Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Environ Sci Technol. Author manuscript; available in PMC 2015 March 18.

Published in final edited form as:

Environ Sci Technol. 2014 March 18; 48(6): 3272–3280. doi:10.1021/es404838k.

An hourly regression model for ultrafine particles in a near-highway urban area

Allison P. Patton[*,†], **Caitlin Collins**[†], **Elena N. Naumova**[†,±], **Wig Zamore**[‡], **Doug Brugge**[±], and **John L. Durant**[†]

[†]Department of Civil and Environmental Engineering, School of Engineering, Tufts University, Medford, Massachusetts 02155, United States

[‡]Somerville Transportation Equity Partnership, Somerville, Massachusetts 02145, United States

[±]Department of Public Health and Community Medicine, School of Medicine, Tufts University, Boston, Massachusetts 02111, United States

Abstract

Estimating ultrafine particle number concentrations (PNC) near highways for exposure assessment in chronic health studies requires models capable of capturing PNC spatial and temporal variations over the course of a full year. The objectives of this work were to describe the relationship between near-highway PNC and potential predictors, and to build and validate hourly log-linear regression models. PNC was measured near Interstate 93 (I-93) in Somerville, MA (USA) using a mobile monitoring platform driven for 234 hours on 43 days between August 2009 and September 2010. Compared to urban background, PNC levels were consistently elevated within 100–200 m of I-93, with gradients impacted by meteorological and traffic conditions. Temporal and spatial variables including wind speed and direction, temperature, highway traffic, and distance to I-93 and major roads contributed significantly to the full regression model. Cross-validated model R^2 values ranged from 0.38–0.47, with higher values achieved (0.43–0.53) when short-duration PNC spikes were removed. The model predicts highest PNC near major roads and on cold days with low wind speeds. The model allows estimation of hourly ambient PNC at 20-m resolution in a near-highway neighborhood.

Keywords

regression; near-highway; ultrafine particles; mobile monitoring; particle number

[*]Corresponding Author: 200 College Ave, Medford, MA 02155, Phone: (617)-627-3211, Fax: (617)-627-3994, Allison.Patton@tufts.edu.

The authors declare no competing financial interest.

Supporting Information

Additional information regarding variables and modeling decisions as described in the text is available free of charge via the Internet at http://pubs.acs.org.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

Ultrafine particles (UFP, <100 nm in aerodynamic diameter) may contribute to increased risks of respiratory and cardiovascular disease for people living near highways and major roadways because they are present at high levels in vehicle exhaust, carry toxic chemicals sorbed to their surface, and rapidly cross biological barriers.[1–3] One challenge in characterizing health risks of near-highway UFP is assessing the spatial and temporal variability of UFP with sufficient accuracy to measure associations with health outcomes. In studies of acute health effects, UFP exposure is estimated using temporal data from centrally-located monitoring sites; however, because the highest concentrations of UFP generally occur <200 m from roadways, central-site measurements may underestimate exposures for the most highly exposed populations.[4,5] Studies of chronic health effects require exposure models that describe intra-urban distance-decay gradients and how gradients change over time because people move between locations of high and low exposures and their integrated exposure may not be directly correlated with annual mean residential concentrations. Therefore, in assigning UFP exposures for epidemiological studies in urban areas, tools are needed that can accurately predict both the temporal and spatial variations of UFP.[6]

Regression modeling based on air pollution measurements and spatial and temporal covariates (land-use regression; LUR) is one approach to estimate traffic-related pollutant exposures.[7–17] This approach is useful in cases where there is a relatively large amount of pollutant data but source emission factors have not been fully characterized and limited microscale meteorological data is available. Covariates including densities of traffic, population, and land use have been used in regression models to estimate annual-average spatial distributions of air pollutants including $PM_{2.5}$, NO_2, and BC at local (<5 km^2) to regional and international (>10,000 km^2) scales.[7,8,10–13] Regional UFP models have been developed[10,18] but only a few studies have reported regression models for UFP in near-road, urban neighborhoods.[11,14–16] In these studies, near-road UFP models were based on particle number concentration (PNC; a proxy for UFP) measurements collected either at fixed locations (*e.g.*, residences) over the course of several days or months,[8,14,15] or by repeated mobile monitoring of an urban area over a single season.[11,16] Intensive mobile monitoring, which can provide ambient measurements with higher temporal and spatial resolution than is typically achieved using only centralized monitoring sites,[19–23] has been used to develop LUR models.[7,11,16] Mobile monitoring in all seasons is necessary to capture the seasonality that has been observed in UFP.[20,23,24] While models have been used to predict *either* temporal or spatial patterns in UFP distribution, epidemiological studies of urban near-highway UFP exposure require models that capture *both* fine-scale temporal changes and spatial gradients to reflect variations observed in UFP measurements.

Our goal was to measure PNC with a mobile-monitoring platform over the course of a year in a near-highway urban area and develop a PNC model based on temporal (1 hr) and spatial (~20 m) covariates to inform exposure estimates. Here we describe (1) the relationship between PNC and potential covariates; and (2) the construction and validation of hourly log-linear PNC regression models.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Methods

Mobile Monitoring

Mobile monitoring was conducted with the Tufts Air Pollution Monitoring Laboratory (TAPL) along a 15.4-km route in a 1.4-km^2 area in Somerville, MA (USA) adjacent to Interstate 93 (I-93; average daily traffic (ADT) ~150,000 vehicles/day,[25] 1–5% diesel;[26] Figure 1). I-93 rises from grade to ~6 m above street level and is filled underneath except at underpasses. A 3-m-high noise barrier runs along ~400 m of the east side of I-93.[23] Two state highways run at grade through the study area: a four-lane highway adjacent to I-93 (Route 38, ADT = 40,000 vpd), and a six-lane highway that crosses underneath I-93 (Route 28, ADT = 50,000 vpd).[25] Local traffic consists of <5% trucks, which travel predominantly on Broadway.[26] The study area is characterized by the presence of many blocks of ~10-m-tall houses with 5–10 m between adjacent houses and ~30 m between blocks of houses; the configuration of the houses likely impacts dispersion of highway-related PNC.[27] The Mystic River bounds the northeast edge of the study area.

The TAPL is a recreational vehicle retrofitted with a suite of rapid-response gas- and particle-phase instruments, as described in detail elsewhere.[23] Monitoring was performed along a fixed route on streets both perpendicular and parallel to I-93. Monitoring was conducted in 3–6–hr shifts on 43 days between September 2009 and August 2010. The TAPL was driven only on non-highway streets at 5–10 m/s to measure local-scale changes in pollutant concentrations. Particular effort was made to monitor during the morning rush hour because high concentrations and wide areas of elevated PNC are typically observed at that time.[28–30] Monitoring was conducted in the morning, afternoon, and evening in winter, spring, summer, and fall on non-consecutive days to maximize meteorological and traffic variability. This strategy spread monitoring times approximately evenly across a full year, which decreased the likelihood of high temporal autocorrelation in the dataset and allowed model development to neglect autocorrelation issues. PNC was measured by a butanol condensation particle counter (CPC 3775, TSI, Shoreview, MN; D_{50} = 4 nm). Spatial coordinates were assigned by matching instrument times to a master clock on a Garmin V GPS (manufacturer-specified accuracy = 3–5 m) mounted in the TAPL.

Quality assurance included side-by-side comparisons in a laboratory in Anderson Hall at Tufts University, flow checks, and lag-time corrections. Based on decision rules to avoid potential self-sampling of exhaust from the mobile laboratory, data were censored for TAPL speeds <5 km/hr and wind directions from behind the TAPL (14% of the data were censored). This generally happened when the TAPL was stopped at traffic signals (up to 90 seconds), the majority of which were on Broadway and may have caused model underestimates within ~20 m of intersections. PNC spikes due to other vehicles were not removed from the dataset.

Regression Model Development

The natural logarithm of PNC, ln(PNC), was used in the model because PNC was approximately log-normally distributed. The equation for a log-linear model of PNC is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\ln(PNC)=\beta_0+\sum_i \beta_i * x_i+\varepsilon \quad \text{(Equation 1)}$$

where β_i is a model coefficient for covariate x_i and ε is the random normally distributed error in the model. Coefficients in this log-linear model can be interpreted as the percent change in PNC per unit change in the covariate.

Explanatory variables related to meteorology (wind, temperature, and precipitation), time (linear and sinusoidal functions of year, day, and hour), highway traffic, and distances from combustion sources were developed. Variables were screened using stepwise regressions to maximize correlations (Pearson's r) and minimize Akaike information criterion (AIC). Following the criteria of Henderson *et al.*,[17] entering variables had $p<0.05$, contributed at least 1% to the adjusted R^2, and had correlations <0.6 with variables already in the model. All modeling was performed in R.[31]

To describe the functional form of the relationship of variables with PNC, generalized additive models (GAM), a type of nonparametric regression model, were produced for each variable identified in the screening process.[32] Loess smoothing windows between 0.1 and 0.75 were tested and windows of 0.25 (smooth using the 25% of values closest to each data point) were chosen as an interpretable balance between over-smoothing and noise. For those relationships that were not well described by linear functions, logarithmic, inverse, square, and exponential transformations were tested. Categorical variables were developed and included as appropriate, and compared to continuous variables.

Temporal Variables—Hourly-averaged temporal variables were assigned to each 1-second PNC measurement to match the expected time-step of major changes in temporal factors. See Supporting Information Section 1 (S-1) for further discussion on the assumption of representativeness within an hour.

Meteorological data collected at nearby stations were expected to explain particle reactivity and transport (see S-2 for a description of the meteorological stations). Temperature was the main particle reactivity parameter because particle numbers generally increase at low temperatures due to increased nucleation or other seasonally-varying factors (*e.g.*, e.g., humidity, engine combustion efficiency, and atmospheric mixing height).[33,34] which decrease the effective air density and in turn decrease overall engine combustion efficiency and atmospheric mixing height.[33,34] PNC dispersion was parameterized by wind speed, wind friction velocity, Monin-Obukhov length, and mixing height. Wind direction was described with both trigonometric functions and wind sectors because circular variables cannot be directly modeled with a linear function. The first continuous variable (*Wind_highway*) was a transformation of wind direction measured in degrees (*wdir*, °) relative to I-93 (*wroad*=140°), and had maximum values when the wind was parallel to I-93.

$$Wind_highway=\cos\left(2*(wdir-wroad)*\frac{\pi}{180}\right). \quad \text{(Equation 2)}$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A second continuous variable (*Wind_SE*) captured the effects of sources other than the nearest highway segment by a transformation of wind direction with maximum concentration for wind coming from the direction of highest concentrations (*wmax*, °) and zero value for wind blowing toward the direction of highest concentrations.

$$if|wdir-wmax \leq 90|,\ Wind_SE = \cos^2\left((wdir-wmax)*\frac{\pi}{180}\right)\ else,\ Wind_SE = 0. \quad \text{(Equation 3)}$$

In addition, wind sectors ranging from 10 to 45 degrees, as have been previously used,[35] were developed to compare to these continuous transformations of wind direction. Interactions of wind direction with wind speed were tested using linear regressions and two-dimensional GAMs.

Temporal traffic variables included hourly speed, total volume, and diesel volume on I-93, as well as transformations and ratios of these measurements. Highway traffic volume and speed were obtained from MassDOT station #8449 southeast of the study area (stakeholder.traffic.com). A similar study of near-highway urban PNC did not include highway traffic in final models due to difficulty identifying the relationship of PNC to traffic.[11] To ensure inclusion of traffic in this model, three traffic categories based on traffic volume and speed were defined: “congestion” (<64 km/hr), “typical” (>64 km/hr and >7000 vehicles/hr), and “low traffic” (>64 km/h and <7000 vehicles/hr). Congested conditions generally occurred during rush hours (07:00–09:00 and 16:00–18:00) and low volume conditions occurred in the early morning (00:00–03:00).

Models were developed both with and without a day of week variable as a proxy for unmeasured weekly variation in fleet mix and local traffic.[36] Each of these models has different strengths and weaknesses. Including the day of week variable may improve generalizability and allow comparison with models from other studies, but may not be valid from a purely statistical point of view because the uneven monitoring on different days of the week may add day-specific bias or spurious effects into the model. The model without a day of week variable avoids over-fitting the model, but may not capture the effects of local traffic that is uncorrelated with highway traffic.

Fixed Site—Fixed site data were obtained for the period during which mobile monitoring was performed and developed as model inputs. Hourly PNC measurements (CPC 3022A, TSI; D_p=7 nm) were obtained from the Harvard stationary monitoring site, which is located in Boston ~6.4 km south of the study area. The monitor is on the roof of Countway Library at the Harvard School of Public Health, and 51-m horizontally and 20-m vertically from the nearest major roadway (~20,000 vehicles/day) and ~3 km from the closest segment of I-93. These data were tested (1) as an additional covariate in the model and (2) as a substitute for all other temporal variables in the model.

Spatial Variables—I-93 and major roads were the main UFP sources expected to affect spatial variability within the study area. Variables were developed for distance, inverse distance, and inverse squared distance from I-93 and major roads (road class ≤4), and for the road type being monitored (see S-3 for more information). Roads within the highway

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

corridor were not included as major roads because inclusion in both categories would lead to unacceptable collinearity. The northeast part of the study area was usually downwind of the highway (59% of the year) and the neighborhoods west of the highway had more local traffic. To account for these factors, both highway side and upwind/downwind Boolean variables were tested. Interactions of distance from the highway corridor with highway side, wind direction relative to highway side, and wind speed were tested using two-dimensional GAMs and linear models with interaction terms.

Because the mobile monitoring platform moved continuously, measurements were not made at any one location for long enough to average out short-term variability in PNC. Short-term (<1 min) increases in PNC an order of magnitude above baseline, referred to here as "spikes", were typically caused by a diesel truck or bus in close proximity to the TAPL and cannot be predicted based on the variables available for regression. To evaluate their effects on the model, we operationally defined spikes as PNC measurements more than two standard deviations above the mean for the monitoring hour.[19] Censoring of data for PNC spikes was done by individual hours to decrease potential bias due to changing background PNC levels. The adjusted R^2 and parameter coefficients were compared for models both with and without spikes.

Results and Discussion

Annual median observed PNC was 50% higher <400 m from I-93 (27,000 particles/cm^3) compared to the background area >1 km from I-93 (18,000 particles/cm^3), with distance-decay gradients varying depending on traffic and meteorology (Figure 2). Median PNC measurements were two-fold higher in the winter (36,000 particles/cm^3) than in the summer (18,000 particles/cm^3). PNC levels were also higher on weekdays and Saturdays compared to Sundays, and higher during morning rush hour than later in the day. PNC distance-decay gradients from I-93 varied due to contributions from local street traffic. A detailed description of the mobile monitoring data is provided in Supporting Information Table S3 and available elsewhere.[23]

Covariate Relationships

The most important factors identified by the variable-screening process were wind direction and speed, temperature, distance to the nearest major road, and distance to the edge of the highway corridor. While linear functions described the relationship of some variables with ln(PNC) well (e.g., temperature, cosines of relative wind directions), the relationships of some other variables with ln(PNC) could not be linearized (*e.g.*, traffic volume and speed on I-93).

Temporal Variables—Parameterization of wind direction was critical because PNC varied by as much as 4-fold for different wind directions. A model of ln(PNC) that used only continuous wind direction variables had an adjusted R^2 of 0.10 and good visual agreement with measurements (Figure 3a). The R^2 using wind direction relative to I-93 (Equation 2) alone was 0.04, and the R^2 with wind direction relative to the southeast (Equation 3) alone was 0.07. Southeast winds were associated with high PNC levels in all four seasons, all days of the week, and all hours of the day.[14,23] The effect of high PNC levels under conditions of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

southeast winds was captured using Equation 3 (*Wind_SE*) with the wind direction of maximum PNC set to *wmax*=125°. Over the range of *Wind_SE*, PNC increased by 51% ± 1% for cold months (November to March) and 69% ± 1% for warm months (April to October). For both warm and cold months, the wind was from the southeast ~10% of the time; however, no southeast winds were captured before 09:00 in cold months. While the variable for winds from the southeast likely captures both temporal and spatial effects (*e.g.*, sea breezes common in coastal locations or major transportation corridors serving downtown Boston), we were unable to separate those effects with the available parameters.

Relative to these continuous variables, wind sectors increased the model R^2 by 0.04 and 0.01 for 25 and 45-degree sectors, respectively. To ensure that all wind sectors had at least one data point, sectors were required to be relatively large (~45 degrees) and therefore had little benefit relative to a physically interpretable function. While wind speed and wind direction had a statistically significant interaction ($p<0.001$), an interaction term was not included in the model because it did not increase the model R^2 or affect the wind speed or wind direction coefficients.

Temperature, wind speed, friction velocity, and mixing height were also strong predictors, with ln(PNC) decreasing linearly with linear increases in these parameters (Supporting Information Figure S1). While solar radiation was a statistically significant predictor of ln(PNC) ($p<0.001$), as shown previously,[35] its inclusion did not improve either the root mean square error (RMSE) or the model R^2 compared to a univariate model including only temperature. Monin-Obukhov length was not statistically associated with ln(PNC) ($p=0.2$). Pearson correlations with wind speed were ~0.7 for both friction velocity and mixing height, so only one of these three variables should be included in the model. When each was individually inserted into the regression model, wind speed, friction velocity and mixing height all resulted in equal values of model R^2; therefore, since it was the most easily obtained of these variables, wind speed was chosen to represent meteorological forcings, along with temperature.

The best-performing traffic variable was a categorization of highway traffic as typical, congested, or low traffic (Figure 3b). Highway traffic total volume, diesel volume, and speed were nonlinearly associated with ln(PNC) ($p<0.001$; Figure S2). While squaring transformations slightly reduced the curvature in functional form, no transformation of traffic speed, total volume, or diesel volume captured the sharp increase of PNC with high volume and speed. Diesel volume on I-93 was not monotonically related to PNC and was unable to predict PNC because it was low compared to gasoline vehicle volume and local diesel vehicles had a larger effect than those on the highway. Spikes were measured by the mobile lab about every 12 minutes, and on average a truck drives down Broadway every 2–3 minutes.[26] Similarly, ratios of traffic volume to wind speed and distance to the highway, as well as transformations of these ratios, were not linearly related to ln(PNC). The relationship of traffic speed and volume suggests this nonlinearity may have resulted from an interaction effect that was not fully captured due to lack of data for low traffic volumes with low travel speeds.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Typical highway traffic conditions had 19.1% higher PNC than low traffic conditions and 11.6% lower PNC than congested traffic flow ($p<0.001$, partial $R^2= 0.007$; Figure 3c). When the day of week was removed from the model, the times of low highway traffic had on average 30.9% lower PNC than typical traffic conditions while the effect of congestion did not change. An alternative to day of week may be a weekday/weekend variable;[36] however, this would not reflect actual conditions because PNC in Somerville was relatively high on Saturdays and low on Sundays. The low coefficient for Fridays likely captures some effects of seasonality because the two Fridays with monitoring data were in August and September. Future studies testing the day-of-week effect should balance the number of weekdays and weekend days to ensure sufficient sampling on each day.

Fixed Site—The Pearson correlation of ln(PNC) between the mobile platform and the central site was 0.52. Adding the central site ln(PNC) measurements as a model covariate increased the model R^2 by 1% and decreased the temperature coefficient by a factor of 1.5. Replacing all temporal variables in our model with the central site measurements resulted in a model R^2 decrease of 6% and a mean square residual increase of 0.1. These results suggest that the fixed site captured most of the temporal variability of PNC, and in particular, the seasonal variability. The lower model R^2 obtained by replacing meteorological and traffic variables with PNC measured at the central site suggests that local sources and wind effects at the central site were not representative of those in Somerville. The fixed site was not included in the model because its inclusion led to high levels of collinearity among temporal variables without sufficiently increasing model predictive power.

Spatial Variables—Spatial variables were also important predictors of ln(PNC). The relationship of ln(PNC) to distance to I-93 was approximately linear, with the highway playing a relatively minor role at greater distances, especially near major roads like Broadway (Figure 3d). All tests of linear distance to road variables within the study area resulted in negative coefficients, consistent with exponential distance-decay gradients that have been reported in the literature.[24] Models built using inverse distance and inverse squared distance had lower adjusted R^2 and higher RMSE than models using a linear treatment of distance to the highway. The decay in PNC with increasing distance to the nearest major road was comparable to the decay with distance upwind of I-93 (−23.0%/km and −20.9%/km, respectively). On average, PNC on major roads was 21% higher than on other roads, likely due to traffic signals at intersections and higher levels of local diesel traffic. Removing either distance to I-93 or distance to major road from the model did not affect the coefficient of the other variable, suggesting that the effect of interactions and collinearity on these spatial variables was negligible.

Gradients of PNC east of I-93 tended to be stronger than those west of I-93. Highway side and upwind/downwind Boolean variables resulted in identical model R^2 and coefficient estimates; therefore, the distance upwind and downwind of I-93 were included in the model to emphasize wind patterns. The gradients measured in Somerville were less pronounced than those reported in other studies (*e.g.*, reference 4) because PNC data from many monitoring days throughout the year, representing different wind directions and speeds, mixing heights, and source strengths, were averaged together.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Final Regression Model

The variables described above were incorporated into a regression model with an adjusted R^2 of 0.43 with day of week or 0.41 without day of week (Table 1). Both temporal and spatial variables contributed significantly to the model. All p-values were <0.001 and the signs of all coefficients in the model are those that were expected a priori (e.g., higher concentrations for higher traffic volumes, colder ambient temperatures, and lower wind speeds). The model predictions have good agreement with measurements across spatial and temporal PNC trends. PNC predictions and measurements were higher both during cold weather and near I-93 for typical morning hours in summer and winter (Figure 4). The mean of model predictions of ln(PNC) for all locations tracked the distribution of measurements by day (Figure 5).

The model can be used to compare ambient PNC at different times and locations. For example, when all other variables were held constant, moving 100 m further downwind of the highway resulted in a 4.49% decrease in PNC while moving the same distance from a major road only resulted in a 2.30% PNC decrease. On average, moving from the downwind side of I-93 in Somerville to the same distance on the upwind side resulted in a 19% decrease in PNC.

Sensitivity Analysis and Validation

The adjusted R^2 and root mean square error (RMSE) were used as measures of model performance (Table 2).[17] Both statistics were stable under leave-one-out cross-validation by iteratively excluding each monitoring day from the final PNC model (1,720–4,718 points per day), suggesting robust PNC estimates with minimal outlier influence. Substituting missing meteorological data with imputed values had little effect on the overall model (ΔAdj-R^2 = −0.002, ΔRMSE = 0.007). Similarly, removing the day of week variable had little effect on validation statistics. Predictions were within a factor of two of measured values 75% of the time, 13% of predictions were less than half the measured PNC, and 12% of predictions were more than twice the measured PNC. The largest under-predictions tended to occur on major roads. While the model residuals were approximately normal for PNC $<10^5$ particles/cm^3, they were not normally distributed (Jarque-Bera Normality Test $p<0.001$) mainly due to under-prediction of PNC spikes (Figure S3). Removing spikes (~4% of all measurements; see descriptive statistics in Table S4) to evaluate the effect of short-term sources increased the adjusted R^2 from 0.43 to 0.49. Coefficients were mostly unchanged (except for a decrease in the coefficient for the categorical variable for whether a measurement was from a major road); therefore, we did not remove spikes from the final model. Spatial autocorrelation of residuals is discussed in S-4.

This model is novel because it was developed using mobile monitoring data from all four seasons. It includes data collected under a wider range in temperature, wind, and traffic conditions compared to previous studies of spatial-temporal UFP.[11,14–16] As a result, the model included temporal factors that were significant in time-series models[35] but not statistically significant in models based on data collected over shorter time-frames. The correlations obtained here are comparable to other hourly, near-highway, urban, PNC regression models. Other researchers achieved adjusted R^2 of 0.22–0.32 in Brooklyn, NY,[11]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

0.36–0.51 in Girona, Spain,[15] 0.43–0.45 in Southern California,[16] and 0.45–0.56 in Somerville, MA.[14] Fuller *et al*[14] based their model on hourly average PNC at 17 homes in the summer and fall. In contrast, we developed our model for a larger study area using 1-second on-road measurements over the course of a year. The differences in the temporal resolution and location of monitoring caused our dataset to have more short-term variability, resulting in lower model R^2. A more detailed comparison of the methods used in these studies is available in Table S5.

Generalizability

An advantage of regression models is that they fit measured values with functions of covariates. Their accuracy is determined by representativeness of sampling and appropriate selection of covariates, not by emissions inputs, which are often unavailable. In this study, mobile monitoring over the course of a year resulted in measurements at more locations than could be monitored by fixed sites and under a wider variety of meteorology and traffic conditions than could be captured by short-term monitoring. Despite the measurement density, the resulting dataset has at least three potential limitations: (1) only three monitoring sessions included precipitation events (two snow and one rain), thus biasing the data towards dry weather conditions; (2) particles emitted directly by vehicles were not distinguished from particles formed as a result of nucleation events; and (3) monitoring was not performed to quantify the effect of the noise barrier or highway elevation, which have been shown to impact the dispersion of PNC.[39–41]

Like other air pollutant regression models, this model will be limited in its transferability. To date, no UFP regression models have been tested over multiple years or locations. Transferability studies of NO_2 models have reported similar explanatory variables with different variable forms (*e.g.*, traffic buffer size) for different cities.[13] Similar results are expected for PNC: while important predictor variables and their magnitude are likely to be similar for similar locations, additional monitoring will be required to validate regression models outside of the model calibration conditions and locations.

The PNC model described here will be used to predict hourly PNC at individual residences for an entire year in the Community Assessment of Freeway Exposure and Health (CAFEH), a community-based participatory research study of the association between UFP exposure and cardiovascular disease risks in adults who live near a highway.[42]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources

This research was funded by the NIEHS (ES015462, PO1 ES-09825), US EPA (FP-91720301, RD-83241601, RD-83479801), Tufts University Tisch College, and a P.E.O. Scholar Award to APP. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the agencies. Further, the agencies do not endorse the purchase of any commercial products or services mentioned in the publication.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Luz Padró-Martínez, Jeffrey Trull, Eric Wilburn, Piers MacNaughton, Kevin Stone, Tim McAuley, Samantha Weaver, Kevin Lane, and Jessica Perkins contributed to data collection and processing. The CAFEH Steering Committee including Ellin Reisner, Baolian Kuang, Michelle Liang, Christina Hemphill Fuller, Lydia Lowe, Edna Carrasco, M Barton Laws, and Mario Davila provided invaluable assistance in planning the data collection effort. Jon Levy and Rex Britter provided model development insight and reviewed manuscript drafts. We are grateful to the anonymous reviewers, whose thoughtful commentary led to improvements in the manuscript.

References

1. Oberdörster G, Oberdörster E, Oberdörster J. Nanotoxicology: An Emerging Discipline Evolving from Studies of Ultrafine Particles. Environmental Health Perspectives. 2005; 113(7):823–839. [PubMed: 16002369]
2. Brugge D, Durant JL, Rioux C. Near-highway pollutants in motor vehicle exhaust: a review of epidemiologic evidence of cardiac and pulmonary health risks. Environmental health : a global access science source. 2007:6. [PubMed: 17291352]
3. Choi HS, Ashitate Y, Lee JH, Kim SH, Matsui A, Insin N, Bawendi MG, Semmler-Behnke M, Frangioni JV, Tsuda A. Rapid translocation of nanoparticles from the lung airspaces to the body. Nat Biotech. 2010; 28(12):1300–1303.
4. Karner AA, Eisinger DS, Niemeier DA. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environmental Science & Technology. 2010; 44(14):5334–5344. [PubMed: 20560612]
5. Morawska L, Ristovski Z, Jayaratne ER, Keogh DU, Ling X. Ambient nano and ultrafine particles from motor vehicle emissions: Characteristics, ambient processing and implications on human exposure. Atmospheric Environment. 2008; 42(35):8113–8138.
6. HEI Review Panel on Ultrafine Particles. HEI Perspectives. Vol. 3. Health Effects Institute; Boston, MA: 2013. Understanding the Health Effects of Ambient Ultrafine Particles.
7. Larson T, Henderson SB, Brauer M. Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. Environmental Science & Technology. 2009; 43(13):4672–4678. [PubMed: 19673250]
8. Hoek G, Beelen R, Kos G, Dijkema M, van der Zee SC, Fischer PH, Brunekreef B. Land Use Regression Model for Ultrafine Particles in Amsterdam. Environmental Science & Technology. 2011; 45(2):622–628. [PubMed: 21158386]
9. Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J, Giovis C. A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Analysis and Environmental Epidemiology. 2005; 15(2):185–204. [PubMed: 15292906]
10. Abernethy R, Allen RW, McKendry IG, Brauer M. A land use regression model for ultrafine particles in Vancouver, Canada. Environmental Science & Technology. 2013; 47(10):5217–5225. [PubMed: 23550900]
11. Zwack LM, Paciorek CJ, Spengler JD, Levy JI. Modeling Spatial Patterns of Traffic-Related Air Pollutants in Complex Urban Terrain. Environ Health Perspect. 2011; 119(6)
12. Dons E, Van Poppel M, Kochan B, Wets G, Int Panis L. Modeling temporal and spatial variability of traffic-related air pollution: Hourly land use regression models for black carbon. Atmospheric Environment. 2013; 74:237–246.
13. Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, Tsai MY, Künzli N, Schikowski T, Marcon A, Eriksen KT, Raaschou-Nielsen O, Stephanou E, Patelarou E, Lanki T, Yli-Tuomi T, Declercq C, Falq G, Stempfelet M, Birk M, Cyrys J, von Klot S, Nádor G, Varró MJ, Dėdelė A, Gražulevičienė R, Mölter A, Lindley S, Madsen C, Cesaroni G, Ranzi A, Badaloni C, Hoffmann B, Nonnemacher M, Krämer U, Kuhlbusch T, Cirach M, de Nazelle A, Nieuwenhuijsen M, Bellander T, Korek M, Olsson D, Strömgren M, Dons E, Jerrett M, Fischer P, Wang M, Brunekreef B, de Hoogh K. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. Atmospheric Environment. 2013; 72:10–23.
14. Fuller CH, Brugge D, Williams P, Mittleman M, Durant JL, Spengler JD. Estimation of ultrafine particle concentrations at near-highway residences using data from local and central monitors. Atmospheric Environment. 2012; 57:257–265. [PubMed: 23645993]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

15. Rivera M, Basagaña X, Aguilera I, Agis D, Bouso L, Foraster M, Medina-Ramón M, Pey J, Künzli N, Hoek G. Spatial distribution of ultrafine particles in urban settings: A land use regression model. Atmospheric Environment. 2012; 54:657–666.
16. Li L, Wu J, Hudda N, Sioutas C, Fruin SA, Delfino RJ. Modeling the Concentrations of On-Road Air Pollutants in Southern California. Environmental Science & Technology. 2013; 47(16):9291–9299. [PubMed: 23859442]
17. Henderson SB, Beckerman B, Jerrett M, Brauer M. Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter. Environmental Science & Technology. 2007; 41(7):2422–2428. [PubMed: 17438795]
18. Gidhagen L, Johansson C, Langner J, Foltescu VL. Urban scale modeling of particle number concentration in Stockholm. Atmospheric Environment. 2005; 39(9):1711–1725.
19. Westerdahl D, Fruin S, Sax T, Fine PM, Sioutas C. Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles. Atmospheric Environment. 2005; 39(20):3597–3610.
20. Bukowiecki N, Dommen J, Prévôt ASH, Weingartner E, Baltensperger U. Fine and ultrafine particles in the Zürich (Switzerland) area measured with a mobile laboratory: An assessment of the seasonal and regional variation throughout a year. Atmospheric Chemistry and Physics. 2003; 3(5):1477–1494.
21. Hagler GSW, Thoma ED, Baldauf RW. High-Resolution Mobile Monitoring of Carbon Monoxide and Ultrafine Particle Concentrations in a Near-Road Environment. Journal of the Air & Waste Management Association. 2010; 60(3):328–336. [PubMed: 20397562]
22. Kolb CE, Herndon SC, McManus JB, Shorter JH, Zahniser MS, Nelson DD, Jayne JT, Canagaratna MR, Worsnop DR. Mobile Laboratory with Rapid Response Instruments for Real-Time Measurements of Urban and Regional Trace Gas and Particulate Distributions and Emission Source Characteristics. Environmental Science & Technology. 2004; 38(21):5694–5703. [PubMed: 15575289]
23. Padró-Martínez LT, Patton AP, Trull JB, Zamore W, Brugge D, Durant JL. Mobile monitoring of particle number concentration and other traffic-related air pollutants in a near-highway neighborhood over the course of a year. Atmospheric Environment. 2012; 61:253–264. [PubMed: 23144586]
24. Zhu Y, Hinds WC, Shen S, Sioutas C. Seasonal Trends of Concentration and Size Distribution of Ultrafine Particles Near Major Highways in Los Angeles. Aerosol Science and Technology. 2004; 38(12 supp 1):5–13.
25. Central Transportation Planning Staff, Average Daily Traffic on Massachusetts Roads. CTPS Geoserver. 2012.
26. McGahan, A.; Quackenbush, KH.; Kuttner, WS. Regional Truck Study. Boston Region Metropolitan Planning Organization, Central Transportation Planning Staff; 2001.
27. Vardoulakis S, Fisher BE, Pericleous K, Gonzalez-Flesca N. Modelling air quality in street canyons: a review. Atmospheric Environment. 2003; 37(2):155–182.
28. Choi W, He M, Barbesant V, Kozawa KH, Mara S, Winer AM, Paulson SE. Prevalence of wide area impacts downwind of freeways under pre-sunrise stable atmospheric conditions. Atmospheric Environment. 2012; 62:318–327.
29. Durant JL, Ash CA, Wood EC, Herndon SC, Jayne JT, Knighton WB, Canagaratna MR, Trull JB, Brugge D, Zamore W, Kolb CE. Short-term variation in near-highway air pollutant gradients on a winter morning. Atmos Chem Phys. 2010; 10(17):8341–8352.
30. Hu S, Fruin S, Kozawa K, Mara S, Paulson SE, Winer AM. A wide area of air pollutant impact downwind of a freeway during pre-sunrise hours. Atmospheric Environment. 2009; 43(16):2541–2549. [PubMed: 25379010]
31. R Core Team. R: A language and environment for statistical computing, 2.13.1. R Foundation for Statistical Computing; Vienna, Austria: 2013.
32. Hastie, T. gam: Generalized Additive Models. 2013. http://CRAN.R-project.org/package=gam
33. Kittelson DB, Watts WF, Johnson JP. Nanoparticle emissions on Minnesota highways. Atmospheric Environment. 2004; 38(1):9–19.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

34. Jamriska M, Morawska L, Mergersen K. The effect of temperature and humidity on size segregated traffic exhaust particle emissions. Atmospheric Environment. 2008; 42(10):2369–2382.
35. Clifford S, Low Choy S, Hussein T, Mengersen K, Morawska L. Using the Generalised Additive Model to model the particle number count of ultrafine particles. Atmospheric Environment. 2011; 45(32):5934–5945.
36. Morawska L, Jayaratne ER, Mengersen K, Jamriska M, Thomas S. Differences in airborne particle and gaseous concentrations in urban air between weekdays and weekends. Atmospheric Environment. 2002; 36(27):4375–4383.
37. MassGIS, EOTMAJROADS. Commonwealth of Massachusetts Executive Office of Energy and Environmental Affairs. 2008. Office of Geographic and Environmental Information (MassGIS).
38. MassGIS, EOTROADS. Commonwealth of Massachusetts Executive Office of Energy and Environmental Affairs. 2008. Office of Geographic and Environmental Information (MassGIS).
39. Ning Z, Hudda N, Daher N, Kam W, Herner J, Kozawa K, Mara S, Sioutas C. Impact of roadside noise barriers on particle size distributions and pollutants concentrations near freeways. Atmospheric Environment. 2010; 44(26):3118–3127.
40. Heist DK, Perry SG, Brixey LA. A wind tunnel study of the effect of roadway configurations on the dispersion of traffic-related pollution. Atmospheric Environment. 2009; 43(32):5101–5111.
41. Hagler GSW, Lin MY, Khlystov A, Baldauf RW, Isakov V, Faircloth J, Jackson LE. Field investigation of roadside vegetative and structural barrier impact on near-road ultrafine particle concentrations under a variety of wind conditions. Science of The Total Environment. 2012; 419:7–15. [PubMed: 22281040]
42. Fuller CH, Patton AP, Lane K, Laws MB, Marden A, Carrasco E, Spengler J, Mwamburi M, Zamore W, Durant JL, Brugge D. A community participatory study of cardiovascular health and exposure to near-highway air pollution: study design and methods. Reviews on Environmental Health. 2013; 28(1):21. [PubMed: 23612527]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

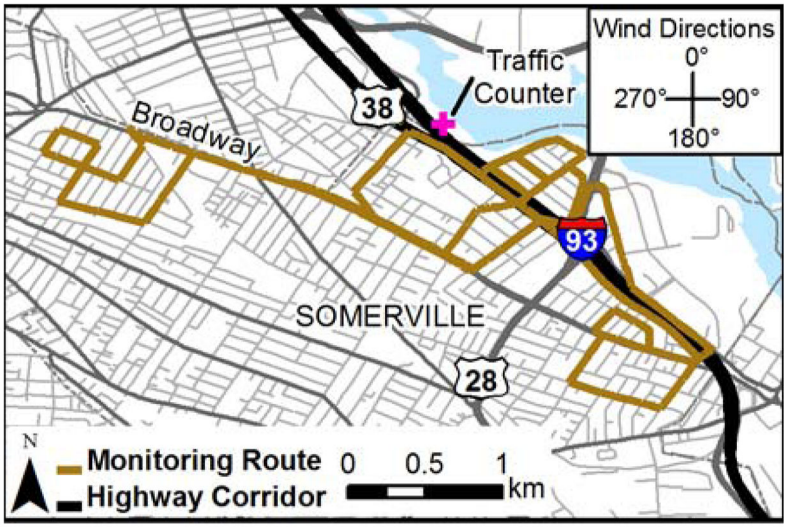

Figure 1.

Map of mobile-monitoring area in Somerville. Continuous traffic counts were recorded by MassDOT at the pink cross on I-93. The inset displays the coordinate system showing where wind is from.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

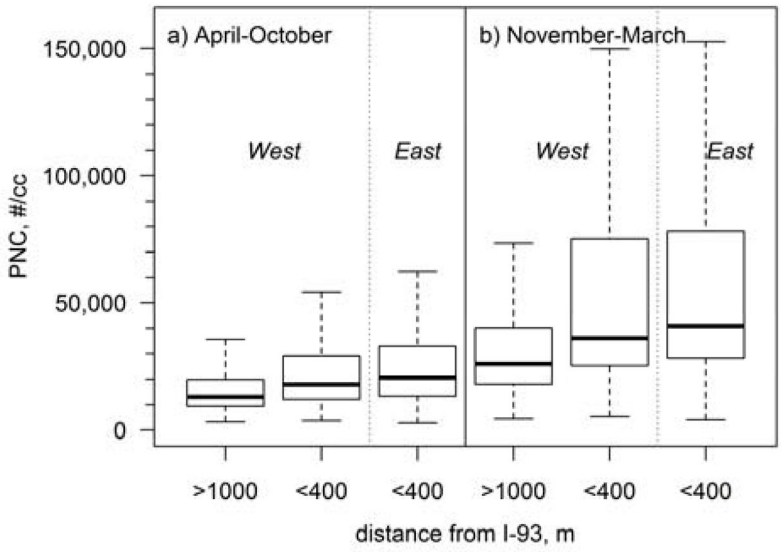

Figure 2.

PNC by distance from the edge of I-93 for (a) April to October (n=154 hr), and (b) November to March (n=129 hr). Dashed vertical lines in each panel represent I-93.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

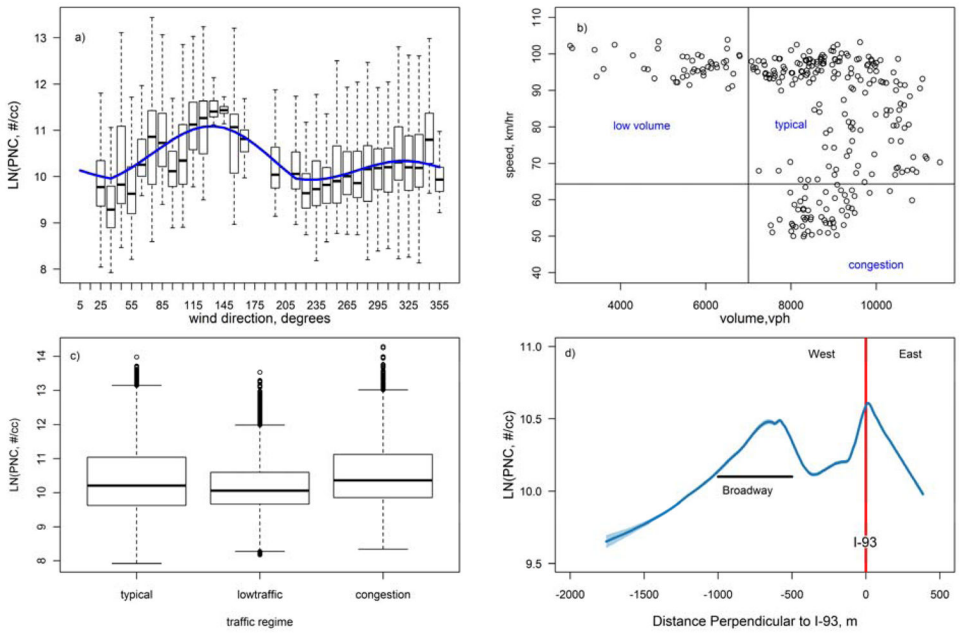

Figure 3.

Graphical exploration of variables used in the regression model. (a) Transformed wind direction variables (dark line) are compared to measurements made by the mobile lab (boxplots). (b) Traffic categories (typical, low traffic, congested) are defined based on both volume and speed measured at station 8449 on I-93. (c) Boxplot of ln(PNC) by the three highway traffic categories defined in (b). (d) GAM of predictions and 95% confidence intervals of ln(PNC) used to verify the linear decrease of ln(PNC) with distance from I-93. Note that Broadway is a major road that cuts through the study area between 400 m and 1000 m from I-93.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

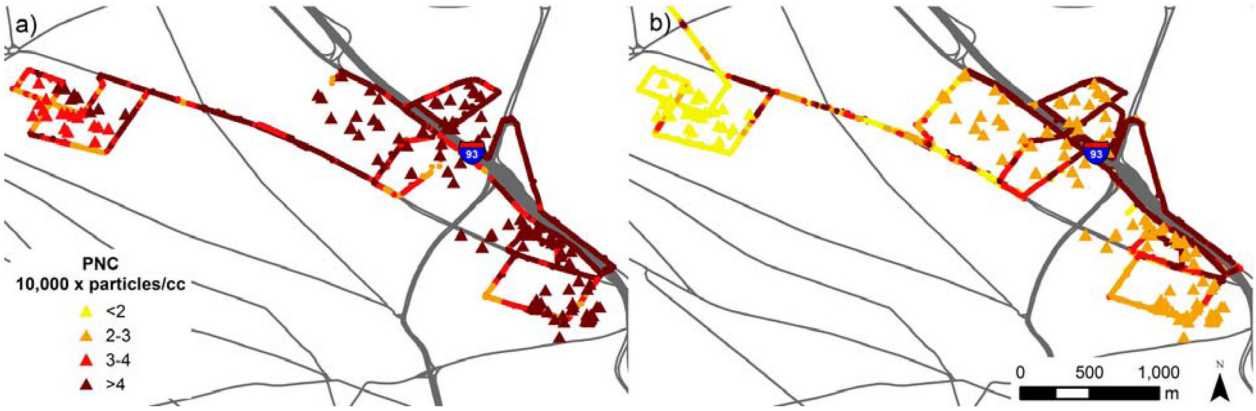

Figure 4.

Comparison of regression model predictions at residences of participants in the CAFEH study (triangles) to mobile monitoring measurements (lines) for one winter and one summer Wednesday morning with typical traffic: (a) January 6, 2010 07:00–08:00 (−6 °C, 4 m/s winds from west-northwest); (b) July 21, 2010 06:00–07:00 (22 °C, <1 m/s winds from south-southwest).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

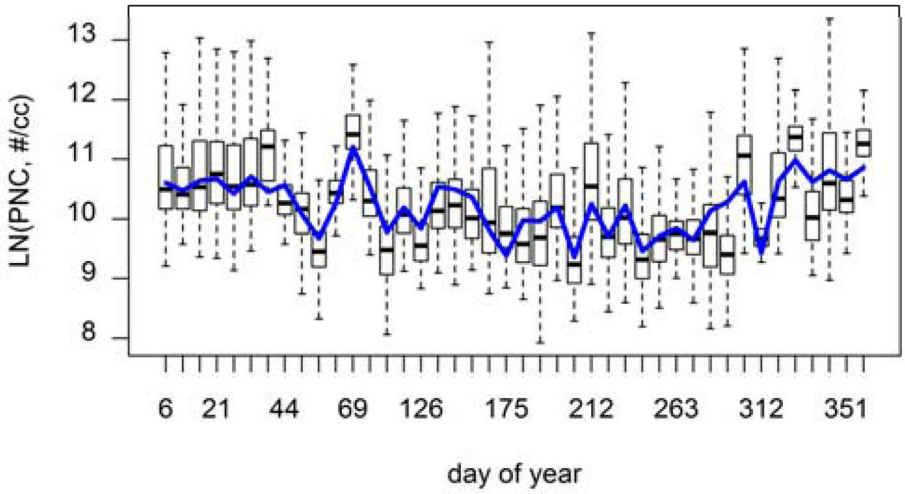

Figure 5.

Mean of modeled daily ln(PNC) for all locations (blue line) as compared to measurements made by the mobile lab (Tukey boxplot, outliers not shown).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Log-linear PNC model summary.

		Model 1			Model 2		
Category	**Variable** [a]	**Coeff.**	**Standard error**	**ΔR^2** [b]	**Coeff.**	**Standard error**	**ΔR^2** [b]
Intercept	(Intercept)	10.68	0.01	NA	11.100	0.007	NA
Meteorology	Wind_highway, cosine transformation of wind direction relative to I-93	0.048	0.003	0.001	0.058	0.003	0.002
	Wind_SE, Square of cosine transformation of wind direction relative to southeast	0.727	0.007	0.04	0.693	0.007	0.04
	Wind speed, m/s	−0.124	0.001	0.05	−0.132	0.001	0.06
	Temperature, °C	−0.0326	0.0002	0.10	−0.0327	0.0002	0.12
Location relative to I-93	Upwind of I-93	−0.185	0.005	0.07 [c]	−0.189	0.005	0.07
	Within highway corridor (Mystic Ave)	0.246	0.006		0.246	0.006	
	Distance upwind of I-93, km	−0.209	0.005		−0.206	0.006	
	Distance downwind of I-93, km	−0.449	0.007		−0.432	0.007	
Location relative to nearest major road	Distance from nearest major road, km	−0.230	0.014	0.001	−0.21	0.01	0.001
	On a major road	0.211	0.005	0.007	0.206	0.005	0.007
Day of Week (relative to Sunday)	Monday	0.335	0.009	0.02 [c]	---	---	---
	Tuesday	0.461	0.008		---	---	---
	Wednesday	0.391	0.008		---	---	---
	Thursday	0.462	0.008		---	---	---
	Friday	0.093	0.011		---	---	---
	Saturday	0.392	0.008		---	---	---
Traffic	Low traffic (<7000 vph)	−0.191	0.006	0.01 [c]	−0.309	0.005	0.02
	Congestion (<64 km/hr)	0.116	0.005		0.108	0.005	

[a] All variables in the model are statistically significant ($p<0.001$). The overall model $R^2 = 0.43$ with the day of week (Model 1) and 0.41 without the day of week (Model 2). All variables are linear with the exception of Wind_highway (cosine transformation of wind direction relative to I-93; Equation 2) and Wind_SE (square of cosine transformation of wind direction relative to southeast; Equation 3) and the categorical variables, which are “upwind of I-93”, “within highway corridor”, “on a major road”, “day of week”, and “traffic”. Temporal variables are input on an hourly basis.

[b] Partial R^2 contributed by this variable after all other variables are introduced into the model. Partial R^2 will generally be lower than the R^2 for a univariate model because multiple variables will pick up some of the same variability.

[c] The partial R^2 for categorical variables is listed in the row for the first value of the variable.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Regression model cross-validation.

Data subset	n (days)[a]	Model 1: includes day of week			Model 2: no day of week		
		Adj-R^2 [b]	RMSE[b]	Prediction RMSE[b]	Adj-R^2 [b]	RMSE[b]	Prediction RMSE[b]
All	39	0.43	0.64	0.67	0.41	0.65	0.66
		0.41, 0.47	0.61, 0.64	0.46,1.36	0.39, 0.44	0.63, 0.65	0.37, 1.15
Substitute missing wind data values [c]	43	0.42	0.65	0.67	0.41	0.66	0.66
		0.40, 0.46	0.62, 0.65	0.48, 1.42	0.38, 0.43	0.64, 0.66	0.37, 1.21
Remove spikes by hour [c, d]	39	0.49	0.56	0.60	0.46	0.57	0.59
		0.46, 0.53	0.54, 0.57	0.33, 1.33	0.43, 0.49	0.56, 0.58	0.29, 1.10

[a] Leave-one-out cross-validation (LOO) was performed by removing one day of measurements at a time, so there are *n* cross-validation models for each data subset.

[b] Each validation result is reported as a mean followed by the minimum and maximum values from the leave-one-out cross-validation. The LOO adjusted R^2 and RMSE are for the model developed on the dataset with one day removed. Prediction RMSE was calculated as the error in hourly predictions for each location on the day that was removed for model calibration.

[c] Effects of missing data and spikes were tested by imputing missing meteorological data and removing spikes from the dataset.

[d] Spikes were defined as PNC > mean(PNC)+2*sd(PNC).[19]