S/V term. For example, for <10 ppm difference in inside versus outside CO_2, and an AER of 2 hr^{-1}, the ((C_{amb} - C_{in}) * AER) term is 20 ppm per hour per unit volume, compared to a typical build-up rate of 1500 ppm per hour per unit volume for two occupants, or less than one per cent.

Under these conditions, Equation 2.1 becomes Equation 2.2:

Equation 2.2

$$\left(\frac{dC_{\text{in}}}{dT}\right) \cong S/V$$

When the vehicle is in motion at a steady speed, eventually the inside concentration will reach an equilibrium value, $C_{eq,}$ when the interior source of CO_2 is balanced by the air exchange of lower-concentration CO_2 outside the vehicle with higher-concentration CO_2 inside the vehicle, and Equation 2.2 becomes

Equation 2.3

$$\left(\frac{dC_{\text{in}}}{dT}\right) = 0 = S/V + (C_{\text{amb}} - C_{\text{eq}})(AER_s)$$

which can be re-written as

Equation 2.4

$$AER_s = (S/V)/(C_{\text{amb}} - C_{\text{eq}})$$

2.1.5 Determination of Source Strength

The source rate of CO_2 was measured by the rate of initial CO_2 increase for a given set of occupants in a given car when the car was not moving and windows were up. (For a fixed source rate of CO2, the build-up rate S/V also varies by interior volume). This initial rate of CO_2 build up approximates the condition of zero CO_2 infiltration since AER is generally low when a car is stationary (i.e., less than 3 hr^{-1}) and the initial difference in CO_2 concentration between inside and outside is zero. For leakier cars, a stationary AER was calculated where possible using the CO_2 concentration when it reached equilibrium, but for most cars, this level exceeded the range of the QTrak (0-5000 ppm), and in those cases only the data from the LI-COR 820 CO_2 monitor were used. Repeated measures of CO_2 build-up were made for 10 vehicles.

2.1.6 Determination of Equilibrium Concentration

AER was determined for steady speeds of 32, 56, and 89 km hr^{-1} with windows closed, ventilation set to air recirculation, and the fan setting set to either 50% or the closest possible to a midway setting. Early in the testing, it was verified that when the ventilation is set to outside air or the windows are open, the AER is extremely high and there are no measurable differences between inside and outside CO_2. Equilibrium CO_2

concentrations were determined when the criterion was met of a maximum fluctuation of 50 ppm for at least the last 10 minutes at each speed. For conditions of closed windows and recirculating air, the fan setting can affect AER, although the contribution to total AER was found to be minor compared to that of speed. The effect of fan setting was tested for a subset of vehicles at several speeds.

2.1.7 Speed

Routes were chosen to allow nearly constant speeds. For speeds of 89 km hr^{-1}, freeways were used during conditions of free-flowing traffic. For speeds of 32 and 48-56 km hr^{-1}, runs were either made in a large cemetery or a continuous loop around the Rose Bowl in Pasadena, depending on the source location of the car being tested. Both of these routes allowed fairly short laps to prevent long duration in one direction, thus canceling any effect of wind direction and velocity on AER. Furthermore, there was minimal vehicular traffic on the roads at both locations during the times the tests were conducted. This minimized changes in outside CO_2 due to the presence of exhaust plumes from other vehicles.

2.1.8 Data Analysis

Time series plots of speed, CO_2, particle number, and fine particulate mass ($PM_{2.5}$, D_p < 2.5 µm) were aligned and adjusted to take into account any differences in instrument clock time or response time. Alignments were made based on events that caused a rapid concentration change, such as an open window rapidly reducing in-vehicle CO_2.

Where the in-vehicle CO_2 concentration met the <50 ppm change criterion for a given speed, the exact equilibrium concentration was determined at the time where CO_2 concentrations showed a less than 2% standard deviation for at least 20 data points (i.e., > 3 minutes of data). Concurrent outside CO_2 concentration was then subtracted. For the 32 and 56 km hr^{-1} speeds, the outside CO_2 concentrations at both the Rose Bowl and the cemetery were very stable, but the outside CO_2 concentrations on freeways for the 89 km hr^{-1} condition were not. Therefore, freeway CO_2 concentrations were averaged over the previous two minutes for each equilibrium value chosen.

The AER is strongly related to speed. However, even after adjusting for speed, repeated measurements of AER on the same vehicle may be correlated (leaky car would have consistently higher AERs and a tight one lower AERs), violating the assumption of independent observations in multiple linear regression (MLR). Generalized estimating equation (GEE) models (23) with an exchangeable correlation structure and robust standard errors were used to account for the correlation and to estimate the average effect of predictors across the population of vehicles. MLR models were also fit to compare results across modeling techniques.

The complete test results for the 59 vehicles tested, generally three AERs per vehicle (i.e., at three different speeds), were modeled to test the predictive power of vehicle characteristics such as vehicle mileage, age, and manufacturer. Squared and cubed terms for mileage, age, and speed were included to account for any non-linear effects. Vehicle characteristics such as interior vehicle volume and frontal area, and fan setting

were also included. Manufacturer variables included specific vehicle manufacturer categories such as Ford, GM, Toyota, Nissan, Honda, and 'other' as well as broader categories such as U.S. and non-U.S. or U.S., Japan, and 'other.' Vehicles were also grouped by the source of the vehicle (i.e., CARB, rental agency, or student volunteers) and tested for differences. Speed was included, both as a predictive variable as well as a stratifying variable, i.e., data were analyzed separately for a given speed. Since AER results had a strong rightward skew, a natural log transformation was used.

Parsimonious GEE and MLR models were obtained by backwards step-wise selection in which variables were retained if they improved R^2 (MLR) or were statistically significant (GEE) at p=0.05 value. Residuals from both models were inspected to check model assumptions. R^2 was calculated for the GEE model by taking the square of the Pearson correlation coefficient between observed and model-predicted values of natural log transformed AER.

2.2. RESULTS AND DISCUSSION

2.2.1 Vehicles Tested

Achieving a representative sample of vehicles for testing was a primary objective of this task since representativeness enhances the utility of any predictive models of AER. We selected 59 vehicles to represent the California fleet in terms of vehicle age and size type based on EPA classes. Vehicles less than 5 years are slightly under represented in the project fleet.

2.2.2 Equilibrium Values and AERs Calculated

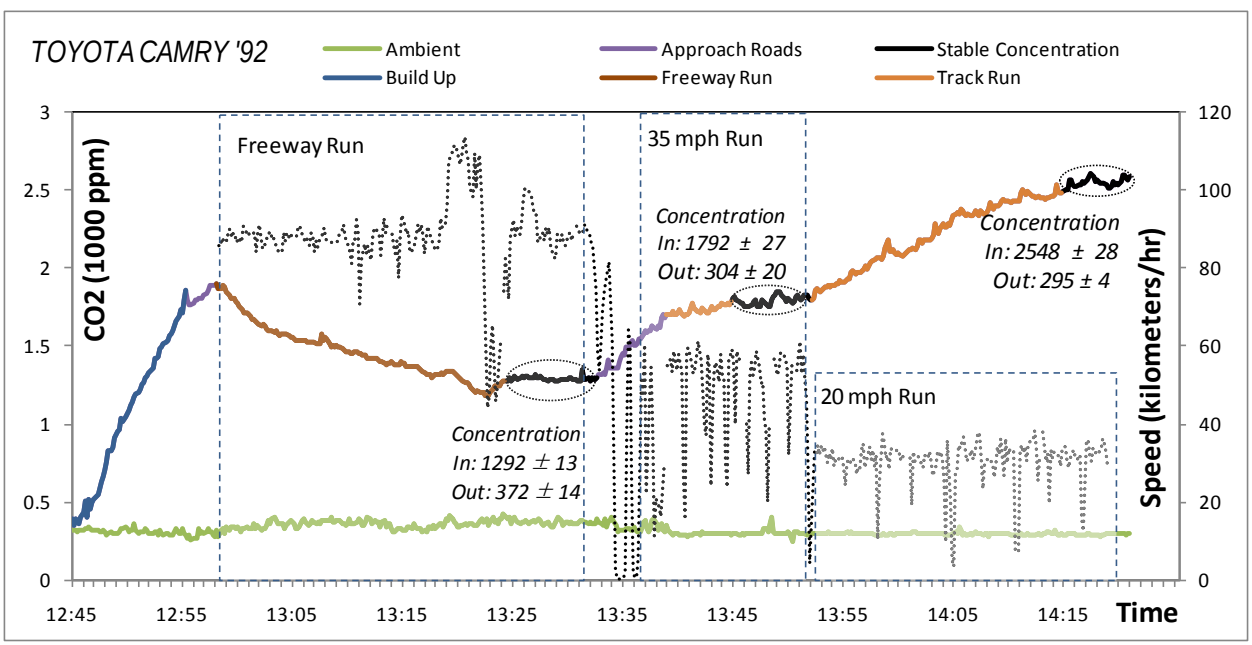


Figure 2.1: Typical Time-series plot for runs conducted at Cemetery along with the initial build up and freeway run. Average speed during Freeway run was 89 ± 10 km hr^{-1} for stable portion highlighted in black). The second black highlight corresponds to stable values during 51.1 ± 9.4 km hr^{-1} and 31.3 ± 5.5 km hr^{-1} speed runs.

A typical time-series plot of in-vehicle and outside CO_2 concentration and speed is shown in Figure 2.1. As shown in this plot, the CO_2 build-up rate at the beginning of the test is quite linear and the various in-vehicle CO_2 concentrations at different speeds show an exponential change that eventually reaches a steady equilibrium concentration despite the small differences in speed. In Figure 2.1, the % standard deviations of the in-vehicle CO_2 concentration were 1.0, 1.5 and 1.1% at 32, 56, and 89 km hr^{-1}, respectively, while the outside CO_2 concentration standard deviations were 4, 7 and 1.4%, respectively. The resulting AER at 89 km hr^{-1} was 13.6 hr^{-1}. If the in-vehicle CO_2 concentration deviated by the 1.1% (13 ppm) standard deviation observed, for example, the AER values ranged from 13.4 and 13.8, or +/- 1.5%. The change in AER values for 48-56 and 32 km hr^{-1} changed +/- 2% and 1%, respectively, if in-vehicle concentrations deviated by the observed standard deviations.

Figure 2.2 shows the results for all cars tested at each speed. The large vehicle-to-vehicle differences are readily apparent, as is the strong dependence of AER on speed for a given vehicle.

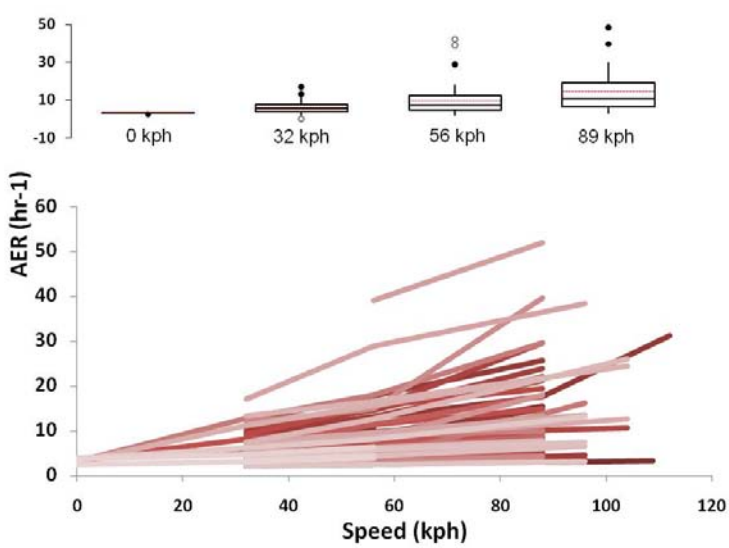


Figure 2.2: AER results for all 59 vehicles tested.

2.2.3 GEE Model Results.

The Generalized Estimating Equation (GEE) model gives the following predictive equation for AER as a function of easily-obtainable parameters related to each vehicle:
Equation 1

$$Ln\,(AER) = 0.63 - (age * 0.066) + (age^2 * 0.0058) + (kilometers * 0.016) - (kilometers^2 * 7.8 * 10^{-5}) + (speed * 0.029) + Manuf\ Adjustment$$

Where 'age' is in years, 'kilometers' is vehicle lifetime mileage in thousands of kilometers, and 'speed' is in kilometers per hour. The manufacturer's adjustment ('Manuf Adjustment', calculated as the regression coefficient using manufacturers as variables) is given in the last four rows of Table 1, with Japanese manufacturers being

the base case (i.e., no adjustment needed). Fan setting, although observed to slightly increase AER, was not significant, nor were vehicle size characteristics such as frontal area. The GEE model R^2 was 0.70. AER is a non-linear function of speed and mileage/age and Figure 2.3, shows how the model-predicted AER strongly increases with speed for the median age and mileage in the study test fleet. Figure 2.3 shows how the model-predicted AER strongly increases with speed for the median age and mileage in the study test fleet, (8 years old and with 138,000 kilometers (about 86,000 miles), respectively). Figure 2.3 also shows how the model predicts AER to increase with each additional year of age assuming 23,000 kilometers per year (about 14,000 miles), the study average mileage change per year.

Table 2.1: AER model coefficients, 95% confidence intervals, and P values.

Source	Value	Standard Error	z	Pr > |t|	95% Confidence Interval	
Intercept	0.63	0.124	5.1	0.000	0.390	0.876
Age (years)	-0.066	0.043	-1.6	0.12	-0.15	0.018
Age^2	0.0058	0.0020	3.0	0.003	0.0020	0.0096
Kilometers (thousands)	0.016	0.0076	2.2	0.025	0.0021	0.032
Kilometers^2	-0.000078	0.000044	-1.7	0.082	-0.000167	-0.000010
Speed (km hr^{-1})	0.029	0.00152	19	0.000	0.026	0.032
Manuf-Japan	0.000	0.000				
Manuf-GM	0.55	0.15	3.7	0.000	0.26	0.85
Manuf-Ford	0.25	0.12	2.0	0.042	0.0085	0.48
Manuf-other	-0.051	0.20	-0.25	0.80	-0.45	0.34

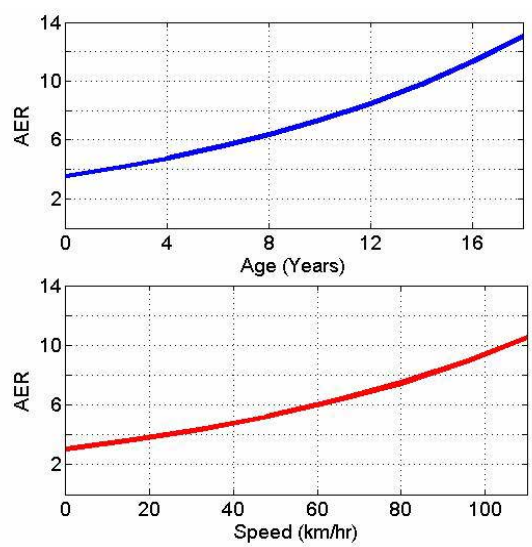


Figure 2.3: Model-predicted AER increase with age and speed for median age study vehicle.

Figure 2.4 below shows the model predictions versus actual measurements, and the normality of the residuals. In addition, the test results, when grouped by source of vehicle (ARB, rental or volunteer), did not show any difference in the pattern of residuals.

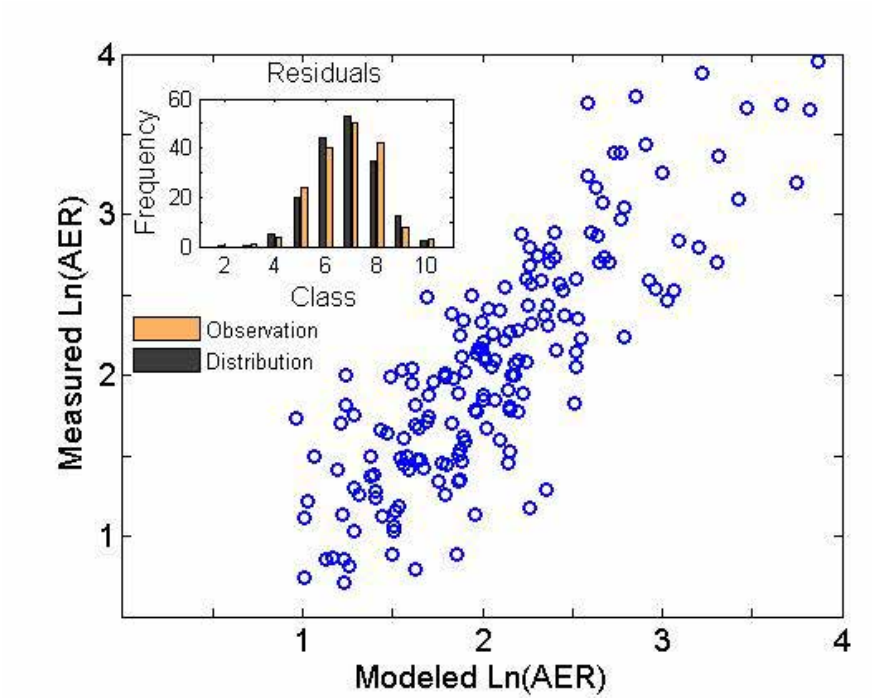


Figure 2.4: Model predictions versus actual measurements, and the normality of the residuals. Each data point represents a measured AER used to populate the predictive model.

As a test of our experimentally-derived and modeled results against other studies, our Equation 5 was used to predict the AER of the vehicles tested in the study by Knibbs et al. (1), the largest AER study conducted before the present study. Our study model slightly under-predicted the AERs measure by Knibbs et al. (1), at low AERs and slightly over-predicted the measured AERs at high AERs, but overall agreement was good, considering that Knibbs et al. (1) conducted their study in Australia with a sample of vehicles selected to span a range of ages rather than be representative of the Australian fleet. The results are shown in Figure 2.5. The R^2 value of 0.83 indicates that our model performed consistently across the vehicles tested by Knibbs et al. (1).

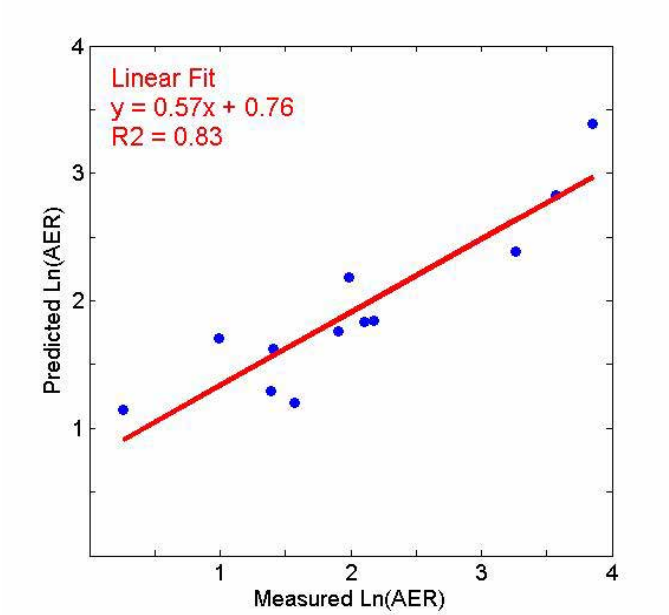


Figure 2.5:. Comparison of model predictions and results from Knibbs et al., 2009.

2.3 SUMMARY AND CONCLUSIONS

For a typical car measured in this study under closed-window and recirculation air conditions, the model-predicted AER is in the range where significant particle number losses would be expected to occur, although this depends somewhat on the size distribution of the aerosol, with fresher aerosols expected to show higher deposition rates. For example, Figure 2.3 shows that for the median age study vehicle, the speed range from 32 to 97 km hr^{-1} would cause AERs to range from about 4 to 9 hr^{-1}, respectively. Measurements of particle number attenuation in these vehicles (see Chapter 3) show that this air exchange rate typically produces 75 to 85% particle number reductions at steady state. Furthermore, this study measured AERs ranging from only 2 or 3 hr^{-1} at zero or low speeds for newer, tight cars to above 50 hr^{-1} for older cars at higher speeds. This range of AER would produce particle number reductions that vary from nearly 0 to nearly 100%. Therefore, AER is a key factor in determining particle number exposure inside vehicles and should be factored into any exposure assessment that predicts in-vehicle particulate matter (PM) exposure.

In summary, the in-vehicle microenvironment is an important route of exposure to traffic-related pollutants, particularly ultrafine particles (UFP, $D_p < 0.1\mu m$). However, significant particle losses can occur in vehicles under conditions of low air exchange rate (AER) when windows are closed and air is recirculating, such as during air conditioning. Despite the importance of AER in affecting in-vehicle exposures, few studies have characterized AER, and of those, all have tested a small number of cars. One reason for this is the difficulty in measuring AER with tracer gases such as SF_6, the most common method. We demonstrated that using vehicle occupants as a source of CO_2 allows an accurate yet simple measure of AER. AER was calculated for three speeds each for 59 vehicles representative of California's fleet, the first time a large and

representative sample of vehicles have been tested for AER. This sample was sufficient to allow the development of robust predictive models that explained 70% of the variation in the observed AERs, from <2 hr^{-1} to >50 hr^{-1}. AER appeared to be primarily driven by speed, along with vehicle age and mileage, and to a lesser extent by manufacturer. These results will therefore be useful in future exposure or epidemiological studies that include commuting and other in-vehicle exposures to ultrafine PM and other air pollutants, since the predictive variables are readily obtainable through questionnaires.

Accurate assessment of risk posed by ultrafine exposure will depend on the ability to characterize exposure during microenvironments like in-vehicle where peak and disproportionate exposure occurs frequently for large sections of the population. To this end, the insights gained on AER in a vehicle will be used to characterize the in-vehicle environment in the following Chapters 2 and 5 (based on Task 2). Since AER is responsible for the influx of pollutants into the vehicles, the in-vehicle concentrations will be highly dependent on AER.

Furthermore, in enclosed environments like the in-vehicle microenvironment, exposure concentration does not always equal the ambient concentration. Pollutant concentrations, especially for particulate species like ultrafine particles (UFP, aerodynamic diameter less than 100 nm) are frequently lower inside vehicles (at multiple ventilation settings) than on roadways because pollutants are often lost during and post infiltration (1). However, they may still exceed ambient concentrations by manifolds (6). Prior to this work, some studies (11,17) have reported in-vehicle UFP concentrations but results cannot be generalized. No study was identified that can systemically quantify the role of roadway concentration variation and vehicle characteristics on in-vehicle exposure concentration.

It is important to identify the most influential parameters and their relative impact in order to design and accurately interpret epidemiological studies. For example frequently at recirculation ventilation mode, inside concentrations is half of (or even lower than) that observed at outside air intake ventilation mode. A two-fold difference in roadway concentration between peak and off-peak traffic hours may be of no consequence for exposure assessment if vehicle is driven in recirculation mode when roadway concentrations are higher. With information of ventilation mode lacking, it will be impossible to draw correct conclusions on in-vehicle concentrations and associate them with any health end points in an epidemiological study.

The next Chapter will examine the influence of various vehicle and driving related factors that influence in-vehicle concentration.

REFERENCES

1. Knibbs, L. D.; de Dear, R. J. and Atkinson, S. E. Field study of air change and flow rate in six automobiles. *Indoor Air* **2009**, 303–313.
2. Ott, W.; Langan, L. and Switzer, P. A time series model for cigarette smoking activity patterns: model validation for carbon monoxide and respirable particles in a chamber and an automobile. *J. Expo. Anal. Environ. Epidemiol.* **1992**, 2 (Suppl. 2), 175–200.
3. Ott, W.; Switzer, P. and Willits, N. Carbon monoxide exposures inside an automobile traveling on an urban arterial highway. *J. Air Waste Manag. Assoc.* **1994**, 44, 1010–1018.
4. Ott, W.; Klepeis, N. and Switzer, P. Air change rates of motor vehicles and in-vehicle pollutant concentrations from secondhand smoke. *J. Expo. Sci. Environ. Epidemiol.* **2008**, 18, 312–325.
5. Park, J.-H.; Spengler, J.D.; Yoon, D.-W.; Dumyahn, T.; Lee, K. and Ozkaynak, H. Measurement of air exchange rate of stationary vehicles and estimation of in-vehicle exposure*. J. Expo. Anal. Environ. Epidemiol.* **1998,** 8, 65–78.
6. Rodes, C.; Sheldon, L.; Whitaker, D.; Clayton, A.; Fitzgerald, K.; Flanagan, J.; DiGenova, F.; Hering, S. and Frazier, C. *Measuring concentrations of selected air pollutants inside California vehicles*. Report prepared for California EPA, **1998**
7. Klepeis, N. E.; Nelson, W. C.; Ott , W. R.; Robinson, J. P.; Tsang, A. M.; Switzer, P.; et al. The national human activity pattern survey (NHAPS): A resource for assessing exposure to environmental pollutants. *J Expos Anal Environ Epidemiol*. **2001**, 11(3), 231-252.
8. Jenkins, P.L.; Phillips, T.J.; Mulberg, E.J.; and Hui, S.P. Activity patterns of Californians: use of and proximity to indoor pollutant sources. *Atmos. Environ.* **1992**, 26A, 2141-2148.
9. Westerdahl, D.; Fruin, S.; Sax, T.; Fine, P.M. and Sioutas, C. Mobile Platform Measurements of Ultrafine Particles and Associated Pollutant Concentrations on Freeways and Residential streets in Los Angeles. *Atmos. Environ.* **2005,** 39, 3597-3610.
10. Chan, C.-C.; Özkaynak, H.; Spengler, J.D. and Sheldon, L. Driver exposure to volatile organic compounds, CO, ozone and NO2 under different driving conditions. *Environ. Sci. Technol.* **1991**, 25, 964–972.
11. Zhu, Y.; Eiguren-Fernandez, A.; Hinds, W. C. and Miguel, A. In cabin commuter exposure to ultrafine particles on Los Angeles freeways. *Environ. Sci. Technol.* **2007**, 41, 2138–2145.
12. Xu, B. and Zhu, Y. Quantitative analysis of the parameters affecting in-cabin to on-roadway (I/O) ultrafine particle concentration ratios. *Aerosol Sci. Technol*. **2009**, 43, 400–410.

13. Gong, L.; Xu, B. and Zhu, Y. Ultrafine particles deposition inside passenger vehicles. *Aerosol Sci. Technol.* **2009**, 43, 544–553.
14. Batterman, S.; Jia, C., Hatzivasilis, G. and Godwin, C. Simultaneous measurement of ventilation using tracer gas techniques and VOC concentrations in homes, garages and vehicles. *J. Environ. Monit.* **2006,** 8, 249–256.
15. Kvisgaard, B. and Pejtersen, P. Measurement of flow in automobile ventilation systems, Innova AirTech Instruments technical document. **1999** http://www.lumasense.dk/Articles.139.0.html
16. Knibbs, L.D.; de Dear, R.J. and Morawska, L. Effect of cabin ventilation rate on ultrafine particle exposure inside automobiles. *Environ. Sci. Technol.* **2010**, 44, 3546–3551.
17. Pui, D. Y. H.; Qi, C.; Stanley, N.; Oberdorster, G. and Maynard, A. Recirculating air filtration significantly reduces exposure to airborne nanoparticles. *Environ. Health Perspect.* **2008**, 116, 863–866.
18. Qi, C.; Stanley, N.; Pui, D. Y. H. and Kuehn, T. H. Laboratory and on-road evaluations of cabin air filters using number and surface area concentration monitors. *Environ. Sci. Technol.* **2008**, 42, 4128-4132.
19. Fletcher, B. and Saunders, C.J. Air change rates in stationary and moving motor vehicles, *J. Hazard. Mater.* **1994**, 38, 243–256.
20. Fruin, S.; Westerdahl, D.; Sax, T.; Sioutas, C. and Fine, P.M. Measurements and predictors of on-road ultrafine particle concentrations and associated pollutants in Los Angeles, *Atmos. Environ.* **2008**, 42, 207–219.
21. *Emissions Factors Model*; California Air Resources Board, Sacramento, CA, 2002; http://www.arb.ca.gov/msei/onroad/latest_version.htm.
22. *California Auto Outlook Market Report*; California New Car Dealers Association, Sacramento, 2006-2010; http://www.cncda.org/publications/publications.html.
23. Liang K. and Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, 73, 13-22.

3. CHAPTER THREE: FACTORS THAT DETERMINE ULTRAFINE PARTICLE EXPOSURE IN VEHICLES

(based on Task 2. Examine the impact of important influential factors that contribute to in-cabin pollutant concentrations.)

Note: Chapter Two discussed how air exchange rate (AER) varies from vehicle to vehicle and by speed, and how AER is the dominant factor in affecting how a given on-road pollutant concentration translates to different in-vehicle concentrations. Chapter Three goes on to explore other factors that can also affect the relationship between on-road concentrations and in-vehicle concentrations, using ultrafine particle (UFP) number concentrations as the primary example. UFP concentrations have the widest range of in-vehicle losses during typical driving conditions. Other particulate pollutants are of larger size and have higher loss rates, but these rates are also a strong function of AER.

3.0 INTRODUCTION

The proximity of vehicles to relatively undiluted emissions from other vehicles on freeways and busy roadways leads to significantly elevated pollutant concentrations in vehicle cabins compared to other indoor environments. Thus, a disproportionate share of total personal exposure can occur while driving, especially for pollutants emitted mostly by vehicles, like ultrafine particles (particles smaller than 100 nm, UFP). Fruin et al. (2007) calculated that 33-45% of UFP exposure occurs while driving in Los Angeles, taking other micro-environmental concentrations and time-activity patterns into account, but ignoring in-vehicle particle losses. In suburban locations of less traffic, Wallace and Ott (2011) estimated a 17% contribution of in-vehicle microenvironment to total UFP exposure.

Despite its importance as a route of exposure, the contribution of the in-transit vehicular microenvironment remains largely uncharacterized, in part due to the difficulty of characterizing the large differences in air exchange rate (AER), which drives particle influx rates and varies not only from vehicle to vehicle but also across different driving conditions. To better characterize AERs, Fruin et al. (2011) tested 59 vehicles and reported that AER under recirculation ventilation conditions can be reliably predicted based on vehicle's age, mileage, driving speed and manufacturer ($r^2 = 0.7$). At outside air intake conditions, for the eight vehicles tested, they found AERs to be an order of magnitude higher than at recirculation settings. They also found that at the outside air ventilation setting, AER was driven by fan speed and not vehicle speed. In a similar but smaller study, Knibbs et al. (2009) reported AER values at higher fan settings during outside air conditions to be 73% higher than at lowest fan setting for six vehicles.

Particle influx and removal rates result from a complicated interaction among multiple factors, including AER, physical characteristics of the vehicle; particle size and in-cabin filter efficiency. Any accurate determination of the relative influence of each factor requires experiments where factors are systemically varied to observe effects under real driving (aerodynamic) conditions; else AERs are not realistic. Several recent studies (Gong et al., 2009; Xu and Zhu, 2009; Xu et al., 2011) addressing in-vehicle particle losses have not used realistic driving, in favor of artificial air movement, or have relied on measurements that are difficult to make outside a laboratory, such as filter efficiency tests. Furthermore, these studies have suffered from small sample sizes, ranging from only one to three vehicles (Gong et al., 2009; Xu and Zhu, 2009).

Of the few studies that have used real driving conditions, Pui et al. (2008) and Qi et al. (2008) demonstrated a dramatic particle number concentration reduction in-cabin during recirculation ventilation settings in two vehicles, although AERs were not reported. Zhu et al. (2007) observed particle losses of about 85% at recirculation settings in three vehicles but AER was not measured and variable speeds during the tests would have resulted in variable AERs. Zhu et al. (2007) was also the only study identified that made size-resolved particle concentration measurements but only up to 217 nm. Furthermore, they reported the fractional losses to be primarily dependent on particle size and vehicle characteristics.

The most useful study from an exposure assessment perspective has been by Knibbs et al. (2010) who measured the inside-to-outside UFP concentration ratios in five vehicles and reported a high correlation between these ratios and AER (r^2 = 0.81). They reported an average particle reduction of 0.69 during recirculation settings at low fan setting and 0.08 at outside air intake, but did not associate the losses with particle size or influence of specific removal mechanisms like cabin filtration.

The goal of this study was to quantify the effects on particle reduction rates due to changes in 1) ventilation setting; 2) measured AER; 3) fan setting; 4) filter condition; 5) driving speed; and 6) easily-obtainable vehicle characteristics such as age and mileage (which affect AER); and to determine the relative importance of each of these variables. It is the first study to combine measurements of AER and particle losses as a function of particle size.

3.1 METHODS

3.1.1 Vehicle Selection and Conditions Tested

Six vehicles were selected such that AERs at recirculation ventilation settings spanned the inter-quartile range of AERs (4.5 - 13 h^{-1}, median 7.4 h^{-1}) measured in Fruin et al. (2011) and Hudda et al. (2012). Two vehicles (a 1999 Ford Contour and 2001 Ford Escort) were more than 10 years old when tested, while four newer

vehicles (a 2010 Toyota Prius, 2010 Scion Xb, 2009 Toyota Matrix and 2009 Honda Civic) were 3-14 months old. All six vehicles were evaluated at both recirculation (RC) and outside air intake (OA) ventilation conditions. At each ventilation condition, experiments were conducted in both stationary and mobile mode at both medium and high fan speed settings.

At RC setting, a total of 42 conditions were evaluated. Experiments measuring particle losses and AER were conducted on six vehicles at seven AERs each (resulting from a combination of ventilation fan setting and driving speed). At OA settings, a total of 34 combinations of vehicle/ speed/ fan strength were evaluated. However, because it has been previously demonstrated that fan setting and not vehicle speed determines AER at OA ventilation settings (Fruin et al., 2011), simultaneous particle loss and AER measurements were conducted for a subset of conditions (10 out of 34), i.e., in stationary vehicles at various fan settings. Particle loss measurements in moving vehicles were assumed to have the similar AER for a given fan speed. Air conditioning was kept on during all experiments, except for those at which the fan was off (6 out of total 76 conditions evaluated). Relevant vehicle characteristics are summarized in the following Table 3.1.

Table 3.1: List of vehicles tested.

Vehicle Model		**Mileage**	**In-Cabin Filter**	**Vehicle Age (Years)**
Ford Contour	1999	115990	Yes	10
Ford Escort	2001	127280	Yes	10
Toyota Prius	2010	3210	Yes	<1 (3 months)
Toyota Matrix	2009	26125	Yes	>1 (16 months
Toyota Scion Xb	2010	24068	Yes	<1 (11 months)
Honda Civic	2009	22000	Yes	>1 (14 months)

3.1.2 Particle Concentration Measurements

Particle number concentration measurements were made using a condensation particle counter (CPC, TSI Inc. Model 3007, size range 10 nm to 1000 nm) and number concentration measurements by size using a Scanning Mobility Particle Sizer (SMPS,TSI Inc. Differential Mobility Analyzer Model 3080 and Model 3022a CPC). The measured mobility diameter (D_p) range was 14-750 nm. However, data are presented only for the 14-400 nm range because above 400 nm, relatively few particles were counted and the resulting concentrations had large uncertainties.

Furthermore, since roadway particles between 14-25 nm are volatile, concentrations in this size range are exceptionally variable in on-road environments. Reported results for this size range should be interpreted with caution.

Experiments were conducted at 0, 20 and 35 miles h^{-1}, with speed recorded each second by a Garmin GPSMAP unit 76CSC. Experiments at 20 and 35 miles h^{-1} were conducted while driving at constant speed around the Rose Bowl Stadium in Pasadena CA, a three mile long loop with little vehicular traffic. Freeway speeds were not evaluated in this study due to the rapidly changing traffic and particle number and size distributions typically present on Los Angeles freeways.

All in-vehicle measurements were made with windows fully closed. Outside vehicle concentrations were assumed to be equal to roadway concentrations, and were measured for 10-20 minutes before and after the in-vehicle measurement period, with windows fully open to allow the outside air to pass freely through the vehicle. Earlier in the study, it was verified that open window conditions allow accurate measurement of roadway particle size distributions and number concentrations by comparing simultaneous measurements with two CPCs, one measuring concentrations with a 1 m long inlet sampling air right outside the vehicle, and the other CPC, sampling in the middle of the backseat of the vehicle. Ambient particle number concentrations and size distributions were also measured using a stationary monitor at a position central to the run loop. All ambient concentrations were stable to within 10%, before, during, and after a run.

Inside-to-outside particle concentration ratio (I/O) was calculated for all measurements, to provide a measure that reflects particle removal indoors. I/O ratio was calculated after concentrations were stable in the vehicle over 15 minutes or more of sampling. I/O ratios reported for specific size ranges are equivalent to the average of mobility-diameter-specific I/O ratios within the range.

3.1.3 Air Exchange Rate Measurements

During both mobile and stationary conditions, air exchange occurs between the vehicle cabin and the outside environment through leaks in the body of the vehicle (door seals, window cracks, etc.) and through the ventilation system, when it is set to draw outside air into the cabin. This air exchange continually replenishes the vehicle cabin with pollutants/particles from the outside environment, hence is an essential measurement in any study of in-vehicle exposure.

AERs were determined at RC conditions using CO_2 as a tracer gas and two occupants as a stable source of CO_2 generation (described in detail in Section 2.1.3-2.1.6, Chapter 2). Build-up rates of CO_2 concentration were used to determine the CO_2 rate of emission, and equilibrium CO_2 concentrations at fixed speeds then allowed calculating the AER at that speed. Under OA conditions, AERs are much higher, requiring higher starting CO_2 concentrations that cannot be readily reached

by occupants. As it was observed that speed plays an insignificant role in affecting AER compared to fan setting at OA conditions, only stationary tests were necessary to characterize AER, and CO_2 from a pressurized cylinder was used to produce the necessary high starting concentrations (described in detail in Section 2.2.7, Chapter 2). Table 3.2 and 3.3 list the AERs for the vehicles at various conditions tested.

Table 3.2: AER rates (h^{-1}) at recirculation setting for the vehicles tested.

Speed (miles h^{-1})	Fan	Ford Contour	Ford Escort	Toyota Prius	Toyota Scion xB	Toyota Matrix	Honda Civic
0	No	2.1	2.20	0.53	0.27	1.70	1.15
0	Medium	3.1	3.50	0.83	0.53	2.40	3.00
0	Full	6.2	5.40	1.50	1.40	3.20	4.20
20	Medium	11.2	8.10	3.00	3.50	4.10	5.50
20	Full	14.1	9.50	3.70	4.80	4.70	6.30
35	Medium	16.0	11.5	3.70	4.50	5.10	7.30
35	Full	19.0	13.5	4.30	5.70	6.30	8.80

Table 3.3: AER rates (h-1) at outside air intake setting for the vehicles tested. 1

Ford Contour		Ford Escort		Toyota Scion xB	
Fan	**AER**	**Fan**	**AER**	**Fan**	**AER**
1/4	36.0	1/3	51.0	1/4	20.0
2/4*	65.0	2/3*	64.0	2/4*	35.0
3/4	94.0	3/3§	83.0	3/4	50.0
4/4§	117.0			4/4§	75.0

[1] *Medium fan setting, §Full fan setting

Table 3.4 continued

Honda Civic		Toyota Prius		Toyota Matrix	
Fan	**AER**	**Fan**	**AER**	**Fan**	**AER**
2/12	57.0	1/7	23.0	1/4	20.0
4/12	72.0	2/7	36.0	2/4*	35.0
6/12*	93.0	3/7	48.0	3/4	-
8/12	112.0	4/7*	59.0	4/4[§]	71.0
10/12	125.0	5/7	69.0		
12/12[§]	141.0	6/7	84.0		
		7/7[§]	97.0		

3.2 RESULTS AND DISCUSSION

3.2.1 Effect of Air Exchange Rate on I/O Ratios

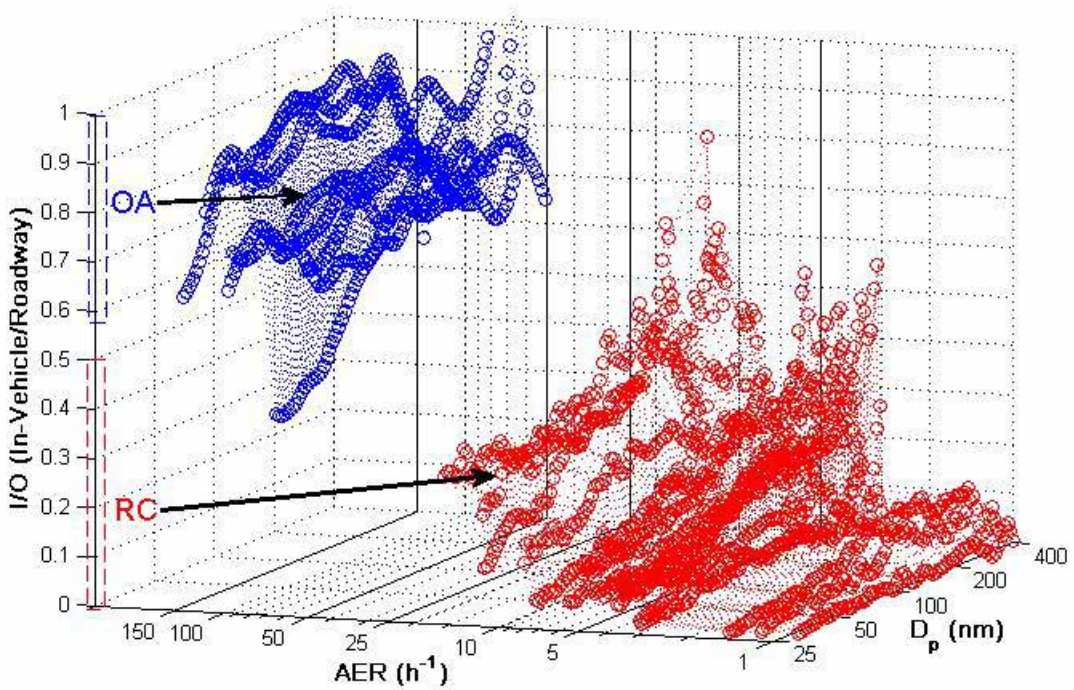


Figure 3.1: I/O ratio dependence on AER for 25-400 nm particles under re-circulation (RC) and outside air (OA) ventilation setting.

At RC ventilation conditions, the 42 AERs across the six vehicles tested varied from less than 1 to 19 h^{-1}. At OA setting, measured AERs varied from 20 to 145 h^{-1}. Any significant increase in AER resulted in an increase in particle influx rate and increased I/O ratios. Figure 3.1 presents I/O ratio results under both ventilation

conditions, and illustrates the strong dependence of I/O ratio on AER, which is exhibited most distinctively in the difference between the I/O ratios at RC and OA ventilation settings. The decrease in I/O ratio with increasing AER was less dramatic at OA conditions compared to RC conditions. Across the size range (25-400 nm) an increase in AER elevated I/O ratio, but the effect was strongest for particles above 200 nm.

The only other study to report both I/O ratio and AER, Knibbs et al. (2010), reported a high correlation between I/O ratio for UFP and AERs (r^2 = 0.81). In the present study, at RC the r^2 between I/O ratio (using a CPC 3007, the same instrument used by Knibbs et al., 2010) and AER was 0.80, indicating that AER is the most significant determinant of I/O ratio at RC ventilation conditions. The average I/O ratio under RC conditions was 0.17 ± 0.13. In contrast, at OA conditions, the I/O ratios averaged 0.67 ± 0.10. The average r^2 between I/O ratios and AER under OA conditions was 0.75 (r^2 values for all but one vehicle were 0.9 or above). On average for the six vehicles tested, a switch in ventilation condition from RC to OA increased I/O ratio by nearly a factor of four.

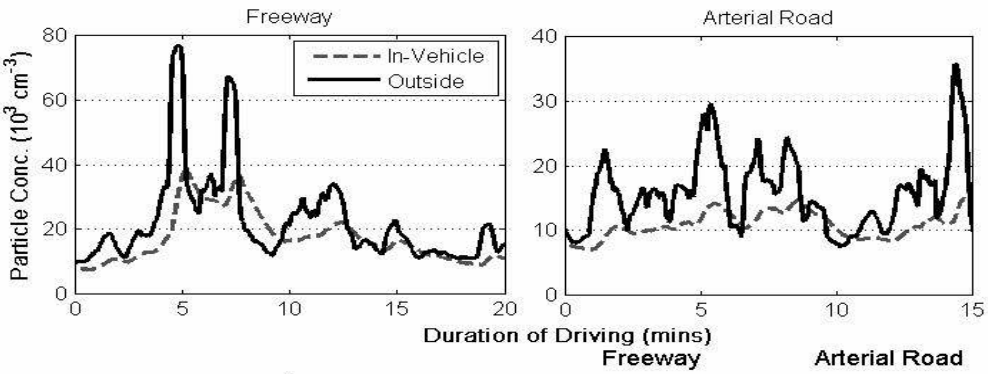
3.2.2 Effect of Vehicle Speed and Age on AER and I/O Ratios

Under RC conditions, an increase in speed increased AER and I/O ratios. On average, a 10 miles h^{-1} increase in speed resulted in 1.65 h^{-1} increase in AER and a 0.035 increase in I/O ratio. Speed affected AER more for older vehicles (+2.4 h^{-1}/10 miles h^{-1}) compared to the newer vehicles (+1.2 h^{-1}/10 miles h^{-1}), similar to the results reported by Knibbs et al. (2009). As a result, I/O ratio, which depends strongly on AER, increased with speed at twice the rate for older vehicles (-0.05 /10 miles h^{-1}, Pearson r^2 = 0.78) than for newer vehicles (-0.025 /10 miles h^{-1}, Pearson r^2 = 0.20).

Despite these differences by vehicle age, an overall strong correlation was observed between AER and speed as well as I/O ratios and speed, as well as AER. In Fruin et al. (2011) an r^2 equivalent of 0.92 was calculated between AER and speed for a much larger fleet of vehicles and for speeds up to 70 miles h^{-1} using a Generalized Estimating Equations (GEE) (Liang et al., 1986). (GEE techniques account for correlated measurements within a vehicle; e.g, a tight vehicle with lower AER will consistently have lower I/O ratios across each speed compared to a leakier vehicle with higher AER). Among the six vehicle data, using GEE technique, 82% of the variation in AER could be accounted for by speed (p value = 2.5×10^{-9}). Furthermore, nearly all variation (r^2 = 0.98, p-value = 6.9×10^{-7}) in I/O ratios at RC setting was explained by changes in AER. Given the consistent and strong correlations between AER and speed (Knibbs et al., 2009, Fruin et al, 2011; this study) and between I/O ratios and AER at RC setting (Knibbs et al., 2009, this study), these results can be expected to extrapolate well to the higher AERs of

vehicles travelling at higher freeway speeds (65-70 miles h^{-1}) from the speeds tested in this study, i.e., 35 miles h^{-1} and less.

In contrast, under OA conditions, no definitive association between speed and I/O ratios could be discerned for the six vehicles tested. Since AER at OA is driven by mechanical ventilation rather than speed-driven pressure differences outside the vehicle shell, this lack of association is not surprising. Knibbs et al. (2009) have previously shown that at OA conditions, for the newer four of total six vehicles tested, the association between AER and speed was weak (linear regression on all 4 vehicles: AER = 0.15 * Speed [miles h^{-1}] + 51, r^2 = 0.06). For the oldest two (10 and 19 year old) of the total six vehicles tested, the relationship was stronger (AER = 0.59 * Speed [miles h^{-1}] + 52, r^2 = 0.52).This suggests that for older vehicles, the AER at OA may increase at speeds higher than those tested in this study, which should elevate I/O ratios. However, for the two oldest vehicles in this study (10 years of age), any increase in AER with speeds between 1 to 35 miles h^{-1} did not cause a noticeable change in I/O ratios.



	Freeway	Arterial Road
Average Speed (mi h^{-1})	55	27
Average Outside Conc. (cm^{-3})	21000	16000
Average In-vehicle Conc. (cm^{-3})	16000	11000
Measured I/O	0.76±0.12	0.70±0.11
Model Predicted I/O	0.78	0.68

Figure 3.2: Agreement between measured I/O during variable speed driving and regression-predicted I/O.

Most real-world driving involves constantly changing speeds due to traffic conditions and widely varying roadway particle number concentrations. To illustrate how well the steady speed/AER and relatively stable roadway concentration condition results apply to such changing speed and concentration conditions, measurements were made during two 15-minute duration trips on a freeway (I-110) and an arterial road (Figueroa Street, downtown Los Angeles) in a Toyota Matrix 2009. No attempt was made to maintain steady speeds during these runs, and average speed was 55 miles h^{-1} and 27 miles h^{-1} on the freeway and arterial road, respectively. The ventilation setting was set to OA, leading to moderately high AER (35 h^{-1}) that would allow rapid influx of roadway concentration into the vehicle cabin and reflect the unsteady

roadway environment, as shown in Figure 3.2. It should be noted that under low AER conditions, in cabin concentrations are fairly steady due to limited influx of particles. The average AF calculated from average in-cabin and roadway particle number concentrations agreed almost perfectly with the I/O ratios predicted from the I/O ratio versus speed regression equation based on measurements at steady speeds.

3.2.3 Effect of Particle Size on I/O Ratios

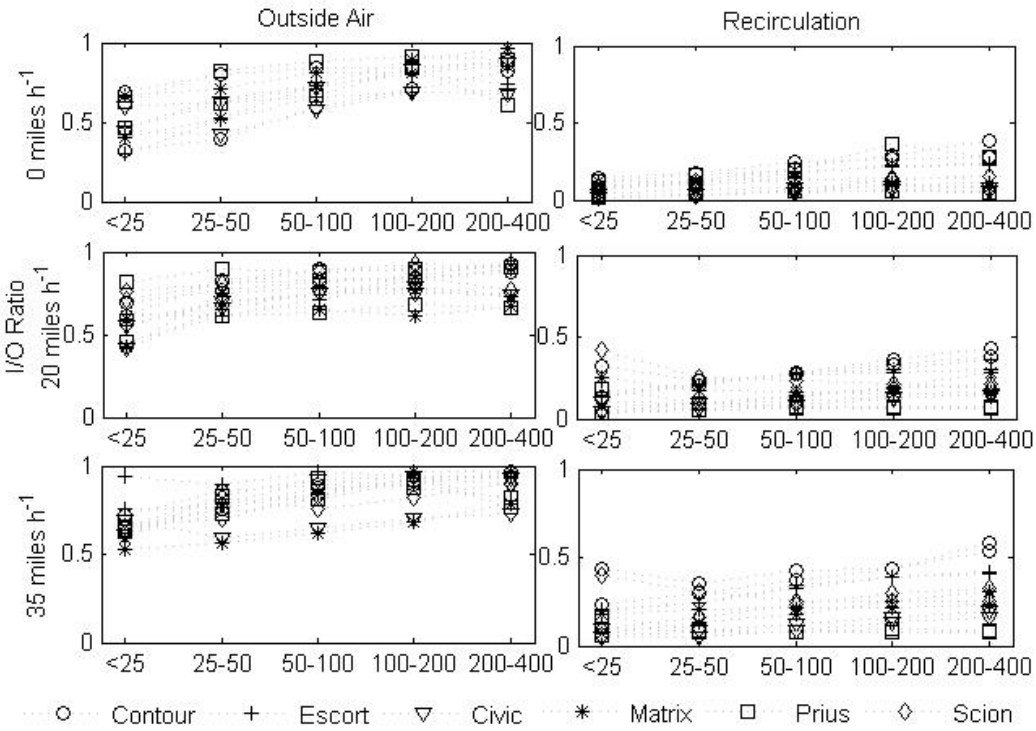


Figure 3.3: Size range specific I/O ratios at three speeds and two ventilation conditions tested. The dashed lines join values from the same vehicle.

For a given ventilation and speed combination, particle size-specific I/O ratios were found to be similar across size for newer vehicles and only moderately different across size for older vehicles. At RC, the I/O ratios for 100-200 and 200-400 nm were respectively 0.04 and 0.07 higher than the average I/O ratio (0.16 ± 0.09) for ultrafine range. At OA, the differences in size specific I/O ratios were more accentuated than at RC. The I/O ratios for both 100-200 and 200-400 nm were 0.06 higher than the average I/O ratio (0.75 ± 0.12) for ultrafine range. As can be observed in Figure 3.3, the difference in I/O ratios across size was much less than the difference in I/O ratios across ventilation conditions.

3.2.4 Effect of Ventilation Fan Setting on I/O Ratios

Figure 3.4 shows that increasing the ventilation fan setting from medium to full elevates the I/O ratios at both RC and OA, but this reduction was far more

pronounced at OA than at RC. Under RC conditions, the attenuation reduction inside the cabin (thus the elevated I/O ratios) is somewhat counter-intuitive, since increasing fan setting might be thought to increase particle removal via greater rates of airflow through the in-cabin filter, when present. Also, the deposition of particles on cabin surfaces increases with increase in in-cabin air velocity (Gong et al., 2009). However, fan setting has been previously demonstrated to increase AER for older vehicles (Fruin et al., 2011; Knibbs et al., 2009), likely due to leaks into the ventilation system, and filtration efficiency decreases at greater air flow velocities associated with higher fan settings as well (Qi et al., 2008). Apparently, these mostly offset the particle reductions expected due to greater volumes of air being filtered at high fan settings under RC conditions. However, higher fan settings resulted in increased losses for particles smaller than 50 nm, perhaps due to the increased turbulence in the ventilation system at higher fan settings.

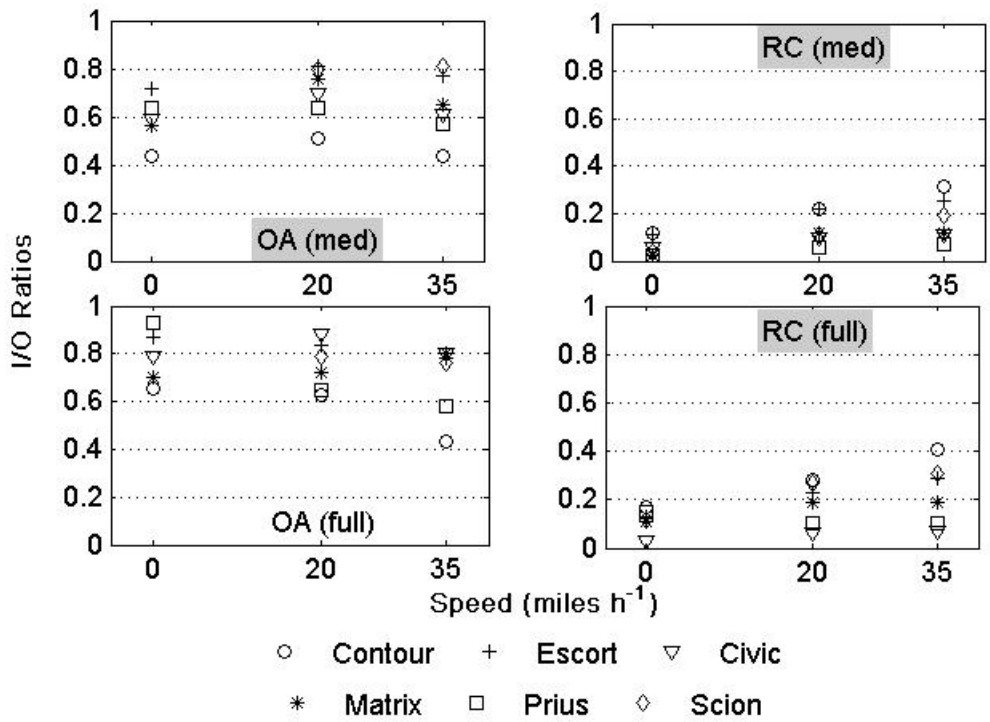


Figure 3.4: Comparison of I/O ratios at different speeds and fan settings.

At OA, an increase in fan speed settings strongly increased AER. The I/O ratios at full fan setting under OA conditions were as much as double that at medium fan setting and AERs at full fan setting were about 65% higher than at medium fan. Thus, ventilation fan setting is a key predictor of I/O ratios under OA ventilation conditions.

3.2.5 Effect of Cabin Air Filter and Loading on I/O Ratios

In order to determine the effect of cabin filters on particle removal and resultant I/O ratios, measurements were conducted under several filter conditions: no filter, used

(loaded), and new, at both OA and RC setting in three stationary vehicles. The Ford Contour's in-use filter had been operational for 36 months at the time of testing and was heavily loaded. The Honda Civic's filter had been in operation for ~14 months at the time of testing and was moderately loaded, and the Prius's filter had been in use for ~3 months, and was lightly loaded. All new filters were standard replacement cabin air filters bought from an auto parts store (all either brand STP or Purolator).

The presence or absence of the filter, or its loading, was observed to have only a small effect on I/O ratios. Used, loaded filters provided only moderately lower or comparable I/O ratios to a new filter. The results in Figure 3.5 for a Toyota Prius are typical and show that overall particle loss is not significantly affected by the presence of a new filter or even a loaded filter at OA settings. Without any filter, I/O ratio was only moderately higher (maximum difference observed was 0.1).

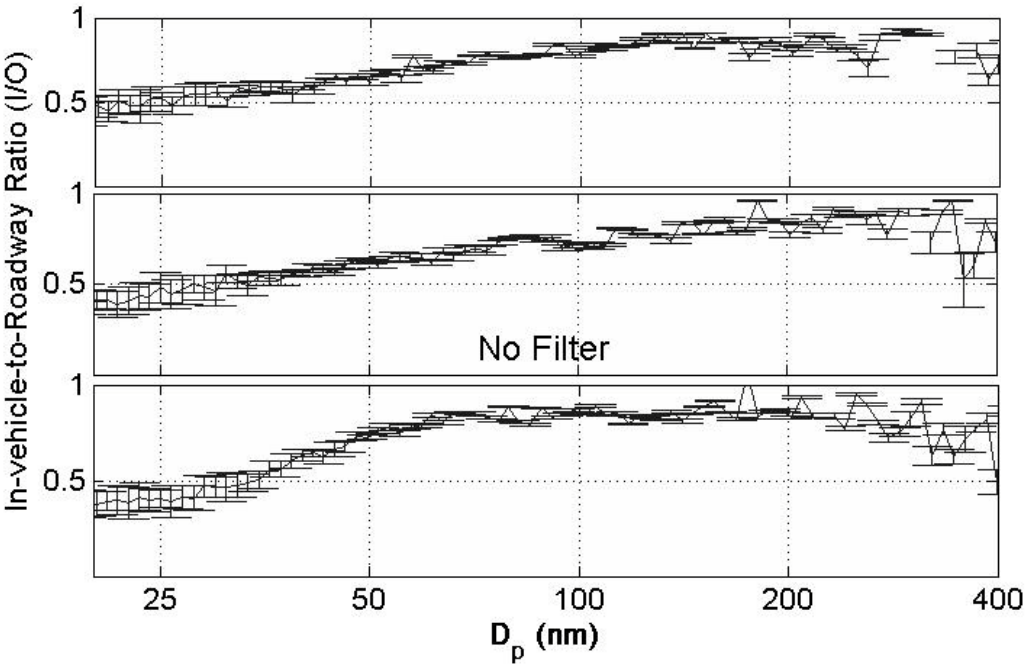


Figure 3.5: I/O ratios by filter condition or absence under OA conditions in a 2010 Toyota Prius.

Furthermore, the effect of several different filter loadings as characterized by in-vehicle pressure drop for the same vehicle was also investigated under OA ventilation conditions. Increased loading and resulting reduced pore size and flow rates, were found to decrease I/O ratios for UFP, but no significant changes were observed for particles exceeding 100 nm in size. Four loaded filters and a brand new filter were placed into Honda Civic 2010 model and tested at OA setting. Experiments were conducted in a stationary vehicle. Highest pressure drop (plotted as a subset in Figure 3.6) was observed for the heaviest loaded filter L4, while values for moderately loaded filters L1, L2 and L3 were comparable (Figure 3.6). Under OA setting, the I/O ratios decreased with filter loading for UFP particles (>100

nm) from 0.36 (± 0.05) for new filter to 0.36 (± 0.04), 0.42 (± 0.05), 0.49 (± 0.05) and 0.54 (± 0.05) for L1, L2, L3 and L4, respectively. A large increase in filter loading resulted in significantly better filtration efficiency for nano-particles (<50 nm), as seen in the abrupt change of I/O values values for L3 and L4 in Figure 3.6.

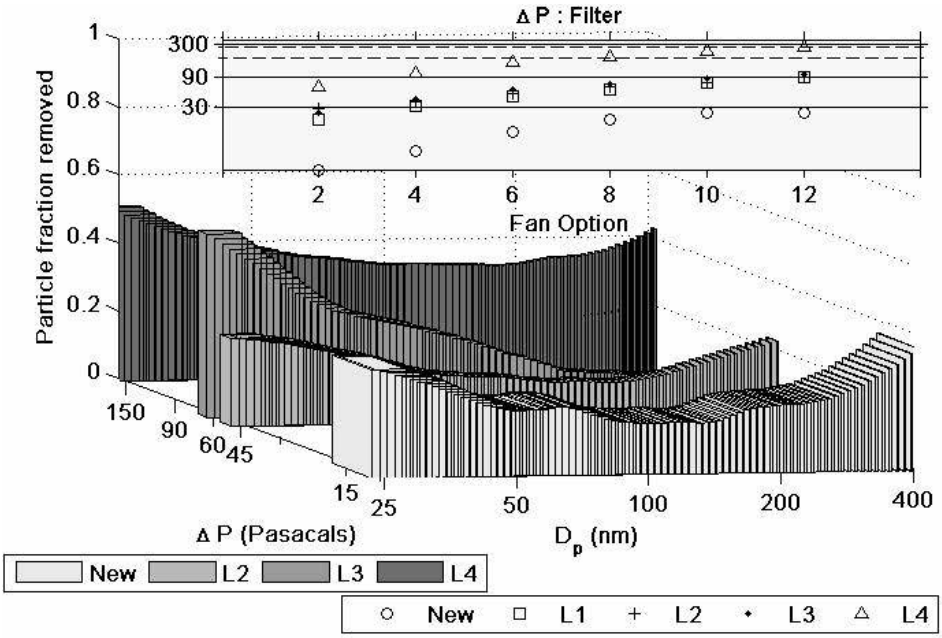


Figure 3.6: Effect of filter loading on particle removal for filters tested in Honda Civic vehicle under OA conditions.

This implies that the particle removal due to filtration is a small fraction of the total particle attenuation, and that most of the attenuation is probably due to turbulent surface deposition in the ventilation system or vehicle surface itself. Pui et al. (2008) reported 19% particle loss in a Toyota Camry in the absence of filter and suggested intrinsic losses in the ventilation system as an attenuation mechanism. In our tests 23, 40 and 40% particle losses were observed for the Honda Civic, Toyota Prius and Ford Contour with no filter, respectively. I/O ratios based on total particle number concentration are reported in Table 3.4 for three vehicles by filter presence and condition. Furthermore under RC settings, the I/O ratios in the absence of a filter were on average only 5% higher than with a filter in place. Although the presence and condition of the filter were insignificant to I/O ratio, it somewhat affected the time that the system requires to achieve the maximum attainable attenuation and stable I/O ratios. Similar increases in the time required to reach maximum attenuation have been reported by Pui et al. (2008).

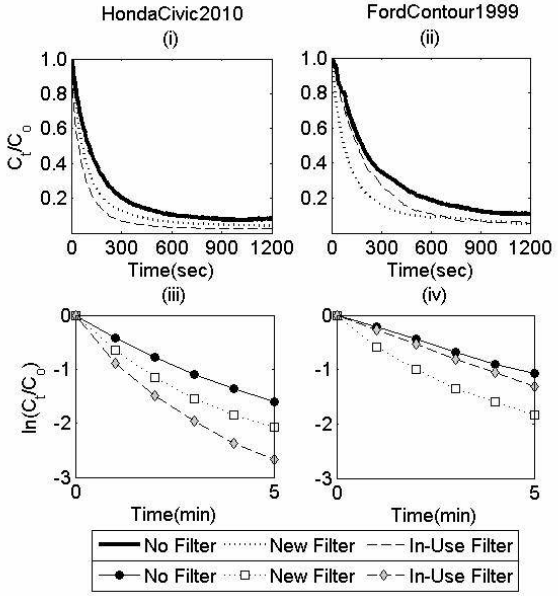


Figure 3.7: Effect of presence of filter in RC ventilation mode. C_0 is the concentration at the beginning of experiment, i.e., at time t = 0 and is equal to the ambient concentration.

Table 3.1: I/O ratios for three filter scenarios.

	Recirculation Setting			**Outside Air Setting**		
	No Filter	**New Filter**	**In-use Filter**	**No Filter**	**New Filter**	**In-use Filter**
Honda Civic	0.08 ± 0.02	0.04 ± 0.01	0.03 ± 0.01	0.77 ± 0.08	0.76 ± 0.09	0.67 ± 0.07
Ford Contour	0.11 ± 0.02	0.07 ± 0.01	0.06 ± 0.01	0.60 ± 0.05	0.64 ± 0.08	0.53 ± 0.01
Toyota Prius - 1	0.05 ±0.01	0.03 ± 0.01	0.02 ± 0.01	0.60 ± 0.08	0.57 ± 0.06	0.52 ± 0.04
Toyota Prius – 2*				(0.51 ± 0.06)	(0.47 ±0.07)	(0.41 ± 0.04)
* Measurements made in two Prius 2010 (vehicle 1 had ~3,200 miles and vehicle 2 had ~11,000 miles) to check for the effects of vehicle make.						

3.3 IMPLICATIONS FOR IN-VEHICLE PARTICLE MODELS

Some recent studies (Xu and Zhu, 2009; Xu et al., 2011) attribute all particle losses in the ventilation system to filtration and incorporate them into models by using filter efficiency as the removal mechanism. If the presence of cabin filters actually plays a

minor role in the attenuation of particles inside the vehicles, as observed in our measurements, a much larger component of attenuation occurs due to losses onto cabin and ventilation system surfaces. These different loss mechanisms should therefore be differentiated in models since they likely differ under different ventilation and driving conditions. Furthermore, any quantitative modeling should account for the intrusion flow of outside air into the ventilation system at RC. Knibbs et al. (2009; 2010) and Fruin et al. (2011) report that at RC, increases in fan setting increase the AER. They also reported that this increase seems to increase with age of the vehicle and could vary considerably by vehicle. Less than 15% increase on average was reported by Fruin et al. (2011) when fan setting was changed from medium to full, but this became as large as 40% for the older vehicles.

3.4 IMPLICATIONS FOR EXPOSURE ASSESSMENT

Predicting particle exposure inside vehicles requires determining ventilation setting first and foremost (i.e., OA or RC), due to its large impact on AER. Under OA conditions, fan setting is the most dominant variable, and I/O ratio was approximately 0.6 and fairly independent of speed. Under RC conditions, I/O ratio has a large range and varies from 0.5 to zero, depending on AER, which can be predicted by speed and vehicle age and mileage (Fruin et al., 2011). Under open window conditions, I/O ratios approach one, i.e., in-cabin concentrations frequently equal roadway concentration.

Difficult to obtain information, such as state of in-cabin filter loading, does not appear to be a crucial factor in assessing I/O ratios and the resulting in-vehicle particle exposures. It also does not appear that changes in on-road size distribution have a large impact on I/O ratios. Figure 3.8 exhibits the I/O ratio differences for four widely different hypothetical size distributions having number concentration mode less than 25 nm (fresh vehicle exhaust plume), 25-50 nm (on-road diluted plumes), 50-100 nm (aged vehicle emissions) and 100-200 nm (aged aerosol observed as urban background). The largest difference in I/O ratio occurs between aged and fresh aerosol, changing the overall I/O ratio by no more than 0.1 at OA and less than 0.05 at RC. Considering that on-road particle size number concentrations are dominated by ultrafine particles (Marawska et al., 1998), I/O ratio measurements based only on total number concentration (e.g., a total particle count from CPC) would be expected to (and were observed to) produce I/O ratio measurements nearly identical to an average I/O ratio resulting from a number-weighted average I/O ratio across multiple size ranges. Therefore, all of the variables needed to estimate I/O ratio within 10% or less can be obtained through questionnaires given to vehicle owners.

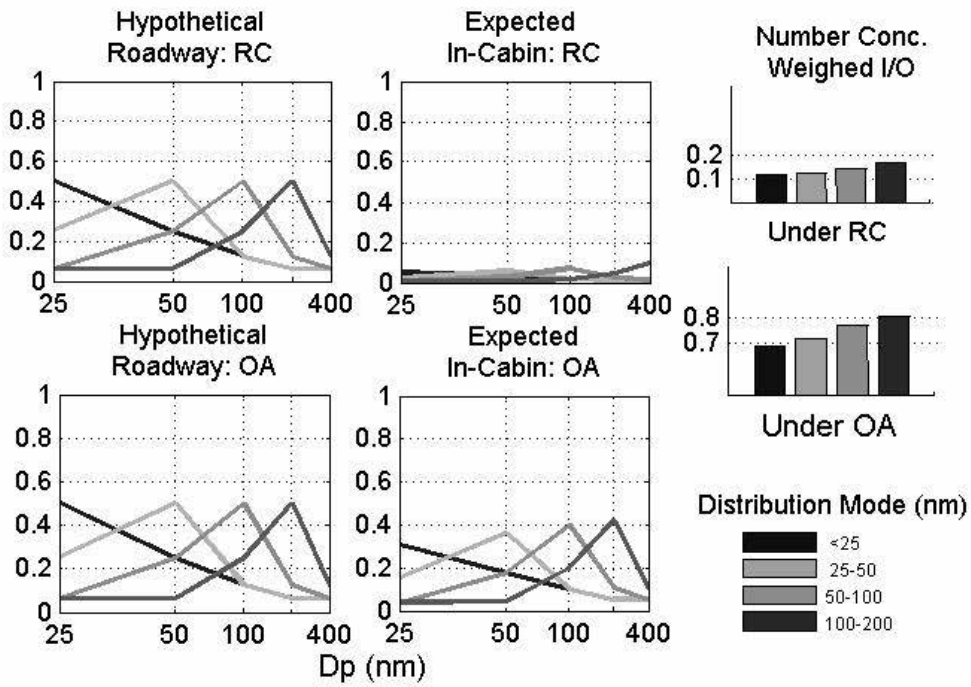


Figure 3.1: Impact of change in particle size distribution on number concentration weighed I/O ratios.

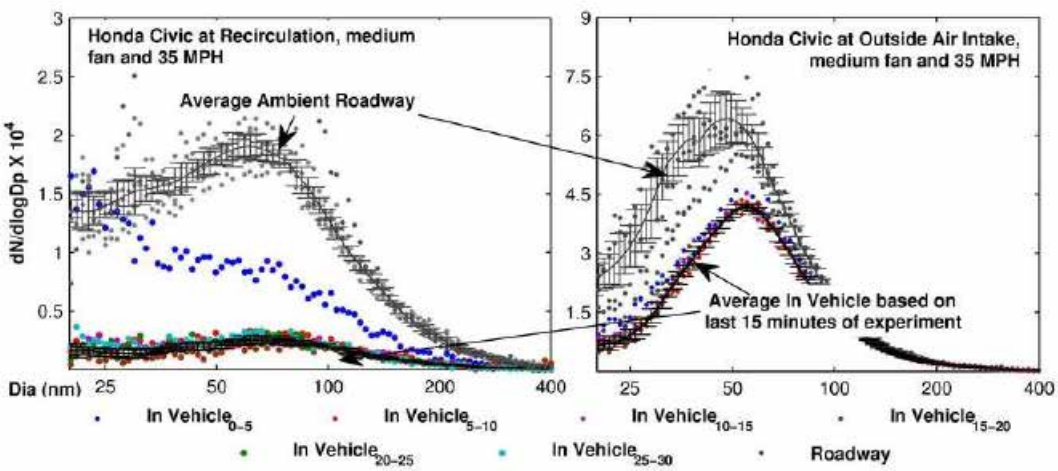


Figure 3.2: Progression of particle loss within vehicles at recirculation and outside air intake condition. The subscripts in the legend indicate the experiment time during which the scan was made. The measured AER at 35 miles h-1, medium fan, recirculation mode was 7.3 h-1 and at 0 miles h-1, medium fan outside air intake mode was 93 h-1.

Lastly, for the case of short duration trips, the equilibration conditions reported in this study for I/O ratio may not be reached. The particle attenuation reported here was attained typically within 5 to 10 minutes at OA conditions and 15 to 20 minutes at RC conditions, as shown in Figure 3.9. The benefit derived from higher reduction in particle exposure with the use of RC setting may be reduced if short trips are

taken repeatedly. The dynamic nature of particle attenuation should be considered in assessing short trip exposures.

3.5 SUMMARY AND CONCLUSIONS

In-vehicle concentrations result from the interaction of on-road concentrations with vehicle characteristics that can reduce or remove the pollutants, depending on the pollutant and the vehicle AER. The actual removal rates are due to a complicated interplay between a vehicle's physical characteristics, ventilation condition, particle size, and changes in air exchange rate (AER) (which increases with speed and vehicle age and increases the particle influx rate). Therefore, accurate determination of losses requires on-road testing under realistic aerodynamic conditions.

For task two, we focused on ultrafine particle (UFP) number concentrations, the particle pollutant with the highest and most widely-varying loss rates. Six vehicles were tested at different driving speeds, fan settings, cabin filter loadings, and ventilation conditions (outside air or recirculation). During outside air conditions, the fraction of particles removed averaged 0.33 ± 0.10 (SD). Fraction removed did not vary with vehicle speed but decreased at the higher ventilation flow rates of higher fan settings. During recirculation conditions, AER was much lower and removal fraction higher, once in-vehicle concentrations were stable (on average within about 10-20 minutes). Removal fraction averaged 0.83 ± 0.13 and was highly correlated with and a strong function of AER. Under both ventilation condition types, particle removal was primarily due to losses unrelated to filtration. Filter condition, or even the presence of a filter, played a minor role in particle fraction removed and most of the losses occur due to losses on to cabin or ventilation system surfaces.

The results from this chapter demonstrated that correct assignment of exposure concentration inside vehicles requires an additional parameter, inside-to-outside ratio (I/O), which can accurately modify roadway concentrations to reflect the combined effects of multiple mechanisms that either limit the pollutant penetration inside vehicles or cause loss therein. For a given vehicle, the I/O ratio depends on ventilation setting preferences and driving speed. Furthermore, the inter-vehicle difference in I/O, due to age is also significant. With the knowledge of the important parameters that influence I/O, more vehicles were tested in this study targeting all the necessary ventilation settings so the variation in I/O (due to ventilation settings) could be captured. The results from this Chapter were helpful in assessing data sufficiency and performing tests that spanned the range of I/O ratios expected under different conditions. This is because it is important to capture the full spread of I/O ratios to develop predictive models. Results from more vehicles and predictive models follow in Chapter 5, Part I. The next Chapter 4 presents results of on-road sampling that were used to develop models in the following Chapter 5, Part II.

REFERENCES

1. Fruin S.A.; Westerdahl, D.; Sax T.; Fine P.M. and Sioutas C. Measurements and predictors of in-vehicle ultrafine particle concentrations and associated pollutants on freeways and arterial roads in Los Angeles, *Atmos. Environ*., **2008**, 42, 207-219.
2. Wallace, L. and Ott, W. Personal exposure to ultrafine particles. *J. Expo. Sci. Environ. Epidemiol.,* **2011**, 21, 20-30.
3. Fruin S.A.; Hudda N.; Sioutas C. and Delfino R. A Predictive Model for Vehicle Air Exchange Rates based on a Large & Representative Sample, *Environ. Sci. Technol.*, **2011**, 45, 3569-3575.
4. Knibbs, L. D.; de Dear, R. J. and Atkinson, S. E. Field study of air change and flow rate in six automobiles. *Indoor Air*. **2009**, 303-313.
5. Gong, L.; Xu, B. and Zhu, Y. Ultrafine particles deposition inside passenger vehicles. *Aerosol Sci. Technol.* **2009**, 43, 544-553.
6. Xu, B. and Zhu, Y. Quantitative analysis of the parameters affecting in-cabin to on-roadway (I/O) ultrafine particle concentration ratios. *Aerosol Sci. Technol*. **2009**, 43, 400-410.
7. Xu, B.; Shusen. L.; Junjie, L. and Zhu, Y. Effects of vehicle cabin filter efficiency on ultrafine particle concentration ratios measured in-cabin and on-roadway. *Aerosol Sci. Technol*. **2011**, 45, 215-224.
8. Pui, D. Y. H.; Qi, C.; Stanley, N.; Oberdorster, G. and Maynard, A. Recirculating air filtration significantly reduces exposure to airborne nanoparticles. *Environ. Health Perspect.* **2008**, 116, 863-866.
9. Qi, C.; Stanley, N.; Pui, D. Y. H. and Kuehn, T. H. Laboratory and on-road evaluations of cabin air filters using number and surface area concentration monitors. *Environ. Sci. Technol.* **2008**, 42, 4128-4132.
10. Zhu, Y.; Eiguren-Fernandez, A.; Hinds, W. C. and Miguel, A. Incabin commuter exposure to ultrafine particles on Los Angeles freeways. *Environ. Sci. Technol.* **2007**, 41, 2138-2145.
11. Knibbs, L.D.; de Dear, R.J. and Morawska, L. Effect of cabin ventilation rate on ultrafine particle exposure inside automobiles. *Environ. Sci. Technol.* **2010**, 44, 3546-3551.
12. Westerdahl, D.; Fruin, S.; Sax, T.; Fine, P.M. and Sioutas, C. Mobile Platform Measurements of Ultrafine Particles and Associated Pollutant Concentrations on Freeways and Residential streets in Los Angeles. *Atmos. Environ*. **2005**, 39, 3597-3610.
13. Morawska, L.; Thomas, S.; Bofinger, N.; Wainwright, D. and Neale, D. Comprehensive Characterization of Aerosols in a Subtropical Urban Atmosphere: Particle Size Distribution and Correlation with Gaseous Pollutants. *Atmos. Environ*. **1998**, 32, 2467-2478.
14. Liu, D. L. and Nazaroff, W. W. Particle Penetration Through Building Cracks, *Aerosol Sci. Technol*. **2003**, 37, 565-573.

15. Nazaroff, W. W. Indoor Particle Dynamics, *Indoor Air.* ***2004.***14,175-183.
16. Liang K. and Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, 73, 13-22.
17. Hinds, W.C. *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles;* Wiley: New York, 1999.

4. CHAPTER FOUR: FREEWAY EMISSION RATES AND VEHICLE EMISSION FACTORS OF AIR POLLUTANTS IN LOS ANGELES

(based on Task 3. Estimate emission factors of PM pollutant concentrations based on roadway and urban background site measurements and CO2-based dilution adjustments.)

Note: This Chapter presents the results from on-road sampling. This data was used to guide the development of models in the following Chapter 5, Part II.

4.0 INTRODUCTION

Mobile emissions are the single largest source of nitrogen oxides (NOx) and carbon monoxide (CO) emissions, and a significant source of fine particulate matter ($PM_{2.5}$, Dp < 2.5 µm) emissions in Southern California (1) and the United States (2). According to the latest national emission inventory for 2011 (2), on-road emissions from mobile sources are responsible for 76% of CO and 30% of NOx emissions. In California, passenger cars and light/medium/heavy duty trucks account for 80% of vehicle miles travelled (VMT) and 70%, 39% and 43%, respectively, of the total emissions inventory for CO, NOx and $PM_{2.5}$ (3).

Vehicle emissions cause not only air quality problems but also adverse health effects. In urban ambient air NO_2, and more typically NO_x, serve as markers of a plethora of toxic air pollutants from fossil fuel combustion sources. Even at low levels, NO_2 has been associated with both cardiovascular and respiratory morbidity and mortality (4-5). Exposure to $PM_{2.5}$ has been linked to pulmonary and cardiovascular disease (6), and $PM_{2.5}$ from diesel combustion has been shown to increase lung cancer risk (7). Furthermore, $PM_{2.5}$ emissions from diesel truck engines are dominated by black carbon (BC) (8-9), which appears to be a better measure of traffic-related morbidity and mortality than $PM_{2.5}$ (10). Further, vehicular emissions emit most of the particles in ultrafine size range (UFP, d_p < 100 nm) that not only have high oxidative potential (owing to their high content of redox active organics and metals, but are also capable of penetrating deep into the lungs and crossing epithelial layers, thereby potentially causing systemic effects (11-13). Near-roadway exposures to vehicular emissions have been associated with adverse health effects by numerous studies (4-5,14).

Monitoring of emissions from mobile sources is not only important from a standpoint of assessing the public health risk that they pose, but also for evaluating the efficacy of regulatory measures and attainment of emission standards. Vehicle tailpipe and evaporative emissions have been the target of multiple regulations including fuel reformulation, state-controlled inspection and maintenance programs, and control-technology requirements. Benefits of improved control technologies have often been apparent; for example, NOx emissions from light-duty vehicles (LDV) have been substantially reduced through the use of three-way catalytic convertors (15).

In recent years, and as a consequence of improvements in LDV emissions control, heavy-duty diesel trucks (HDDT) have been reported as disproportionate contributors to on-road emissions, taking their low engine population and activity rates into consideration (15). In Los Angeles County, on-road emissions from passenger cars, which are responsible for 53% of VMT, contribute only 15% to NOx and 30% to $PM_{2.5}$ emissions from mobile sources. Light- to heavy-duty trucks, though responsible for only 28% of VMT, emit 25 % of NOx and 18% of $PM_{2.5}$ (3). Furthermore, previous studies (16-17) in Los Angeles have demonstrated a strong link between on road and near-roadway levels of pollutants, like BC and UFP, and truck density. However, effects of recent initiatives and regulations for HD trucks in California will erode the dominance of HD contributions to NOx and BC as the fleet turns over. For example, lower diesel PM mass regulations are requiring diesel particulate filter (DPF) controls, that dramatically reduce PM mass and BC emissions, but not always with commensurate reductions in particle number that dominate UFP (18). Selective catalytic reduction measures to meet NOx standards are also proving effective (19-20).

In addition to greatly exceeding per-mile emissions from gasoline vehicles, emissions from HDVs are often concentrated on certain truck routes. For example, goods movement to and from the Port of Los Angeles and Port of Long Beach results in high localized air pollution impacts (17). As a result, port-related diesel-engine activity has been the focus of recent regulations in Los Angeles, such as the San Pedro Bay Ports Clean Air Action Plan (21) (expected to eliminate more than 47% of diesel particulate matter and more than 45% NOx from port-related sources by 2014. Additionally, the California Air Resources Board (CARB) implemented Drayage truck regulations and a ban on pre-1993 engines, with the expected benefit of accelerated fleet turn-over and 85% reduction in $PM_{2.5}$ by 2014 (22).

Recent studies (23-25) have demonstrated the benefits of using a mobile monitoring platform (MMP), a vehicle equipped with real time instrumentation to determine on-road pollutant concentrations and emission factors during actual driving conditions. This study builds on those developments and uses an MMP to investigate real-world emissions factors on Los Angeles freeways. Measurements for gas and particulate phase pollutants were made during summer 2011 on several freeways. Fuel based emission factors and fractional contribution of diesel powered engines to total emissions was calculated. The goal of this study was to estimate freeways based emission rates based on vehicle emission factors and engine activity (quantified using vehicle miles traveled) per mile per hour of the freeway to investigate freeway-to-freeway variability, owing to differences in HDV fractions and total vehicular activity on freeways. Despite a factor of two difference in HDV fractions between freeways, and up to a magnitude difference in HDV and LDV emission factors, total freeways emissions were found to be comparable underscoring the importance of

considering total vehicle activity (in terms of miles travelled) over just vehicle or HDV counts.

4.1 METHODS

4.1.1 Mobile Measurement Platform (MMP) and continuous measurement instruments

A hybrid vehicle (2010 Honda Insight) was used as a mobile measurement platform. All the continuous instruments listed in Table 4.1 below drew air samples from a common sampling duct installed across the rear windows. The response time of the fastest instrument (Condensation Particle Counter Model 3007, (CPC), TSI Inc., MN, USA) to an on-road plume, including the residence time for the air in the sampling duct, was less than two seconds. The concentration time-series recoded by all other instruments were aligned with respect to the fastest instrument to adjust for the delayed response. Time-lag remained constant (to within a second) over the campaign due to fixed instrument response behavior and flow rates. Instruments were periodically calibrated and time was synced to be within 1 second with the Global Positioning System (GPS) device (Garmin GPSMAP 76CSC). Further, data quality assurance comprised regular flow and zero reading checks. MMP was driven in the central freeway lane, when possible, through this study.

A limitation of the Dust Trak DRX (Model 8533, TSI, USA) was its potential lack of sensitivity to smaller particles, such as those found in diesel exhaust, which could decrease the accuracy of the measurements of emissions. Since the instruments were not calibrated to accurately reflect fresh emissions, its data may be used for quantitative inter-comparisons within this study. In this campaign, black carbon (BC) mass concentrations were determined by using an Aethelometer (Magee Scientific, Model AE51), which measures the optical attenuation (ATN) of a light beam transmitted through a sample collected on a filter. At low filter loadings, there is a linear relationship between BC and ATN; however, as particles accumulate on the filter this linearity breaks down, and a correction (described in Wang et al. (24) was applied to obtain an accurate BC concentration. UFP concentration reported by CPC 3007 was corrected using the method described in Westerdahl et al. (23) when levels exceed 10^5 particles cm^{-3}.

Table 4.1: Instruments used in this study

Instrument	Parameter measured	Instrument Flow Rate (lpm)	Response Time (s)	Resolution	Detection Limit
TSI portable CPC (butanol-based) model 3007	UFP count, 10 nm - 1 um	0.8	1	1 particle/cm^3	10 nm, <0.01 particles/cm^3
TSI DustTrak DRX, model 8533	PM2.5 mass	1.7	5	+/- 0.001 mg/m^3	0.001 - 100 mg/m^3, 0.1 - 2.5 um size range
Magee Scientific Aethalometer AE 51	Black carbon	150 mL/min	60	0.001 µg BC/m3	±0.1 µg BC/m3 , 1 min avg., 150 mL/min flow rat
LI-COR model LI-820	CO2	1	<1	>4% of the reported value	3.0 ppm
2-B Technology Nodel 408	NO	1	8	Greater of 3 ppb or 3% of reading	
2-B Technology Nodel 401-410	NOX	1	8	Higher of 1.5 ppb or 2% of reading	
EcoChem PAH analyzer, model PAS 2000	Particulate matter-phase PAH	2	< 10	~ 0.3 -1 g /m3 PAH per picoamp	3 ng/m^3
TSI Q-Trak Plus monitor, model 7565	CO, Temperature, humidity		20		1 ppm
Garmin GPSMAP 76CSx	GPS location, speed	N/A	1	3m	

4.1.2 Sampling Routes

Emissions from motor vehicles were measured on five Los Angeles freeways – I-110, I-405, I-710, CA-60 & CA-91 (See Figure 4.1). While freeway 110's northern segment (110N) is closed to HDV, they are allowed on the southern segment (110S). These two segments of 110 have been discussed separately throughout this study. Based on California Department of Transportation (Caltrans) 2009 Annual Average Daily Traffic (AADT) and truck data (counts trucks 2-axle or higher), trucks constitute less than 1% of the total vehicle flow on 110 N and 5.0 ± 1.5 % on 110S segment (26). Freeway I-405 has a mixed-fleet but is mostly dominated by LDV (the Caltrans data based truck fraction was 3.8 ± 0.58 %). The other three freeways, I-710, CA-60, CA-91, have a relatively higher fraction of HDVs, i.e., 12 ± 5.7 %, 6.9 ± 1.6 % and

7.6 ± 1.1, respectively. Date and time of sampling, along with meteorological conditions during sampling period are listed in Table 4.2.

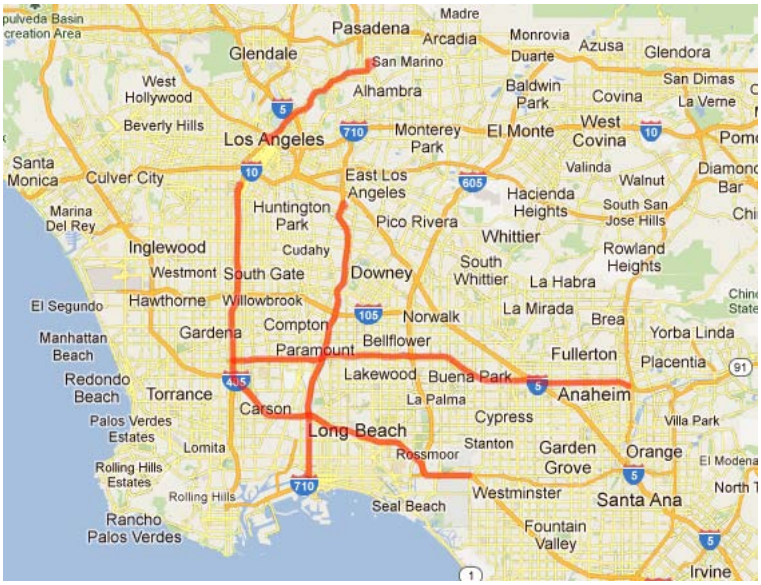


Figure 4.1: Freeway segments where measurements were conducted (generated using Google Maps).

4.1.3 Mathematical calculations and equations

<u>4.1.3.1 Emission Factor (EF)</u>

Fuel-based EFs were calculated for each freeway segment each day using carbon balance approach, shown in Equation 4.1. Pollutant emission was normalized by total carbon emissions on the freeway to compute emission factors in units of mass of pollutant emitted per unit mass of fuel burned. Carbon combustion products that were accounted were carbon dioxide (CO_2), carbon monoxide (CO) and black carbon or soot (BC).

Equation 4.1: Pollutant emission factor

$$EF_P = 10^3 \left(\frac{\Delta[P]}{\Delta[CO_2]+\Delta[CO]+\Delta[BC]} \right) \times w_c$$

Table 4.2: Sampling days, hours and meteorological conditions

710 (North)

Date	Hours	Temp (deg C)	Relative Humidity (%)	Wind Speed (m/s)	Wind Direction (degrees)
17-May	10:00-11:00	15.6	56	1.8	76
19-May	10:00	13.9	86	0.9	296
1-Jun	20:00	13.9	71	3.1	261
2-Jun	08:00	20.0	51	1.3	353
3-Jun	13:00-15:00	23.1	46	4.0	201
4-Jun	12:00	21.7	47	3.1	207
8-Jun	06:00-08:00	16.1	79	2.7	161
14-Jun	11:00	26.1	52	2.8	252

710 (South)

Date	Hours	Temp (deg C)	Relative Humidity (%)	Wind Speed (m/s)	Wind Direction (degrees)
17-May	09:00-10:00	13.9	87	2.2	79
19-May	09:00-10:00	14.4	81	1.3	275
1-Jun	20:00	13.9	71	3.1	261
2-Jun	07:00-08:00	18.9	56	1.3	186
3-Jun	12:00, 14:00	23.3	43	4.0	199
4-Jun	12:00	21.7	47	3.1	207
8-Jun	05:00-08:00	16.3	79	2.6	162
14-Jun	10:00	25.6	51	2.2	155

110 N (North)

17-May	08:00	11.7	84	4.9	75
1-Jun	19:00	15.6	61	3.1	179
2-Jun	09:00	21.1	45	4.5	264
3-Jun	13:00	24.4	46	4.0	228
4-Jun	13:00	22.2	48	3.6	187
8-Jun	06:00	16.1	77	2.2	159
14-Jun	11:00	26.1	53	3.6	210

110 N (South)

1-Jun	19:00-20:00	15.0	63	2.5	172
2-Jun	09:00	15.0	63	2.5	172
3-Jun	13:00-14:00	24.2	47	4.0	236
4-Jun	13:00	22.2	48	0.0	187
8-Jun	07:00	16.1	75	2.2	174
14-Jun	12:00	25.0	56	4.5	258

110 S (North)

1-Jun	18:00	16.1	59	3.6	263
2-Jun	11:00, 13:00	21.7	41	4.7	256
4-Jun	11:00	20.6	51	3.6	228
14-Jun	14:00	24.4	57	3.1	278
15-Jun	18:00	18.3	74	2.7	281

110 S (South)

1-Jun	16:00-17:00	18.6	49	4.5	256
2-Jun	10:00-12:00	21.1	45	4.5	263
4-Jun	10:00	20.0	54	3.1	225
14-Jun	12:00-13:00	25.3	54	2.9	271
15-Jun	16:00	22.2	63	2.7	260

405 (North)

1-Jun	18:00	16.1	49	4.5	275
2-Jun	10:00-11:00	20.0	47	3.4	253
4-Jun	11:00	18.3	57	2.7	216
14-Jun	13:00-14:00	23.3	61.5	3.4	278

405 (South)

17-May	12:00	18.3	65	1.8	357
2-Jun	12:00	20.0	66	4.9	275
15-Jun	16:00	19.4	77	2.7	200

60 (West)

19-May	11:00-12:00	20.6	48	3.1	180
3-Jun	08:00-09:00	20.6	51.5	1.6	201

60 (East)

17-May	14:00	18.3	52	3.1	148
19-May	11:00-12:00	20.6	48	3.1	180
8-Jun	09:00	16.7	65	2.2	194

91 (West)

19-May	11:00-12:00	20.8	53	3.4	148
2-Jun	12:00-13:00	21.7	40	4.9	260
15-Jun	17:00-18:00	19.4	71	2.5	281

91 (East)

1-Jun	17:00-18:00	16.9	55	4.0	261
2-Jun	10:00	21.1	48	4.0	277
4-Jun	10:00-11:00	20.3	53	3.4	227
14-Jun	13:00	25.0	54	3.1	261

where *EF*p is the emission factor (g emitted per kg fuel burnt) for pollutant P, Δ[*P*] is the increase in the concentration of pollutant P (g m^{-3}, or #/m^3 for ultrafine particle number concentration (PNC)) above the background concentration, Δ [CO2], Δ [CO] and Δ [BC] are the increases in the concentrations of carbon combustion products (g m^{-3}). w_c, the mass fraction of carbon in fuel was used as 0.85 for gasoline fuel and 0.87 for diesel fuel (27). Background values for use in Equation 4.1, were estimated as the 5th percentile of pollutant concentration observed on the freeway segment, while the median value was used as an estimate of elevated pollutant concentration due to vehicular emissions. Since the goal of this study was to estimate fleet average EFs, the use of median value allowed for excluding any bias (which would otherwise be present in averages) due to capture of specific high-emitting vehicle plumes.

Pollutant concentrations were partitioned to estimate LDV and HDV emission factors. The approach used was similar to that used by Ban-Weiss (2008) (9). Pollutants measured on other freeways were apportioned using pollutant-to-CO_2 emission ratios measured on 110N, where emissions were assumed to be solely from gasoline-fueled engines. The following equation summarizes the technique.

Equation 4.2: Elevation in pollutant concentration due to diesel fuel combustion

$$\Delta\,[P]_{f,d} = \Delta\,[P]_f - \Delta\,[CO_2]_{f,g}\left(\frac{\Delta\,[P]_{110N}}{\Delta\,[CO_2]_{110N}}\right)$$

where $\Delta\,[CO_2]_{f,g}$ is the fraction of CO_2 attributed to gasoline, and is apportioned using Equation 4.3 (below), which takes into account the difference in gasoline and diesel vehicle fuel economies.

Equation 4.3: CO2 Apportionment

$$\Delta\,[CO_2]_{f,g} = \Delta\,[CO_2]_f\left(\frac{(1-f_d)\times(\frac{1}{FE_g})\times\rho_g\times w_g}{(f_d)\times(\frac{1}{FE_d})\times\rho_d\times w_d + (1-f_d)\times(\frac{1}{FE_g})\times\rho_g\times w_g}\right)$$

where f_d represents the fraction of vehicles using diesel fuel, FE is the fuel economy (mile L^{-1}), ρ Is the density of fuel (kg L^{-1}), and w_g and w_d are the mass fraction of carbon in gasoline and diesel. Former studies (9, 27-28) have used an expression similar to Equation 4.3. An underlying assumption in use of this equation is that distances travelled by both gasoline and diesel vehicles are equal. While in former studies conducted in tunnel environments this assumption is valid, it should be used with caution in studies that employing mobile platforms. If gasoline-fueled (mostly LDV) and diesel-fueled (mostly HDV) engines vehicles are segregated in different lanes with significant differences in lane speeds, this will lead to significantly different distances travelled (and fuel burnt) by gasoline and diesel vehicles on that freeway. The MMP's lane location and the direction of wind are also a consideration if lanes are segregated. Using a fraction based solely on vehicle counts might therefore lead to erroneous results. A more accurate estimate of f_d can be obtained by breaking down total vehicle miles travelled (if available) into those by gasoline and diesel vehicles over the measured span of the freeway, which inherently accounts for differences due to speed and thereby distance travelled and fuel burnt. This approach was used in the current study and CO_2 was apportioned based on f_d calculated using Equation 4.4, where VMT_d and VMT_g are vehicle-miles traveled

by diesel and gasoline vehicles. Using this modified approach, Equation 4.3 can be re-written as Equation 4.5.

Equation 4.4: Fraction of fuel consumed that was diesel

$$f_d = \frac{VMT_d}{VMT_d + VMT_g}$$

Equation 4.5: Revised CO2 Apportionment for mobile monitoring

$$\Delta\,[CO_2]_{f,g} = \Delta\,[CO_2]_f \left(\frac{VMT_g \times (1/FE_g) \times \rho_g \times W_g}{VMT_d \times (\frac{1}{FE_d}) \times \rho_d \times W_d + VMT_g \times (\frac{1}{FE_g}) \times \rho_g \times W_g} \right)$$

The values for fuel economy used in this study were 5.8 miles L^{-1} and 1.6 miles L^{-1} for gasoline and diesel fuel engines, respectively. Similar to other studies (9) fuel density values were 0.74 kg L^{-1} and 0.84 kg L^{-1} for gasoline and diesel fuel, respectively.

4.1.3.2 Traffic Characterization

The total and break-down of vehicle miles travelled (VMT) by HDV and LDV were obtained from aggregate data over all lanes of the freeway reported by the California Department of Transport (CALTRANS) Performance Measuring System (PeMS) (29), which is publicly available. Further, PeMS classifies VMT into those traveled by light-duty vehicles (LDV) and truck or heavy-duty vehicles (HDV). The PeMS dataset cannot account for VMT travelled by medium duty vehicles (MDV), and they are by default attributed to LDV. The 1.45% fraction of LDV VMT in Los Angeles County resulting from diesel powered engines were neglected. We assumed that all LDV VMT are travelled using gasoline fuel. Based on EMFAC 2011 (3) estimates for Los Angeles County, only 10% of VMT associated with MDV are travelled using diesel engines. Attributing all these to gasoline LDV VMT in addition to 1.45 % of diesel LDV VMT assumed to be gasoline, would lead to 2.2 % overestimation of LDV VMT. All VMT traveled by HDV were attributed to diesel fuel. PeMS estimates truck traffic volume to within 5.7% of the values reported by weight-in-motion sensors (30). However, use of PeMS estimates does offer the advantage of obtaining HDV estimates at a much finer spatial resolution (than limited weight-in-motion truck sensors), which was required in this study for relating real-time pollutant measurements to real-time traffic estimates.

4.1.3.3 Freeway emission rate calculations

Partitioned EFs were used to calculate freeway emission rates (ER), i.e., pollutant mass/number emitted per mile of freeway per unit time (kg or g or # $mile^{-1}$ h^{-1}) using Equation 4.6 for six freeways. The subscripts *p* and *f* are for pollutant and freeway, respectively.

Equation 4.6: Freeway emission rate estimation

$$ER = VMT_{HDV} \times EF_{HDV} \times \left(\frac{1}{FE_{HDV}}\right) \times \rho_d + VMT_{LDV} \times EF_{LDV} \times \left(\frac{1}{FE_{LDV}}\right) \times \rho_g$$

4.2 RESULTS AND DISCUSSION

4.2.1 Pollutant Concentrations

Comparison across the freeways suggests that pollutant concentrations on gasoline or LDV dominated freeways were lower than those with higher fraction of HDVs. Concentrations appear to have dropped in the last decade as well. Median concentrations on I-710 BC, UFP NO were dramatically lower, i.e., 37.5 %, 32.1% and 37.2%, respectively, of their median values reported for measurements conducted in spring of 2003 by Fruin et al. (16). Comparison to summer 2005 concentrations reported by Fujita et al. (33) for the morning (AM) period on similar segments of freeways, i.e., when the proportion of trucks was high/highest also suggest a reduction. Comparison to values for afternoon (PM) hours, during which traffic is dominated by gasoline vehicles, as reported by Fujita et al. (33), did not suggest any clear trend. Average concentrations for various pollutants and freeways are listed in Table 4.3

4.2.2 LDV and HDV emission factors

Emission factors for light -and heavy-duty vehicles, calculated using Equations 4.2, 4.3 and 4.6 are shown in Table 4.4. It should be noted that standard deviations for HDV are much higher compared to LDV due to propagation of uncertainties related to apportioning HDV fractions, but also reflect the relatively larger variation in HDV fleet freeway-to-freeway. However, freeway-to-freeway differences in vehicle emission factors were not significant enough to merit separate discussion, except for I-710, where slightly lower EFs were observed. Consistent with the trend reported by Ban-Weiss et al. (9) and Bishop and Stedman (20) lower NOx and NO emission factors were found for HDV, though the spread was relatively high. Specifically for I-710, NOx and NO emission factors were 21 ± 5 g/kg-fuel and 12 ± 4 g/kg-fuel (n=14). LDV NOx and NO emission factors were found to be comparable to the most recent studies – Ban Weiss et al. (9) and Bishop and Steadman (19). This decrease could be interpreted with some caution as a consequence of the Clean Air Action Plan (21) and CARB regulations (2).

Table 4.1: Comparison of Emission factors from current study to previous studies

Reference	Year of measurement	Vehicle Type	NO	NOx†	BC	$PM_{2.5}$	PNC
			g kg^{-1}	g kg^{-1}	g kg^{-1}	g kg^{-1}	# kg^{-1}
This work	2011	LDV	1.8 ± 0.01	3.0 ± 0.01	0.02 ± 0.009	0.16 ± 0.04	(2.9 ± 0.53) x 10e14
		HDV	13.1 ± 6.6	24.4 ± 10.7	0.53 ± 0.39	0.60 ± 0.54	(5.7 ± 2.5) x 10e15
Bishop et al., 2008,2009	2008	LDV	3.2				
		HDV	17.7	30.5			
Park et al., 2011*	2007	LDV		9.4 (2.5-5.7)	0.06 (0.01-0.03)	0.15 (0.04-0.07)	6.0 × 10e14 (1.5-5.2) × 10e14
		HDV		34 (6.8-17.6)	0.5 (0.07-0.17)	0.73 (0.08-0.33)	4.5 × 10e15 (0.71-1.4) × 10e15
Ban-Weiss et al., 2008, 2010	2006	LDV		3.0 ± 0.2	0.026 ± 0.004	0.07 ± 0.02	(3.9 ± 1.4) x 10e14
		HDV		40 ± 3.0	0.92 ± 0.07	1.4 ± 0.3	(3.3 ± 1.3) x 10e15
Kirchestetter et al., 1999a	1997	LDV		9.0 ± 0.4	0.035 ± 0.004	0.11 ± 0.01	
		HDV		57 ± 7	1.4 ± 0.6	2.7 ± 0.3	
Geller et al., 2005	2004	LDV				0.07 ± 0.03	(2.5 ± 1.4) x 10e15
		HDV				1.02 ± 0.06	(8.2 ± 6.3) x 10e15

†expressed as NO_2 equivalents, * mean (median values across various driving modes)

Table 4.2: Pollutant concentrations on freeways

Route	Statistic	CO2	CO	BC	NO	NOx	PNC	PM2.5	PB-PAH	Speed
		ppm	ppm	ng/m3	ppb	ppb	#/cm3	mg/cm3	ng/m3	miles h-1
I-110N	Mean	495	1.4	3190	88	69	29929	0.03	34	44
	SD	48	1.2	3482	53	49	15697	0.032	83	18
	Median	485	1.35	3550	78	54	25900	0.03	17	53
	5th Percentile	436	0.5	1373	23	12	14735	0.012	6	12
	IQR	(463 - 516)	(0.9 - 1.9)	(1930 - 4130)	(52 - 115)	(30 - 93)	(19600 - 35000)	(0.016 - 0.073)	(10 - 32)	(44 - 57)
I-110S	Mean	474	1.6	3642	119	104	42772	0.021	47	49
	SD	30	1.9	5816	71	74	25268	0.014	61	21
	Median	470	1.4	2350	103	82	38000	0.02	30	58
	5th Percentile	432	0.6	718	43	31	13641	0.01	10	0
	IQR	(455 - 489)	(0.9 - 1.8)	(1540 - 3480)	(76 - 142)	(56 - 133)	(26900 - 52900)	(0.01 - 0.02)	(19 - 54)	(44 - 63)
I-405	Mean	497	1.3	6049	121	98	43493	0.024	75	46
	SD	40	0.6	6894	60	54	31335	0.015	99	24
	Median	488	1.2	4490	114	91	34000	0.02	51	59
	5th Percentile	445	0.6	986	42	28	13454	0.008	11	1
	IQR	(470 - 516)	(0.8 - 1.5)	(2500 - 7490)	(81 - 144)	(60 - 124)	(22500 - 53800)	(0.01 - 0.03)	(29 - 94)	(24 - 65)
I-710	Mean	493	1	8665	194	152	72255	0.024	82	54
	SD	37	1	20719	85	74	61994	0.017	67	16
	Median	491	1.05	4554	170	131	57298	0.02	59	60
	5th Percentile	438	0	835	73	53	20643	0.009	15	16
	IQR	(466 - 516)	(0 - 1.5)	(2483 - 8045)	(132 - 244)	(99 - 189)	(43387 - 85635)	(0.015 - 0.025)	(38 - 104)	(50 - 65)
SR-91	Mean	482	1.5	6295	158	140	54014	0.024	98	51
	SD	34	1.2	4940	86	90	40303	0.014	123	20
	Median	478	1.3	4750	147	123	46900	0.02	65	61
	5th Percentile	429	0.7	1203	43	30	12803	0.01	12	7
	IQR	(502 - 459)	(1.6 - 0.9)	(8180 - 2790)	(211 - 93)	(187 - 73)	(67600 - 25800)	(0.03 - 0.02)	(119 - 32)	(65 - 40)
SR-60	Mean	509	1.8	10411	190	170	68289	0.027	122	35
	SD	46	0.7	5537	105	100	41468	0.009	115	24
	Median	501	2	9950	193	178	60994	0.025	102	32
	5th Percentile	449	1.2	3342	47	24	19197	0.015	9	0
	IQR	(533 - 475)	(1.9 - 1.4)	(15072 - 5050)	(250 - 101)	(226 - 84)	(84759 - 39627)	(0.031 - 0.02)	(169 - 39)	(60 - 17)

Direction in parentheses indicates lane of travel. Note: I-110 is divided into a LDV-only section (110 N) and a mixed fleet section (110 S).

4.2.3 Fraction contribution of HDV to total emissions

LDV and HDV emission factors can differ by up to a magnitude or more. Coupled with the much lower HDV fuel economy (miles travelled per kg of fuel burnt), it results in highly disproportionate contributions from HDV to total emissions. Fractional contribution of HDV (Ef_{HDV} or emission fraction) to total NO_X, BC and PNC emissions was calculated using Equation 4.7.

Equation 4.7: Fractional contribution of HDV to total emissions

$$Ef_{HDV} = \frac{EF_{HDV} \times \left(\frac{1}{FE_{HDV}}\right) \times \rho_d}{EF_{LDV} \times \left(\frac{1}{FE_{LDV}}\right) \times \rho_g + EF_{HDV} \times \left(\frac{1}{FE_{HDV}}\right) \times \rho_d}$$

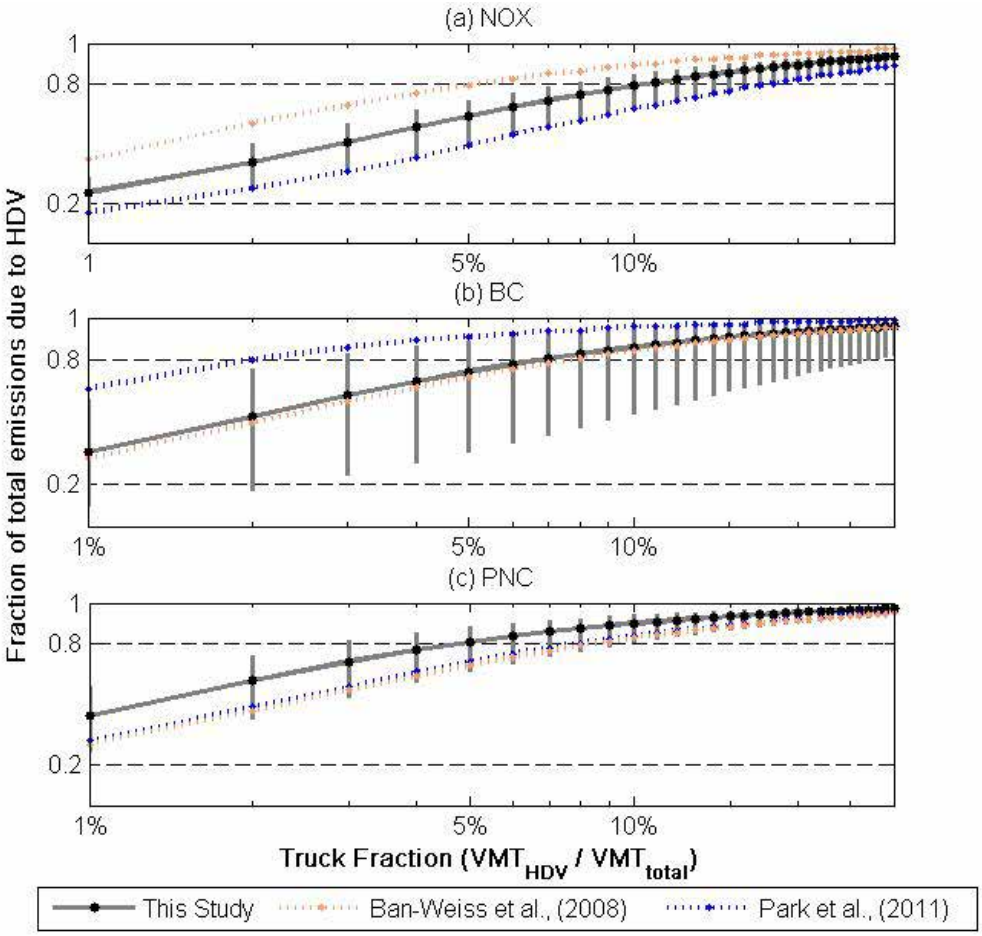


Figure 4.1: Contribution of HDV to total emissions

Results suggest that HDV contributing a mere 1 % to VMT would be responsible for 25% of NOx, 44% of PNC and 36% of total BC emissions. Concordantly, for a mile of travel, HDV emissions were 33, 55 and 79 times that of LDV emissions for NOX, BC and PNC, respectively. The latest studies conducted in Caldecott tunnel 4 years ago (9, 18), which has a 4% up-grade that puts engines under higher load, showed

much higher HDV emissions for NOX and BC, but lower for PNC. Specifically, based on vehicle emission factors and fuel efficiency values reported by Ban-Weiss et al. (9), HDV emissions were 73, 193 and 46 times of LDV emissions for NOx, BC and PNC, respectively. These differences (in part due to both emission technology improvements and real-freeway driving) convey the need for estimation of real-driving based emission factors in Los Angeles to trace the effects of regulation and emission improvements.

Figure 4.2 shows the fraction of total emissions attributable to HDV based on this study's EF, and also compares them to previous studies. Accounting for the standard deviation, the fractional contribution of HDV to NOx emissions was lower in 2011 compared to measurements conducted in 2006, by Ban-Weiss et al. (9), but higher than the mean value reported by Park et al. (25) for measurements conducted in 2007. It should be noted that this difference is due to a much higher LDV NOx emission value reported by Park et al., (25) (about three times than other recent studies). Contribution of HDV to BC and UFP emissions in the present study was comparable to previous studies

4.2.4 Freeway Pollutant Emission Rates

4.2.4.1 Annual average emission rates

Pollutant emission rates for four freeways were computed using Equation 4.6. Daily VMT (and fraction due to HDV) on the entire segment of freeway in Los Angeles County during the 215 working days from Dec 1, 2010 – Nov 30, 2011 were used to generate daily emissions from freeway during normal working days. Further normalized by 24 hours, annual average hourly value for emission rates are plotted in Figure 4.3 and daily time series for VMT and HDV VMT is shown in Figure 4.4. Except for summer to fall increase in port related goods activity on I-710, there were no significant seasonality aspects to consider. (Standard deviation in VMT was < 2% and trucks was < 5%.) Despite lower VMT attributable to HDV on I-110S and CA-91, ER on I-110S were comparable to I-710 and ER for CA-91 was significantly higher than I-710. Conventionally, I-710 has often been studied as a high end of freeway emissions. The present results suggest that 'truck counts' may not be a sufficient or the best indicator of line emission sources like freeways. Emissions are better correlated with vehicle activity, including speed, than vehicle counts. Using an indicator like 'vehicle miles travelled,' which accounts both count and speed, may be better suited for modeling techniques predicting concentrations near roadways.

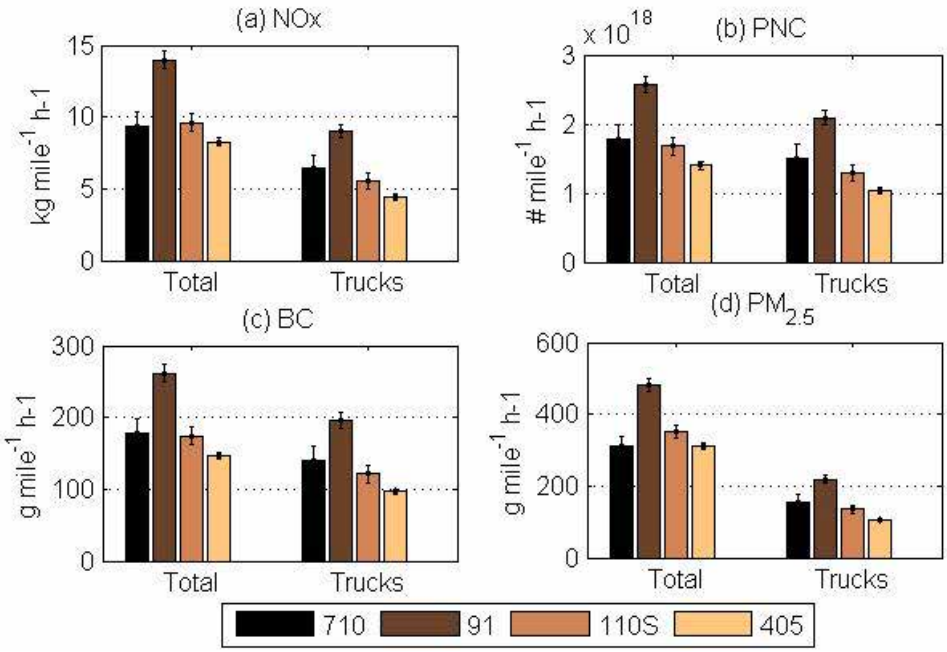


Figure 4.2: Annual average hourly freeway emission rates

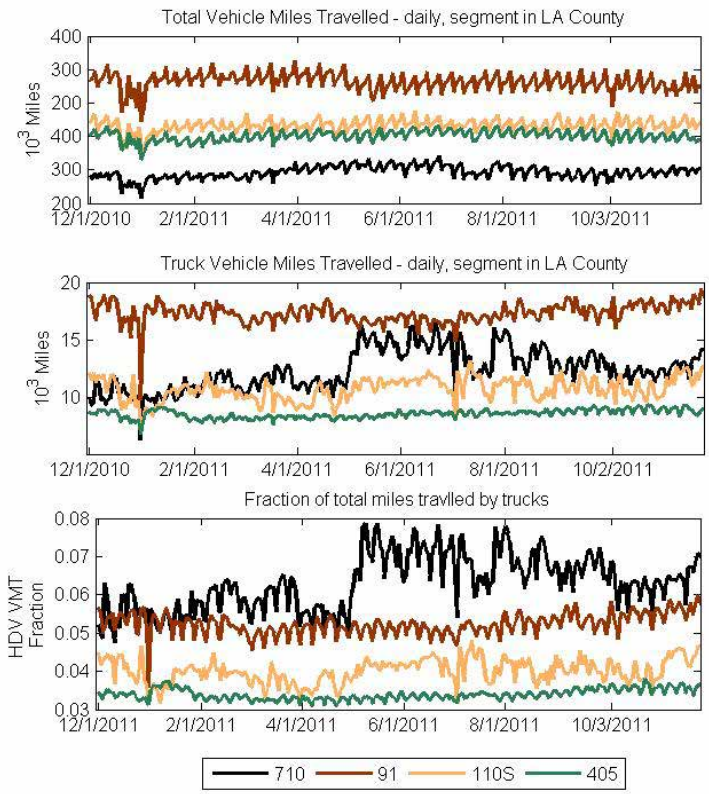


Figure 4.3: Vehicle miles travelled, truck vehicle miles travelled and fraction of total miles traveled by truck on four Los Angeles freeways in LA County during 12/1/2010 – 30/11/2011.

4.2.4.2 Diurnal variation in freeway emission rates

Mean hourly VMT (and fraction due to HDV) on the freeway segments for an archetypal month, May 2011, were used to generate diurnal profiles. Pollutant emission rates have been plotted in Figure 4.5 (a) – (d). Several important observations can be made from these figures. Firstly, as expected, diurnal profiles of emission rates are shaped similarly to the vehicle activity profiles on each freeway (See Figure 4.6). Secondly, unlike concentration profiles, which tend to be bi-modal (32-34) with peaks during the morning and evening commute hours, the emission rate profile has a single mode in the middle of the day. During midday (10:00 – 13:00 hours) a drop in vehicle activity is often observed, more strongly in vehicle flow (number of vehicles per hour) than in vehicle miles travelled. However, HDV activity increases and peaks during midday hours. The emissions from increased HDV activity seem to compensate for- and in fact outweigh any reductions in emission due to fewer vehicles, thereby producing distinctly uni-modal profiles for emission rates. As noted above, this contrasts the pollutant concentration profiles that have often been reported to be bi-modal for multiple pollutants near roadway locations and urban areas impacted by traffic sources (32-34). The reduction in pollutant concentrations reported during mid-day hours (often half or lower) is likely due to the increase in atmospheric dispersion (increase in mixing height) and in wind speed (34) during this time period, despite the apparent increase in emission rates, and underscores the role of meteorology, in addition to traffic sources, on the observed concentrations of air pollutants even in urban areas impacted by vehicular emissions.

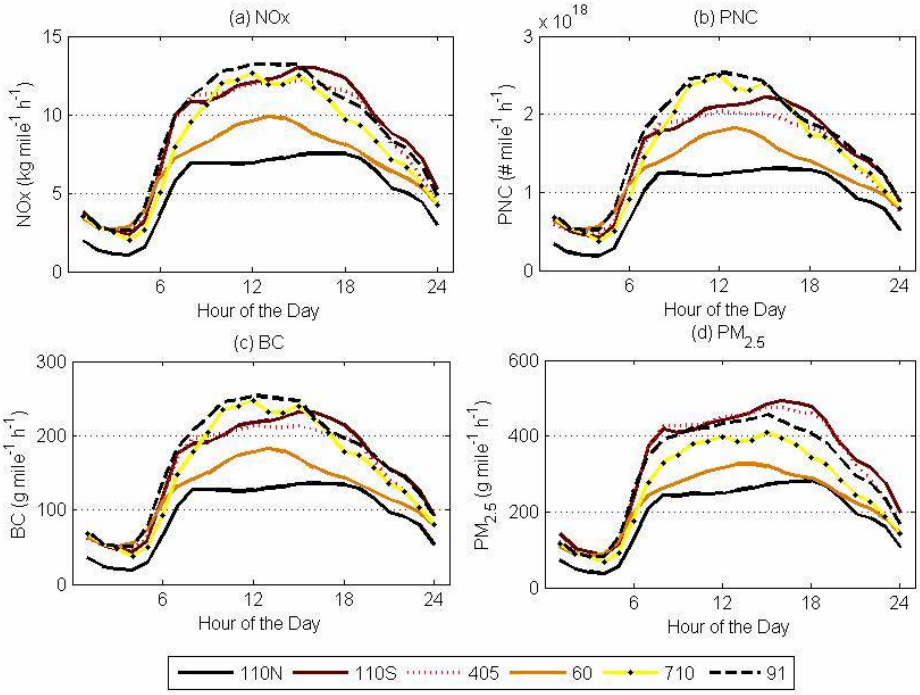


Figure 4.4: Diurnal profiles for freeway emission rates

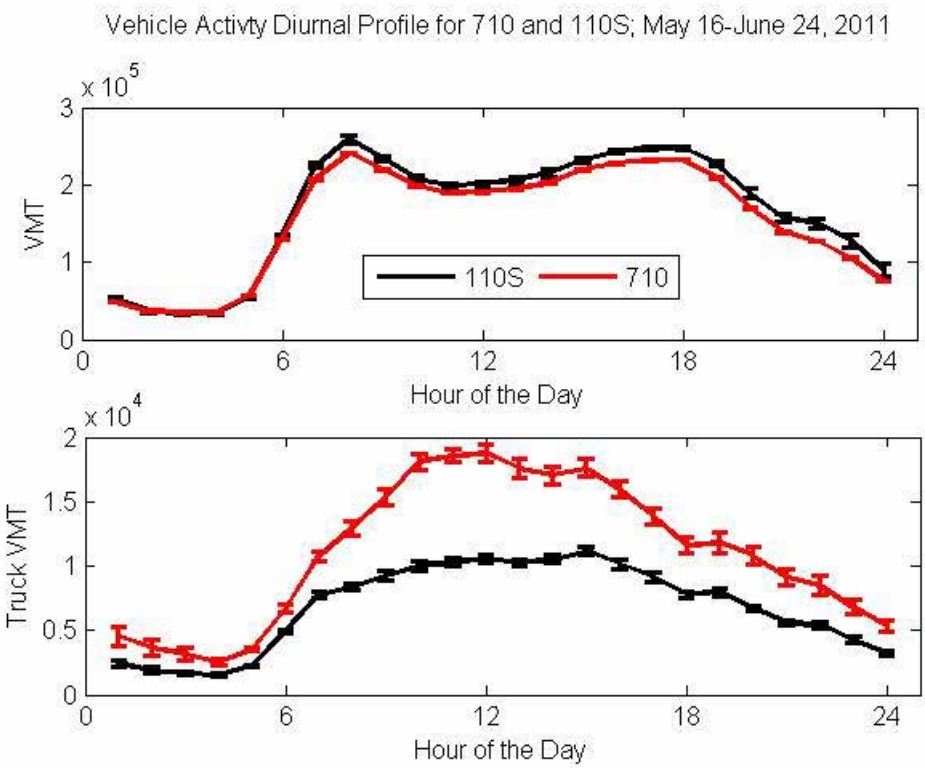


Figure 4.5: Diurnal vehicle activity trend on two Los Angeles freeways, representative of general trend on all freeways.

4.2.4.3 Freeway-to-freeway variability in emission rates

Since freeway emissions are often dominated by HDV fraction, a variation in freeway emissions can be expected due to considerable variation in HDV fractions on Los Angeles freeways and differences VMT (and HDV fraction) at different times of the day. To test the freeway-to-freeway variability in hourly emission rates, Kruskal-Wallis test (nonparametric version of the classical one-way analysis of variance, not requiring normal population distributions or equal variance (35), was conducted for the hypothesis that the mean of hourly emission rates on all freeways was the same. The distributions were considered statistically different if $p < 0.05$. The results are presented in Table 4.5. Generally, emission rates for UFP and BC were not significantly different across most freeways, though greater variation was found in NOx,NO and $PM_{2.5}$ hourly emission rates. Hourly emission rates on freeway I-110N were significantly different (and lower) from all other freeways for all pollutants, except for NOx on CA-91 and $PM_{2.5}$ on CA-60. Emission rates for I-110S, I-405 and I-710 had similar distribution. CA-60 had a different distribution than both 110N and 110S, but was most similar to the I-710.

Table 4.3: P-values for non-parametric ANOVA analysis of freeway-to-freeway differences in hourly emission rates (p-value < 0.05 for freeways having different distribution of hourly emission rates)

Pollutant	NOx						UFP					
Freeway	I-110N	I-110S	I-405	CA-60	I-710	CA-91	I-110N	I-110S	I-405	CA-60	I-710	CA-91
I-110N	1						1					
I-110S	**0.00**	1					**0.00**	1				
I-405	**0.00**	0.59	1				**0.00**	0.45	1			
CA-60	**0.02**	**0.01**	**0.03**	1			**0.02**	**0.04**	0.09	1		
I-710	**0.01**	0.30	0.65	0.18	1		**0.00**	0.80	0.54	0.14	1	
CA-91	0.18	**0.00**	**0.01**	0.36	**0.04**	1	**0.00**	0.32	0.17	**0.02**	0.41	1
Scale	<0.10	<0.2	<0.4	<0.6	<0.80	1.00	<0.10	<0.2	<0.4	<0.6	<0.80	1.00
Pollutant	BC						$PM_{2.5}$					
Freeway	I-110N	I-110S	I-405	CA-60	I-710	CA-91	I-110N	I-110S	I-405	CA-60	I-710	CA-91
I-110N	1						1					
I-110S	**0.00**	1					**0.00**	1				
I-405	**0.00**	0.43	1				**0.00**	0.87	1			
CA-60	**0.03**	**0.02**	0.06	1			0.07	**0.00**	**0.01**	1		
I-710	**0.01**	0.98	0.71	0.14	1		**0.01**	**0.03**	**0.04**	0.11	1	
CA-91	**0.00**	0.48	0.30	**0.01**	0.33	1	**0.00**	0.32	0.29	**0.01**	0.09	1

4.3 SUMMARY AND CONCLUSIONS

Hourly-aggregate vehicular activity on several freeways was related to vehicle emission factors to estimate pollutant emission rates from these freeways (mass/number mile^{-1} h^{-1}) and their hourly distribution. Three parameters drive these distributions – total VMT per mile on the freeway, fraction of VMT due to HDV, and the ratio of LDV to HDV emission factors. It was found that the hourly distributions of emission rates may vary significantly among different freeways. This implies that vehicle fleet mix on the freeway in addition to total vehicle activity should be taken

into consideration as a significant variable that characterizes the emissions of air pollutants from that freeway. Further, VMT as well as truck (or HDV) fraction of VMT is a better surrogate for characterizing emission source strength than total vehicle and or truck counts. This is corroborated by the lack of mid-day reduction in emission rates in contrast to a drop in mid-day traffic counts. The use of traffic counts as a surrogate of emission source strength of a given roadway is therefore not a reliable predictor of the pollutant concentrations in the vicinity of this freeway, and the role of meteorological parameters, such as atmospheric dilution and wind speed, is equally crucial in influencing the values of these concentrations downwind a freeway.

REFERENCES

(1) CARB. 2009 Estimated Annual Average Emissions; California Air Resources Board: Sacramento, 2009.

(2) EPA, Current Emissions Trends Summaries from the NEI, **2011**

(3) CARB. California Motor Vehicle Emission Factor/Emission Inventory Model. *EMFAC V 2.3.* California Air Resources Board, Research Division: Sacramento, **2011**.

(4) Brook RD, Rajagopalan S, Pope CA 3rd, Brook JR, Bhatnagar A, Diez-Roux AV, Holguin F, Hong Y, Luepker RV, Mittleman MA, Peters A, Siscovick D, Smith SC Jr, Whitsel L, Kaufman JD; American Heart Association Council on Epidemiology and Prevention, Council on the Kidney in Cardiovascular Disease, and Council on Nutrition, Physical Activity and Metabolism. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association, *Circulation*. **2010**, 121, 2331-78.

(5) Salam MT, Islam T, Gilliland FD. Recent evidence for adverse effects of residential proximity to traffic sources on asthma, *Curr Opin Pulm Med*. **2008**, 14, 3-8.

(6) Pope, C. A. and Dockery, D.W. Health effects of fine particulate air pollution: Lines that connect, *J. Air Waste Manage. Assoc.* **2006**, 56, 709–742.

(7) Lloyd, A. C. and Cackette, T.A. Diesel engines: Environmental impact and control. *J. Air Waste Manage. Assoc.* **2001**, 51, 809–847.

(8) Sawyer, R. F.; Harley, R.A.; Cadle, S.H.; Norbeck, J.M.; Slott, R. and Bravo, H.A. Mobile sources critical review. *Atmos. Environ.* **2000**, 34, 2161–2181.

(9) Ban-Weiss, G.A.; McLaughlin, J.P.; Harley, R.A.; Lunden, M.M.; Kirchstetter, T.W.; Kean,A.J.;Strawa, A.W.; Stevenson, E.D. and Kendall, G.R. Long-Term Changes in Emissions of Nitrogen Oxides and Particulate Matter from On-Road Gasoline and Diesel Vehicles. *Atmos. Environ*. **2008**, 42, 220-232.

(10) Janssen, N.A.; Hoek, G.; Simic-Lawson, M.; Fischer, P.; van Bree, L.; Ten Brink, H.; Keuken, M.; Atkinson, R.W.; Anderson, H.R.; Brunekreef, B. and Cassee, F.R. Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM10 and PM2.5. *Environ. Health Perspect.* **2011**, 119, 1691-1699.

(11) Li, N.; Wang, M.; Bramble, L.A.; Schmitz, D.A.; Schauer, J.J.; Sioutas, C.; Harkema,

J.R. and Nel, A.E. The Adjuvant Effect of Ambient Particulate Matter Is Closely Reflected by the Particulate Oxidant Potential. *Environ. Health Perspect*. **2009**, 117, 1116-1123.

(12) Delfino, R.J.; Malik, S. and Sioutas, C. Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health. *Environ. Health Perspectives*. **2005**, 113, 934-946.

(13) Morgan, T.E.; Davis, D.A.; Iwata, N.; Tanner, J.A.; Snyder, D.; Ning, Z.; Kam, W.; Hsu, Y.; Winkler, J.W; Chen, J; Petasis, N.A.; Baudry, M.; Sioutas, C. and Finch, C.E. Glutamatergic Neurons in Rodent Models Respond to Nanoscale Particulate Urban Air Pollutants in Vivo and in Vitro. *Environ. Health Perspect.* **2011**, 119, 766-772.

(14) Gan, W.; Tamburic, L.; Davies, H.W.; Demers, P.A.; Koehoorn, M. and Brauer, M. Long-Term Exposure to Traffic-Related Air Pollution and the Risk of Coronary Heart Disease Hospitalization and Mortality. *Environ. Health Persp.* **2011**, 119, 501–507.

(15) Harley, R.A.; Marr, L.C.; Lehner, J.K. and Giddings, S.N. Changes in motor vehicle emissions on diurnal to decadal time scales and effects on atmospheric composition. *Environ. Sci. Technol.* **2005**, 39, 5356–5362.

(16) Fruin, S.; Westerdahl, D.; Sax, T.; Sioutas, C. and Fine, P.M. Measurements and Predictors of On-Road Ultrafine Particle Concentrations and Associated Pollutants in Los Angeles. *Atmos. Environ.* **2008**, 42, 207-219.

(17) Kozawa, K.H.; Fruin, S.A.; Winer, A.M. Near-road air pollution impacts of goods movement in communities adjacent to the Ports of Los Angeles and Long Beach. *Atmos. Environ.* **2009**, 43, 2960-2970.

(18) Ban-Weiss, G.A.; Lunden, M.M.; Kirchstetter, T.W. and Harley, R.A. Size-resolved particle number and volume emission factors for on-road gasoline and diesel motor vehicles. *Journal of Aerosol Science.* **2010**, 41, 5-12.

(19) Bishop, G.A. and Stedman, D.H. A decade of on-road emissions measurements. *Environ. Sci. Technol.* **2008**, 42, 1651–1656.

(20) Bishop, G.A.; Holubowitch, N.E.; Stedman, D.H. Remote measurements of on-road emissions from heavy-duty diesel vehicles in California; Year, 1, 2008. Report, Univ. of Denver, Denver, Colo. **2009**.

(21) San Pedro Bay Ports Clean Air Action Plan, Port of los Angeles and Port of Long Beach, http://www.cleanairactionplan.org/

(22) CARB, ARB's Drayage Truck Regulatory Activities, http://www.arb.ca.gov/msprog/onroad/porttruck/porttruck.htm, 2011

(23) Westerdahl, D.; Fruin, S.; Sax, T.; Fine, P.M. and Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmos. Environ*. **2005**, 39, 3597–3610.

(24) Wang X.; Westerdahl, D.; Wu, Y.; Pan, X. and Zhang, K.M. On-road emission factor distributions of individual diesel vehicles in and around Beijing, China. *Atmos. Environ.* **2011**, 45, 503-513.

(25) Park, S.S.; Kozawa, K.; Fruin, S.; Mara, S.; Hsu, Y.; Jakober, C.; Winer, A. and Herner, J. Emission Factors for High-Emitting Vehicles Based on On-Road Measurements of Individual Vehicle Exhaust with a Mobile Measurement Platform. *J. Air Waste Manage. Assoc.* **2011**, 61, 1046-1056.

(26) California Department of transportation, Annual Average Truck Counts 200.

(27) Kirchstetter, T.W.; Harley, R.A.; Kreisberg, N.M.; Stolzenburg, M.R. and Hering, S.V. On-road measurement of fine particle and nitrogen oxide emissions from light- and heavy-duty motor vehicles. *Atmos. Environ.* **1999**, 33, 2955–2968.

(28) Geller, M.D.; Sardar, S.B.; Phuleria, H.; Fine, P.M. and Sioutas, C. Measurements of particle number and mass concentrations and size distributions in a tunnel environment. *Environ. Sci. Technol.* **2005**, 39, 8653–8663.

(29) Performance E Measuring System (PemS),Caltrans

(30) Kwon J.; Varaiya P. and Skabardonis, A. Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation. *Freeways, High-Occupancy Vehicle Systems, and Traffic Signal Systems.* **2003**, 1856, 106-117.

(31) Fujita, E. M., David E. C., Zielinska B., Arnott W.P., and ChowJ.C., Concentrations of Air Toxics in Motor Vehicle-Dominated Environments, Health Effects Institute, **2011**

(32) Hudda N., Cheung K., Moore K. F., Sioutas C., Inter-community variability in total particle number concentration in eastern Los Angles air basin. *Atmos. Chem. Phys.,* **2010**, 10, 11385-11399

(33) Moore K.F., Krudysz M., Pakbin P., Hudda N., Sioutas C. Intra-Community Variability in Total Particle Number Concentrations in the San Pedro Harbor Area (Los Angeles, California). *Aerosol Sc.Technol.*, **2009**, 43, 587-603

(34) Ntziachristos L., Ning Z., Gellar M., Sioutas C. Particle concentration and characteristics near a major freeway with heavy-duty diesel traffic. *Environ. Sci. Technol.* **2007**, 41, 2223-2230

(35) Gibbons, J. D. Nonparametric Statistical Inference. New York: Marcel Dekker, **1985**.

5. CHAPTER FIVE, PART I. LINKING IN-VEHICLE ULTRAFINE PARTICLE EXPOSURES TO ON-ROAD CONCENTRATIONS
(based on Task 4: Develop and validate in-vehicle exposure models for BC, UFP number, $PM_{2.5}$, particle-bounded PAH, and NO_x.)

Note: this task and chapter has two parts. Part I describes the development of an in-vehicle air exchange and particle loss model using particle number concentrations. Part II describes the on-road concentration models for all pollutants. These two components are both necessary to determine in-vehicle concentrations.

The overarching goal of Task 4 was to predict exposures of human subjects while in vehicles. Therefore, although some important characterization data were used in Tasks 1-3 such as AER, we are not aiming to perform a detailed characterization of the complex dynamics of exposure for all vehicle-related pollutants. However, we laid a strong foundation from which to do this by: 1) establishing the importance of AER in affecting the differences between on-road and in-vehicle concentrations; 2) developing predictive AER models that use easy-to-obtain information that can be made available to epidemiologists; and 3) showing that other particle-related pollutants are also highly dependent on AER. The modeling effort here will include statistical analyses of many serially correlated predictors and will use an approach that incorporates predictors anticipated to be available in exposure models in epidemiologic studies.

5.0 INTRODUCTION

The particular components of traffic emission responsible for causing adverse health effects are not known (1,2), but ultrafine particles (UFP), defined as particles having aerodynamic diameter less than 100 nm, are of particular interest due to their high surface area and the ability to trans-locate through epithelium as well as their high proportion of organic and metals content and resulting high oxidative potential (1,3,4).

Numerous studies (e.g., 5,6) have shown that UFP concentrations on or in the vicinity of roadways are frequently almost one order of magnitude higher than ambient levels. This has important implications for exposure assessment. For example, less than 10% of daily time spent in vehicular transit microenvironments (7) has been estimated to contribute 35-50% of total UFP exposure by Fruin et al. (8) for Los Angeles residents under open window conditions and 17% by Wallace and Ott (9) for more suburban locations. However, large variations in exposure incurred inside vehicles are expected to occur not only due to differences in roadway environments but also because inside-to-outside ratios (i.e., in-vehicle to roadway

concentration ratios) (I/O) vary from vehicle to vehicle due to differences in ventilation conditions and other vehicle characteristics that affect air exchange rate (AER), which is defined as the number of times per hour vehicle cabin air is replaced by roadway/outside air. In general, UFP I/O ratios in vehicles can range from nearly zero to nearly one.

Recent studies have shown that I/O is strongly dependent on AER. Knibbs et al. (10) reported an r^2 of 0.81 (Pearson correlation coefficient) between AER and I/O. This study, as the present, performed measurements under real driving conditions (multiple speed and ventilation conditions) and found that ventilation preference (windows open, outside air intake or in-cabin air recirculation) and ventilation fan setting strongly influences AER and the resulting I/O ratio.

As it is impractical to measure either the I/O ratio or AER for large numbers of subjects' vehicles required in an epidemiological study addressing drive-time exposure, predictive models are needed for estimating AER and I/O ratios. If these models could be based on information that can be collected via questionnaire, they can be useful tools for accurately estimating personal UFP exposures and their associated health effects. The purpose of the part of the study presented here was to measure UFP I/O ratios and AER in a sufficiently large number of vehicles to develop accurate predictive models for assessing drive-time UFP exposure based on easy-to-obtain information.

5.1 METHODS

5.1.1 Vehicle selection and ventilation conditions tested

Vehicles were selected to provide a wide distribution of age and mileage, which are both important factors affecting AER, albeit highly correlated. Measurements were performed in 73 vehicles that were selected from different size categories (sub-compact, compact, mid-size, etc.) in proportions similar to their presence in U.S. fleet. To expand our data and include outside air ventilations test conditions at non-zero speeds, in addition to other settings, we also added data from Knibbs et al. (10,11) measurements in Sydney (Australia) to Los Angeles measurements, where the bulk of the measurements were performed.

Measurements were made with the air conditioning system operating at both ventilation setting options: recirculation (RC), where the in-cabin air is re-circulated, and outside air intake (OA), where outside air is drawn into the vehicle cabin and passed through a filter (if present). Overall, AER was measured in 73 vehicles for 453 unique combinations of vehicle, speed, and ventilation fan setting (308 at RC and 145 at OA setting) and I/O ratio was measured in 43 vehicles for 241 combinations (110 at RC and 131 at OA setting). Table 5.1 list the vehicles tested for this study.

Table 5.1: List of Vehicles tested in the study.

Manufacturer	Model	Year	Odometer	Manufacturer	Model	Year	Odometer
Mazda	Mazda121	1989	98955	Chevrolet	Tahoe	2002	157776
Toyota	Cressida	1990	286872	Ford	E Series	2003	196778
Buick	Sabre	1991	191198	Chevrolet	Silverado	2004	123634
Toyota	Camry	1992	332878	Chevrolet	Cavalier	2004	109509
Buick	Sabre	1992	197576	Chevrolet	Traverse	2004	43235
Cadillac	DeVille	1993	265342	Toyota	Corolla	2005	53234
Honda	Accord	1993	134521	Honda	Accord	2005	106656
Saab	Saab	1994	137600	Chevrolet	Silverado	2005	152896
Chevrolet	Astro van	1995	280046	Mazda	Mazda-6	2005	141003
Jeep	Cherokee	1995	225835	VW	Golf	2005	10663
Jeep	Cherokee	1995	306675	Toyota	Scion xB	2006	70538
Toyota	Corolla	1996	300774	GMC	Sierra	2006	118160
Nissan	Sentra	1996	428982	Toyota	Corolla	2006	139461
Toyota	RAV4	1997	295957	GMC	Colorado	2007	65238
Hyundai	Accent	1997	164968	Honda	Accord	2007	67800
Honda	Civic	1998	137430	Toyota	Yaris	2007	67141
Mitsubishi	Magna	1998	85677	Toyota	Matrix	2007	65501
Honda	Civic	1999	122336	Subaru	Outback	2007	6777
Toyota	Lexus	1999	80528	Chrysler	Crysler300	2008	66955
Ford	Expedition	1999	86061	Toyota	Hilux	2005	6992
Honda	Accord	1999	244470	Toyota	Prius	2009	16713
Ford	Taurus	1999	156619	Toyota	Scion xD	2009	11200
Ford	Contour	1999	183632	Toyota	Scion XB	2009	63894
Honda	Accord	2000	148626	Toyota	Matrix	2009	42078
Subaru	Liberty	2000	58648	Ford	Explorer	2010	1893
Toyota	Corolla	2000	126176	Toyota	Prius	2010	4733
Toyota	Camry	2000	167235	Toyota	Prius	2010	26741
Chevrolet	Tahoe	2001	142445	Toyota	Scion XB	2010	39934
Ford	Escort	2001	202419	Honda	Insight	2010	45771
Nissan	Pathfinder	2001	258848	Honda	Insight	2010	39740
Chevrolet	Cavalier	2001	80184	Honda	Civic	2010	10245

Chevrolet	Express 2500SL	2002	24882	Smart	SmartCar	2010	2339
Nissan	Infinity G35	2002	98219	Toyota	Prius	2010	1893
Toyota	Camry	2002	173610	Honda	Civic	2010	35520
Toyota	RAV4	2002	387434	Hyundai	Elantra	2011	660
Audi	A4	2002	166126				

5.1.2 Speed and routes driven

In the Los Angeles measurements, in order to maintain a steady AER, measurements were made while driving at near constant speeds that ranged from 20- 65 miles h^{-1}. Experiments at speeds up to 35 miles h^{-1} were conducted around the Rose Bowl, Pasadena, a 3.3 mile loop where vehicular traffic was light. Measurements at speeds ranging from 55-70 miles h^{-1} were made on Freeways I-10, CA-60 and I-605 during free flowing traffic conditions. An on-board Global Positioning System (GPS) device (Garmin GPSMAP 76CSC) recorded the location and speed of the car at 1-s intervals.

In Sydney, measurements were performed during trips through a 2.5 mile long road tunnel and on above-ground roads in its vicinity. AER measurements were performed with test vehicles stationary and when driving on open roads at 37 and 68 miles h^{-1}. The measurements are described in detail by Knibbs et al. (11). Each I/O sampling session in a given vehicle involved multiple trips through the tunnel (reported in Knibbs et al. (10) interspersed with above-ground travel. Average vehicle speeds on each segment of the sampling route were calculated based on known distance and time taken, because poor satellite reception impeded GPS based measurements of speed.

5.1.3 Particle concentration measurement, I/O and AER determination

Air exchange rates were determined using Carbon Dioxide (CO_2) as a tracer gas and measurements were made using either TSI Q-Trak model 7565 (TSI Inc., MN, USA) or LI-COR Li-820 units (LI-COR Biosciences, NE, USA) for the vehicles that were tested in Los Angeles. The AER determination procedure is detailed in publications from the present study: Fruin et al. (12) and Hudda et al. (13). Sulfur Hexafluoride (SF_6) was used as a tracer gas for the vehicles tested in Sydney using an Innova type 1412 (Lumasense Technologies, Ballerup, Denmark) photo-acoustic field gas monitor and Innova type 1303 multipoint sampler and doser. Further details for these measurements can be found in Knibbs et al. (11). We (12) have demonstrated very good agreement with predictions reported by Knibbs et al. (11) (Pearson r^2 = 0.83), despite the use of different tracer gases.

Particle measurements at both locations were performed with a TSI Model 3007 Condensation Particle Counter (CPC) with a 50% lower size detection limit of 10 nm. In Los Angeles, in-vehicle to roadway concentration ratios (I/O) were determined as the average value observed for at least 10 minutes of measurement after a stable value had been attained (i.e., a standard deviation less than 5%). To harmonize the I/O measurements across the two study locations, those performed in Sydney (10) were weighted by their respective durations. Data quality assurance comprised regular flow and zero reading checks. In Los Angeles, all instruments used were run simultaneously before and after test runs to check for consistency of response and ambient concentrations. All instruments were synced to within 1 s of the time recorded by GPS. Further details on measurements are available in our publication (13).

5.1.4 Predictive models

Models were developed to predict AER and I/O for UFPs under both RC and OA conditions, using the following as candidate independent variables: ventilation fan (fraction of maximum setting), vehicle age (years), mileage (thousands of miles), speed (miles h^{-1}), manufacturer (United States, Japan or Germany/Other), interior volume (ft^3), and the product of coefficient of drag (C_d) and frontal area (A, m^2) along with pair-wise interactions between vehicle speed, age, and fan setting, and between C_d and frontal area of the vehicle. Ventilation fan fraction was defined as the ratio of the selected fan setting to the total number of options for fan setting. For example, if a vehicle had seven fan setting options and was operated at the third strongest option, the fan setting was set to 3/7(or, = 0.43) in the models. Since AER was positively skewed, natural log transformed AER (lnAER) was used as the outcome in AER prediction models. For I/O ratios, which varied between 0 and 1, a logit transformation (the natural log of [I/O]/[1-(I/O)]) was used, often more appropriate for fractions. Predicted values on the original scales can be recovered using the equations AER = exp(LnAER) and I/O = exp(logitIO)/(1+exp(logitIO)).

Multiple measurements of I/O and AER were performed in each vehicle at different speeds and/or ventilation settings and these repeated measurements were sometimes correlated. This correlation violates the assumption of completely independent observations in multiple linear regression (MLR) models, and MLR models fit to correlated data have unbiased regression coefficients but incorrect standard errors (14). To account for correlated observations, we present results from Generalized Estimating Equations (GEE) models (15) for continuous outcomes, with an exchangeable correlation structure and robust sandwich estimates of regression coefficient standard errors. MLR estimated regression coefficients were similar to those from GEE and are provided for comparison. Model fit was assessed by adjusted R^2 and by leave-one-vehicle-out cross-validated adjusted R^2, which

provides a more reliable estimate of the predictive ability of the same model fit to a new dataset containing information on different cars.

All-subset MLR was used to identify the most important set of predictor variables. From this set, a parsimonious GEE model was developed that included all lower-ordered terms of any interactions or squared variables, had high cross-validated adjusted R^2, statistically significant predictor effects ($\alpha = 0.10$), and satisfied linear model assumptions. For each model, significance of an indicator variable for Sydney data was also evaluated.

5.2 RESULTS AND DISCUSSION

5.2.1 In-vehicle-to-roadway concentration ratios

The I/O ratios measured under RC conditions were far lower than those under OA conditions due to lower AERs under RC (13). The median I/O value at RC was 0.11 (inter-quartile range: 0.07-0.22) compared to 0.66 at OA (inter-quartile range: 0.53-0.80). The median AER value at RC was 6.0 h^{-1} (inter quartile range: 3.6-10 h^{-1}) compared to 63 h^{-1} for OA (inter quartile range: 47-83 h^{-1}). The maximum uncertainty associated with AERs was 7.5% and I/O ratio was 7 % (using root mean square error propagation accounting for both instrument accuracy and stability of continuous measurements for AER measurements and only the stability of continuous measurements for I/O.) Figure 5.1 shows the distributions of AER and I/O results and their transformed values, under both RC and OA ventilation mode. The measurements in Los Angeles and Sydney (10) have been differentiated in the sub-figures.

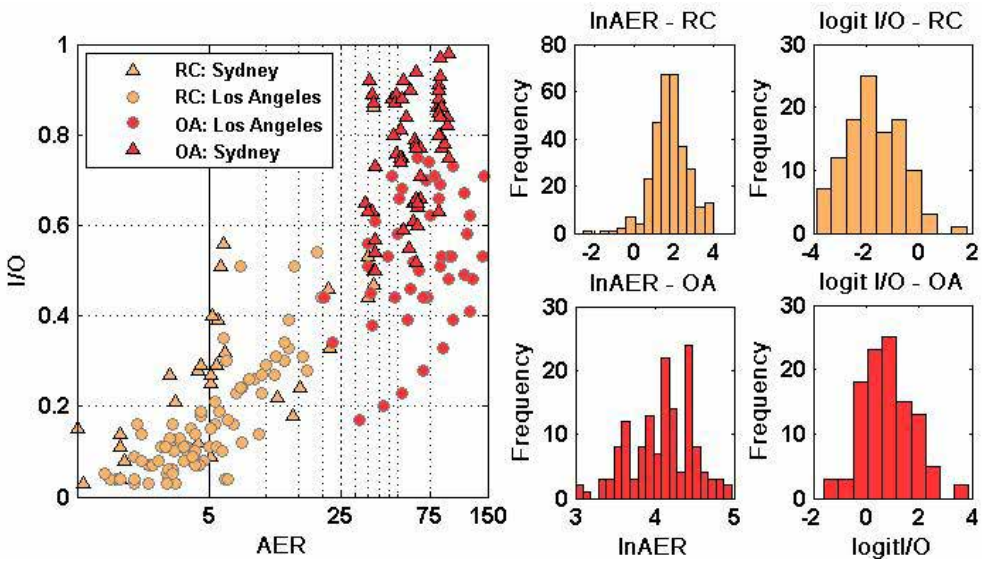


Figure 5.1:Distribution of Dependent Variables.

5.2.2 Predictive model for ln(AER) at RC and OA setting

The GEE model gave the following Equations 5.1 and 5.2 for predicting lnAER under RC and OA condition, respectively.

Equation 5.1: ln(AER) under RC conditions

$$\boldsymbol{ln(AER) = 2.79 + (0.019 \times speed) + [0.015 \times age - 3.3 \times 10^{-3}\, age^2] + [-0.023 \times vol - 6.6 \times 10^{-5}\, vol^2] + Manuf\ Adjustment}$$

where the manufacturer adjustment is -0.71 for German vehicles and -0.39 for Japanese vehicles. If the speed is zero, a -0.51 factor should be added.

Equation 5.2: ln(AER) under OA conditions

$$ln(AER) = 4.20 + [(1.88 \times fan\ strength) + (-0.92 \times fan\ strength^2)] + (0.0048 \times speed) + (-0.0073 \times vol)$$

where the coefficients for fan strength and fan strength^2 should be 0.40 and 0.13, respectively, at zero speed, and the speed term should be -0.32 at zero speed.

GEE and MLR models for predicting AER were able to account for 68 % of the variability in observed AER under RC conditions and 79% under OA conditions. Cross validated R^2 was 0.60 for RC conditions and 0.73 for OA conditions. The GEE confidence intervals can be seen to be roughly a third larger than MLR intervals (Tables 5.1 and 5.2). In all AER model runs, an indicator variable for Sydney data was not significant.

Table 5.1: AER under RC Model Coefficients, Confidence Intervals, and P Values

GEE	Estimate	Std.err	Wald	Pr(>|W|)	Confidence Intervals	
					2.5%	97.5%
Intercept	2.79	0.36	62	4.10E-15	2.1	1.1
speed > 0 (miles h^{-1})	0.019	0.0013	223	< 2e-16	0.017	0.0038
speed = 0	-0.51	0.12	19	1.60E-05	-0.75	0.36
age (yr)	0.015	0.031	0.24	0.62	-0.046	0.092
age^2 (yr^2)	0.0033	0.0017	4.0	0.045	-0.000032	0.0050
vol (ft^3)	-0.023	0.0049	21	4.00E-06	-0.033	0.015
vol^2 (ft^6)	0.000066	0.000015	18	1.90E-05	0.000037	0.000044
Manuf: Japan	-0.39	0.12	11	0.00091	-0.63	0.36
Manuf: Germany	-0.71	0.25	8.1	0.0045	-1.2	0.74
MLR	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Confidence Intervals**	
					2.5%	**97.5%**
Intercept	2.97	0.28	10.75	< 2e-16	2.43	3.52
speed > 0 (miles h^{-1})	0.018	0.0020	8.93	< 2e-16	0.014	0.022
speed = 0	-0.49	0.11	-4.34	2.00E-05	-0.71	-0.27
age (yr)	0.010	0.019	0.53	0.59313	-0.027	0.047
age^2 (yr^2)	0.0039	0.0011	3.67	0.00029	0.0018	0.0060
vol (ft^3)	-0.025	0.0037	-6.77	6.80E-11	-0.032	-0.018
vol^2 (ft^6)	0.000074	0.000013	5.69	3.00E-08	0.000048	0.000099
Manuf: Japan	-0.34	0.070	-4.86	1.90E-06	-0.48	-0.20
Manuf: Germany	-0.88	0.12	-7.26	3.30E-12	-1.11	-0.64

Table 5.2: AER under OA Model Coefficients, Confidence Intervals, and P Values

GEE	Estimate	Std.err	Wald	Pr(>|W|)	Confidence Intervals	
					2.5%	**97.5%**
Intercept	4.2	0.24	295	< 2e-16	3.7	0.7
fan strength	1.88	0.14	170	< 2e-16	1.6	0.43
fan strength^2	-0.92	0.11	70	< 2e-16	-1.1	0.33
speed = 0	-0.32	0.09	11	0.0007	-0.50	0.28
speed > 0 (miles h^{-1})	0.0048	0.0013	14	0.0002	0.0023	0.0038
vol (ft^3)	-0.0073	0.0019	15	0.0001	-0.011	0.0056
fan strength at speed = 0	0.40	0.26	2.5	0.12	-0.10	0.76
fan strength^2 at speed = 0	0.13	0.20	0.45	0.50	-0.26	0.59
MLR	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	**Confidence Intervals**	
					2.5%	**97.5%**
Intercept	4.2	0.14	29	< 2e-16	3.9	4.5
fan strength	2.3	0.40	5.7	0	1.5	3.1
fan strength^2	-1.3	0.36	-3.7	0.00040	-2.1	-0.61
speed = 0	-0.34	0.15	-2.3	0.023	-0.63	-0.047
speed > 0 (miles h^{-1})	0.0043	0.0014	3.0	0.0028	0.0015	0.0071
vol (ft^3)	-0.0074	0.00080	-8.8	0	-0.0090	-0.0057
fan strength at speed = 0	0.077	0.56	0.14	0.89	-1.0	1.2
fan strength^2 at speed = 0	0.45	0.49	0.91	0.36	-0.5	1.4

The predicted AER under RC conditions is plotted against the two most significant determinants of AER, speed and age, in Figure 5.2 (a). Model results suggest that an 11 year old vehicle (~ 75th age percentile) has an AER that is about 1.5 times higher than that of a 4 year old vehicle (~25th age percentile). Furthermore, AER during typical freeway driving speed (65 miles h^{-1}) is expected to be 1.8 times higher than under typical arterial driving conditions (35 miles h^{-1}).

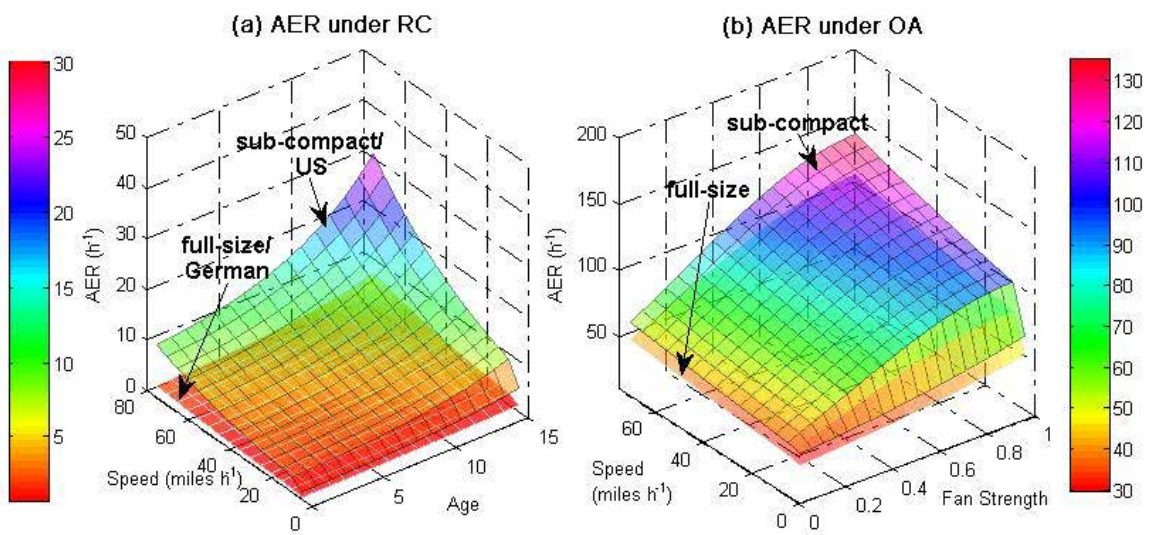


Figure 5.2: Predicted values for lnAER plotted against the two most significant variables under RC and OA ventilation modes

The two surfaces plotted in 2 (a) represent the extremes of other model inputs under RC conditions: a U.S. manufactured sub-compact vehicle (85 ft^3 cabin) and a German manufactured large vehicle (120 ft^3 cabin), the range of AER variation that can be expected due to manufacturer and volume. A U.S. manufacturer's vehicle is expected to have an AER nearly 50% higher than a Japanese vehicle and about twice as high as a German manufactured vehicle, for given cabin volume, age and speed. It is interesting that cabin volume was found to be negatively correlated with AER when expressed in units of air changes per hour under RC conditions (Similar relationship was found with I/O, discussed in following section). For example, an 85 ft^3 vehicle is expected to have an AER, which is 2.2 times that in a 120 ft^3 vehicle (or 1.6 times higher if AER units are $ft^3\ h^{-1}$). To provide a typical AER value under RC conditions for reference, a seven-year-old vehicle (50th age percentile, U.S. manufactured, and 110 ft^3 , the average U.S. fleet cabin volume) would have an AER of 3.7 h^{-1} at 35 miles h^{-1} and 6.7 h^{-1} at 65 miles h^{-1}.

Under OA conditions, fan strength explained the most variability in lnAER, followed by speed. For example, increasing the fan setting from low (0.25) to medium (0.5) to highest (1.0) increased the AER by a factor of 1.3 and 1.7, respectively. In comparison, increasing the driving speed, the second most significant variable, from arterial to freeway speeds only increased the AER by 1.2. Vehicle cabin volume was also found to be significant, with higher volume vehicles having lower predicted AER (h^{-1}). An 85 ft^3 sub-compact vehicle had 1.3 times higher AER than a 120 ft^3 large sedan. Fan strength and zero speed interaction terms, while not significant individually, were significant as a pair, so included in the OA model.

Figure 5.2 (b) above shows the model predictions plotted against the two most significant determinants of AER under OA conditions; ventilation fan strength and vehicle speed, for a sub-compact (85 ft^3) and large sedan vehicle (120 ft^3), thus capturing the full range of AERs that can be expected under OA condition. For the previously mentioned reference vehicle travelling at 35 and 65 miles h^{-1}, AER would be 72 h^{-1} and 83 h^{-1}, respectively, at the middle fan setting, roughly an order of magnitude higher than under RC conditions.

5.2.3 Predictive model for logit(I/O) under RC and OA setting

I/O UFP number concentration ratios under both ventilation conditions were modeled together, using a binary indicator variable for RC setting, (i.e., variable RC = 1 under RC setting and zero otherwise). The resultant Equations 5.3 and 5.4 from the GEE model for predicting I/O under RC and OA conditions are as follows:

Equation 5.3: Logit(I/O) under RC conditions

$$\text{logit}(I/O) = -(0.29 + 2.93) + 0.54 \times fan\ strength + 0.025 \times speed + (0.017 + 0.086) \times age$$

Equation 5.4: Logit(I/O) under OA conditions

$$\text{logit}(I/O) = -0.29 + 0.54 \times fan\ strength + 0.025 \times speed + 0.017 \times age$$

The GEE coefficients, confidence intervals and p values are listed in Table 5.3. GEE models required only vehicle age, speed and ventilation fan strength to account 79 % of the variability in I/O across RC and OA conditions. Cross validated R^2 was 0.76. The indicator variable for Sydney was significant for the I/O modeling, but did not appreciably change the results. Therefore, this variable was omitted in the final model since that factor would not be useful to other users of the model.

Figure 5.3 (a) shows the full range of I/O that can be expected in vehicles up to 20 years old and travelling at speeds up to 75 miles h^{-1}, age and speed being the most important predictors. Under RC ventilation conditions, I/O can be expected to vary from less than 0.1 to nearly 0.8 in the leakiest cars (old and travelling at high speeds). The two surfaces mark the upper (full fan) and lower limits (low fan setting, equal to 0.33) of variation that can be expected due to the third most significant variable, fan strength. Under RC conditions, fan setting was relatively unimportant. For the entire range of age/speed plotted in Figure 5.3 (a), fan strength made an average difference of only 0.07 ± 0.02 in I/O ratio.

Table 5.3: I/O GEE Model Coefficients, confidence intervals and p-values.

GEE Model						
	Estimate	**Std.err**	**Wald**	**Pr(>|W|)**	**Confidence Intervals**	
					2.5%	**97.5%**
Intercept	-0.29	0.19	2.4	1.20E-01	-0.65	0.078
fan strength	0.54	0.21	6.8	9.10E-03	0.14	0.95
speed (miles h^{-1})	0.025	0.0028	81	< 2E-16	0.019	0.030
age (yr)	0.017	0.02	0.84	3.60E-01	-0.020	0.055
RC	-2.95	0.14	468	< 2E-16	-3.2	-2.7
RC X age (yr)	0.086	0.019	20	7.00E-06	0.048	0.12
MLR Model						
	Estimate	**Std.err**	**Wald**	**Pr(>|W|)**	**Confidence Intervals**	
					2.5%	**97.5%**
Intercept	-0.31	0.14	-2.2	3.30E-02	-0.59	-0.026
fan strength	0.30	0.17	1.8	7.40E-02	-0.029	0.62
speed (miles h^{-1})	0.033	0.0027	12	< 2e-16	0.028	0.039
age (yr)	0.042	0.011	3.9	1.30E-04	0.021	0.063
RC	-2.93	0.13	-22	< 2e-16	-3.19	-2.67
RC X age (yr)	0.06	0.02	3.9	1.30E-04	0.031	0.093

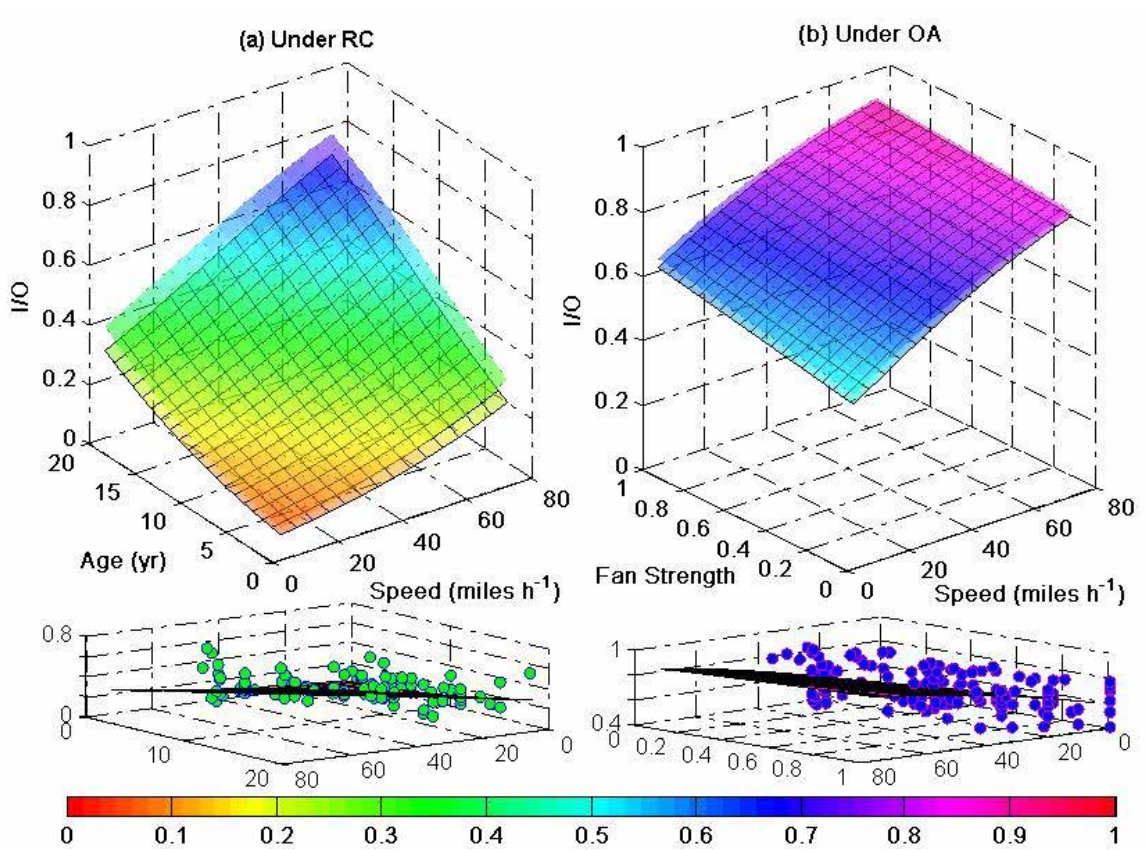


Figure 5.3:Predicted values for I/O under RC and OA ventilation mode versus two most important model variables for each mode. Bottom subsets show actual measurements versus surface of median model predictions.

In contrast, under OA conditions, I/O ratios were most strongly dependent on vehicle speed and fan speed. I/O ratios were higher but had a smaller range compared to RC conditions, varying from 0.5 – 0.9 (Figure 5.3 (b)). The plotted surfaces in Figure 5.3 (b) mark the lower (25th age percentile) and upper bounds (75th age percentile) of predicted I/O due to variation in the third variable, vehicle age, though the distinction is barely discernible. Age (by itself) under OA was not significant and made a maximum difference of 0.03 in I/O ratios predicted using Equation 4. The lower subset figures show measured I/O ratios plotted along with median predicted surface to show modeled data fit.

We performed a sensitivity analysis based on maximum expected variable measurement uncertainty: 5 miles h^{-1} uncertainty in speed, 1 year in vehicle age, and a 10% uncertainty in fan speed based on fraction of maximum. These uncertainties led to 8.0, 5.6, and 3.4% difference in I/O ratio, respectively. These relatively modest changes reflect maximum uncertainties. Typical uncertainties would tend to be smaller, and since independent, would tend to cancel each other out (i.e., they are just as likely to be positive as negative). However, the model

predictions may be less accurate for vehicles older than 15 years and at speeds exceeding 60 miles h^{-1} due to limited coverage of measured data for such conditions.

We also assessed the sensitivity of our findings regarding the generalizability of our observed model fit by implementing a modified version of approximately 10-fold cross-validation. For each model, we first determined the number of cars contributing data to the estimation of the model (AER under RC conditions: 308 measurements from 70 cars, AER under OA conditions: 144 measurements from 16 cars, and I/O ratios under RC and OA conditions: 241 measurements from 42 cars). For the AER under RC conditions model, we randomly partitioned the data into 10 subsets of data from 7 cars, and then successively fit the model to 9 of 10 subsets and used this model to predict the outcome in the 10th hold-out subset of 7 cars. Finally, the adjusted R^2 was calculated using these predictions. Since the partitioning into 10 subsets of 7 cars was not unique, we performed the modified 10-fold cross-validation 10 times to get a sense of the variability in the estimate of the modified 10-fold cross-validation adjusted R^2. We report the minimum, mean and max of these 10 values below. Since the number of cars in the other two models was not divisible by 10, we chose to use a modified 8-fold and 7-fold cross-validation in the AER under OA and I/O ratios under OA and RC condition models, respectively.

The leave-one-car-out value was similar to the mean value from the approximately 10-fold approach. In the adjusted R^2 calculated using all data is somewhat larger than we would expect in a new dataset because the leave-one-car out cross-validated value of adjusted R^2 was smaller than the value calculated using all data. However, the adjusted R^2 we would expect in a new dataset is still good, providing evidence that our models did not greatly overfit to the observed data.

Log(AER) under RC conditions**:**
Adjusted R2 from GEE models:

All data	Leave 1 car out CV	10 fold CV - Min	10 fold CV - Mean	10 fold CV - Max
0.6773	0.6039	0.5700	0.593	0.628

Log(AER) under OA conditions:
Adjusted R2 from GEE models:

All data	Leave 1 car out CV	8 fold CV - Min	8 fold CV - Mean	8 fold CV - Max
0.792	0.729	0.694	0.723	0.747

Logit(IO) combined model for OA
Adjusted R2 from GEE models:

All data	Leave 1 car out CV	7 fold CV - Min	7 fold CV - Mean	7 fold CV - Max
0.789	0.762	0.754	0.764	0.774

5.2.4. Fleet-wide distributions of AER and I/O

To calculate individual in-vehicle UFP exposures, the models presented in previous sections for prediction of AER and I/O require six inputs: (a) ventilation setting; (b) fan setting; (c) manufacturer; (d) vehicle age; (e) speed; and (f) vehicle volume. For conducting a large epidemiological study, these variables can be gathered directly through a questionnaire or generated from vehicle-related information like age, vehicle identification number (which holds information on model, make and manufacturer), and driving/trip related information like ventilation setting choice, fan setting, and trip time and destination.

To calculate population-size distributions of in-vehicle UFP exposure, the distribution of predicted AER and I/O ratios in a fleet of vehicles can be computed if required input variable distributions for the fleet are known. As an example, probability distributions for AER and I/O (predicted using Equations 5.1, 5.2, 5.3 and 5.4) were computed for a fleet of sedan type vehicles using input distributions based on the U.S. fleet.

Vehicles were divided into three categories based on average cabin volume for three size categories: compact (99 ft^3), mid-size (112 ft^3) or full-size (135 ft^3). The frequency of each size was determined from the fraction of passenger cars in each volume category for the years 1990-2010 (16). For fan setting, it was arbitrarily assumed that an equal fraction of vehicles were being driven at three fan settings, low (fan setting = 0.33), medium (0.67) and highest (1.0). The current fractions of manufacturer share were used: (44.5% U.S., 42% Japanese and 13.5% German/other). For age and speed, EPA Motor Vehicle Emission Simulator (MOVES) default age and speed inputs for gasoline passenger cars were used (17). Different speed distributions were used for arterial and collector roads and freeway and its ramps, which also varied with respect to time of day (i.e., rush hour or non-rush hour). The weighted average speeds were 33 and 37 miles h^{-1}, respectively, for arterial roads during rush hours and non-rush hours, and 45 and 55 miles h^{-1}, respectively, for freeways during rush and non-rush hours. Further details on all these input distributions have been presented in the Figures 5.4-5.6 and Table 5.5

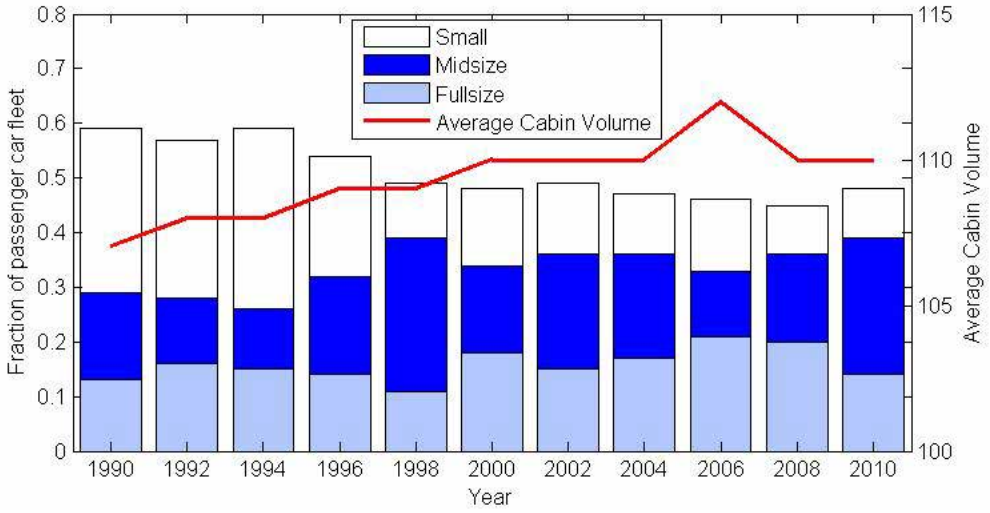


Figure 5.4: Distribution of passenger cars in various volume-based size classes

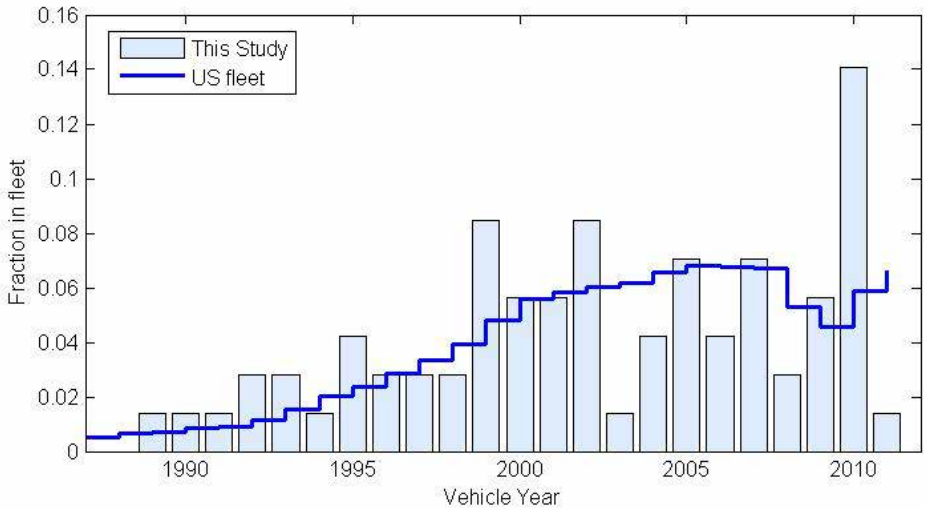


Figure 5.5: Age distribution of US passenger car fleet and comparison with study fleet

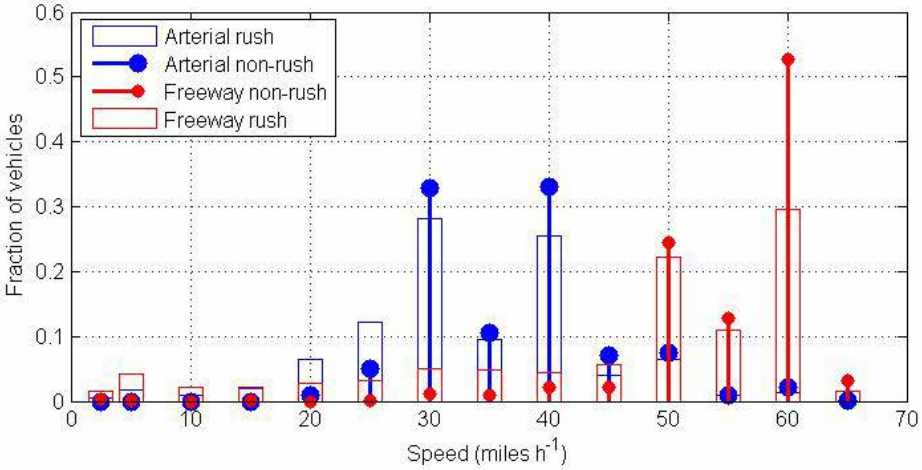


Figure 5.6: Typical distribution of driving speed on urban freeways and arterial roads during congested or peak and not congested or off-peak traffic conditions.

Table 5.4: Manufacturer share of the vehicles in operation in U.S.

Manufacturer	% of vehicle currently in operation	Manufacturer	% of vehicle currently in operation
Ford	20	Hyundai	2.9
GM	28.5	Honda	8.1
Chrysler	12.8	Nissan	5.3
Toyota	11.8	Other	10.6

The fractions of vehicles having a specific AER or I/O are plotted in Figure 5.7 for both ventilation choices (RC and OA). Several important observations can be made from Figure 5.7. First and foremost, though roadway type and associated speed differences affect AER and I/O, the most significant difference occurs due to ventilation setting choice. Under RC conditions, 80% of the fleet is expected to have I/O ratio between 0.15 and 0.5—significant protection—under all road types and speeds, but for OA conditions, 80% all vehicles are expected to have I/O ratios from 0.65 to 0.85, only moderately reduced concentrations. Looked at another way, under RC conditions, the fraction of vehicle fleet that will experience cabin concentrations lower than half of on-road concentrations exceeds 80%, but virtually none of the fleet is expected to have I/O ratios less than 0.5 under OA conditions. Furthermore, the difference between rush hour and non-rush hour speed distributions leads to a far

more significant difference in AER and I/O distribution for freeway driving than arterial driving.

5.2.5 Expected in-cabin concentrations for given roadway concentrations

The ultimate goal of generating predictive models for I/O is to be able to predict in-cabin concentrations from roadway concentrations (calculated as $\text{Concentration}_{\text{in-cabin}}$ = I/O x $\text{Concentration}_{\text{roadway}}$). To illustrate, representative probability distributions of UFP concentrations were generated from 10 hours of sampling on arterial and 12 hours of sampling on Los Angeles freeways and are shown in Figure 5.8. In turn, these distributions were joined in a Monte Carlo-type sampling method with the I/O distributions in Figure 5.7 to generate distributions of UFP concentrations inside the U.S. vehicle fleet if driven on Los Angeles roads.

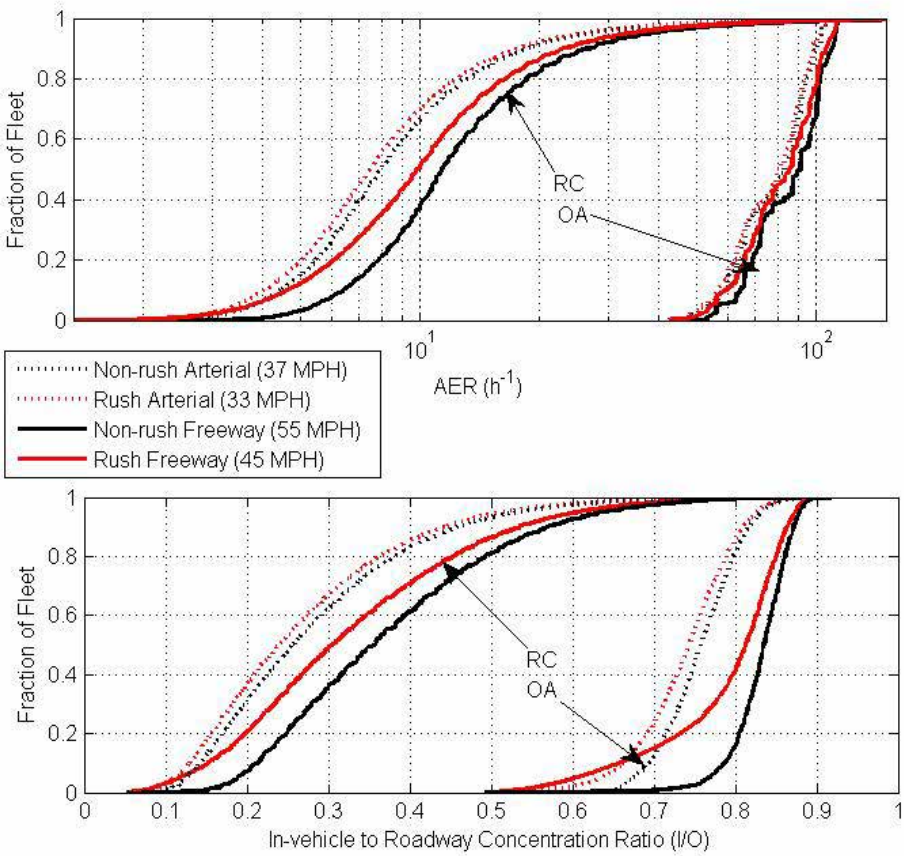


Figure 5.7:Distribution for AER and I/O for a fleet similar to U.S. passenger car fleet in terms of manufacturer's market share, vehicle volume and age.

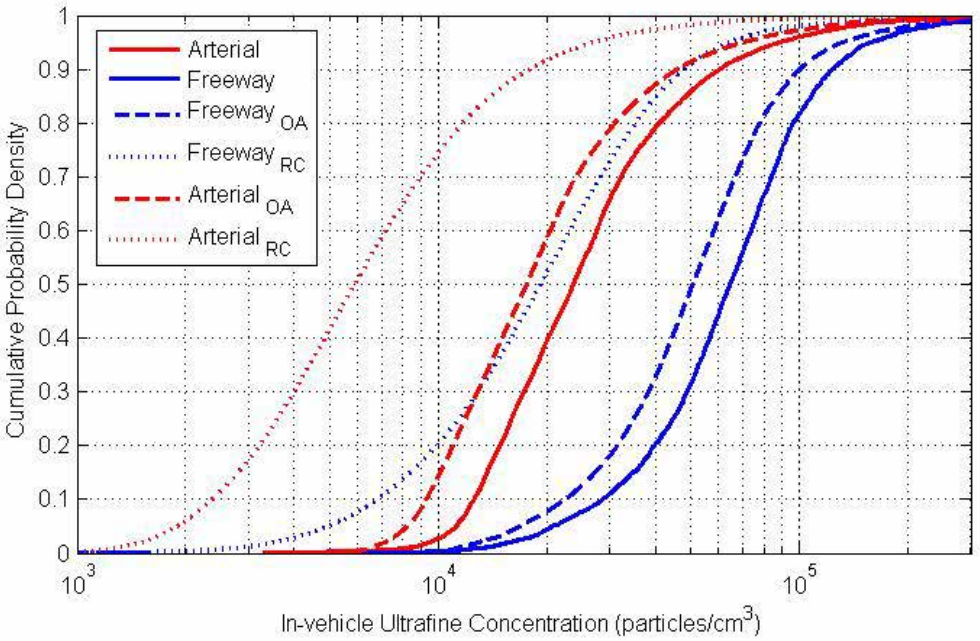


Figure 5.8:Expected in-cabin concentration for U.S. vehicle fleet travelling on Los Angeles arterial roads and freeway

Comparison of the measured roadway concentrations and the predicted in-cabin concentrations under RC and OA conditions shown in Figure 5.8 suggests that for the range of fleet vehicle characteristics such as age and mileage (e.g., 25th to 75th percentile differences for a ventilation setting and road type), we would expect a two to three-fold range in in-vehicle UFP exposures, while the differences due to ventilation mode selection alone for a given vehicle on either road type was larger, with factors ranging from two to four. The increase in speed going from arterial to freeway speeds, however, along with increase in on-road concentrations on freeways, only increased in-vehicle UFP exposure for a given vehicle at either ventilation mode by a factor of 1.5. Overall, while ventilation choice is still the dominant factor, it is interesting that once the variability of on-road UFP concentrations are taken into account, the spread of in-vehicle UFP concentrations between RC and OA conditions overlaps, unlike the spread of I/O distributions.

5.3. SUMMARY AND CONCLUSIONS

Models have been presented for predicting UFP in-vehicle to roadway concentration ratios (I/O) based on simple driving preferences and vehicle characteristics. Scalability of these models was demonstrated at a fleet-wide level and in dynamic roadway environments. In general, factors that increase air exchange rates (AER) increase UFP I/O. Age was significant and positively correlated with both AER and I/O under recirculation ventilation setting (RC), but age was not significant under fresh air intake setting (OA). Under OA conditions, fan strength was also a strong determinant and positively correlated with I/O ratio. Under both ventilation settings, an increase in vehicle volume decreased I/O. Overall, combining these results with

on road UFP concentration distributions measured on Los Angeles roadways, in-cabin UFP exposure concentrations during freeway driving were up to 1.5 times that of arterial driving, but the switch from OA ventilation condition to RC dropped the in-vehicle concentration on either road type two to four fold.

This chapter presented results that were the capstone of the work presented in Chapters 2 and 3. Accurate assessment of risk posed by ultrafine exposure will depend on the ability to characterize exposure in microenvironments like that in-vehicle where peak and disproportionate exposures occur frequently for large sections of the population. A novel contribution of this study is the empirical model for predicting AER and UFP I/O ratios. This work makes it possible for exposure scientists or epidemiologists to predict drive-time UFP exposure based on estimated on-road concentrations and the collection of survey data regarding the characteristics of vehicles used by the study population. The use of these prediction models would bypass the reliance on time-consuming measurements and mechanistic models (such as those described above), which are unfeasible in an epidemiologic study, and makes it possible to generate predictions for large cohorts based on simple information such as vehicle make, age and ventilation choice.

These results of the present chapter were be used in work presented in Chapter 5 – Part II to develop predictive models for exposure.

REFERENCES

1. Delfino R.J., Malik S. and Sioutas C., 2005. Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health. Environ Health Perspectives 113(8), 934-946
2. Sioutas C., Delfino R.J., Singh M., 2005. Exposure assessment for atmospheric ultrafine particles (UFP) and implications in epidemiological research. Environ Health Perspect 113(8), 947-955
3. Li, N., Wang, M., Bramble, L.A., Schmitz, D.A., Schauer, J.J., Sioutas, C., Harkema, J.R. and Nel, A.E., 2009. The Adjuvant Effect of Ambient Particulate Matter Is Closely Reflected by the Particulate Oxidant Potential. Environ. Health Perspect 117, 1116-1123.
4. Morgan T.E., Davis D.A., Iwata N., Tanner J.A., Snyder D., Ning Z., et al., 2011. Glutamatergic Neurons in Rodent Models Respond to Nanoscale Particulate Urban Air Pollutants in Vivo and in Vitro. Environ Health Perspect 119(7), 766-772.
5. Leung P.L. and Harrison R.M., 1998. Evaluation of personal exposure to monoaromatic hydrocarbons, Occup. Environ. Med. 55, 249–257.
6. Westerdahl D., Fruin S., Sax T., Fine P.M., and Sioutas C., 2005. Mobile platform measurements of ultrafine particles and associated pollutant

concentrations on freeways and residential streets in Los Angeles. Atmos Environ 39, 3597–3610.
7. Klepeis, N. E., Nelson, W. C., Ott , W. R., Robinson, J. P., Tsang, A. M., Switzer, P., et al., 2001. The national human activity pattern survey (NHAPS): A resource for assessing exposure to environmental pollutants. J Expos Anal Environ Epidemiol. 11(3), 231-252.
8. Fruin S., Westerdahl D., Sax T., Sioutas C., and Fine P.M., 2008. Measurements and predictors of on-road ultrafine particle concentrations and associated pollutants in Los Angeles. Atmos Environ 42, 207–219.
9. Wallace L., and Ott W. Personal Exposure to Ultrafine Particles., 2011. J Expo Sci Environ Epidemiol 21, 20–30.
10. Knibbs, L.D., de Dear, R.J. and Morawska, L., 2010. Effect of cabin ventilation rate on ultrafine particle exposure inside automobiles. Environ. Sci. Technol. 44, 3546-3551.
11. Knibbs, L. D., de Dear, R. J. and Atkinson, S. E., 2009. Field study of air change and flow rate in six automobiles. Indoor Air 19(4), 303-313.
12. Fruin S.A., Hudda N., Sioutas C., Delfino R., 2011 A Predictive Model for Vehicle Air Exchange Rates based on a Large & Representative Sample, Environ. Sci. Technol. 45, 3569-3575.
13. Hudda N., Kostenidou E., Delfino R., Sioutas C. and Fruin S., 2011 Factors that Determine Ultrafine Particle Exposure in Vehicles. Environ. Sci. Technol. 45, 8691-8697.
14. Diggle, Peter J.; Patrick Heagerty, Kung-Yee Liang, Scott L. Zeger (2002). Analysis of Longitudinal Data. Oxford Statistical Science Series.
15. Liang K. and Zeger S., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13-22
16. Light-Duty Automotive Technology, Cabon Dioxide Emissions. and Fuel Economy Trends: 1975-2010, Office of Transportation and Air Quality, U.S. Environmental Protection Agency, 2010.
17. MOtor Vehicle Emissions Simulator (MOVES). Office of Transportation and Air Quality, U.S. Environmental Protection Agency. http://www.epa.gov/otaq/models/moves/index.htm

CHAPTER FIVE, PART II. DEVELOP AND VALIDATE THE ON-ROAD EXPOSURE MODELS FOR PARTICLE-BOUNDED PAH, PNC, $PM_{2.5}$, NO_X, AND BC (based on Task 4: Develop and validate in-vehicle exposure models for BC, UFP number, $PM_{2.5}$, particle-bounded PAH, and NO_x.)

5.4. INTRODUCTION

In the following work we used measurements of on-road air pollutants acquired by USC during work to accomplish the above Tasks 1-3. Models to predict in-vehicle UFP particle number concentration (PNC) (also part of Task 4) were presented above (Part I of Chapter Five). Those results and the present results were used to help develop the predictive model for Task 5 in the next section for in-vehicle personal PAH collected by study subjects. No other in-vehicle data were collected for BC, NO_x or $PM_{2.5}$ and therefore, in-vehicle models cannot be developed for these pollutants. Due to these limits in the data collected, we focus herein on the development of predictive models for on-road BC, PAH, PNC, $PM_{2.5}$ and NO_x. The purpose is to provide on-road models to predict air pollutants that could be combined with the models developed in Tasks 1-2 for AER to then predict in-vehicle exposures in human subjects. No models were developed for CO and CO_2 since they are generally not of primary interest in epidemiologic research. Although CO can be a marker of fossil fuel combustion, the measurements selected are believed to be more directly representative of pollutant components involved in oxidative stress, inflammation and damage to macromolecules and other cell constituents.

The primary goal of Tasks 4-5 is to predict exposures of human subjects while in vehicles. We start this effort here with the prediction of on-road concentrations. Characterization data useful in predicting in-vehicle exposure to PN was presented in Tasks 1-3, such as AER for a given OA and RC setting (OA refers to the time when the car's ventilation system injects fresh Outside Air into the vehicle; RC refers to the time when the car's ventilation system ReCirculates the air with the vehicle cabin). In Tasks 4-5 we are not aiming to perform a detailed characterization of the complex dynamics of exposure. This is because some variables used in the above characterization (e.g., real-time OA and RC) would typically not be available in an epidemiologic study.

Human exposure prediction data would be averaged over long periods as compared with the real-time nature of the characterization data. Nevertheless, given the amount of data available and on model fit, we chose one-minute average air pollutant concentration and also tested its temporal autocorrelation. The modeling effort included statistical analyses of many serially-correlated predictors and used an approach that incorporates predictors of on-road pollutant concentrations anticipated to be available for exposure models in epidemiologic studies. For in-vehicle human exposures, this on-road data would be combined with the detailed characterization data already discussed (vehicle type and age, AER based on OA and RC condition,

etc.). An initial limited effort in Task 5 makes this union of on-road exposure predictors with subject-reported in-vehicle conditions based in part on findings using Task 1-3 characterization data.

For the development and validation of the predictive models for on-road pollutants, we first conducted an exploratory data analysis to obtain a normal transform function for the air pollutant data, and then performed initial screening of the predictor variables and the linear or non-linear relationship between them and the on-road air pollutant concentrations. Finally, we built linear regression models and non-linear generalized additive models to explore the influence of predictor variables and modeling methods. We also constructed the models with incorporation of temporal autocorrelation. Using cross validations and independent holdout tests, we tested the effects of predictor variables, modeling methods and temporal autocorrelation for the prediction of on-road air pollutant concentrations.

5.5. MATERIALS

The study domain mainly covered most of several major freeways, partial arterials and local roads (Figure 5.9).

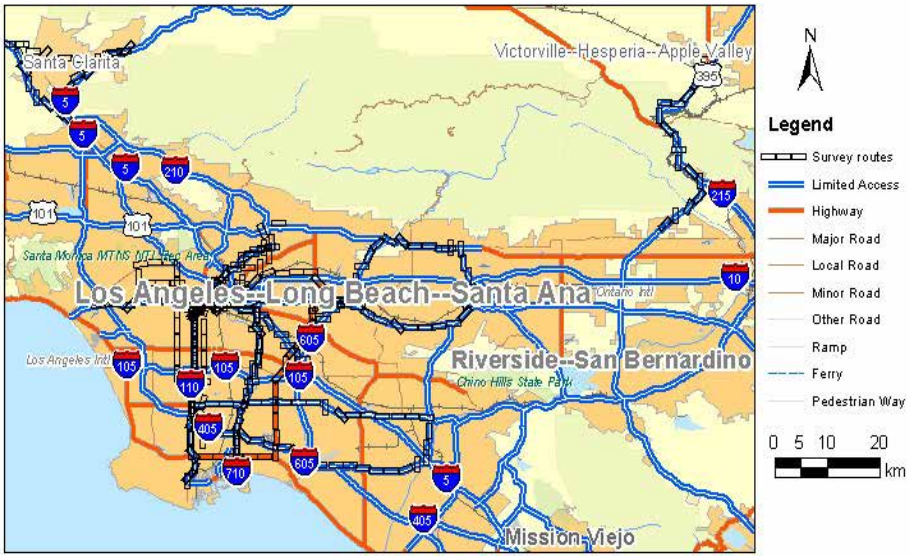


Figure 5.9: Routes of on-road pollutant measurements from Task 4

5.5.1 Mobile Measurement Platform and Concentrations Measured

Vehicle information such as vehicle type, age, model, production and speed are important for in-vehicle concentrations. However, our concentrations were measured on-road using one 2010 Honda Insight hybrid vehicle. This was selected as a mobile measurement platform due to its limited level of self-pollution, which is more relevant for in-vehicle air pollutant concentrations.

On-road pollutant measurements were obtained from the USC field work for five air pollutants (PAH, PNC, $PM_{2.5}$, NO_x and BC). We averaged the original data at 10-second intervals to one-minute averages for model development. Figure 5.9 represents the routes for USC's measurements of pollutant concentrations. Most measurements were performed on freeways (over 60% of the data).

5.5.2 Road and Traffic Classification

We compiled a comprehensive traffic database for freeways in the study region based on both 5 minute total traffic measurements and estimated truck counts. Five-minute total traffic counts and estimated hourly truck counts were based on aggregated count and occupancy data for freeways and highways during the study period and were obtained from the California Department of Transportation (Caltrans) Performance Measurement System (PeMS) (http://pems.dot.ca.gov/). The PeMS data provided a high temporal resolution; however, they were limited by spatial coverage (mostly freeways and highways), limited sampling sites, and sometimes missing data probably due to malfunction of detectors.

Methods were developed in Geographical Information System (GIS) using ArcGIS v10.0 (ESRI, Redland, CA) and PostGIS v1.5 (Refractions Research, British Columbia, Canada) software to assign traffic volume data to roadway segments, map GPS-based on-road measurement data, and classify GPS data into different categories based on roadway type and traffic volumes. Five-minute total traffic and hourly estimated truck counts at Caltrans sampling locations were linked and assigned to adjacent roadway segments (within 5 km along the roadway) with matching names. Although we extended the PeMS point measurements to 1 km along the roadway, corresponding to the measurement time of on-road concentrations, the PeMS data only covered 66.5% (in length) of the freeways and highways of the routes in this task.

To overcome the limitations in the PeMS data, we also obtained the annual average daily traffic (AADT) count data from the Caltrans, which had continuous coverage for all freeway/highway and major arterial segments. The AADT dataset was produced by Caltrans staff based on a combination of measurements (e.g. continuous measurements on freeways and highways and tri-annual measurements on surface streets) and modeled values.

5.5.3 Meteorological Parameters

We obtained meteorological data from both on-road measurements and from the nearest weather monitoring stations operated by National Weather Service and South Coast Air Management District. The on-road measurements of temperature (dew point and wet bulb) and humidity were collected simultaneously with pollutant measures and at the same temporal resolution (every ten seconds). Hourly temperature, relative humidity, and wind (wind speed and direction) were obtained

from meteorological monitoring stations. The shortest distance was used to assign the monitoring station data to each GPS point. In addition to the surface meteorological data, we obtained Pasquill atmospheric stability class data every three hours at approximately 40 km by 40 km spatial resolution during each sampling period from the National Oceanic and Atmospheric Administration (NOAA) AIR Resources Laboratory archive of the Eta 4-D Data Assimilation System (EDAS) (http://www.arl.noaa.gov/ready.html). For each GPS point, we assigned the atmospheric stability from the nearest modeling grid and time. We classified stability classes E, F and G as stable, whereas classes A, B, C, and D were classified as unstable and neutral. Stable atmospheric condition is usually associated with lower mixing height and higher pollutant concentrations in the atmosphere than unstable conditions.

5.5.4 Independent and Dependent Variables

The following variables were used to construct on-road exposure models.

Dependent variables: on-road pollutant concentrations of particle-bound PAH (ng/m^3), UFP particle number concentration (PNC, number of particles/cm^3), $PM_{2.5}$ (g/m^3), NO_x (ppbv) and black carbon (BC, ng/m^3, measured at 30 second averaging time). We used raw, natural log, or square root transformed pollutant concentrations averaged over one minute.

Independent variables:

(1). Roadway type: categorical variable (merging to freeways, freeways, major arterials, and minor surface streets or local roads). Both the USC roadway classification and ESRI street data were used to classify roadway type. In the ESRI data, A2x refers to primary roads without limited access, non-interstate roads (A2); A3x refers to smaller, secondary or connecting roads, usually with more than two lanes, and A4 refers to local, neighborhood and rural roads, usually with a single lane of traffic in each direction. We classified A1x-A2x as freeways/highways, A3x as major arterials and A4x as local roads.

(2). Vehicle speed of the mobile measurement platform: continuous independent variable (miles/hour).

(3). PeMS total traffic counts: continuous variable averaged by selected roadway segments. The data were limited both spatially and temporally by the available PeMS sampling sites as we described above.

(4). Diesel truck counts: continuous variable averaged by selected roadway segments. Limited only to certain freeways and highways.

(5). On-road temperature: including dew point and wet bulb (Celsius, ℃).

(6). Ambient temperature: continuous variable averaged over selected periods from the nearest meteorological site (Celsius, ℃).

(7). Ambient wind speed and direction: wind speed (WS, meters/second, abbreviated as m/s) and direction (WD) from the nearest meteorological site. They were also used as categorical variables. Wind speed was classified into five types: calm condition referred to wind speed lower than 1 m/s; light wind's speed was between 1 m/s and 3 m/s; moderate wind's speed was between 3 m/s and 5 m/s; strong wind's speed was between 5 m/s and 8 m/s; high wind speed was greater than 8 m/s. Wind directions were classified into four types: north-east (0-90°), north-west (90°-180°), south-west (180°-270°) and south-east (270°-360°). We also combined wind speed and wind direction according to their classifications (20 combinations of five levels of wind speeds by four types of wind directions).

(8). On-road relative humidity: continuous variable averaged over selected periods (%).

(9). Ambient relative humidity: continuous variable averaged by selected periods from the nearest meteorological site (%).

(10). Atmospheric stability from the nearest EDAS modeling grid. There are 7 levels from unstable to stable situation: A-Extremely unstable conditions; B: Moderately unstable conditions; C: Slightly unstable conditions; D: Neutral conditions; E: Slightly stable conditions; F: Moderately stable conditions; G: Extremely Stable. We combined A-D as unstable to neutral and E-G as stable as described above.

(11). AADT and VMT_AADT: AADT is annual average daily traffic counts estimated by CalTrans. Vehicle miles travelled AADT (VMT_AADT) was derived by multiplying AADT by road length within 500 m of a measurement point.

(12). Day period: categorical variable. The classification was done as: early morning: 12:0am - 06:00am; morning rush hour: 06:00am - 09:00am; mid morning: 09:00am -12:00pm; noon: 12:00pm - 02:00pm; afternoon: 02:00pm - 05:00pm; evening rush hour: 04:00pm - 07:00pm; night: 07:01pm -12:00pm.

(13). Lanes: This is the number of lanes at a measurement site from the Caltrans data.

(14). Lagged variable of vehicle speed and GPS_leg_length from one minute to ten minutes corresponding to the time of pollutant concentration measurement. GPS_leg_length was defined as the distance traveled in 1 sec or distance traveled in X seconds if GPS was recording at every X second.

5.6. METHODS

There were five steps in the construction of predictive models of on-road concentrations of air pollutants. First, exploratory data analysis was used to produce summary statistics and correlation analyses. Box and scatter plots were used to detect outliers and initial relationships between dependent and independent variables and to decide upon grouping statistics. Second, significant predictors were selected according to initial screening with correlation analyses and scatter plots and then further selected according to the variance inflation factor (VIF) and statistical significance. Third, we selected variables and constructed the models using linear regression with categorical variables (factors) and generalized additive model (GAM) regression with linear variables and categorical factors. Fourth, serial residuals from the linear models were examined to check whether there was statistically significant temporal autocorrelation among residual errors. If statistically significant, autocorrelation factors were used to adjust the bias in prediction. Fifth, we used independent holdout validation using 2/3 of the data as training data and 1/3 of the data as test data and 3-time x 3-fold cross validation to test the general linear and generalized additive models. For linear and GAM models, 2/3 of measurement data were used to train the models, but for models adjusted for autocorrelation, 3/4 of measurement data were used to train the models. The remaining part of the data were used to test the model for validation purposes (holdout data).

We constructed the spatial database with concentrations, relevant independent variables and corresponding GPS coordinates using POSTGIS 1.5 (Refractions Research, British Columbia, Canada) and used R 2.11.1(Bell Laboratories, New Jersey, USA) to conduct exploratory data analysis, construct the statistical models, and validate the models.

5.6.1 Exploratory Data Analysis

Exploratory data analysis is an initial analysis to evaluate the summary statistics across different groups as well as correlations, to find a suitable transform for a normally distribution dependent variable, to identify possible outliers, etc. We have conducted the following nine operations below:

(1). Summary statistics: to give an initial evaluation of the measured concentrations of PAH, PNC, $PM_{2.5}$ and NO_x.
(2). Box plots and identification of outliers.
(3). Histogram of the original data and transformed variables.
(4). Correlation analysis and scatter plots. Pearson and Spearman's correlations were used to evaluate the correlations between variables.
(5). Lagged correlation analysis. Some covariates such as vehicle speed may have a lagged relation to pollutant average concentrations (1). Therefore, lagged correlation analysis was necessary to detect the potential relationship between the pollutant concentration and the lagged variable.

(6). Grouping statistics by roadway type. There are four types of roads, namely, local roads, arterial roads, single freeway, and merging freeways. Roadway type is expected to be important for concentrations of traffic-related air pollutants. We present summary statistics across four roadway types and explored the changes in their distributions across roadway types.
(7). Grouping statistics by time of day. Time of day is a categorical variable that includes seven categories, i.e. early morning, morning rush hour, mid morning, noon, afternoon, evening rush hour and night. We examined the changes of concentrations by time of day.
(8). Grouping statistics by stability. More stable atmospheric conditions are usually associated with higher pollutant concentrations than unstable conditions. We examined the influence of atmospheric stability on pollutant concentrations using grouping statistics of concentrations based on the two combined levels of atmospheric stability (A-D as one group for unstable and neutral conditions and E-G as another group for stable conditions).
(9). Student t and Wilcox statistics were used to compare the differences in concentrations across two different groups of samples: freeways vs. non-freeways, morning vs. non-morning, and stable atmospheric stability vs. unstable atmospheric stability.

Due to the limited amount of measurement data, some variables such as summer vs. winter seasons, weekdays vs. weekends could not be examined in the models.

5.6.2 Selection of Predictor variables

Correlation analysis was used as the first step for variable screening. A variable was dropped from further analysis if the absolute value of its Pearson and Spearman's correlation with measured air pollutant concentrations were less than 0.1 with their scatter plots showing no obvious or regular patterns.

Then, we checked the multi-collinearity of independent variables and their statistical significance. First, to avoid multi-collinearity, we used variance inflation factors (VIFs) to help divide the covariates into several groups as follows:

1) one group of weakly correlated covariates (VIF<10);

and the following 3 groups of remaining highly correlated covariates (VIF≥10)

2) a traffic group, including traffic count, truck count, lanes, freeways and AADT; vehicle characteristic group including vehicle speed, gps_leg_length;

3) a meteorological group, including on-road dew point, wet bulb, humility, ambient temperature, ambient humility and atmospheric stability; and

4) a temporal group, including hour, day time, weekday/weekend and season etc.).

We selected one variable from each group of the highly correlated covariates at a time and combined them with all the weakly correlated covariates to construct a combination of covariates for the prediction model. All of the variables were tested in the model. Then, Akaike's information criterion (AIC) or R^2 was used to further backward-select the variables in each combination: the covariates with p values ≥0.1 were removed until R^2 remained the same, improved, or decreased least when all possible combinations of the remaining covariates were considered. Finally, the covariate combination with the maximum R^2 or minimum AIC was selected as optimal input in the model. If the VIFs of all the independent variables were smaller than 10, we would select those with statistical significance ($p < 0.1$).

5.6.3 General Linear and Non-Linear Models with Inclusion of Factor Variables

5.6.3.1 Basic model: linear regression with factor variables

Our independent variables include both continuous variables such as vehicle speed, ambient temperature, and on-road dew point, as well as categorical variables such as roadway type, time of day, and atmospheric stability etc. Linear regression with factor variables is the most basic prediction regression model that was often used to predict the concentrations (2-8). Given a data set $\{y_i, x_{i_1}, x_{i_2}, ..., x_{i_p}; x_{i_{(p+1)}}, ..., x_{i(p+m)}\}$ of n statistical units with p continuous variables (such as vehicle speed, wind speed, on-road dew point, ambient temperature) and m categorical variables as factors (such as roadway type, time of day, and atmospheric stability), we assumed for the linear regression model that the relationship between the dependent variable y_i (air pollutant concentration) and the p predictor variables x_i (i=1,…,p) is linear. In our model, each categorical variable as a factor, was transformed into multiple continuous dummy variables with a value of 0 or 1 indicating their status. In the linear model, these transformed multiple dummy variables were also assumed to be linearly related to the target variable, y_i (concentration). This relationship is modeled through a so-called "disturbance term". ε_i is an unobserved random variable that describes the random error to the linear relationship between the dependent variable and predictor variables.

Equation 5.5

$$y_i = \beta_0 + \beta_1 x_{i_1} + ... + \beta_p x_{i_p} + factor(x_{i_{p+1}}) + ... + factor(x_{i_{p+m}}) + \varepsilon_i, \; i=1,\ldots,n$$

In practice, we used the least squares approach to fit the linear regression models.

5.6.3.2 Non-linear model: generalized additive model with factor variables

Generalized additive model (GAM) can incorporate both linear and factor variables. The model specifies a distribution (such as a normal distribution, or a binomial distribution) of the dependent variable and a link function, g relating the expected

value of the distribution to the m predictor variables, and attempts to fit functions $f_i(x_i)$ to satisfy:

Equation 5.6

$$g(\hat{\mu}_u) = \beta_0 + \sum_{i=1}^{q} f_i(x_u^i, df) + \sum_{j=q+1}^{q+p} \beta_j x_u^j + \sum_{k=p+1}^{p+m} factor(x_u^k) + \varepsilon$$

where $\hat{\mu}_u$ is the estimate of the expected value of concentration at the location, u ($\hat{\mu}_u = E(y_u)$), $\hat{a}_0$ is the model's intercept, x_u^i, x_u^j or x_u^k are independent variables among which , x_u^i is q continuous variables with non-linear relationships, x_u^j is p continuous variables with linear relationship and x_u^k is k categorical variables as factors. $f_i(\ldots)$ is the non-parameter smooth function used to construct the non-linear relationship between x_u^i and $g(\hat{\mu}_u)$, df is degrees of freedom that controls the smooth degree of the curve fit, $\hat{a}_i$ are the linear parameters used to construct the linear relationship between x_u^j and $g(\hat{\mu}_u)$, and $g(\ldots)$ is the link function of expected value and the independent variables. For normally distributed air pollutants, the link function is $g(\hat{\mu}_u) = \hat{\mu}_u$. Similarly, in GAM, each categorical variable (e.g. roadway type, time of day and atmospheric stability), as a factor, would be transformed into multiple continuous dummy variables with a value of 0 or 1 indicating their status and each of these transformed multiple dummy variables is also assumed to be linearly relative with concentration. ε is the random error term $\varepsilon \in N(0, \sigma^2)$

The functions $f_i(x_i)$ may be fit using parametric or non-parametric means, thus providing the potential for a better model fits to the data than other methods. The method hence is general – a typical GAM might use a scatterplot smoothing function such as a locally weighted mean for $f_1(x_1)$ to model the non-linear relationship such as between ambient temperature and concentration of $PM_{2.5}$, and use a factor model for $f_2(x_2)$ such as roadway type and atmospheric stability. By allowing nonparametric fits, well designed GAMs allow good fits to the training data with relaxed assumptions on the actual relationship.

Overfitting can be a problem with GAMs. The number of smoothing parameters can be specified, and this number should be reasonably small (well under the degrees of freedom of the modeled data). Cross-validation can be used to detect and/or reduce overfitting problems with GAMs. Other models such as GLMs may be preferable to

GAMs in prediction of on-road concentrations unless GAMs improve predictive ability substantially for the application in question.

In practice, we used correlation analysis and scatter plots to determine the linear or non-linear relationships between continuous variables and pollutant concentrations. For variables that showed simple linear relationship in their scatter plot with good correlation they were used as linear regressors in GAM (no need for smoothing parameters). For those with complicated regular non-linear relationship and a better non-linear R^2 in their scatter plots, we set the smoothing parameter to a higher degree of freedom *df*. Thus, we could decrease the overfitting problem while improving the accuracy.

5.6.4 Time series model with temporal autocorrelation and factor variables

In the general linear regression and GAM regression, it was assumed that the contiguous measurements between two continuous time slices are independent. But there may be significant temporal autocorrelation between every two or more subsequent minutes due to continuous measurements during such a short period (every one minute). Without consideration of temporal autocorrelation, predictions from the linear or GAM models may be biased. Thus, we developed a time series model that incorporated temporal autocorrelation and factor variables:

Equation 5.7

$$y = X^{(p)}\beta + F(X^{(m)}) + \varepsilon$$

Equation 5.7 is similar to 5.5 or 5.6 but here ε includes serially correlated errors and is not random error (white noise). Here we assume that the errors from a regression model are unlikely to be independent in the time series data, where the observations represent different moments or intervals of time (i.e. measurements between every one minute), usually equally spaced. The process generating the regression errors is assumed to be stationary. That is, all of the errors have the same expectation and the same variance (σ^2), and the covariance of two errors depends only upon their separation *s* in time:

Equation 5.8

$$C(\varepsilon_t, \varepsilon_{t+s}) = C(\varepsilon_t, \varepsilon_{t-s}) = \sigma^2 \rho_s$$

under this model, $\rho_1 = \varphi$, $\rho_s = \varphi^s$, and $\sigma^2 = \sigma_v^2 / (1 - \phi^2)$ .Temporal autocorrelation is calculated as:

Equation 5.9

$$r_s = \frac{\sum_{t=s+1}^{n} e_t e_{t-s}}{\sum_{t=1}^{n} e_t^2}$$

where s is the number of lags, t is time slice, e_t is the error or residual at t, n is the total number of time slices. If the residuals were independently distributed, the standard error of each r_s would be approximately $1/\sqrt{n}$, a quantity that can be used as a rough guide to the statistical significance of the residual autocorrelations.

In practice, we used GLS with inclusion of categorical variables to get preliminary estimates of residuals and then calculated their Dubin-Watson statistics to check which lag's temporal autocorrelation is statistically significant. That residual autocorrelation were used to adjust the final predictions from the model.

We used the most common term for temporally auto-correlated regression errors, the first-order auto-regression process, AR(1):

Equation 5.10

$$\varepsilon_t = \phi \varepsilon_{t-1} + v_t$$

where the 'random shocks' v_t are assumed to be Gaussian white noise, $v_t \in N(0, \sigma_v^2)$.

In tests, we compiled the data from all the dates to model temporal autocorrelation and refined the model. Although there were some missing values between minutes and between days, the number of such missing values was limited and we assumed that such few missing values had limited influence upon temporal autocorrelation and the prediction model. We evenly and randomly divided the data into two parts according to the date: 75% of the data were used to train the data and the remaining 25% were used to independently test the model. We selected the explanatory variables in final linear regression models for each air pollutant.

5.6.5 Model validation

We used holdout or 3-fold cross validation to validate the fitted models.

5.6.5.1 Holdout validation as an independent test and validation

For linear and GAM models, 2/3 of measurement data were used to train the models, but for model adjusted for autocorrelation, 3/4 of measurement data were used to train the models. The remaining part (1/3 or 1/4) of the data was used to test the model for validation purposes (holdout data). The training data were selected by

stratified random sampling. Strata were defined by roadway types (merging freeways, freeways, arterials and local roads) and time of day (early morning, morning rush hour, noon, afternoon non-rush hour, afternoon rush hour, and night time non-rush hour). Strata were chosen to avoid bias due to an uneven distribution across roadway types and time of commute so that every roadway and commute time during each day period was represented by at least 2/3 or 3/4 of the values in the training group and 1/3 or 1/4 in the holdout group.

5.6.5.2 3x3 cross-validation

The original samples were randomly partitioned equally into 3 groups of subsamples; one of the groups was treated as the validation data for testing the model while the remaining two groups were used as training data. The process was repeated 3 times so that each of 3 subsamples has been used once for validation. The final validation results were the averages of the three model runs.

5.6.5.3 Measurement criteria

After model selection using the training sample was completed, predicted values for the testing subsamples were generated from the prediction equations. We then calculated squared multiple correlations equal to the squared univariate correlation between the sample's observed and predicted values as follows:

Equation 5.11

$$R^2(1) = R^2(Y|X_1, X_2, ..., X_p) = r^2(Y_1, \hat{Y}_1)$$

where $\hat{Y}_1$ is the set of predicted particulate concentrations from the *p* variables, and Y_1 is the set of observed concentrations for the training group.

The prediction equation built from the training group were then used to predict concentrations $\hat{Y}_2$ for the 1/3 holdout group and 3x3 cross validation. These predicted concentrations $\hat{Y}_2$ were then correlated to the observed concentrations in the holdout group Y_2 to give the "cross-validation correlation":

Equation 5.12

$$R^2(2) = r^2(Y_2, \hat{Y}_2)$$

A residual analysis was then performed to check for additional outliers in the holdout group and to describe those observations that did not fit the prediction equations. The cross-validation correlation was then used to derive the "shrinkage on cross-validation:"

Equation 5.13

$$R^2(1) - R^2(2)$$

Since $R^2(2)$ is less positively biased than $R^2(1)$, the shrinkage is usually positive. A reliable model is clearly suggested when the shrinkage is < 0.10 (9). In practice, if

the model was considered reliable, all of the observations could be pooled to estimate regression coefficients for the final prediction equation.

In this task, we validated models for the exposures of five concentrations, i.e. PAH, PNC, $PM_{2.5}$, NO_x, and BC collected in Tasks 1-2.

5.7. RESULTS AND DISCUSSION

5.7.1 Dependent variable concentrations

On-road pollutant measurements were obtained from the USC field work for five air pollutants (PAH, PNC, $PM_{2.5}$, NO_x, BC). We averaged the original data at 10-second intervals to one-minute averages for model development. Figure 5.9 represents the routes taken for the measurements of pollutant concentrations. Most measurements were performed on freeways (over 60% of the data). Table 5.6 gives the summary statistics for concentrations, and correspondingly, Figure 5.10 presents box plots of the concentrations. There were some missing data (9.2% for PAH, 61% for PNC, 46% for $PM_{2.5}$, and 0.8% for NO_x and 40.5% for BC) during the process of measurements due to device failure. BC was measured with considerable noise by the MicroAeth AE51 and was very weakly predictive, possibly due to lack instrument precision. We removed outliers from BC to clean the data before modeling and validation. The missing data may impact the accuracy of the models for different air pollutants, especially PNC, $PM_{2.5}$ and BC.

Table 5.6. Summary Statistics for the One-Minute Average On-Road Air Pollutants

Air Pollutant	Samples	Min	Max	Mean	Median
PAH (ng/m^3)	4638	0.5659	927.8	56.44	31.95
PNC (number of particles $/cm^3$)	2161	16360.0	319500.0	37510.0	28690.0
$PM_{2.5}$ (g/m^3)	3992	6.0	135.2	23.01	19.83
NO_X (ppbv)	5337	0.62	499.8	119.3	99.09
BC (ng/m^3)	4130	2.674	14580	3798	5369

Figure 5.11 shows histograms after removing several outliers. After exploratory analysis, we used a log transformation of PAH and $PM_{2.5}$ and used a square root transformation of PNC, NO_X and BC to reduce the skewness (Figure 5.12) and normalize the distribution.

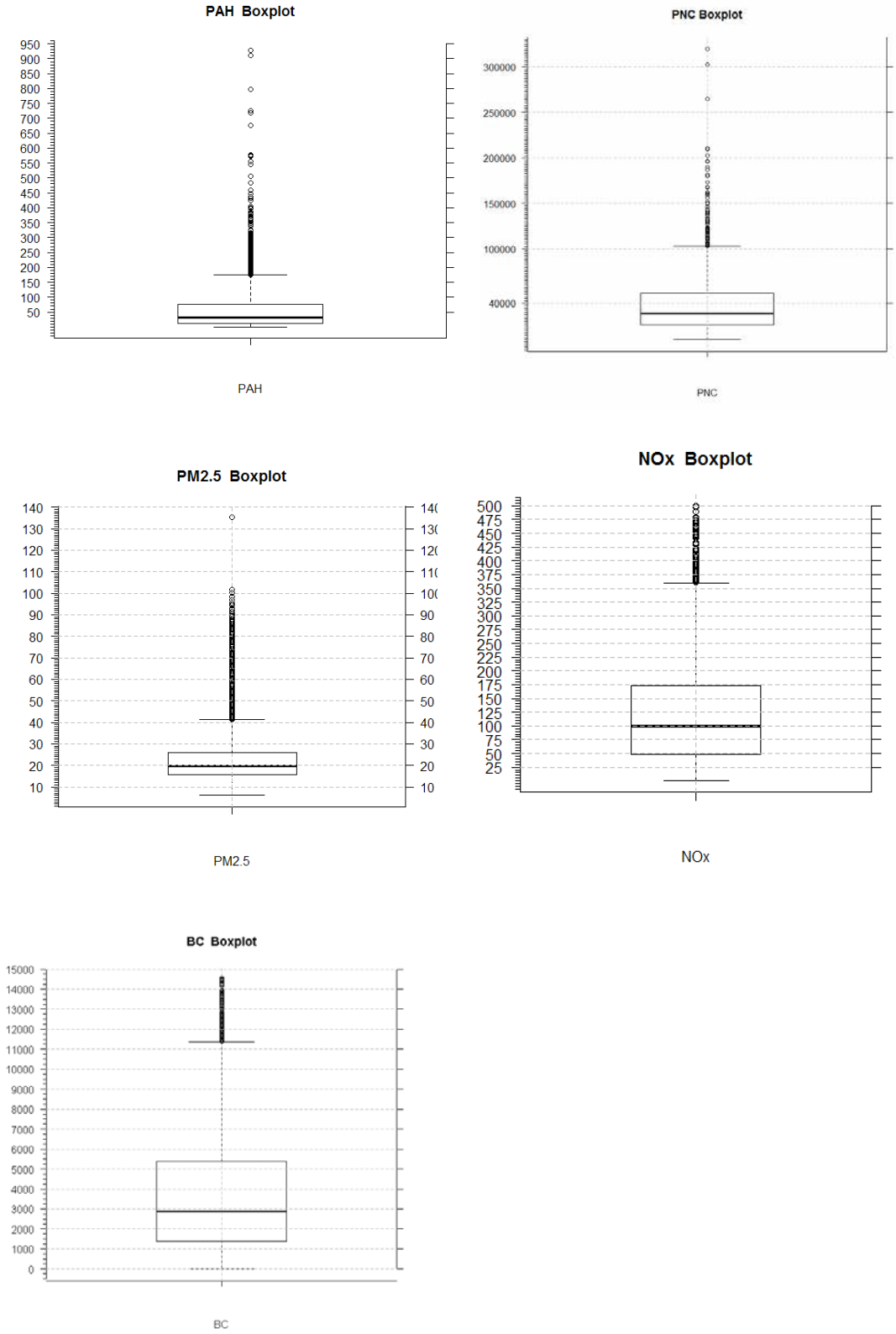
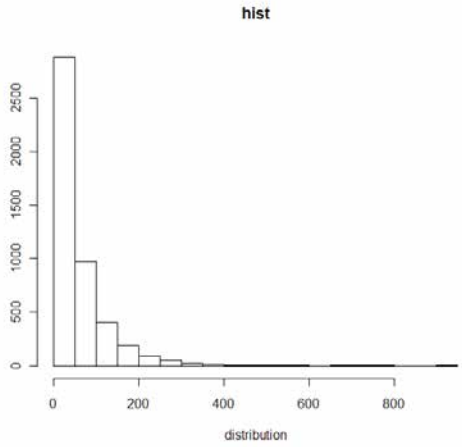
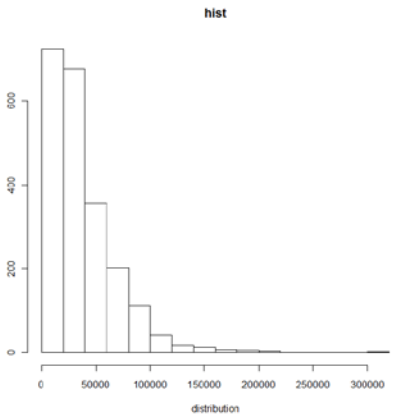


Figure 5.10: Box plots for four concentrations, PAH, PNC, $PM_{2.5}$, NO_X and BC

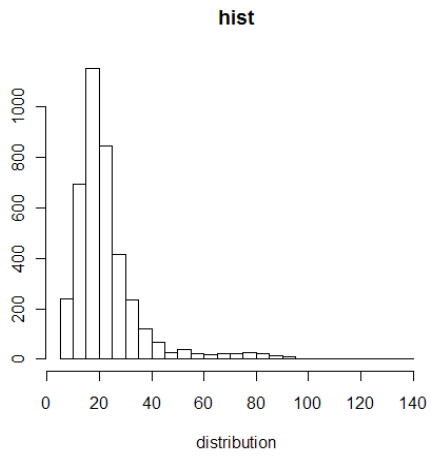
5.7.2 Transformation and correlation analysis



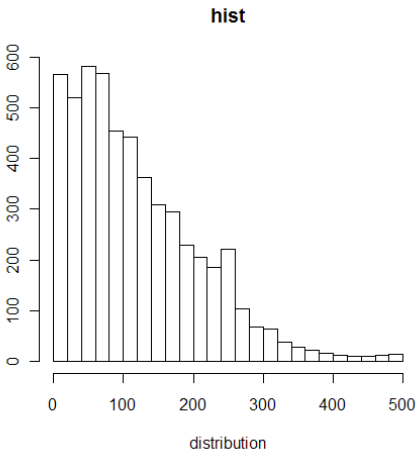
a. PAH (skewness=3.53)



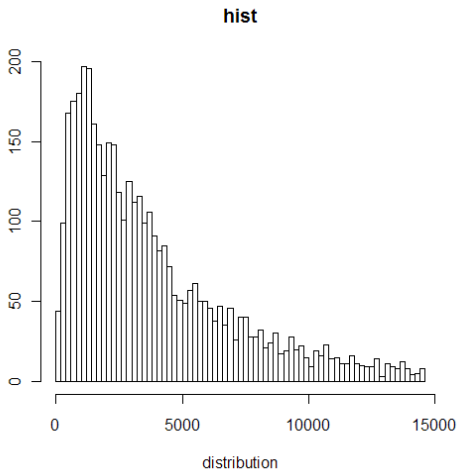
b. PNC (skewness=2.16)



c. $PM_{2.5}$ (skewness=2.68)



d. NO_x (skewness=1.05)

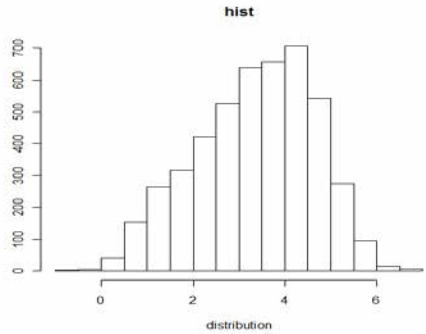


f. BC (skewness=1.24)

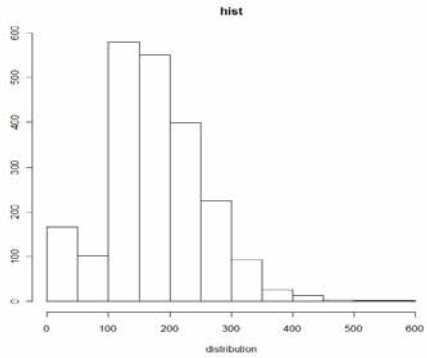
Figure 5.11: Histograms for raw air pollutant concentrations without transformation

The Pearson and Spearman's correlation of independent covariates and the

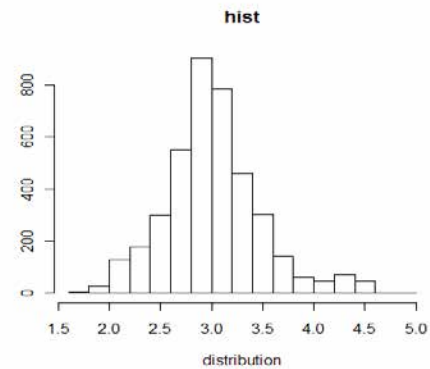
transformed (log or square root) dependent variable of concentrations (PAH, PNC, $PM_{2.5}$, NO_x, BC) are shown in Table 5.7 and 5.8. Typical scatter plots of covariates significantly correlated with the dependent variables are shown in Figure 5.13- 5.17. We paired typical linear or non-linear relationship scatter plots with their linear and non-linear regression lines as shown in these scatter plots. To be selected for regression modeling these variables had to have linear or non-linear relationships with concentrations.



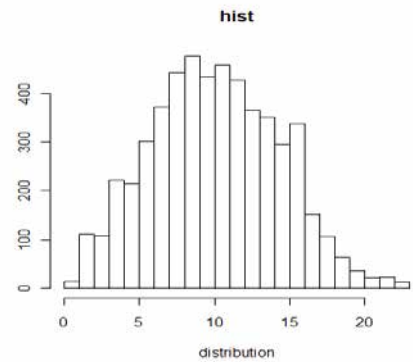
a. log(PAH) (skewness=-0.28)



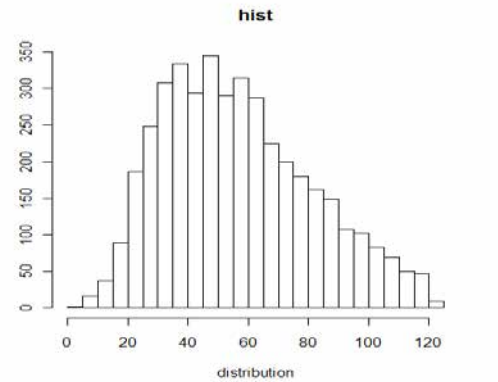
b. square root(PNC) (skewness=0.2)



c. log($PM_{2.5}$) (skewness=0.59)



d. square root(NO_x) (skewness=0.12)



e. square root(BC) (skewness=0.47)

Figure 5.12: Normal histograms for the transformed values of air pollutant concentrations.

Table 5.7. Correlation of predictor variables with dependent air pollutant variables PAH, PNC, $PM_{2.5}$ and NO_X.

	PAH			PNC			$PM_{2.5}$			NO_X		
	N	P.c	S.c	N	P.c	S.c	N	P.c	S.c	N	P.c	S.c
Vehicle speed	4638	0.395*	0.381*	2161	0.37*	0.38*	3992	0.074	0.045	5337	0.401*	0.406
Amb. temp	3899	0.037	0.063*	1807	0.075*	0.135*	3354	0.389*	0.299*	4481	0.024	0.049*
#Traffic	2558	0.159*	0.173*	1229	0.139*	0.142*	2227	0.224*	0.25*	3117	0.17*	0.18*
AADT	4616	0.482*	0.422*	2128	0.458*	0.442*	3956	0.164*	0.162*	5282	0.459*	0.404*
VMT_AADT	4616	0.483*	0.508*	2128	0.481*	0.557*	3956	0.119*	0.128*	5282	0.463*	0.518*
#Truck	4638	0.19*	0.264*	2161	0.138*	0.229*	3992	0.102*	0.147*	5337	0.21*	0.23*
Lanes	4616	0.428*	0.396*	2142	0.364*	0.402*	3970	0.119*	0.113*	5301	0.413*	0.374*
Freeways	4638	0.464*	0.446*	2161	0.487*	0.516*	3992	0.088*	0.093*	5337	0.478*	0.446*
Wind speed	3630	-0.181*	-0.196*	1756	-0.093*	-0.101*	3119	-0.156*	-0.183*	5337	-0.15*	-0.19*
Wind direction	3630	-0.123*	-0.118*	1756	-0.155*	-0.140*	3119	-0.019	-0.051*	4158	-0.07*	-0.095*
Wind_sd_com	3630	-0.123*	-0.151*	1756	-0.155*	-0.149*	3119	-0.02	-0.108	4158	-0.071*	-0.123*
On-road DP	2821	0.098*	0.078*	1636	0.146*	0.155*	2754	0.547*	0.553*	3040	0.067*	0.064*
On-road WB	2821	0.096*	0.093*	1636	0.112*	0.189*	2754	0.635*	0.588*	3040	0.1*	0.092*
On-road RH	2821	-0.007	0.052*	1636	0.041	0.095*	2754	0.045*	0.20*	3040	-0.044	0.048
Amb RH	3899	0.098*	0.078*	1807	0.044	0.008	3354	0.02	0.095*	4481	0.063*	0.066*
Gps_leg_length	723	0.416*	0.379*	310	0.087	0.084	316	-0.119	-0.135	1015	0.049	0.501*
Hour	4664	-0.301*	-0.320*	2161	-0.368*	-0.364*	3992	-0.308*	-0.339*	5337	-0.182*	-0.278*
Roadway type	4638		0.443*	2161		0.499*	3992		0.101*	5337		0.447
Time of day	4638		-0.323*	2161		-0.359*	3992		-0.331*	5337		-0.257*
Stability	43		0.155*	2005		0.203*	3706		0.093*	4941		0.116*

* indicates statistical significance at $\alpha = 0.1$; N: number of valid samples for the covariate and dependent variable; P. c: Pearson correlation; S. c: Spearman's correlation; Amb temp: temperature of the nearest meteorological site to the measurement; #Traffic: traffic counts derived from the PeMS; #Truck: estimated truck counts derived from the PeMS; Wind_sd_com: combination of four levels of wind speed and four wind directions; DP: dew point; WB wet bulb; Amb RH: ambient relative humility; On-road RH: on-road relative humility.

Table 5.8. Correlations of predictor variables with BC Concentration Measurements

	N	P.c	S.c
Vehicle speed	4130	0.30*	0.28*
Temp	3433	0.142*	0.199*
#Traffic	2632	0.085*	0.102*
AADT	4094	0.352*	0.303*
VMT_AADT	4075	0.355*	0.368*
#Truck	4130	0.121*	0.200*
Lanes	4094	0.321*	0.287*
Freeways	4130	0.284*	0.261*
Wind speed	3204	-0.092*	-0.121*
Wind direction	3204	-0.084*	-0.103
Wind_sd_com	3204	-0.085*	-0.121
Dew point	2574	0.215*	0.209*
Wet bulb	2574	0.248*	0.256*
In hum.	2574	-0.002	0.074*
Ambient humility .	3433	0.005	-0.038*
Gps_leg_length	675	0.112*	0.505*
Hour	4130	-0.29*	-0.323*
Road type	4130		0.259*
Daytime	4130		-0.304*
Stability	4130		0.034*

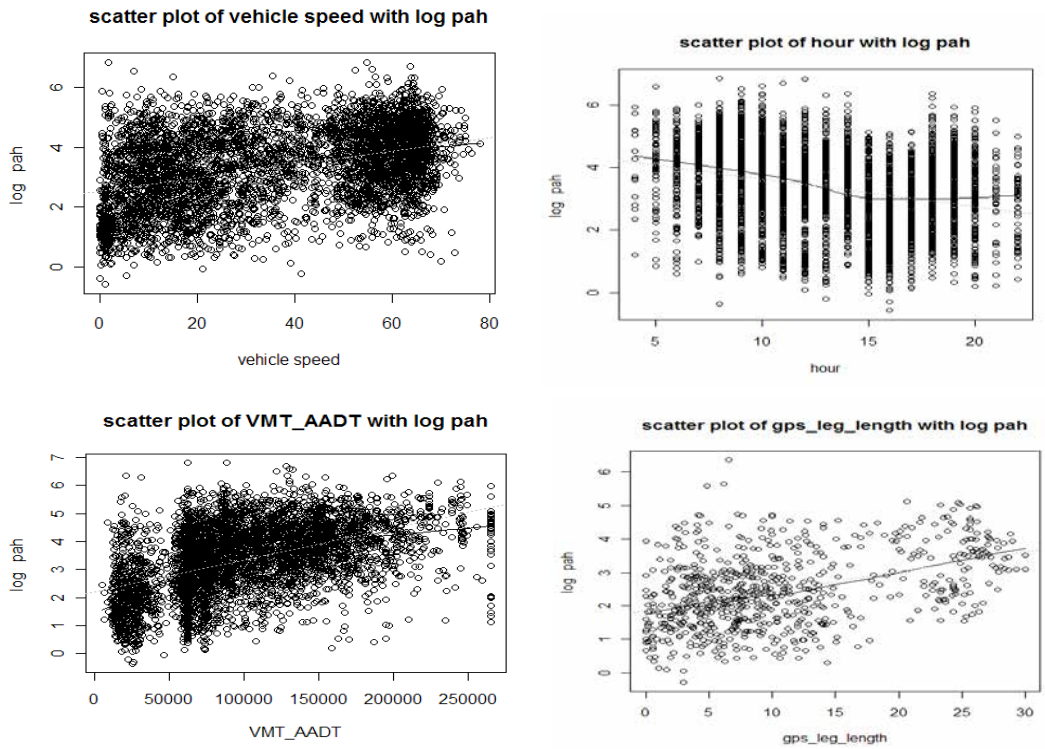


Figure 5.13: Scatter plots of several covariates with the log dependent variable of PAH

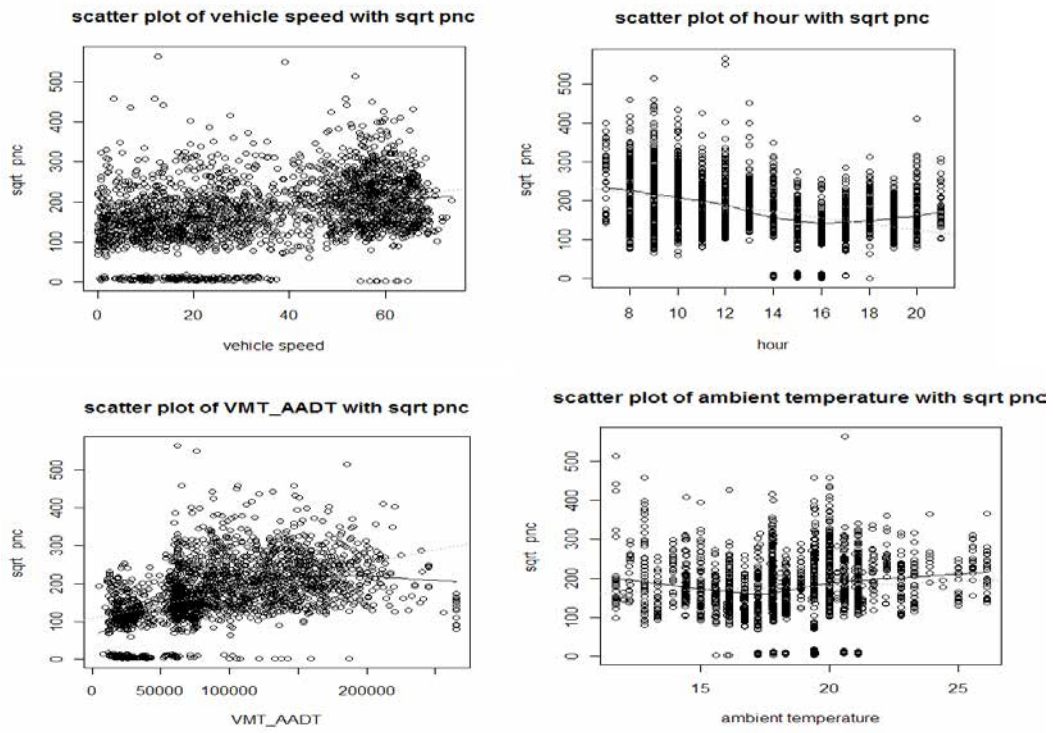


Figure 5.14: Scatter plots of several covariates with the square root dependent variable of PNC

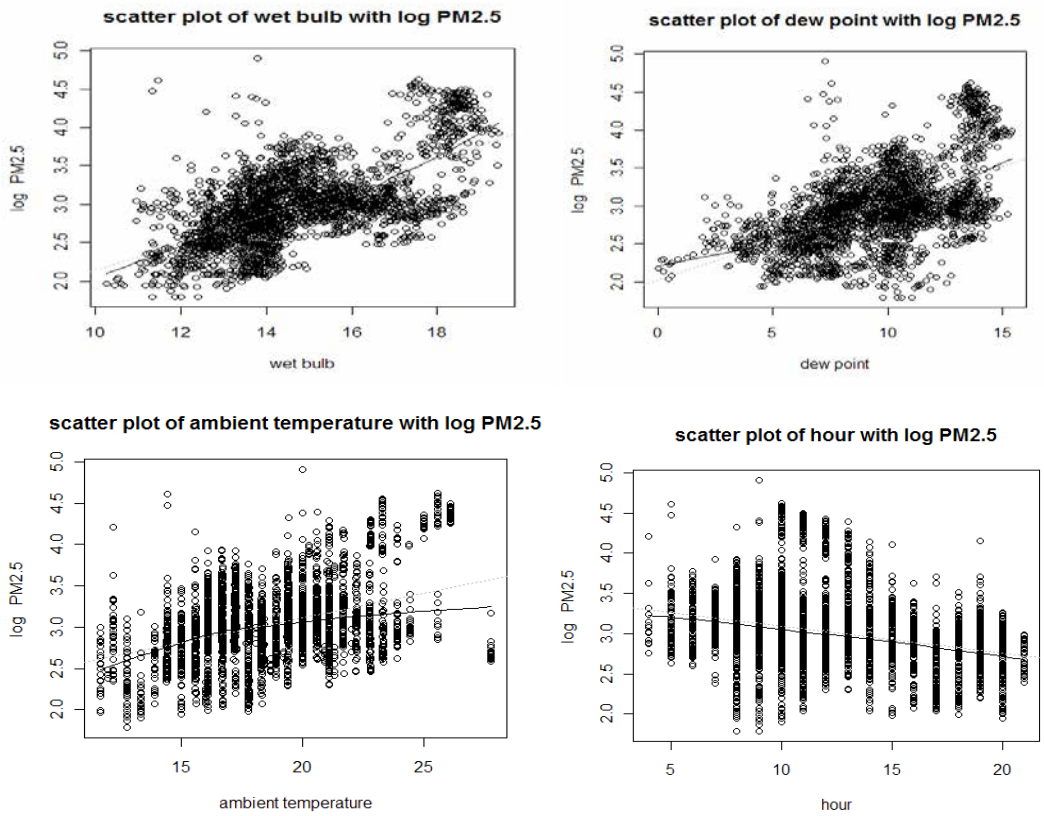


Figure 5.15: Scatter plots of several covariates with the log dependent variable of $PM_{2.5}$

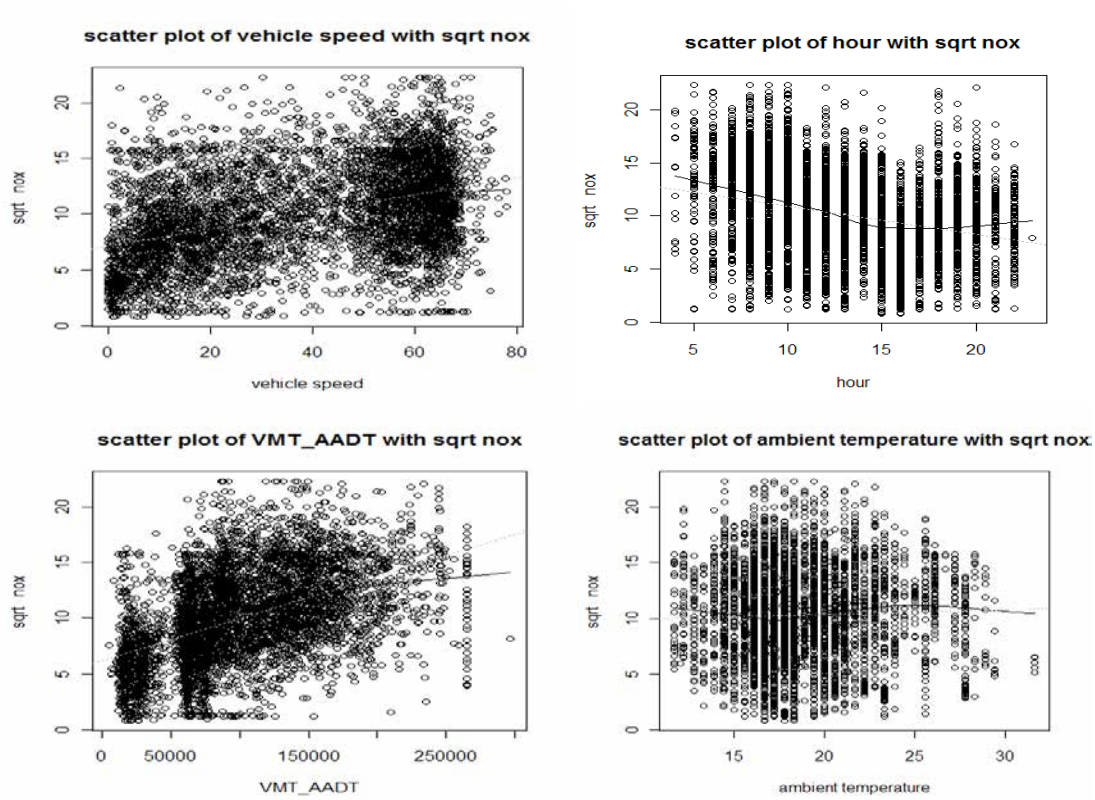


Figure 5.16: Scatter plots of several covariates with the log dependent variable of NO_x

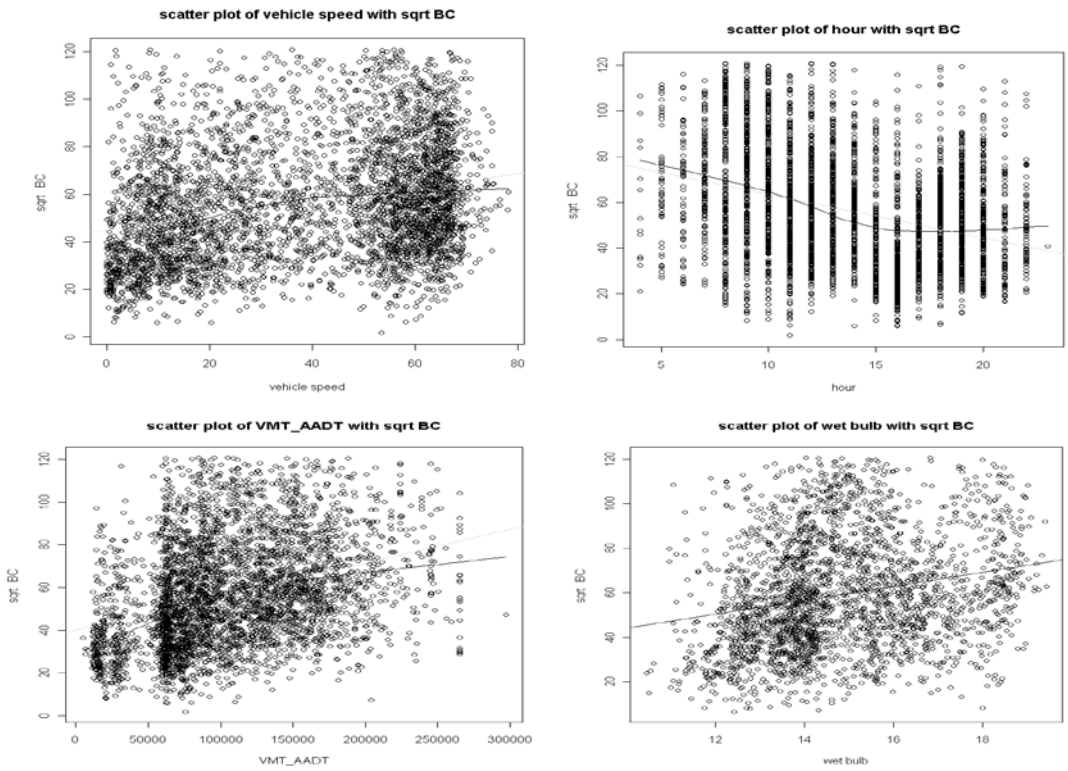


Figure 5.17: Scatter plots of several covariates with the log dependent variable of BC

5.7.3 Grouping Comparison

From correlation analysis, we found that roadway type, time of day and stability may have significant influence on concentrations of some pollutants. We made grouping

statistics accordingly and checked how the measurement values varied across different categories.

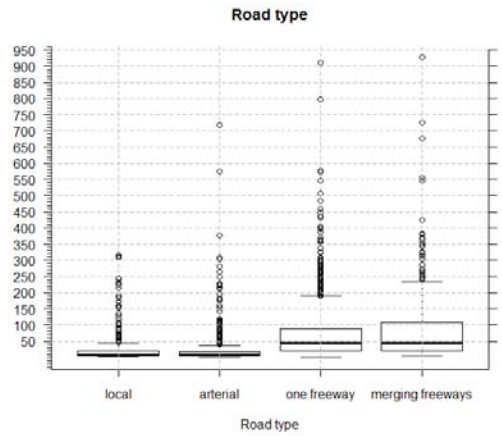
5.7.3.1 Roadway types

Table 5.9 shows the grouping statistics by roadway type and statistical significance for the differences between freeways and non-freeways groups. Figure 5.18 presents the grouping box plots of concentrations. There was a significantly higher concentration on freeways than on non-freeways as expected.

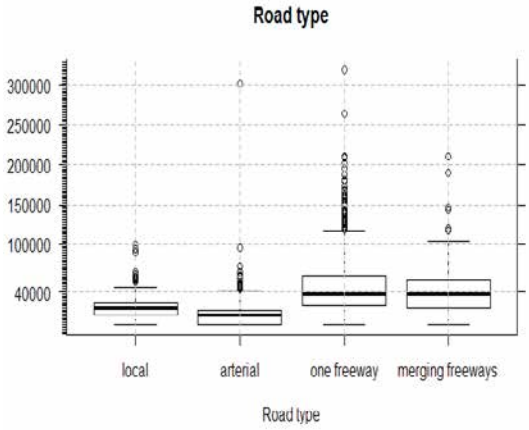
Table 5.9. Grouping Statistics for Air Pollutants by Roadway Types

Pollutant	Roadway Type	N	Mean	Median	P. c of spd	S. c of spd	Student t	Wilcox
PAH	Local	442	24.98	6.14	0.21*	0.41*	t=-21.51*; p=2.2e-16; 22.89 vs. 66.23	W=644737*; p=2.2e-16
	Arterial	606	22.83	7.32	0.24*	0.20*		
	Freeways	3338	64.33	43.96	0.17*	0.25*		
	M. Freeways	252	91.29	43.99	0.18*	0.15*		
PNC	Local	112	24115.23	19791.92	0.22*	0.18*	t=-26.31*; p=2.2e-16; 15477.6 vs. 45701.6	W=129699.5*; p=2.2e-16
	Arterial	474	13436.66	10675.30	0.20*	0.032		
	Freeways	1462	45744.65	36938.38	0.17*	0.19*		
	M. Freeways	113	45143.89	37643.87	0.181	0.19		
$PM_{2.5}$	Local	419	19.75	18.0	0.07	0.09	t=-6.68*; p=2.9e-11; 20.74 vs. 23.72	W=1267091*; p=9.85e-10
	Arterial	543	21.51	19.33	0.05	0.02		
	Freeways	2805	23.67	20.0	0.06*	0.016		
	M. Freeways	225	24.27	21.00	0.002	0.05		
NO_X	Local	503	43.16	19.60	0.29*	0.31*	t=-38.8*; p=2.2e-16; 49.41 vs. 137.95	W=738666*; p=2.2e-16
	Arterial	620	54.49	35.62	0.30*	0.20*		
	Freeways	3933	137.56	120.56	0.27*	0.27*		
	M. Freeways	281	143.48	114.27	0.15*	0.16*		
BC	Local	366	2118.1	1229.47	0.25*	0.31*	t=-17.72*; p=2.2e-16; 2076. vs. 4087	W=535772*; p=2.2e-16
	Arterial	230	2011.16	978.62	0.36*	0.31*		
	Freeways	3308	4078.41	3238.52	0.17*	0.20*		
	M. Freeways	226	4223.8	2966.5	0.06	0.05		

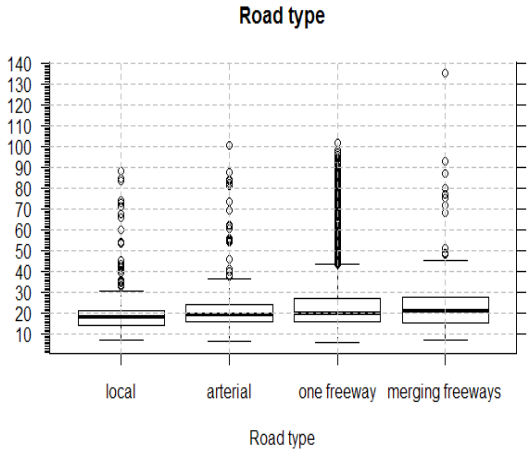
Notes: N: number of samples; * indicates statistical significance at p-value<0.1; P. c of spd: Pearson's correlation of vehicle speed with concentration within a group; S. c of spd: Spearman's correlation of vehicle speed with concentration within a group; student t statistics used to check whether the differences in concentration between groups of freeways vs. non freeways is statistical significant assuming the normal distribution; Wilcox statistics used to check whether the differences in concentration between groups of freeways vs. non freeways is statistical significant without normal assumption.



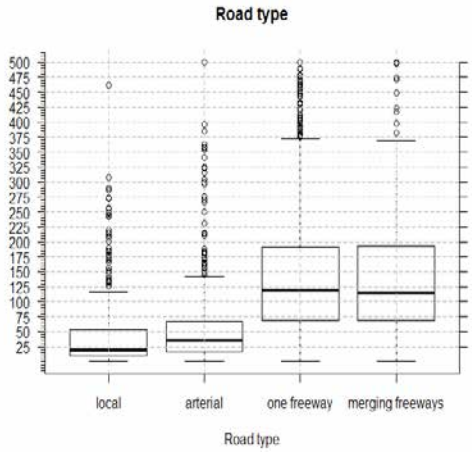
a. PAH by roadway type



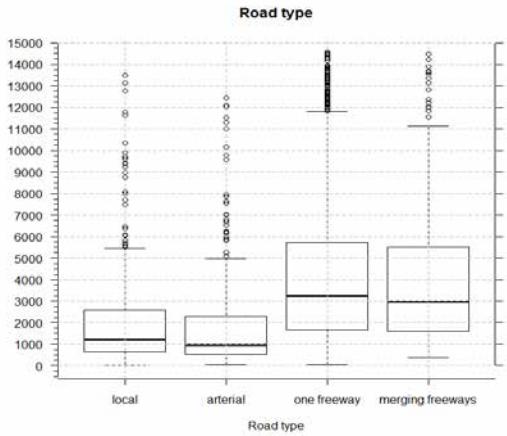
b. PNC by roadway type



c. $PM_{2.5}$ by roadway type



d. NO_x by roadway type



f. BC by roadway type

Figure 5.18: Box plots of pollutant concentrations across roadway types

5.7.3.2 Time of day

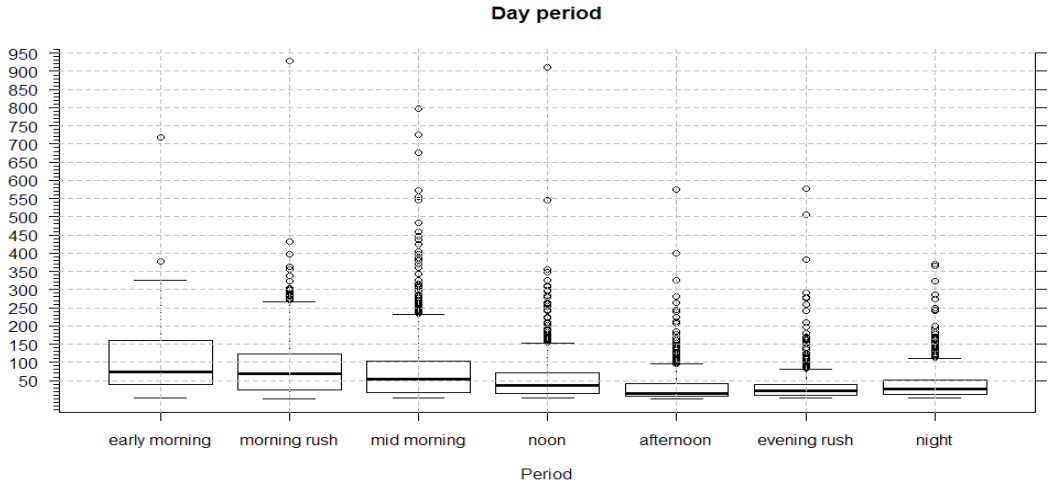
Table 5.10 shows the grouping statistics by time of day and statistical significance for the differences between morning and non-morning groups. Figure 5.19 presents

such grouping box plots. There was a statistically higher concentration of PAH, PNC, $PM_{2.5}$, NO_X, BC in the morning than at noon, in the afternoon or at night.

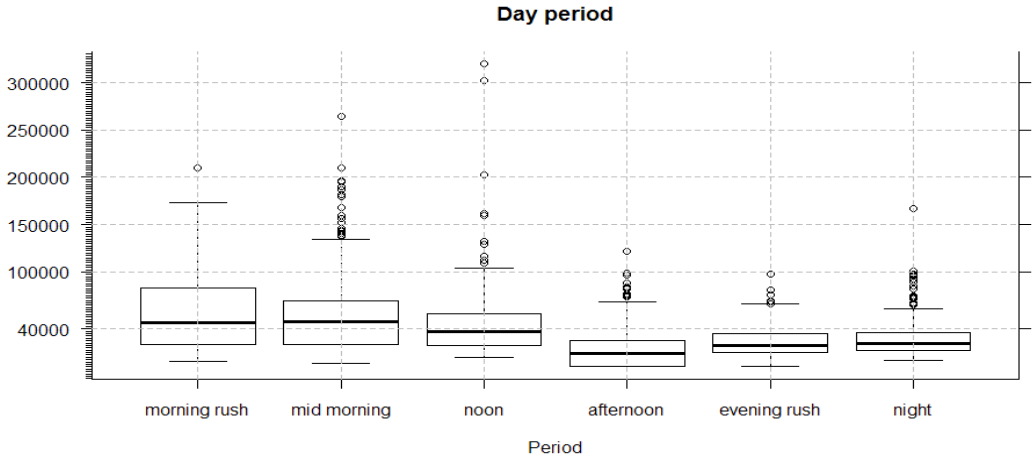
Table 5.10. Grouping Statistics for Air Pollutants by Time of Day

Type	Time of day	N	Mean	Median	P. c of spd	S. c of spd	Student t	Wilcox
PAH	Early morning	95	107.5	73.83	0.046	0.042	t=-18.73*; p=2.2e-16; 80.67 vs. 39.63	W=3567512*; p=2.2e-16
	Morning rush hour	728	84.82	68.01	0.16*	0.21*		
	Mid-morning	1076	75.50	52.9	0.29*	0.38*		
	Noon	635	54.99	35.66	0.08*	0.32*		
	Afternoon	901	31.42	13.92	0.21*	0.34*		
	Evening rush hour	592	34.07	20.56	0.18*	0.31*		
	Night	611	41.19	26.58	0.038	0.02		
PNC	Early morning	0	NA	NA	NA	NA	t=16.74*; p=2.2e-16; 53602.0 vs. 28487.1	W=776111*; p=2.2e-16
	Morning rush	219	57741.2	46630.1	0.28*	0.36*		
	Mid-morning	557	51974.6	47142.9	0.42	0.50		
	Noon	322	43963.9	36766.2	0.14*	0.15*		
	Afternoon	456	18309.8	13685.3	0.23*	0.16*		
	Evening rush	298	25989.6	22758.9	0.13*	0.16*		
	Night	309	29786	24807.7	0.07*	0.04*		
$PM_{2.5}$	Early morning	95	25.30	21.33	-0.003	-0.15	t=12.07*; p=2.2e-16; 25.72 vs. 20.54	W=2581573*; p=2.2e-16
	Morning rush	728	24.36	23.33	-0.022	-0.08		
	Mid-morning	1076	26.67	21.0	0.05*	0.1*		
	Noon	635	26.43	20.33	-0.09	-0.143		
	Afternoon	630	20.11	18.25	0.17*	0.083*		
	Evening rush	340	15.67	16.55	-0.34*	-0.33*		
	Night	488	16.82	16.69	-0.43*	-0.43*		
NO_X	Early morning	95	174.76	165.3	0.13	0.21*	t=21.65*; p=2.2e-16; 152.77 vs. 96.74	W=4506686*; p=2.2e-16
	Morning rush	755	162.7	166.7	0.31*	0.31*		
	Mid-morning	1301	145.32	135.73	0.38*	0.30*		
	Noon	721	116.85	109.69	0.37*	0.41*		
	Afternoon	1074	86.15	67.94	0.43*	0.43*		
	Evening rush	799	99.16	87.71	0.35*	0.38*		
	Night	799	99.16	87.71	0.078	0.071		
BC	Early morning	68	5163.8	4198.3	0.11	0.11	t=19.69*; p=2.2e-16; 5100.2 vs. 3030.1	W=2685479*; p=2.2e-16
	Morning rush	424	5777.44	5148.9	0.17*	0.19*		
	Mid-morning	1039	4819.7	3900.9	0.24*	0.21*		
	Noon	624	4022.2	3405.5	0.13*	0.22*		
	Afternoon	743	2596.0	1770.6	0.27*	0.38*		
	Evening rush	516	2616.6	2116.8	0.20*	0.26*		
	Night	716	2914.1	2296.9	0.04	0.02		

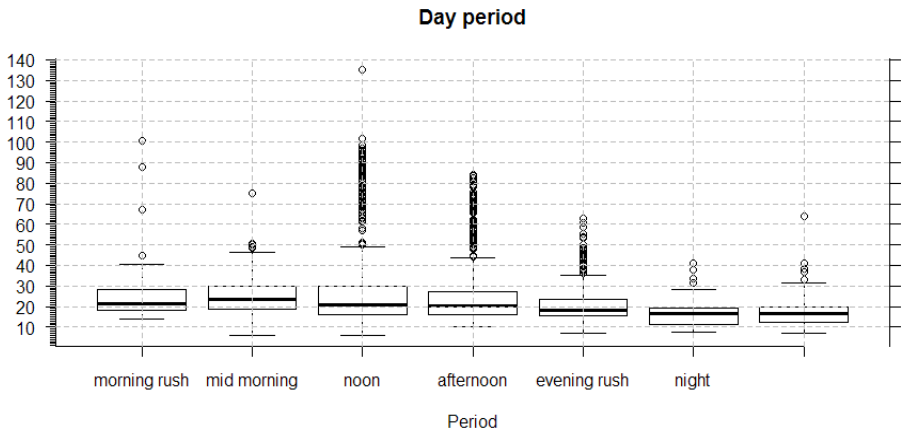
Note: N: number of samples; * indicates statistical significance at p-value<0.1; P. c of spd: Pearson's correlation of vehicle speed with concentration within a group; S. c of spd: Spearman's correlation of vehicle speed with concentration within a group; student t statistics used to check whether the differences in concentration between groups of morning vs. non morning is statistical significant assuming the normal distribution; Wilcox statistics used to check whether the differences in concentration between groups of morning vs. non morning is statistical significant without normal assumption.



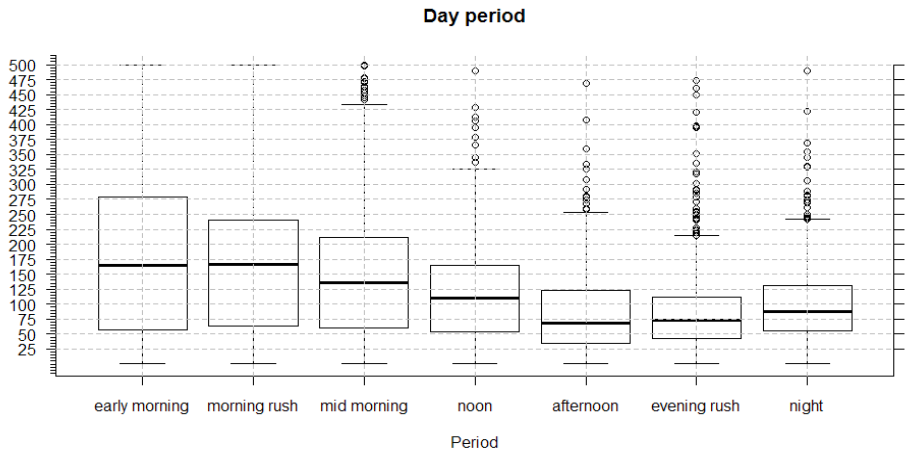
a. PAH box plots by time of day



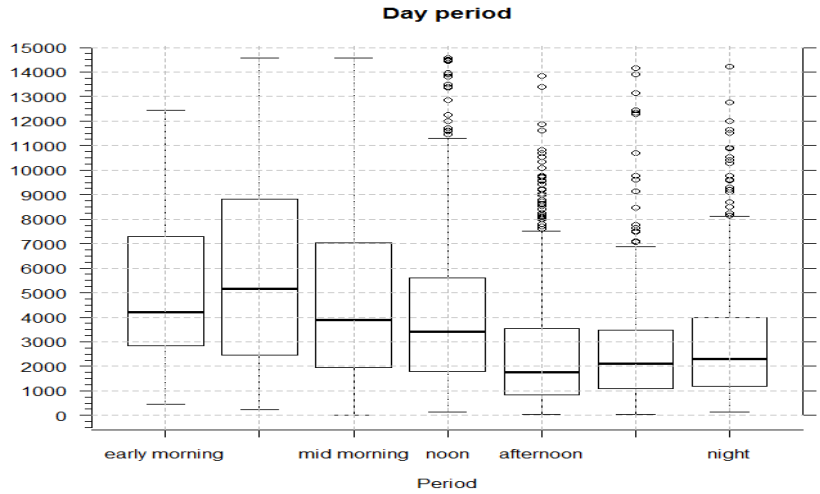
b. PNC box plots by time of day



c. $PM_{2.5}$ box plots by time of day



d. NOx box plots by time of day



e. BC box plots by time of day

Figure 5.19: Box plots of pollutant concentrations by time of day

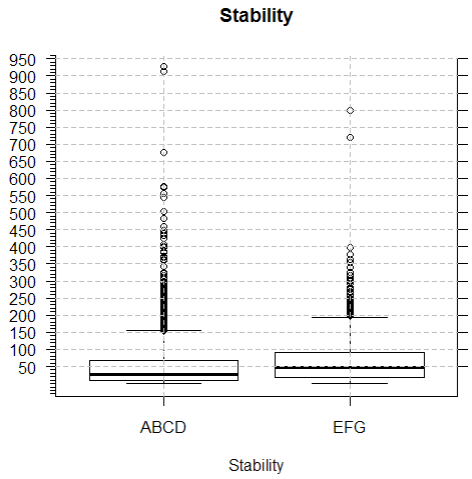
5.7.3.3. Atmospheric Stability

Table 5.11 shows the result of grouping statistics by atmospheric stability and statistical significance for the difference between stable and non stable groups. Figure 5.20 presents the grouping box plots. Under the stable atmospheric situation, there were a statistically significant higher concentrations of PAH, PNC, $PM_{2.5}$ and NO_X, but only borderline significant differences for BC.

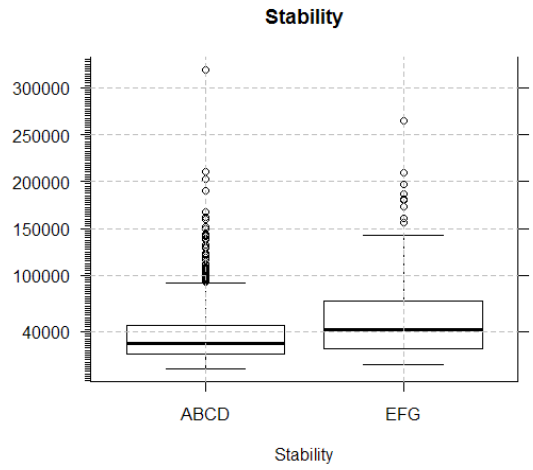
Table 5.11. Grouping Statistics for Air Pollutants by Modeled Atmospheric Stability

Type	Stability class	N	Mean	Median	P. c of spd	S. c of spd	Student t	Wilcox
PAH	A,B,C,D	2910	53.13	27.59	0.25*	0.37*	t=-5.68*; p=1.44e-8; 53.13 vs. 65.76	W=1639320*; p=2.2e-16
	E,F,G	1393	65.76	47.15	0.18*	0.25*		
PNC	A,B,C,D	1584	34479.3	26909.40	0.27*	0.33*	t=-8.44*; p=2.66e-16; 34479 vs. 50824	W=247671*; p=2.2e-16
	E,F,G	421	50824.2	42658.7	0.33*	0.31*		
$PM_{2.5}$	A,B,C,D	2517	22.73	19.5	0.1*	0.06*	t=-1.98*; p=0.048; 22.74 vs. 23.6	W=1325143*; p=1.8e-8
	E,F,G	1189	23.61	20.5	-0.05	-0.12*		
NO_X	A,B,C,D	3262	114.13	92.07	0.40*	0.43*	t=--7.61*; p=3.5e-14; 114 vs. 135	W=2350250*; p=3.0e-16
	E,F,G	1679	134.97	118.8	0.26*	0.25*		
BC	A,B,C,D	2488	3725.06	2784.1	0.24*	0.28*	t=-7.83; p=0.068; 3725 vs. 3907	W=1960054*; p=0.027
	E,F,G	1642	3907.3	3029.6	0.23*	0.27*		

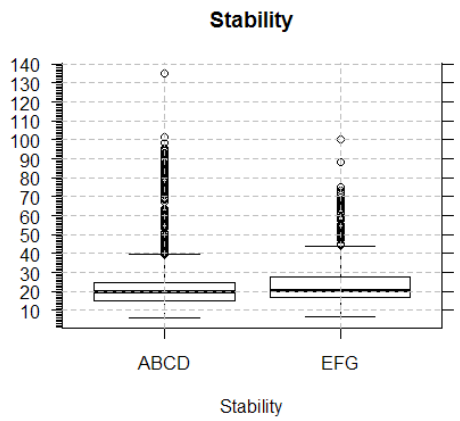
Note: N: number of samples; * indicates statistical significance; P. c of spd: Pearson's correlation of vehicle speed with concentration within a group; S. c of spd: Spearman's correlation of vehicle speed with concentration within a group; student t statistics used to check whether the differences in concentration between groups of un-stable atmospheric situation (ABCD) vs. stable atmospheric situation (EFG) is statistical significant assuming the normal distribution; Wilcox statistics used to check whether the differences in concentration between groups of un-stable atmospheric situation vs. stable atmospheric situation is statistical significant without normal assumption.



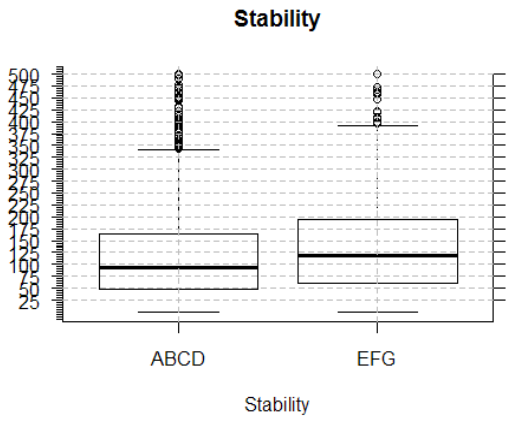
a. PAH grouping by stability



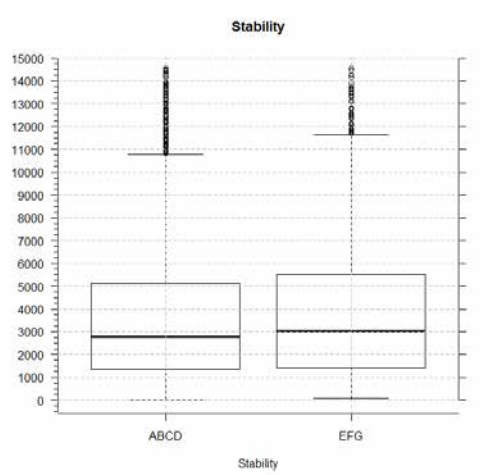
b. PNC grouping by stability



c. $PM_{2.5}$ grouping by stability



d. NOx grouping by stability



e. BC grouping by stability

Figure 5.20: Box plot of pollutant concentrations by stability groups

5.7.4 Regression models for prediction

We used linear regression and non-linear generalized additive models to build the predictive equations. The concentrations were log- or square root-transformed to produce normal distributions.

5.7.4.1 PAH modeling

In these models, PAH has been log transformed. In the modeling of grouped variables, these variables were selected by their higher R^2 and their statistical significance at p-value<0.1.

Model 1: grouping regression by roadway type

Independent variables for selection in linear regression or GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, temperature) and two factor variables (stability, time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences in the independent variables selected in the models of different roadway types and such differences had limited influence upon the prediction accuracy. Table 5.12 lists the regression results by roadway type with their optimal models.

Table 5.12. Prediction Performance for Grouping PAH by Roadway Types

Roadway type	Samples	R^2 for linear regression	R^2 for GAM
Local road	442	0.41	0.51
Arterial road	606	0.30	0.41
One freeway	3338	0.28	0.31
Merging of 2 or more freeways	252	0.34	0.53

Model 2: Grouping regressions by time of day

Independent variables for selection in linear regression or GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature) and two factor variables (stability, roadway type). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences in the independent variables selected in the models of different times of day and such differences had limited influence upon the prediction accuracy. Table 5.13 lists the grouping regression results with their optimal models.

Table 5.13 Prediction Performance for Grouping PAH by Time of day

Time of day	N Samples	R^2 for linear regression	R^2 for GAM
Early morning	89	0.35	0.52
Morning rush hour	622	0.37	0.44
Mid-morning	869	0.51	0.55
Noon	495	0.47	0.46
Afternoon	762	0.39	0.41
Evening rush hour	498	0.31	0.31
Night	544	0.27	0.33

Model 3: grouping regression by estimated stability

Independent variables for selection in linear regression or GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature) and two factor variables (time of day, roadway type). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences in the independent variables selected in the models of different stability class and such differences had limited influence upon the prediction accuracy. Table 5.14 listed the grouping regression results with their optimal models.

Table 5.14. Prediction Performance for Grouping PAH by Stability

Stability	Samples	R^2 for linear regression	R^2 for GAM
A,B,C,D	2580	0.41	0.43
E,F,G	1299	0.24	0.36

Model 4: Final regression models and model validation

Predictor variables in the linear regression model included four continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature) and two factor variables (time of day, roadway type);

Predictor variables in GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, temperature) and two factor variables (time of day, roadway type).

Table 5.15 gives the coefficients in the linear model, degree of freedom in GAM and the variances explained. Coefficients of the linear model indicate the influence of each variable and degrees of freedom in GAM indicates the fit degree of smooth for each coefficient. Table 5.16 presents the results of independent holdout test (2/3 for training; 1/3 for test) and 3x3 cross validation.

Table 5.15. Coefficients Regressed and Variance Explained for the Prediction of PAH.

	Coefficients in linear model	Degree of freedom in GAM	Variance explained (%)	
			Linear model	GAM
Intercept	2.58			
Vehicle speed	0.004	9.96*	15.41	2.17
VMT_AADT	0.0000035	13.63*	8.72	2.56
Lanes	0.041	6.5*	2.15	1.41
Truck counts		6.2*		6.9
Ambient temperature	0.02	24.21*	0.13	5.89
Time of day			7.44	23.25
Roadway type			9.2	9.60
Total variance Explained			43.05	51.8

Note: * indicates degree of freedom; gray color indicates categorical variables as factors.

We also conducted a multi-collinearity diagnostic analysis: VIF of each predictor was < 10 (maximum VIF=2.12), so we can safely use the model: with vehicle speed: 1.58; VMT_AADT:1.89; Lanes: 2.04; hour: 1.24; roadway types: 2.12; truck count: 1.2; stability: 1.33; temperature: 1.19.

Table 5.16. Independent 1/3 Holdout and 3x3 Cross Validation of Predictive Models for PAH.

	Linear regression			Generalized additive model		
	General	3 times CV	1/3 test	General	3 times CV	1/3 test
Samples	3879	3879	2596 vs. 1283	3879	3879	2596 vs. 1283
R^2	0.43	0.42	0.42	0.52	0.43	0.46
P. cor.	0.65*	0.65	0.65*	0.73*	0.65*	0.68*
S. cor.	0.64*	0.64*	0.64*	0.71*	0.64*	0.66*

Note: General: no cross validation; 3 times CV: 3-times 3-fold cross validation; 1/3 test: two thirds used for training and one third used for test; P. cor.: Pearson correlation between the observed values and the predicted values; S. cor.: Spearman's correlation between the observed values and the predicted values; * indicates statistical significance.

5.7.4.2 PNC modeling

In these models, PNC has been square root transformed to normalize its distribution. In the modeling of grouped variables, these variables were selected by their higher R^2 with their statistical significance at p-value<0.1.

Model 1: Grouping regression by roadway type

Independent variables for selection in linear regression or GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature) and two factor variables (time of day, stability). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different roadway types and such differences had limited influence upon the prediction accuracy. Table 5.17 listed the grouping regression results with their optimal models.

Table 5.17. Prediction Performance for Grouping Particle Number Concentrations by Roadway Type

Roadway type	N Samples	R^2 for linear regression	R^2 for GAM
Local road	86	0.33	0.46
Arterial road	387	0.35	0.52
One freeway	1151	0.29	0.34
Merging of 2 or more freeways	39	0.48	0.40

Model 2: grouping regression by time of day

Independent variables for selection in linear regression or GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature) and one factor variable (roadway type). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different times of day and such differences had limited influence upon the prediction accuracy. Table 5.18 lists the grouping regression result with their optimal models.

Table 5.18. Prediction Performance for Grouping Particle Number Concentrations by Time of Day

Time of day	N Samples	R^2 for linear regression	R^2 for GAM
Early morning	0	NA	NA
Morning rush	151	0.46	0.43
Mid-morning	457	0.41	0.44
Noon	239	0.19	0.22
Afternoon	374	0.54	0.55
Evening rush	251	0.29	0.32
Night	191	0.21	0.19

Model 3: grouping regression by stability

Independent variables for selection in linear regression or GAM include five continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature) and two factor variables (roadway type, time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different stability classes and such differences had limited influence upon the

prediction accuracy. Table 5.19 lists the grouping regression results with their optimal models.

Table 5.19. Prediction Performance for Grouping Particle Number Concentration by Atmospheric Stability

Stability	Samples	R^2 for linear regression	R^2 for GAM
A,B,C,D	1311	0.38	0.43
E,F,G	352	0.41	0.52

Model 4: Final regression models and model validation

Predictor variables in the linear regression model include four continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature) and two factor variables (roadway type, time of day);

Predictor variables in GAM include four continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature) and two factor variables (roadway type, time of day).

Table 5.20 gives the coefficients in linear model, degree of freedom in GAM and the variances explained for PNC modeling. Table 5.21 presents the results of independent holdout test and cross validation.

Table 5.20. Coefficients Regressed and Variance Explained for the Prediction of Particle Number

	Coefficients Linear model	Degree of freedom GAM	Variance explained (%)	
			Linear model	GAM
Intercept	162.0			
Vehicle speed	0.29	5.8*	15.44	2.64
VMT_AADT	0.000317	14.33*	10.28	4.21
Lanes		1*		0.21
Ambient temperature		28.7*		13.21
Time of day			17.83	26.60
Roadway type			3.9	15.80
Total variance explained			47.45	62.67

Note: * indicates degree of freedom; gray color indicates the categorical variables as factors.

Table 5.21. Independent 1/3 Hold-out and 3x3 Cross Validation of Predictive Models for Particle Number.

	Linear regression			Generalized additive model		
	General	3 times CV	1/3 test	General	3 times CV	1/3 test
N Samples	1789	1789	1199 vs. 590	1789	1789	1199 vs. 590
R^2	0.47	0.46	0.47	0.63	0.54	0.52
P. cor.	0.69*	0.67*	0.66*	0.78*	0.73*	0.77*
S. cor.	0.70*	0.68*	0.67*	0.76*	0.73*	0.76*

Note: general: no cross validation; 3 times CV: 3 times 3-fold cross validation; 1/3 test: two thirds used for training and one third used for test; P. cor.: Pearson correlation between the observed values and the predicted values; S.

cor.: Spearman's correlation between the observed values and the predicted values; * indicates statistical significance.

We also conducted a multi-collinearity diagnostic analysis: VIF of each predictor was < 10 (maximum VIF=4.51), so we can safely use the model: with vehicle speed: 1.64; VMT_AADT:2.33; Lanes: 2.31; hour: 1.37; roadway types: 4.56; truck count: 1.2; stability: 1.3; temperature: 1.14

5.7.4.3 $PM_{2.5}$ modeling

In these models, $PM_{2.5}$ has been log transformed to normalize its distribution. In the modeling of grouped variables, these variables were selected by their higher R^2 and their statistical significance at p-value<0.1.

An optimal scheme was to just use wet bulb temperature rather than dew point temperature to decrease multi-collinearity.

Model 1: grouping regression by roadway type

Independent variables for selection in linear regression or GAM include four continuous variables (ambient temperature, AADT, wet bulb temperature, wind speed) and one factor variable (time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different roadway types and such differences had limited influence upon the prediction. Table 5.22 lists the grouping regression results with their optimal models.

Table 5.22. Prediction Performance for Grouping $PM_{2.5}$ by Roadway Type

Roadway type	Samples	R^2 for linear regression	R^2 for GAM
Local road	419	0.66	0.79
Arterial road	543	0.69	0.84
One freeway	2805	0.68	0.75
Merging of 2 or more freeways	225	0.63	0.76

Model 2: grouping regression according to time of day

Independent variables for selection in linear regression or GAM include four continuous variables (ambient temperature, AADT, wet bulb temperature, wind speed) and one factor variable (roadway type). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different time of day and such differences had limited influence upon the prediction accuracy. Table 5.23 lists the grouping regression results with their optimal models.

Table 5.23. Prediction Performance for Grouping $PM_{2.5}$ by Time of Day

Time of day	Samples	R^2 for linear regression	R^2 for GAM
Early morning	72	0.20	0.14
Morning rush hour	406	0.47	0.55
Mid-morning	64	0.73	0.76
Noon	346	0.77	0.84
Afternoon	153	0.78	0.79
Evening rush hour	173	0.75	0.82
Night	268	0.78	0.82

Model 3: grouping regression according to stability

Independent variables for selection in linear regression or GAM include four continuous variables (ambient temperature, AADT, wet bulb temperature, wind speed) and one factor variable (time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different stability classes and such differences had limited influence upon the prediction accuracy. Table 5.24 lists the grouping regression results with their optimal models.

Table 5.24. Prediction Performance for Grouping $PM_{2.5}$ by Stability Class

Stability	Samples	R^2 for linear regression	R^2 for GAM
A,B,C,D	1212	0.66	0.73
E,F,G	850	0.39	0.53

Model 4: Final regression models and model validation

Predictor variables in the linear regression model include three continuous variables (temperature, wet bulb temperature, wind speed) and one factor variable (time of day);

Predictor variables in GAM include four continuous variables (temperature, aadt, wet bulb temperature, wind speed) and one factor variables (time of day).

Similarly, Table 5.25 gave the coefficients in linear model, degree of freedom in GAM and variance explained for $PM_{2.5}$ modeling. Table 5.26 presents the results of independent holdout test and cross validation.

Table 5.25 Coefficients Regressed and Variance Explained for the Prediction of $PM_{2.5}$.

	Coefficients in Linear model	Degree of freedom in GAM	Variance explained (%)	
			Linear model	GAM
Intercept	0.0243			
Ambient temperature	0.0502	8.74	34.34	7.87
AADT		5.428		1.01
Wet bulb temperature	0.188	8.72	11.84	31.08
Wind speed	-0.088		0.93	2.56
Time of day			20.59	30.31
Total variance explained			67.7	72.83

Note: * indicates degree of freedom; gray color indicates the categorical variables.

Table 5.26. Independent 1/3 Holdout and 3x3 Cross Validation of Predictive Models for $PM_{2.5}$.

	Linear regression			Generalized additive model		
	General	3 times CV	1/3 test	General	3 times CV	1/3 test
Samples	2062	2062	1385 vs. 677	2062	2062	1385 vs. 677
R^2	0.68	0.67	0.67	0.73	0.72	0.72
P. cor.	0.82*	0.82*	0.82*	0.86*	0.84	0.85*
S. cor.	0.77*	0.77*	0.76*	0.79*	0.78*	0.77*

Note: general: no cross validation; 3 times CV: 3 times 3-fold cross validation; 1/3 test: two thirds used for training and one third used for test; P. cor.: Pearson correlation between the observed values and the predicted values; S. cor.: Spearman's correlation between the observed values and the predicted values; * indicates statistical significance.

We also conducted a multi-collinearity diagnostic analysis: VIF of each predictor was < 10 (max VIF=9.73), so we can safely use the model: with temperature: 3.42; AADT:7.38; Lanes: 7.26; hour:1.82; wet bulb: 9.72.

5.7.4.4 NO_X modeling

In these models, NO_x has been square root transformed to normalize its distribution. In the modeling of grouped variables, these variables were selected by their higher R^2 with their statistical significance at p-value<0.1.

Model 1: grouping regression according to roadway type

Independent variables for selection in linear regression or GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature, wind speed) and one factor variable (time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different roadway types and such differences had limited influence upon the prediction accuracy. Table 5.27 lists the grouping regression results with their optimal models.

Table 5.27. Prediction Performance for Grouping NOx by Roadway Type

Roadway type	N Samples	R^2 for linear regression	R^2 for GAM
Local road	503	0.31	0.62
Arterial road	620	0.31	0.66
One freeway	3933	0.25	0.34
Merging of 2 or more freeways	281	0.23	0.54

Model 2: grouping regression by time of day

Independent variables for selection in linear regression or GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature, wind speed) and one factor variable (roadway type). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different time of day and such differences had limited influence upon the prediction accuracy. Table 5.28 lists the grouping regression results with their optimal models.

Table 5.28. Prediction Performance for Grouping NOx by Time of Day

Time of day	N Samples	R^2 for linear regression	R^2 for GAM
Early morning	72	0.38	0.65
Morning rush hour	460	0.50	0.58
Mid-morning	814	0.51	0.66
Noon	392	0.30	0.32
Afternoon	160	0.44	0.63
Evening rush hour	204	0.41	0.48
Night	396	0.35	0.43

Model 3: grouping regression by stability

Independent variables for selection in linear regression or GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature, wind speed) and two factor variables (roadway type, time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different stability classes and such differences had limited influence upon the prediction accuracy. Table 5.29 lists the grouping regression results with their optimal models.

Table 5.29. Prediction Performance for Grouping NOx by Stability

Stability	Samples	R^2 for linear regression	R^2 for GAM
A,B,C,D	1508	0.47	0.49
E,F,G	990	0.43	0.49

Model 4: Final regression models and model validation

Predictor variables in the linear regression model include five continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature, wind speed) and two factor variables (roadway type, time of day). Predictor variables in GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, truck count, ambient temperature) and two factor variables (roadway type, time of day). Table 5.30 gives the coefficients in linear model, degree of freedom in GAM and variance explained for NO_X modeling. Table 5.31 presents the results of independent holdout test and cross validation.

We also conducted a multi-collinearity diagnostic analysis: VIF of each predictor was < 10 (max VIF=2.02), so we can safely use the model: with vehicle speed: 1.46; VMT_AADT:1.617; Lanes: 2.02; hour: 1.18; roadway types: 1.48; truck count: 1.153; wind speed 1.154; temperature: 1.088.

Table 5.30. Coefficients Regressed and Variance Explained for the Prediction of NO_x.

	Coefficients in Linear model	Degree of freedom in GAM	Variance explained (%)	
			Linear model	GAM
Intercept	7.218			
Vehicle speed	0.018	12.355	16.22	2.33
VMT_AADT	0.0000154	17.4	8.29	3.77
Lanes		7.2		0.50
Ambient temperature	0.0358	27.8	0.0078	5.6
Wind speed	-0.3223		0.25	6.42
Truck count				2.23
Time of day			5.19	16.92
Roadway type			10.25	11.93
Total variance explained			40.21	49.7

Note: * indicates degree of freedom; gray color indicates the categorical variables as factors.

Table 5.31. Independent Holdout and 3x3 Cross Validation of Predictive Models for NO_x.

	Linear regression			Generalized additive model		
	General	3 times CV	1/3 test	General	3 times CV	1/3 test
N Samples	4446	4446	2974 vs. 1472	4446	4446	2974 vs. 1472
R^2	0.40	0.40	0.36	0.50	0.47	0.44
P. cor.	0.634*	0.63*	0.60*	0.66*	0.67*	0.66*
S. cor.	0.635*	0.63*	0.59*	0.66*	0.68*	0.66*

Note: 3 times CV: 3 times 3-fold cross validation; 1/3 test: two thirds used for training and one third used for test; P. cor.: Pearson correlation between the observed values and the predicted values; S. cor.: Spearman's correlation between the observed values and the predicted values; * indicates statistical significance.

5.7.4.5 BC modeling

In these models, BC has been square root transformed to normalize its distribution. In the modeling of grouped variables, these variables were selected by their higher R^2 and their statistical significance at p-value<0.1.

Model 1: grouping regression according to roadway type

Independent variables for selection in linear regression or GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature, wet bulb, wind speed) and one factor variable (time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different roadway types and such differences had limited influence upon the prediction accuracy. Table 5.32 lists the grouping regression results with their optimal models.

Table 5.32. Prediction Performance for Grouping BC by Roadway Type

Roadway type	N Samples	R^2 for linear regression	R^2 for GAM
Local road	366	0.23	0.47
Arterial road	230	0.48	0.72
One freeway	3208	0.32	0.42
Merging of 2 or more freeways	226	0.37	0.61

Model 2: grouping regression by time of day

Independent variables for selection in linear regression or GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature, wind speed, wet bulb) and one factor variable (roadway type). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different time of day and such differences had limited influence upon the prediction accuracy. Table 5.33 lists the grouping regression results with their optimal models.

Table 5.33. Prediction Performance for Grouping BC by Time of Day

Time of day	N Samples	R^2 for linear regression	R^2 for GAM
Early morning	65	0.52	0.85
Morning rush hour	362	0.38	0.54
Mid-morning	705	0.40	0.54
Noon	326	0.12	0.34
Afternoon	135	0.33	0.68
Evening rush hour	183	0.15	0.26
Night	327	0.23	0.43

Model 3: grouping regression by stability

Independent variables for selection in linear regression or GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, ambient temperature, wind speed, wet bulb) and two factor variables (roadway type, time of day). They were selected by the higher R^2 and statistical significance at p-value<0.1. There were small differences for the independent variables selected in the models of different stability classes and such differences had limited influence upon the prediction accuracy. Table 5.34 lists the grouping regression results with their optimal models.

Table 5.34. Prediction Performance for Grouping BC by Stability

Stability	Samples	R^2 for linear regression	R^2 for GAM
A,B,C,D	1249	0.39	0.45
E,F,G	854	0.34	0.47

Model 4: Final regression models and model validation

Predictor variables in the linear regression model include five continuous variables (vehicle speed, VMT_AADT, ambient temperature, wind speed, wet bulb) and two factor variables (roadway type, time of day);

Predictor variables in GAM include six continuous variables (vehicle speed, VMT_AADT, number of lanes, wet bulb, wind speed, ambient temperature) and two factor variables (roadway type, time of day)

Table 5.35 gives the coefficients in linear model, degree of freedom in GAM and variance explained for BC modeling. Table 5.36 presents the results of independent holdout test and cross validation.

Table 5.35. Coefficients Regressed and Variance Explained for the Prediction of BC.

	Coefficients in Linear model	Degree of freedom in GAM	Variance explained (%)	
			Linear model	GAM
Intercept	-6.88			
Vehicle speed	0.022	2.65	4.41	1.23
VMT_AADT	0.000051	4.74	4.06	2.72
Ambient temperature	1.81	6.6	1.87	5.31
Wind speed	-3.365		1.76	3.37
Wet bulb	2.89	6.7	7.35	3.52
Time of day			12.99	19.81
Roadway type			4.18	6.64
Total variance explained			36.63	42.6

Note: * indicates degree of freedom; gray color indicates the categorical variables as factors.

Table 5.36. Independent Holdout and 3x3 Cross Validation of Predictive Models for BC.

	Linear regression			Generalized additive model		
	General	3 times CV	1/3 test	General	3 times CV	1/3 test
N Samples	2103	2103	1410 vs. 693	2103	2103	1410 vs. 693
R^2	0.37	0.30	0.34	0.43	0.31	0.39
P. cor.	0.61*	0.56*	0.58*	0.67*	0.56*	0.62*
S. cor.	0.62*	0.57*	0.61*	0.67*	0.58*	0.63*

Note: 3 times CV: 3 times 3-fold cross validation; 1/3 test: two thirds used for training and one third used for test; P. cor.: Pearson correlation between the observed values and the predicted values; S. cor.: Spearman's correlation between the observed values and the predicted values; * indicates statistical significance.

We also conducted a multi-collinearity diagnostic analysis: VIF of each predictor was < 10 (max VIF=1.89), so we can safely use the model: Vehicle speed: 1.42; VMT_AADT: 1.60; Lanes: 1.88; hour: 1.27; roadway types: 1.50; wind speed 1.18; temperature: 1.83; wet bulb 1.89.

5.7.5 Time series analysis

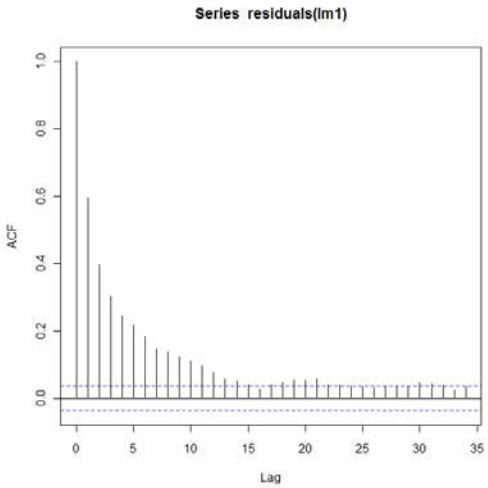
Table 5.37 lists temporal autocorrelations by temporal lag using the training data to train the model. Lag refers to minutes, i.e. Lag *n* indicating *n* minutes lagged. Figure 5.21 presents temporal and partial temporal autocorrelations for PAH, PNC, $PM_{2.5}$, NO_X, and BC.

Given the temporal autocorrelation shown in Table 5.37 and Figure 5.21, we controlled for temporal autocorrelation at lag 1 in models and made final predictions of concentrations for PAH, PNC, $PM_{2.5}$, NO_X and BC. The result (Table 5.38) shows that the prediction with incorporation of first order autocorrelation (AR1) resulted in a significant improvement in the model's R^2 (ranging from 20% to 48.9%), and Pearson or Spearman's correlation for PAH, PNC, NO_X and BC, but not for $PM_{2.5}$.

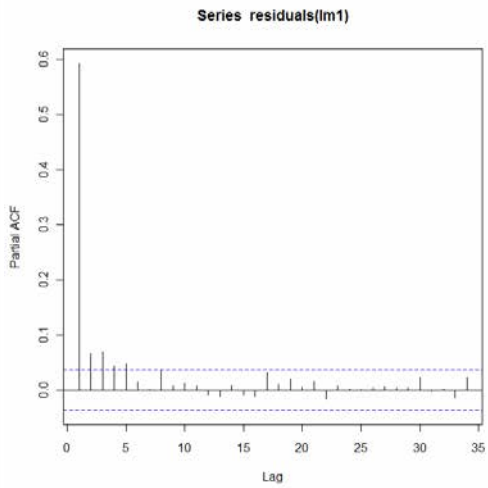
Table 5.37. Temporal Autocorrelation among Different Daily Lags

Type	#	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
PAH	3879	0.59*	0.40*	0.30*	0.25*	0.22*	0.18*	0.15*	0.14*	0.12*	0.11*
PNC	2161	0.69*	0.56*	0.48*	0.41*	0.33*	0.29*	0.24*	0.24*	0.21*	0.21
PM2.5	2062	0.75*	0.62*	0.56*	0.45*	0.41*	0.38*	0.36*	0.35*	0.35*	0.33*
NOX	5337	0.68*	0.47*	0.37*	0.31*	0.24*	0.22*	0.20*	0.18*	0.14*	0.13*
BC	4130	0.73*	0.50*	0.40*	0.33*	0.27*	0.25*	0.23*	0.17*	0.16*	0.15*

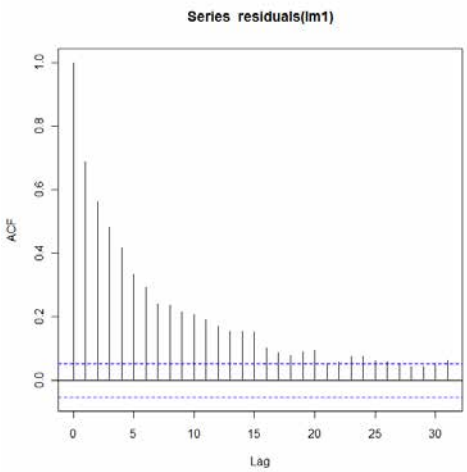
Note: # number of samples; * indicates statistical significance



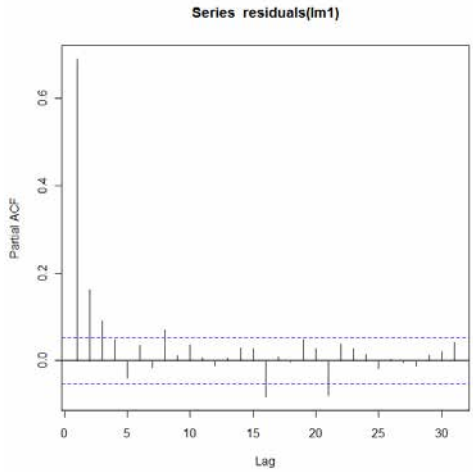
a. Autocorrelation for PAH



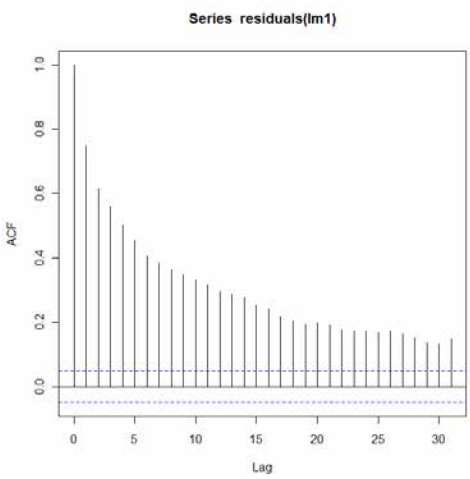
b. Partial autocorrelation for PAH



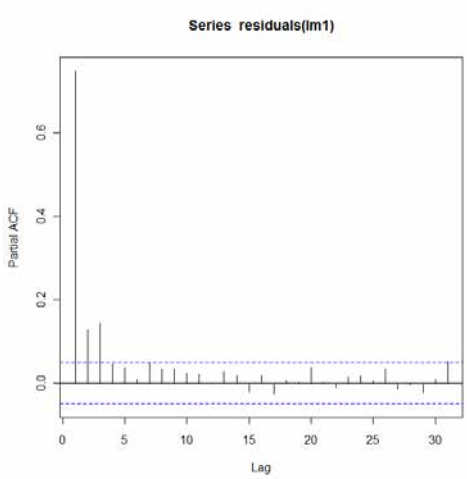
c. Autocorrelation for PNC



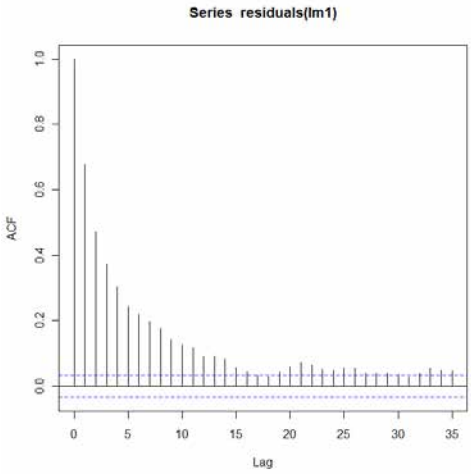
d. Partial autocorrelation for PNC



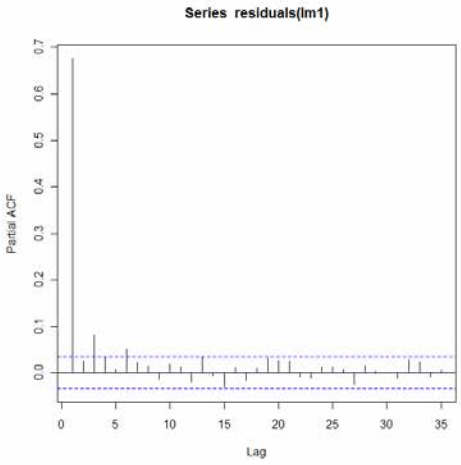
e. Autocorrelation for PM2.5



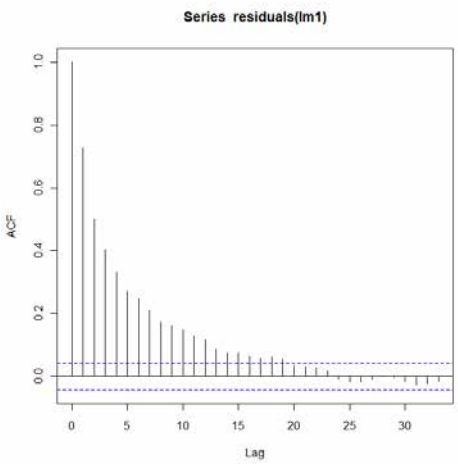
f. Partial autocorrelation for PM2.5



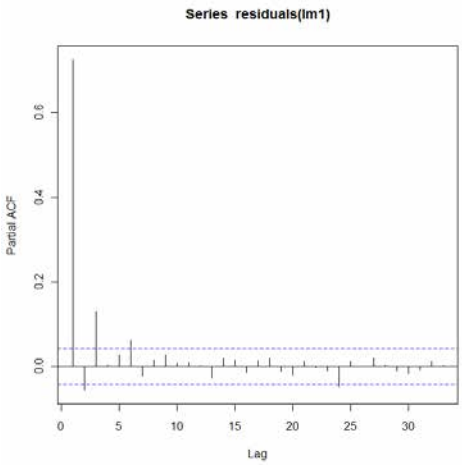
g. Autocorrelation for NOX



h. Partial autocorrelation for NOX



i. Autocorrelation for BC



j. Partial autocorrelation for BC

Figure 5.21: Autocorrelation and partial-autocorrelation autocorrelogram for the residuals from the ordinary least squares (OLS) regression of concentrations

Table 5.38. Evaluation of the Time Series Models Constructed.

Type	Samples	Model's output			Independent test (25% of data)		
		R^2	P.Cor	S.Cor	R^2	P.Cor	S.Cor
PAH	3879	0.62	0.8*	0.80*	0.59	0.78*	0.76*
PNC	2161	0.68	0.85*	0.84*	0.67	0.84*	0.84*
$PM_{2.5}$	2062	0.78	0.89*	0.88*	0.73	0.88*	0.87*
NO_X	5337	0.66	0.83*	0.84*	0.60	0.82*	0.82*
BC	4130	0.64	0.80*	0.81*	0.57	0.77*	0.79*

With all the results of linear regression and GAM, we also calculated the shrinkage on 3x3 cross validation using equation 5.13 in Table 5.39.

Table 5.39. Shrinkage on 3x3 Cross Validation of Predictive Time Series Models for the Air Pollutants.

Concentration	Model	R^2	CV R^2	shrinkage on cross-validation
PAH	LM	0.43	0.42	0.01<0.1
	GAM	0.51	0.43	0.08<0.1
PNC	LM	0.48	0.46	0.02<0.1
	GAM	0.63	0.54	0.09<0.1
PM2.5	LM	0.68	0.67	0.01<0.1
	GAM	0.73	0.72	0.01<0.1
NOx	LM	0.40	0.40	0<0.1
	GAM	0.50	0.47	0.03<0.1
BC	LM	0.37	0.30	0.07<0.1
	GAM	0.43	0.39	0.04<0.1

5.7.6 Discussion

5.7.6.1 Correlation analysis and scatter plots

Our results show that each independent variable had varying correlations with the dependent air pollutant variables (Table 5.7 and 5.8). Overall, vehicle speed, roadway type, AADT were moderately or highly positively correlated with PAH, PNC and NO_x and BC. For PAH, PNC, and NO_x, meteorological factors such dew point, wet bulb temperature, relative humility, wind speed and direction had little relation to them. For $PM_{2.5}$, traffic-related factors had less influence than meteorological parameters. This is expected since local traffic emissions were a major source of PAH, PNC, NO_x and BC while $PM_{2.5}$ has been shown to be more of a regional pollutant with less local traffic contributions (10,11).

Scatter plots with linear and non-linear regression lines (Figure 5.13-5.17) helped us examine the linear or non-linear relationships between the independent and dependent variables. For example, although time of day (hour) represents a negative correlation with concentration, the variation of concentrations is non-linear along a day's timeline as illustrated in the scatter plots (Figure 5.13-5.17 scatter plots of hour with transformed measured values of concentrations). Thus, using non-linear smooth function to fit the term of day-time can improve the model's prediction.

5.7.6.2 Influence of roadway types

As expected, we observed much higher concentrations for PAH, PNC, NO_X and BC (not $PM_{2.5}$) on freeways than arterials and local roads, mainly due to more vehicles and higher speed of vehicles on freeways. The grouping statistics (*t* student and Wilcoxon statistics) also showed that the difference between freeways and non-freeways was statistically significant for PAH, PNC, NO_X and BC. Furthermore, such a difference in concentrations between freeways and non-freeways were great as shown in Table 5.9. In the final regression models, the R^2 for roadway type was 9.2% (linear regression) or 9.6% (GAM) for PAH (Table 5.15), 3.9% (linear regression) or 15.8% (GAM) for PNC (Table 5.20), 10.25% (linear regression) or 11.93% (GAM) for NO_X (Table 5.30) and 4.18% (linear regression) or 6.64% (GAM) for BC (Table 5.35). However, we found that roadway type was not an effective predictor variable for $PM_{2.5}$.

5.7.6.3 Influence of time of day

Time of day had a negative correlation with concentrations of PAH, PNC, NO_x, BC and $PM_{2.5}$ (Table 5.7 and 5.8, -0.182 to -0.368 for Pearson's correlation and -0.287 to -0.364 for Spearman's correlation). Scatter plots also showed the regular pattern of concentration along a day's timeline (Figure 5.13-5.17). Furthermore, in the final regression models, the R^2 for time of day explained 7.44% (linear regression) or 23.25% (GAM) for PAH, 17.83% (linear regression) or 26.60% (GAM) for PNC, 20.59% (linear regression) or 30.31% (GAM) for $PM_{2.5}$, 5.2% (linear regression) or 16.92% (GAM) for NO_X and 12.99% (linear regression) or 19.81% (GAM) for BC. Time of day showed the greatest R^2 among all predictor variables for any air pollutant.

The grouping statistics (*t* student and Wilcoxon statistics) for the concentrations in the morning vs. non-morning showed statistically significant differences in concentrations (Table 5.10 and Figure 5.19). Our analysis showed the highest concentration of PAH, PNC, $PM_{2.5}$, NO_X or BC in the morning. Time of day is a significant variable for the prediction of all five air pollutant concentrations in the present study. This is expected because time of day reflects diurnal variations in both traffic activity patterns and meteorological parameters.

5.7.6.4 Influence of traffic variables

Traffic variables are expected to be a critical influential variable since it is a major emission source for PAH, PNC, NO_X and BC. The PeMS five-minute traffic counts and estimated truck counts was not a sufficient predictor of traffic because it only covered a small part of the study routes and periods.

Annual average daily traffic (AADT) or VMT_AADT is the total volume of vehicle traffic of a highway or road for a year divided by 365 days. Given its greater completeness, we found AADT or VMT_AADT, as an alternative to 5-minute PeMS traffic counts, was an effective predictive variable. VMT_AADT was selected for final regression models given its somewhat stronger correlation and had an R^2 of 8.72% (linear regression) or 2.56% (GAM) for PAH, 10.28% (linear regression) or 4.21% (GAM) for PNC, 1.01% (GAM) for $PM_{2.5}$, 8.29% (linear regression) or 3.77% (GAM) for NO_X and 4.06% (linear regression) or 2.72% (GAM) for BC. This suggests that VMT_AADT is a moderately predictive variable.

5.7.6.5 Influence of meteorological factors

For PAH, PNC and NO_X, meteorological factors such as wind speed, wind direction, temperature, dew point temperature, and relative humility had slight or moderate influence on their prediction. Among these variables, temperature plays a more significant role, with an R^2 of 0.13% (linear regression) or 5.89% (GAM) for PAH, 13.21% (GAM) for PNC, 34.34% (linear regression) or 7.87% (GAM) for $PM_{2.5}$, and 0.0078% (linear regression) or 5.60% (GAM) for NO_x, and 1.87% (linear regression) or 5.31% (GAM) for BC.

However, meteorological factors such as wet bulb or dew point temperatures were good predictor variables for $PM_{2.5}$ since log-transformed $PM_{2.5}$ was highly correlated with both (>0.5, Table 5.7). Furthermore, meteorological factors improved the predictability of $PM_{2.5}$ better than the other pollutants (overall model R^2: 0.67 for linear regression and 0.72 for GAM, Table 5.26). Our results also showed that meteorological factors such as wet bulb temperature had a moderate influence on BC, illustrating that both were simultaneously affected by traffic and meteorological factors.

Wind speed and direction was weakly correlated with the air pollutant concentrations. Wind speed always showed a negative correlation with air pollutant concentrations and was an effective predictor. The combination of wind speed and wind direction as a categorical variable (20 categories) did not improve the prediction. This is likely because we relied on hourly wind data from distant monitoring stations rather than real-time wind data nearby the sampling locations.

Stable atmosphere is more favorable to higher concentrations of air pollutants than unstable atmosphere. Our student *t* and Wilcoxon statistics tests also showed that the difference in concentrations of PAH, PNC, $PM_{2.5}$ and NO_X between stable atmospheres and unstable atmospheres is statistically significant (Table 5.11). However, in our final models we did not include stability due to lack of statistical significance in the predictive model. This may be due to the inaccuracy of the stability data (e.g. modeled data with large uncertainty, every 3 hour and 40 km by 40 km resolution). The other possible reason is that the stability data may correlate with time of day that did remain in the model, e.g. early morning more stable and mid afternoon less stable. However, the variance inflation factor (VIF) for stability and time of day in the model was not high (1.19 for PAH, 1.4 for PNC, 2.02 for $PM_{2.5}$, 1.24 for NO_X, 1.37 for BC).

5.7.6.6 Linear vs. non-linear models

GAM can incorporate both linear and non-linear relationships as well as factor variables in the model. If the scatter plot presented a clear non-linear relationship, we used the smooth function of GAM to fit such a non-linear relationship (illustrated in Figure 5.13-5.17 for temperature's relationship with air pollutant concentrations). Using GAM, we were able to model the complex non-linear relationships while keeping the linear predictive and categorical variables as factors in the model. In our final prediction model, although linear regression and GAM did not use the same set of variables and thus could not be compared directly, GAM still provided a better prediction accuracy than linear regression (R^2: GAM's 0.52 vs. LM's 0.43 for PAH in Table 5.16; GAM's 0.63 vs. LM's 0.48 for PNC in Table 5.21; GAM's 0.73 vs. LM's 0.68 for $PM_{2.5}$ in Table 5.26; GAM's 0.50 vs. LM's 0.40 for NO_X in Table 5.31; GAM's

0.43 vs. LM's 0.37 for BC in Table 5.36). Generally GAM could improve the prediction over linear regression by approximately 5% to 10%. Even though linear regression in R also incorporates factor variables and the same variables without consideration of statistical significance, GAM achieved somewhat better predictive accuracy. In order to avoid overfitting in GAM, we needed to first use scatter plots to detect possible non-linear relationships. For a more complex non-linear relationship, we set a higher degree of freedom to simulate the practical relationship. Thus, a more precise linear or non-linear relationship can be established in GAM and the overfitting can be minimized.

5.7.6.7 Validation of predictive models

We conducted 3x3-fold cross validation and holdout independent tests (Tables 5.14, 5.19, 5.24, 5.29 and 5.34) for linear regression and GAM. For linear regression with inclusion of temporal autocorrelation, the independent holdout test was done (¾ data used for training and ¼ data for test). Table 5.39 shows the shrinkage on 3x3 cross validation. This table shows that the difference in R^2 is less than 0.1, demonstrating that the model was valid and all of the observations for PAH, PNC, $PM_{2.5}$, NO_X and BC can be used to estimate regression coefficients for the final prediction equations.

5.7.6.8 Consideration of temporal autocorrelation

We explored temporal autocorrelation using autoregression for the continuous time data of PAH, PNC, $PM_{2.5}$, NO_X and BC (Table 5.37 and Figure 5.21). The results show that, as expected, there was significant temporal autocorrelation of residuals. The improvement of models with adjustment of temporal autocorrelation (AR1) was considerable for all air pollutants. In our independent holdout tests (using ¾ data for training and ¼ data for test, Table 5.38), R^2 improved by about 28.3% from 0.46 to 0.59 for PAH, by about 42.5% from 0.47 to 0.67 for PNC, by about 9.0% from 0.72 to 0.73 for $PM_{2.5}$, by about 36.4% from 0.34 to 0.60 for NO_X, and by about 46.2% from 0.39 to 0.57 for BC.

With inclusion of an AR1 parameter, the improvement in the predictions from linear regression was great for PAH, PNC, NO_X or BC, ranging from 28.3% to 46.2%. However, the application of the time-series model may be limited in epidemiological studies since we need to know the concentration of the last one or few minutes to predict the current concentration and such information is usually difficult to acquire. In practical applications, we may use the average of concentration over more minutes to construct the regression model to decrease the influence of temporal autocorrelation on the model. For our study, the measured values were limited, and averaging over five minutes resulted in fewer samples and consequently models with lower R^2. This might also be due to more variance of roadways and meteorological factors within the interval of five minutes since the measurement was done mainly on freeways. After a series of tests, we finally chose one minute as the averaging interval to generate the samples. For our regression models, averaging over one minute generated more samples, decreased the variance of predictor variables within the one-minute interval and had a good predictive performance in comparison with averaging over five minutes. With incorporation of one-minute temporal autocorrelation, the prediction was further improved.

5.8 SUMMARY AND CONCLUSIONS

In this section, we have examined the influence of a variety of explanatory variables, continuous or categorical, including traffic variables, meteorological factors and time of day, modeling methods and temporal autocorrelation upon the prediction of the on-road concentration of five pollutants, i.e. PAH, PNC, $PM_{2.5}$, NO_X, and BC. Final prediction models showed the variance explained ranged from 37% to 73% depending on the pollutant and modeling method (linear or nonlinear). The missing data did not clearly impact the accuracy of the models for different air pollutants, since the variance explained was not greater for NOx, with the least missing data, compared with pollutants with the most missing data (PNC, $PM_{2.5}$ and BC). Nevertheless, considerably more data on a larger number of roadway types could be have improved the models.

Our study found that on-road concentrations of any of the five pollutants usually peaked in the morning, gradually lowered in the noontime and afternoon hours, and returned to a moderate value at night. The time of a day was one of the most important influential factors for the prediction of the five pollutants, explaining a considerable part (5.2%-30.3%) of the total variance. Traffic-related factors such as roadway type, AADT or VMT_AADT, lanes and freeways had a moderate and significant influence on prediction of the traffic-derived pollutants (i.e. PAH, PNC, NO_X and BC), but not for $PM_{2.5}$, which was much more affected by meteorological factors such as wet bulb, dew point and ambient temperature.

In terms of modeling methods, linear regression and non-linear regression (GAM) had different predictive performance. Given the non-linear relationship between partial explanatory variables and the target variable of air pollutant concentration demonstrated in the scatter plots, GAM is preferred to linear model. After being trained, GAM can also be used to make predictions of future data.

Using the time series of one-minute averages of measured concentrations over multiple dates, we examined temporal autocorrelation of residuals and incorporated it in the final prediction. Statistically significant temporal autocorrelation was clearly observed (Table 5.37 and Figure 5.21) and adjustment of such residual autocorrelation improve prediction considerably for PAH, PNC, NO_X and BC (Table 5.38). Although this improved the predictions, such methods may be impractical in epidemiological studies since it is difficult to obtain minute-by-minute measured values.

For validation, we used 3x3 cross validation or/and independent holdout tests for the different models. For linear regression and GAM without consideration of temporal autocorrelation, we used cross validation and holdout test. The results showed that the difference between cross validation and holdout test was small (<0.1) and our predictive model was valid for the prediction of other data. For linear regression with incorporation of temporal autocorrelation, the independent holdout test (¾ data for training and ¼ data for the test) showed that the temporal autocorrelation contributed to the improvement in prediction.

5.9 REFERENCES

1 Finkenstadt, B., Held, L., Isham, V., 2007. Statistical Methods for Spatio-Temporal Systems. Chapman & Hall/CRC, New York.

2 Chen, L., Bao, Z., Kong, S., Han, B., You, Y., Ding, X., Du, S., Liu, A., 2010. A land use regression for predicting NO2 and PM10 concentrations in different seasons in Tianjin region, China. Journal of Environmental Science 22, 1364-1373.

3 Clougherty, E.J., Wright, J.R., Baxter, K.L., Levy, I.J., 2008. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. Environmental Health 7, 1-14.

4 Crouse, L.D., Goldberg, S.M., Ross, A.N., 2009. A prediction-based approach to modelling temporal and spatial variability of traffic-related air pollution in Montreal, Canada. Atmospheric Environment 43, 5075-5084.

5 Hoek, G., Beelen, R., Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric Environment 42, 7561–7578.

6 Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Analysis and Environmental Epidemiology 15, 185-204.

7 Su, G.J., Jerrett, M., Beckerman, B., Wilhelm, M., Ghosh, K.J., Ritz, B., 2009. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy Environmental Research 109, 657-670.

8 Wu, J., Wilhelm, M., Chung, J., Ritz, B., 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. Environmental Research, doi: 10.1016/j.envres.2011.1003.1008.

9 Kleinbaum, D.G., Kupper, L.L., Muller, K.E., 1988. Applied Regression Analysis and Other Multivariate Methods, 2 ed. Kent Pub Co., Boston.

10 Gomiscek, B., Hauck, H., Stopper, S., Preining, O., 2004. Spatial and temporal variations Of PM1, PM2.5, PM10 and particle number concentration during the AUPHEP-project. Atmos. Environ. 38, 3917–3934.

11 Russell, M., Allen, D.T., Collins, D.R., Fraser, M.P., 2004. Daily, seasonal, and spatial trends in PM2.5 mass and composition in Southeast Texas. Aerosol Sci. Technol. 38, 14–26.

6. CHAPTER SIX. TASK 5: VALIDATE THE IN-VEHICLE EXPOSURE MODEL FOR PAH AGAINST MEASUREMENTS FROM REPRESENTATIVE SUBJECTS.

6.1. MATERIALS AND METHODS

Overview: Data from a group of human subjects who carried personal PAH samplers were used as a first test of the predictive ability of variables identified from the models developed in Task 4. These data are from a study of working subjects in real world driving conditions and the data are collected after one week of sampling that was unattended by research technicians (NIH, NIEHS R21 ES016379, Wu). It serves as pilot study to determine the approach needed to validate the models developed in Tasks 1-4 for use in human subjects.

The vehicles that subjects usually used are listed in Table 6.1.

Table 6.1. Subject Vehicles.

Year	Make	Model
1994	Nissan	Altima
2007	Honda	Accord
1999	Chevrolet	Camero
2000	Nissan	Maxima
2007	Ford	Edge
2001	Honda	Accord EX
2001	Mercedes	430E
2002	Chevy	Tahoe
2007	Toyota	Camry
2003	Honda	CRV
2009	Honda	Civic
2005	Mazda	3
2000	Dodge	Avenger
1999	Mercury	Sable
2001	Toyota	4Runner
2004	Volkswagon	Jetta
2003	Nissan	Altima
2006	Lexus	IS 250
2002	Daewoo	Leganza
2004	Saturn	Ion 2
2004	Chevy	Avalanche
1999	Mitsubishi	Diamente
1999	Mercury	Mystique
1999	Mitsubishi	Galant
2006	Toyota	Prius
2008	Jeep	Cherokee
unknown	VW	Jetta
unknown	Mitsubishi	Eclipse
1998	Lexus	ES 300
2009	Ford	Focus

Subjects: We acquired personal particle-bound PAH data (EcoChem PAS) for 25 women. Subjects were part of an NIH-funded pilot study of time-activity and air pollution exposure assessment during pregnancy among 90 women using GPS (NIEHS R21 ES016379, Wu, with supplemental funding from NICHD National Children's Study LOI3-ENV-4-A #14, Delfino and Wu) (1,2). The goal was to collect up to three weeks of personal GPS and PAH data at different periods (trimesters or the postnatal period). Sixteen subjects had one week of data, 7 subjects had 2

weeks of data, and 2 subjects had 3 weeks of data (36 total weekly series). Primarily due to equipment malfunction and noncompliance, not all subjects had a full 7 days of GPS and PAH data. In total, there were 175 person-days of personal PAH data with concurrent GPS measurements.

The average age of subjects was 28 (± 5 years) ranging from 18-36 years. Nine subjects were Hispanic, 7 white non-Hispanic, 6 Asian, and 3 other race. 24% had family incomes < $30,000 per year, 56% were married, 32% had a college diploma and 16% had a graduate school diploma.

Analysis

Two types of regression models were used to assess the consistency of the methods, autoregressive models and generalized estimating equations (GEE). Since analyses involved time series, autoregressive models were tested using the SAS Autoreg procedure to estimate the autoregressive parameter(s). Autocorrelation in the residuals was investigated using autocorrelation functions and partial autocorrelation functions to lag 15 minutes and with the Durbin-Watson statistic (3). We found that controlling for first order autocorrelation (AR1) was sufficient to adjust for autocorrelated error terms. GEE models were tested in the SAS Genmod procedure, which additionally allows for an autoregressive within-subject correlation structure (4).

Two model approaches were used to predict PAH data. Approach 1 assessed the relation of one-minute PAH to time-variant predictor data (one-minute if possible) (8785 observations) whereas Approach 2 additionally assessed the relation of one-minute PAH to each subject's reported vehicle data (36 weekly exposure assessment series as described below).

For Approach 1 we tested regression models to predict the subject's one-minute average in-vehicle PAH exposure. Predictor data for Approach 1 is largely described above in Task 4 including total traffic counts, roadway type (merging to freeways, freeways, major arterials, and minor surface streets or local roads), truck route, number of roadway lanes, VMT_AADT, time of day, meteorological data, and vehicle speed. Time of day was categorized as morning rush hour: 06:00am -- 09:00am; mid morning: 09:00am --12:00pm; noon: 12:00pm -- 02:00pm; afternoon: 02:00pm -- 05:00pm; evening rush hour: 05:00pm -- 07:00pm; and all other times: 07:01pm -- early 06:00am. The time-stamped GPS data were linked with the TeleAtlas roadway map to derive roadway data. Vehicle speed (Km/h) was estimated from GPS data as described above. Meteorological data included stability class, ambient temperature, ambient relative humidity and wind speed. Stability class was from A to G with A indicating most unstable and G most stable. The above data were then used to construct regression models based on the one minute data.

For Approach 2 we added largely time-invariant data on vehicle type, mileage, and use of vehicle by the subject to assess average recirculation/outside air (RC/OA) conditions and fan use. Based on above results (Table 5.1) we classified vehicles into vehicle manufacturer region: Asian, German/other, and the referent category is US manufacturers, which had the highest AER. For missing odometer mileage in 7 subjects, we replaced the missing data with the US annual average for females of

12,000 miles (http://www.fhwa.dot.gov/ohim/onh00/bar8.htm) times the age of the car. Questionnaire data on vehicle type, use of vehicle by the subject to assess average recirculation/outside air (RC/OA) conditions and fan use was coded as follows.

RC/OA conditions: We used a fixed ordinal classification at each weekly series when the same questionnaire was administered. The coding from the questionnaire is given below. Basically, we gave each response an equal weight of +1 if exposure is expected to increase as a result of increased AER as described in above sections, or weight of -1 if exposure is expected to decrease as a result of decreased AER. The questionnaire below shown that if a subject would change one or more of their answers for the different weather conditions if they were in traffic, then a single positive or negative point would be scored.

Ventilation Fan Setting: We did the same as above with 8 specific responses. The coding from the questionnaire is given below.

The following questions ask how you typically cool and heat the inside of your car. Please check the boxes to indicate the usual way you or the driver cools or heats the car you use for your main mode of car transportation (check all that apply down the column).

Please check ALL that apply.				
	When it is hot outside?	***When it is cold outside?***	***When the weather is mild?***	***When I am in busy traffic or on a freeway?*** ☐ Same as answers to left for when it is hot or cold outside (leave boxes below blank) ☐ Different than answers to left for when it is hot or cold outside (fill in appropriate boxes below)
Coding for RC/OA				
Recirculation	☐= -1	☐= -1	☐= -1	☐= +1
Outside air, no recirculation	☐= +1	☐= +1	☐= +1	☐= -1
Open windows	☐= +1	☐= +1	☐= +1	☐= -1
Close window	☐= -1	☐= -1	☐= -1	☐= +1
Coding for Ventilation Fan Setting				
Lower fan setting	☐= -1	☐= -1	☐= -1	☐= +1
Higher fan setting	☐= +1	☐= +1	☐= +1	☐= -1

6.2. RESULTS AND DISCUSSION

Univariate Statistics

One-minute PAH data were positively skewed (skew = 2.06) (Figure 6.1) and were normalized using a square root transformation (skew = 0.48). The untransformed PAH mean was 100 ng/m^3 (±108), median 67, range 0-1078. The subject weekly session average PAH data was normally distributed (Figure 6.2). The PAH mean was 102 ng/m^3 (± 53), median 88, range 14-240.

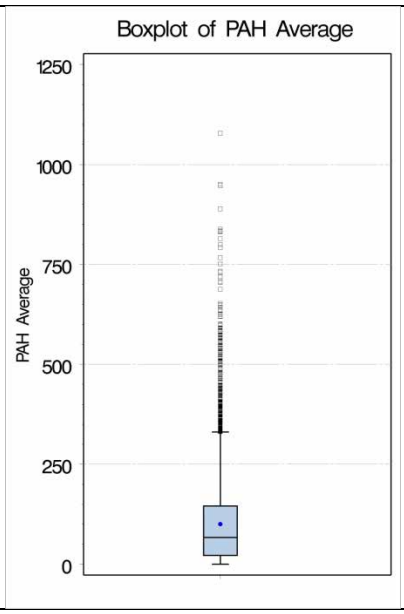


Figure 6.1. Box plot for 1-minute average PAH concentrations (N=8785 in 25 subjects).

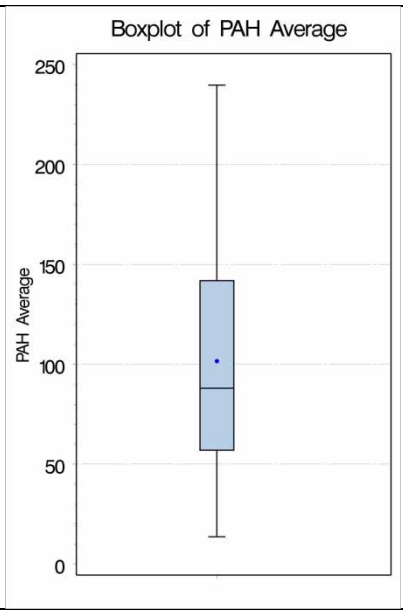


Figure 6.2. Box plot for series average PAH concentrations (N=36 weekly series in 25 subjects).

Of the 8785 one-minute PAH measurements, subjects were on freeways or major highways for 1733 minutes (19.24%), on a freeway or highway ramp for 536 minutes (5.95%), on a major surface street for 3321 minutes (36.87%) and on a minor surface street for 3417 minutes (37.94%). This is distinctly different than proportion of time on various routes used for Tasks 1-4 where a majority of data was collected on freeways. The average speed was 42 ± 36 km/hr.

Regression Models

Table 6.2 shows results of the full model for Approach 1 including all variables and both the GEE model and the autoregressive model (both fit with an AR1 parameter). Results were similar for the two modeling approaches. In autoregression models, stability class, VMT for 1000 m buffers, number of roadway lanes, wind speed, ambient temperature and truck route were not significantly associated with in-vehicle PAH. This also included VMT 500m ($p < 0.35$, not shown). These results were the same in the GEE models except for wind speed, which was inversely associated with PAH. Significantly higher predicted PAH was observed during the early morning and late afternoon to early evening rush hours. A linear increase in exposure with vehicle speed was observed relative to slow speeds (< 5 Km/h). An autoregression model estimate for continuous speed alone was 0.0141 ng/m^3 per 1.0 increase in Km/h. Roadway type was significantly associated with PAH with higher PAH for larger highways compared with the referent category of minor surface streets.

The overall R^2 for the ordinary least squares (OLS) model was small (0.10) whereas for the autoregressive model the R^2 increased to 0.51, but the R^2 due to regression dropped to 0.02. This suggests that autocorrelation explained most of the variance. A regression of predicted versus observed PAH from the full GEE model in Table 6.2, adjusted for autocorrelation, revealed an R^2 of 0.10.

We then added the individual-level fixed predictors to the regression analysis of one-minute data (Approach 2). As discussed, the fixed predictors were those related to vehicle characteristics and questionnaire-reported typical use of vehicle ventilation (RC/OA) and fan. We found that none of these variables were significantly associated with PAH in GEE models (Table 6.3). However, in autoregressive models we found that a higher score for the RC/OA variables (indicating increased AER) was significantly positively associated with PAH as expected. The adjusted R^2 for the OLS models was 0.12 whereas for the autoregressive models the regression R^2 was 0.03 and total R^2 was 0.51, which is nominally better than the model without the time invariant predictors described above (Table 6.2). Furthermore, in autoregressive models we found that vehicle type was significantly associated with PAH in the expected direction with lower exposures in German and Asian vehicles than in US vehicles, which is again consistent with the results in the above tasks. Odometer mileage was also significant in autoregressive models, however, not in the expected direction since higher mileage was associated with lower PAH.

Table 6.2. Multivariate regression models for the prediction of particulate PAH: continuously measured or estimated predictors

Parameter	GEE Model				Autoregression Model			
	Regression Coefficient	Standard Error	Z value	*p*-value	Regression Coefficient	Standard Error	T value	*p*-value
Intercept	6.8432	2.8123	2.43	0.0150	7.1521	1.0220	7.00	<.0001
Time of Day:[a]								
6:00-9:00	1.8075	0.6532	2.77	0.0057	1.7826	0.3303	5.40	<.0001
9:00-12:00	1.0002	0.7380	1.36	0.1753	0.9801	0.3853	2.54	0.0110
12:00-14:00	0.8913	0.6307	1.41	0.1576	0.8538	0.4041	2.11	0.0346
14:00-17:00	1.4719	0.4977	2.96	0.0031	1.5117	0.3484	4.34	<.0001
17:00-19:00	1.6616	0.4896	3.39	0.0007	1.6626	0.3728	4.46	<.0001
Speed 5-15 Km/h	0.4438	0.2153	2.06	0.0393	0.5268	0.1386	3.80	0.0001
Speed ≥15 Km/h	0.9266	0.2546	3.64	0.0003	1.0484	0.1243	8.43	<.0001
Stability E,F,G[a]	-0.0306	0.3267	-0.09	0.9253	-0.0541	0.2268	-0.24	0.8115
VMT 1000m	0.0033	0.0026	1.27	0.2058	0.0039	0.0026	1.50	0.1331
No. of Lanes	-0.0233	0.0480	-0.49	0.6266	-0.0182	0.0348	-0.52	0.6012
Wind Speed (m/s)	-0.4484	0.3299	-1.36	0.1741	-0.4947	0.1260	-3.93	<.0001
Roadway Type:[a]								
Major Street	0.3392	0.1165	2.91	0.0036	0.3633	0.1114	3.26	0.0011
Freeway Ramp	0.4289	0.1608	2.67	0.0077	0.4690	0.1861	2.52	0.0118
Freeway/Major Highway	0.5483	0.1388	3.95	<.0001	0.5705	0.1619	3.52	0.0004
Amb. Temp (°C)	0.0349	0.1051	0.33	0.7400	0.0200	0.0338	0.59	0.5539
Truck Route (Y/N)	0.1138	0.3310	0.34	0.7310	0.1378	0.2344	0.59	0.5567

[a] Referent categories were: Time of Day, 19:00 – 6:00; Stability, A-D; Roadway Type, minor surface streets.

Table 6.3. Multivariate regression models for the prediction of particulate PAH: continuously measured or estimated predictors plus time-invariant subject-reported vehicle characteristics.

Parameter	GEE Model				Autoregression Model			
	Regression Coefficient	Standard Error	Z value	*p*-value	Regression Coefficient	Standard Error	t value	*p*-value
Intercept	7.9428	2.7245	2.92	0.0036	8.1933	1.1107	7.38	<.0001
Time of Day:[a]								
6:00-9:00	1.9688	0.6351	3.10	0.0019	1.9640	0.3300	5.95	<.0001
9:00-12:00	1.1556	0.7310	1.58	0.1139	1.1544	0.3850	3.00	0.0027
12:00-14:00	0.9900	0.5893	1.68	0.0929	0.9693	0.4048	2.39	0.0167
14:00-17:00	1.6218	0.4320	3.75	0.0002	1.6749	0.3509	4.77	<.0001
17:00-19:00	1.9108	0.4777	4.00	<.0001	1.9551	0.3736	5.23	<.0001
Speed 5-15 Km/h	0.4382	0.2150	2.04	0.0415	0.5180	0.1386	3.74	0.0002
Speed ≥15 Km/h	0.9178	0.2528	3.63	0.0003	1.0308	0.1244	8.29	<.0001
Stability E,F,G[a]	0.0528	0.3061	0.17	0.8631	0.0438	0.2295	0.19	0.8485
VMT 1000m	0.0036	0.0024	1.51	0.1315	0.004379	0.002600	1.68	0.0922
No. of Lanes	-0.0177	0.0454	-0.39	0.6959	-0.0119	0.0348	-0.34	0.7327
Wind Speed (m/s)	-0.5329	0.3352	-1.59	0.1119	-0.5773	0.1321	-4.37	<.0001
Roadway Type:[a]								
Major Street	0.3422	0.1155	2.96	0.0031	0.3650	0.1115	3.27	0.0011
Freeway Ramp	0.4105	0.1667	2.46	0.0138	0.4424	0.1861	2.38	0.0174
Freeway/Major Highway	0.5174	0.1465	3.53	0.0004	0.5285	0.1618	3.27	0.0011
Amb. Temp (°C)	0.0513	0.1106	0.46	0.6429	0.0375	0.0362	1.04	0.3001
Truck Route (Y/N)	0.0764	0.3266	0.23	0.8149	0.0929	0.2344	0.40	0.6918
RC_OA score	0.3240	0.3073	1.05	0.2917	0.3102	0.0985	3.15	0.0016
Fan Use score	0.1680	0.3666	0.46	0.6468	0.1619	0.1289	1.26	0.2092
Vehicle type[a]								
Asian	-1.2051	1.0098	-1.19	0.2327	-1.2306	0.2595	-4.74	<.0001
German	-1.4601	1.5316	-0.95	0.3404	-1.4390	0.5684	-2.53	0.0114
Odometer (miles)	-0.0027	0.0016	-1.64	0.1012	-0.002534	0.001189	-2.13	0.0331

[a] Referent categories were: Time of Day, 19:00 – 6:00; Stability, A-D; Roadway Type, minor surface streets; Vehicle type, U.S. manufacturers.

6.3 SUMMARY AND CONCLUSIONS

In Task 5, using data from 25 subjects during one to several weeks of exposure assessment, we examined the predictive ability of model variables also tested in the other tasks. Only PM PAH was tested for subjects mainly due to the instrument portability and the importance of PAH as markers of traffic emissions. Future work with separate funding is envisioned to test the other modeled pollutants in similar epidemiologic settings.

For the one-minute models to predict personal in-vehicle PM PAH, although many predictors were significant and in the direction anticipated, the overall predictive power of the PAH models was low (only 12% for OLS models) compared to the models from the technician-administered in-vehicle testing for Tasks 1-4 (37% to 73% depending on the pollutant and model). This is likely due to the fact that the on-road and in-vehicle measurements for Tasks 1-4 were done by field technicians and much of the predictive data were collected in real time by staff who planned and well-documented all travel routes and procedures. In addition, most of the data for Tasks 1-4 were collected on freeways, with higher and perhaps more predictable emission sources. In contrast, for Task 5 subjects were relied upon to carry the personal samplers and in-vehicle exposure times were extracted using GPS data from the total personal exposure dataset. Other predictors such as vehicle characteristics, the use of RC/OA and fans were not scripted as in Tasks 1-4 but were derived from baseline questionnaires collected before each of a subject's weekly exposure series.

Also, we did not specifically collect in-vehicle exposure data only. Although our time-activity model performed reasonably well in identifying in-vehicle travel points based on personal GPS data (approximately 88% sensitivity, 99% specificity, and 86 % precision) (1), there were still misclassified in-vehicle travel points, especially for points with a relatively low speed. To reduce the impact of misclassified in-vehicle travel points, we visually checked and removed apparently erroneous points (e.g. scattered low-speed points that lasted for at least 10 minutes) before the modeling. Furthermore, the women may or may not have used the vehicle she usually drove during the sampling week, especially if families have more than one vehicle at home or if she carpooled with others. In addition, their actual driving patterns during the sampling week may also be different from answers in the baseline questionnaire data.

In Task 4, we developed and validated multiple models for in-vehicle PN (Part 1) and on-road BC, UFP number, $PM_{2.5}$, particle-bounded PAH, and NO_x (Part 2). Fitted models for each species were validated against measurement data collected in Tasks 1 and 2 using validation methods. The extensive model validations provided us with useful and rich information on model performances in predicting in-cabin concentrations of the selected pollutants under normal driving conditions in our study region. Predictions were strong in general.

However, our ultimate goal is to use the validated models to estimate in-vehicle exposure for a large number of subjects in epidemiological studies where personal exposure measurements are not feasible. The limited sample of Task 5 using only PM PAH in 25 subjects was informative, but exposure predictions were less than desired due to data limitations, which can be improved in future epidemiological studies. For example, we could monitor only in-vehicle exposures. We also could obtain vehicle and driving pattern information after or during the one-week of sampling, which would be more accurate than the information we obtained before the sampling in the present study.

6.4 References

1. Wu J, Jiang C, Houston D, Baker D, Delfino RJ. 2011. Automated Time Activity Classification Based on Global Positioning Systems (GPS) Tracking Data. Environmental Health 10:101.
2. Wu J, Jiang C, Jaimes G, Bartell S, Dang S, Baker D, Delfino RJ. Travel Patterns During Pregnancy: Comparison Between Global Positioning System (GPS) Tracking and Questionnaire Data, Submitted.
3. Shumway, R.H. 1988. Applied Statistical Time Series Analysis. Englewood Cliffs, NJ: Prentice Hall.
4. Diggle P, Heagerty P, Liang, KY and Zeger S. 2002. Analysis of Longitudinal Data. 2nd Edition. New York:Oxford University Press.

7. CHAPTER 7. STUDY LIMITATIONS

The model predictions (for AER and I/O ratios) may be less accurate for vehicles older than 15 years and at speeds exceeding 60 miles h^{-1} due to limited coverage of measured data for such conditions. Moreover, during real world driving, speeds are often unsteady and vary mainly due to differences in roadway type (unrestricted access freeways versus restricted access arterial roads) and traffic intensity (free flowing during off-peak hours versus congested during peak traffic hours). I/O ratios are dependent on and positively correlated with vehicle speed and this relation is stronger under recirculation (RC) ventilation setting than outside air intake condition (OA). Therefore, variable speed during trips results in a distribution of I/O ratios for the trip (which follow from the unsteady and varying AER), and will reflect the variation in speed if ventilation conditions are unchanged during the trip. (Predictive models presented in Chapter 5 quantify the effect of speed on AER and I/O ratios.)

To investigate the influence of unsteady speeds on I/O ratios during real world driving, I/O ratio distributions were calculated for typical speed frequency distribution during peak (4-5 PM) and off-peak traffic hours (4-5 AM) for both roadway types – freeways and arterial roads. The probability distributions for speed are plotted in Figure 7.1 below and are the same as input distributions in the EPA Motor Vehicle Emission Simulator (MOVES), with 16 speed bins covering a 0-75 miles h^{-1} range (http://www.epa.gov/otaq/models/moves/tools/averagespeedconverter_mobile6.xls). The hours plotted had the highest contrast. The weighted average speeds were 27 and 30 miles h^{-1}, respectively, for arterial roads during peak and off-peak traffic hours, and 27 and 52 miles h^{-1}, respectively, for freeways.

The resulting I/O ratio distributions are plotted in Figure 7.1 for a vehicle of median age (7 years for U.S. fleet, EPA 2010) operating at medium fan speed (fan strength = 0.5). The difference in average I/O ratio (frequency weighed) between peak and off-peak hours was less than 0.05 on arterial roadways, under both RC and OA ventilation modes. This difference was up to or less than 0.15 for freeways. Greater difference in peak and off-peak traffic speed distribution on freeways results in greater difference in I/O ratios. Generally, higher speeds during off-peak traffic hours resulted in higher I/O ratios. However, this does not imply higher resultant in-cabin concentrations. Higher pollutant concentration during peak traffic hours may more

than compensate for lower I/O ratios. In absolute terms, the difference in I/O ratios during peak and off-peak hours was greater for older vehicles, higher fan settings and under OA ventilation condition. However, in terms of percentage change in I/O ratio, the trend was reversed. The most significant change in I/O ratios was observed under RC ventilation mode for new vehicles operation at low fan settings.

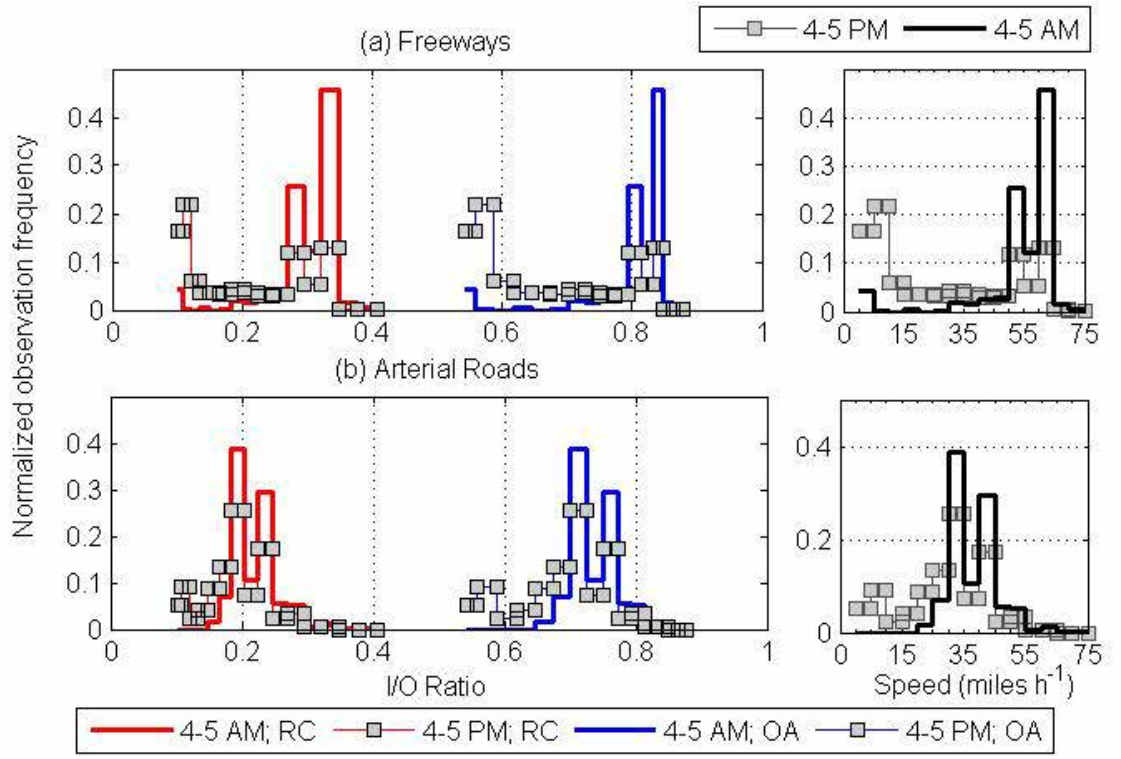


Figure 7.1: I/O ratio distribution in a median aged vehicle operating at medium fan setting corresponding to typical distribution of driving speed on urban freeways and arterial roads during peak (4-5 AM) and off-peak (4-5 PM) traffic conditions.

I/O ratio for the same vehicle (ventilation settings being constant) will be higher if vehicle is driven on freeways in comparison to that on arterial roads because of higher driving speeds. This effect of roadway type/increased speed is greater under recirculation setting than outside air setting, which follows from the stronger dependence and positive correlation of I/O ratios under RC mode on vehicle speed. Any other pattern, for example, the disparity between freeway and arterial I/O ratios being higher for newer cars compared to older cars, was not significant. The effect of difference in roadway type on I/O ratios seemed homogeneous for vehicles up to 15 years old and full range of fan settings under a ventilation mode for a speed distribution.

Furthermore, the difference between average I/O on freeways and arterial roads during peak traffic hour was negligible due to little to no difference in speed distribution on these two roadways during rush hours, about 0.01. However, during off-peak traffic hours the difference was significant, about 0.07-0.08. This has important implications for exposure assessment. During commute hours, which generally correspond to peak traffic hours, the aspect of roadway type that is of greater consequence and will drive the exposure concentration variability is the difference in pollutant concentration and not the travel speed associated with

different roadway types. However, during off-peak hours, the differences in travel speed on different roadway types will significantly contribute to the variability in exposure concentrations.

We conducted an extensive exposure measurement and modeling study. However, it was not feasible in the proposed work to cover all possible exposure conditions in vehicles. We consider this the main limitation of the study. For instance, we were not able to assess every day of the week (Mon vs. Friday) or every season and climatic condition. Model predictive power thus varied by species investigated (e.g., PN vs. PAH vs. BC). The different measurements were also not always highly correlated and so the different predicted concentrations may have varying associations with health outcomes in future epidemiologic analyses. This is anticipated based on the degree to which each measurement represents underlying toxicity or is subject to exposure error in the model predictions. We cannot predict that at this time. Finally, the limited sample of Task 5 using only PM PAH in 25 subjects was informative, but exposure predictions were less than desired due to data limitations, which can be improved in future epidemiological studies that test the models produced in this study.

8. CHAPTER 8. OVERALL SUMMARY AND CONCLUSIONS

For traffic-related pollutants (like ultrafine particles D_p < 100 nm), a significant fraction of overall exposure occurs within or close to the transit microenvironment because on-road concentrations of traffic-related pollutants are typically much higher than concentrations measured at ambient monitoring stations. Therefore, understanding exposure to these pollutants in such microenvironments is crucial to accurately assessing overall exposure. However, correct assignment of in-vehicle exposures requires a parameter the inside-to-outside ratio (I/O), which can be used to accurately modify roadway concentrations to reflect the combined effects of multiple mechanisms that either limit the pollutant penetration inside vehicles or cause in-cabin losses. For a given vehicle, the I/O ratio depends on parameters like ventilation setting preferences and driving speed that influence air exchange rate (AER). Air exchange rate (AER) is the dominant factor in affecting how close in-vehicle concentrations of traffic-related particulate pollutants come to equal on-road concentrations. Furthermore, the inter-vehicle differences in I/O (and AER), due to aging and make/manufacturer type, themselves are just as significant. As a result, I/O ratios can vary from nearly zero to one both within a singular vehicle (depending on ventilation mode and driving speed) and within a fleet. Despite this importance of AER in affecting in-vehicle particle exposures, few studies have characterized AER and all have tested a small number of cars.

We developed a simplified yet accurate method for determining AER using the occupants' own production of CO_2 (Task 1). By measuring initial CO_2 build-up rates and equilibrium values of CO_2 at fixed speeds, AER was calculated for 59 vehicles representative of California's fleet, thus producing the first accurate and representative characterization of vehicle AERs. AER measurements correlated and agreed well with the largest other study conducted (R^2 = 0.83). Multivariate models captured 70% of the variability in observed AER using only vehicle age, mileage, manufacturer and speed, all easily ascertainable information suitable for studies with

large numbers of subjects. We found that AER increases strongly with increasing vehicle age and mileage and increasing speed, and AER is high if windows are open or outside air ventilation settings are chosen. We suspect that vehicle age is acting as a surrogate for aging and failing seals and other factors allowing progressively greater AER as well as newer technology that may contribute to increased vehicle tightness and thus lower AER and lower in-cabin particle concentrations.

In-vehicle concentrations result from the interaction of on-road concentrations with vehicle characteristics that can reduce or remove the pollutants, depending on the pollutant and the vehicle AER. A sufficiently high AER results in in-vehicle concentrations equaling on-road particle mass and number concentrations while low AER tends to reduce particle mass and number concentrations. The actual pollutant removal rates are due to a complicated interplay between a vehicle's physical characteristics, ventilation condition, particle size, and changes in AER. To address this interplay, we made determinations of losses using on-road testing under realistic aerodynamic conditions (Task 2). We focused on ultrafine particle number concentrations, the particle pollutant with the highest and most widely-varying loss rates. Six vehicles were tested at different driving speeds, fan settings, cabin filter loadings, and ventilation conditions (outside air or recirculation). During outside air conditions, the fraction of particles removed averaged 0.33 ± 0.10 (SD). The fraction removed did not vary with vehicle speed but decreased at the higher ventilation flow rates of higher fan settings. During recirculation conditions, AER was much lower and removal fraction higher. Removal fraction averaged 0.83 ± 0.13 and was highly correlated with and was a strong function of AER. Under both ventilation condition types, particle removal was primarily due to losses unrelated to filtration. Filter condition, or even the presence of a filter, played a minor role in particle fraction removed. Based on these results, predictive models for in-vehicle ultrafine particle number concentration were developed, as described below.

To characterize on-road concentrations for Tasks 4-5 below, extensive on-road measurements were made in Tasks 1-2 on two arterial and three freeway routes covering 39 to 57 and 77 to 86 miles, respectively. Measurements of real-time black carbon, UFP, $PM_{2.5}$, NO, NO_2, CO, CO_2, and particle-bound PAH were made, with GPS and video to capture time, location, and surrounding traffic conditions.

For Task 3, fuel-based emission factors (EF) were calculated based on simultaneous on-road pollutant and CO_2 measurements (Task 3). Partitioned EFs were used to calculate freeway emission rates (ER). EFs for light-duty vehicles (LDV) were generally in agreement with the most recent studies but lower for heavy-duty vehicles (HDV), and significantly lower only for oxides of nitrogen (NO_x), probably resulting from the newer NOX emission standards for trucks operating at the ports in LA. Annually on I-710, a major truck route, 6.5% fraction of total vehicle miles travelled (VMT) is associated with HDV, but HDV were estimated to contribute 69% to total NO_x emissions. These differences in EFs by freeway segment and traffic mix were incorporated into our on-road concentration prediction models discussed below.

For Task 4, we first developed models for predicting in-cabin UFP concentrations if roadway concentrations are known, taking into account vehicle characteristics, ventilation settings, driving conditions and air exchange rates (AER). Particle

concentrations and AER were measured in 43 and 73 vehicles, respectively, under various ventilation settings and driving speeds. Multiple linear regression (MLR) and generalized estimating equation (GEE) regression models were used to identify and quantify the factors that determine inside-to-outside (I/O) ratios and AERs across a full range of vehicle types and ages. AER was the most significant determinant of UFP I/O ratios, as I/O was most strongly influenced by ventilation setting (decreased with recirculation vs. increased with outside air intake). Additional inclusion of ventilation fan speed, vehicle age or mileage, and driving speed explained greater than 80% of the variability in measured UFP I/O ratios across a representative sample of the California fleet and across the expected range of normal driving conditions. The I/O ratios measured under RC conditions were far lower than those under OA conditions due to lower AERs under RC. The median I/O ratio value at RC was 0.11 (inter-quartile range: 0.07-0.22) compared to 0.66 at OA (inter-quartile range: 0.53-0.80). The median AER value at RC was 6.0 h^{-1} (inter quartile range: 3.6-10 h^{-1}) compared to 63 h^{-1} for OA (inter quartile range: 47-83 h^{-1}).

We also developed and validated predictive models for on-road concentrations of particle-bound PAH, UFP, $PM_{2.5}$, NO_X and BC (Task 4) that can be combined with subject information on vehicle use and our predictive models for AER to evaluate exposure to in-vehicle pollutants. The on-road measured data were one-minute averaged and compiled with traffic variables (traffic volume, roadway type, number of lanes), on-road or ambient meteorological factors (dew point, wet bulb, relative humility, ambient temperature, atmospheric stability) and time of day to develop linear regression models and non-linear generalized additive models. We found that time of day was a statistically significant predictor of the six pollutants, accounting for a considerable part of the variance explained (5.2%-30.3%). Traffic variables such as VMT_AADT, roadway type, and number of lanes were significant for the traffic-derived pollutants but not $PM_{2.5}$. $PM_{2.5}$ is a regional pollutant and meteorological factors were stronger predictors than the traffic variables. Final prediction models showed the variance explained ranged from 37% to 73% depending on the pollutant and modeling method (linear or nonlinear). Adjustment for temporal autocorrelation of residuals led to a modest improvement in prediction. Models were shown to be valid using 3x3 cross validation and independent holdout validation.

We examined the predictive ability of model variables tested in above tasks for human subjects using personal particle-bound PAH exposure for 25 subjects during their time in the in-vehicle environment (derived from personal GPS data). Although many predictors from Task 4 were significant and in the direction anticipated, the overall predictive power of models for our human subjects data was low (R^2 around 0.1) compared to the models from the technician-administered testing for Tasks 1-4. Perhaps the most important factors that limited our predictive power were that the time on freeways was limited (unlike work in Tasks 1-4) and predictors such as ventilation conditions that determine AER were not scripted but were derived from baseline questionnaires on general preferences collected before each of a subject's weekly exposure series. Because our studies demonstrated the overwhelming importance of AER on in-vehicle exposure, our limited ability to predict our human subject exposures indicates the importance of accurately characterizing AER and ventilation setting preference. Our baseline questionnaires were not designed to do this, although this would be easily remedied in future work. The limited sample size

of Task 5 using only PM PAH in 25 subjects was informative, but exposure predictions were less than desired due to these data limitations.

Exportability of the Findings:
Most of the previous in-vehicle exposure studies conducted in California (1-3) are characterization studies, which provide helpful information on in-cabin concentrations under various conditions but are difficult to quantitatively apply to other exposure and epidemiological studies. One of our major goals and the strength in this study is to generalize the measurement data to a larger population by developing models based on information of vehicle (age and type), roadway (type), traffic activity (total and diesel traffic count), emissions, driving pattern (AC use, window position), meteorology (mixing height, temperature, relative humidity), seasonality, and time of day. These models will provide helpful information to other researchers on the impact potency of various parameters. They can also be adapted and applied to other regions after validation using local measurement data.

To this end, prediction of particle exposure inside vehicles requires determining ventilation setting first and foremost (i.e., OA or RC), due to its large impact on AER. Under OA conditions, fan setting is the most dominant variable, and I/O ratio was approximately 0.6 and fairly independent of speed. Under RC conditions, I/O ratio has a large range and varies from 0.5 to zero, depending on AER, which can be predicted by speed and vehicle age. Under open window conditions, I/O ratios approach one, i.e., in-cabin concentrations frequently equal roadway concentration. Difficult to obtain information, such as the state of in-cabin filter loading, does not appear to be a crucial factor in assessing I/O ratios and the resulting in-vehicle particle exposures. It also does not appear that changes in on-road size distribution have a large impact on I/O ratios. Therefore, all of the variables needed to estimate I/O ratio within 10% or less can be obtained through questionnaires given to vehicle owners.

Other studies that have measured UFP I/O ratios include Pui et al. (2008) (4) and Qi et al. (2008) (5), who investigated in-cabin air filter efficiency, and found large UFP reductions inside two new vehicles under recirculating conditions, but neither AER nor speed were reported. Zhu et al. (2007) (1) also reported large reductions in in-vehicle UFP concentrations in three cars, but under conditions of variable speeds, and AER was variable and only estimated indirectly from concentration change time lags. Therefore, prior to Knibbs et al. (2010) (6), no in-vehicle UFP I/O ratio results have been reported in a form that can be generalized. Using a CPC 3007 (the same instrument used by us) Knibbs et al. (2010), reported a high correlation between I/O ratio for UFP and AERs ($r^2 = 0.81$). In the present study, at RC the r^2 between I/O ratio and AER was 0.80, indicating that AER is the most significant determinant of I/O ratio at RC ventilation conditions. Additional comparisons of our study with the results of Knibbs et al. (2010) (6) and others are extensively discussed in Chapters 2-5. Other recent studies (Gong et al., 2009; Xu et al., 2009) (7-8) that have produced mechanistic models of particle losses inside vehicles relied on parameters like particle penetration through cracks and surface deposition rates, which while useful, are not obtainable outside of a laboratory setting.

However no study until this one has made it possible for a researcher to *estimate* in-vehicle AER or UFP I/O ratios instead of having to *measure* them directly. A novel contribution of this study is the empirical model for predicting AER and UFP I/O ratios. The ability to estimate AER and UFP I/O using models from the present work makes it possible for exposure scientists or epidemiologists to predict drive-time in-vehicle UFP exposure based on estimated on-road UFP concentrations and the collection of survey data regarding the characteristics of vehicles used by the study population. It bypasses the reliance on time-consuming measurements and mechanistic models (such as those described above), which are unfeasible in an epidemiologic study, and makes it possible to generate predictions for large cohorts based on simple information such as vehicle make, age and ventilation choice.

Furthermore, it must be recognized that we did not develop total personal exposure models. Exposures in other microenvironments were not assessed in the study, but could be addressed in future epidemiologic research that may use the models developed in this study. Given the potentially large impact of in-vehicle exposures, the impact of that type of exposure alone on health outcomes should be investigated. This is similar to the usual approach of assessing exposure only at the home address but ignoring exposure at other locations. Nevertheless, there is great potential to incorporate the predicted in-vehicle exposures using the proposed data with other microenvironmental exposure data to predict total personal exposure.

We envision that time-activity data (GPS, questionnaire, etc.) will also assist us in assessing exposure at other locations, including commuting routes as discussed. It is possible that among actual study subjects in an epidemiologic study, travel route data will not accurately reflect exposures during all times. For example, a subject quits work after the time-activity assessment and this information is not made available, or a subject uses mass transit that cannot be assessed using the proposed model data. Although this may be a limitation in a cohort or other epidemiologic study, we have not proposed to test model accuracy in actual study subjects followed over long periods. Nevertheless, the next stage will be to test models in a sample of subjects in a cohort study using the available study tools as compared with a gold standard (e.g., long-term GPS not otherwise feasible in the overall cohort). This type of validation study is well beyond the scope of the present study and it would require extensive resources to conduct. We anticipate that such a study will be conducted given the supportive data obtained in the present study.

References

1. Rodes, C., Sheldon, L., Whitaker, D., Clayton, A., Fitzgerald, K., Flanagan, J., et al. 1998. Measuring Concentrations of Selected Air Pollutants Inside California Vehicles. Final Report. Contract No. 95-339. Sacramento, CA: California Air Resources Research Division Board.
2. Westerdahl, D.; Fruin, S.; Sax, T.; Fine, P.M. and Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. *Atmos. Environ*. 2005, 39, 3597–3610.

3. Zhu, Y., Eiguren-Fernandez, A., Hinds, W.C., Miguel, A.H., 2007. In-Cabin Commuter Exposure to Ultrafine Particles on Los Angeles Freeways. *Environ Sci Technol* 41(7): 2138-2145.
4. Pui, D. Y. H.; Qi, C.; Stanley, N.; Oberdorster, G. and Maynard, A. Recirculating air filtration significantly reduces exposure to airborne nanoparticles. Environ. Health Perspect. 2008, 116, 863-866.
5. Qi, C.; Stanley, N.; Pui, D. Y. H. and Kuehn, T. H. Laboratory and on-road evaluations of cabin air filters using number and surface area concentration monitors. Environ. Sci. Technol. 2008, 42, 4128-4132.
6. Knibbs, L.D., de Dear, R.J. and Morawska, L., 2010. Effect of cabin ventilation rate on ultrafine particle exposure inside automobiles. Environ. Sci. Technol. 44, 3546-3551.
7. Gong, L.; Xu, B. and Zhu, Y. Ultrafine particles deposition inside passenger vehicles. Aerosol Sci. Technol. 2009, 43, 544-553.
8. Xu, B. and Zhu, Y. Quantitative analysis of the parameters affecting in-cabin to on-roadway (I/O) ultrafine particle concentration ratios. Aerosol Sci. Technol. 2009, 43, 400-410.

9. CHAPTER 9. RECOMMENDATIONS

Results of this study can be used to conduct epidemiologic investigations of the health effects of in-vehicle exposures to air pollutants, but additional testing of exposure models may be needed. In particular, additional subjects with varying commute behaviors would be needed to validate predictive models for the range of exposures assessed in this study.

Additional research is needed in consort with health studies so that the relevance of this potentially important source of traffic-related air pollution exposure can be assessed. Given the long duration of commuting for the average Californian, adverse health impacts could be great and completely unrecognized to date. Therefore, it is important at this early stage to make the public aware of the potential for high air pollutant exposure while traveling in vehicles and what are the basic characteristics of the vehicle and the occupants' use of the ventilation system that most affect this exposure. The present data combined with previous literature is sufficient to provide this information.

10. LIST OF PUBLICATIONS PRODUCED

Fruin S, Hudda N, Sioutas C, Delfino RJ. Predictive model for vehicle air exchange rates based on a large, representative sample of vehicles in California. Environ Science Technol, 2011;45:3569-3575.

Hudda N, Kostenidou E, Delfino RJ, Fruin SA, Sioutas C. Vehicle and driving characteristics that influence in-cabin particle number concentrations. Environ Science Technol 2011;45:8691-8697.

Hudda N, Eckel SP, Knibbs LD, Sioutas C, Delfino RJ, Fruin SA. Linking in-vehicle ultrafine particle exposures to on-road concentrations. Atmos Environ, in press 5-2012.

Hudda N, Fruin SA, Delfino RJ, Sioutas C. Vehicle emission factors using on-road measurements: a new approach. Journal of Aerosol Science, submitted 03-2012.

Kam W, Liacos J, Schauer JJ, Delfino RJ, Sioutas C. On-road emission factors of PM pollutants for light-duty vehicles (LDVs) based on real-world urban street driving conditions. Submitted to Atmospheric Environment, March 2012.

Kam W, Liacos J, Schauer JJ, Delfino RJ, Sioutas C. Size-segregated composition of particulate matter (PM) in major roadways and surface streets. *Atmospheric Environment*, in press.

Li L, Wu J, Hudda N, Sioutas C, Fruin SA, Delfino RJ. Modeling the concentrations of on-road air pollutants in southern California. Submission in preparation.

Liacos J, Kam W, Schauer JJ, Delfino RJ, Sioutas C. Concentrations of organic, metal and trace element PM species and derivation of freeway-based emission rates in Los Angeles, CA. Submitted to Science of Total Environment, March 2012.