Atmos Environ (1994). Author manuscript; available in PMC 2015 April 01.

Published in final edited form as:

Atmos Environ (1994). 2014 April 1; 86: 84–92. doi:10.1016/j.atmosenv.2013.11.077.

Determinants of the Spatial Distributions of Elemental Carbon and Particulate Matter in Eight Southern Californian Communities

Robert Urman[a], **James Gauderman**[a], **Scott Fruin**[a], **Fred Lurmann**[b], **Feifei Liu**[a], **Reza Hosseini**[c], **Meredith Franklin**[a], **Edward Avol**[a], **Bryan Penfold**[b], **Frank Gilliland**[a], **Bert Brunekreef**[d], and **Rob McConnell**[a]

[a]Division of Environmental Health, Keck School of Medicine, University of Southern California, 2001 N. Soto Street, Los Angeles, CA 90089, USA [b]Sonoma Technology, Inc., 1455 N. McDowell Blvd. #D, Petaluma, CA 94954-6503, USA [c]Department of Mathematical Informatics, University of Tokyo, Japan [d]University of Utrecht, Netherlands Institute for Risk Assessment Sciences, Utrecht University, The Netherlands and Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

Abstract

Emerging evidence indicates that near-roadway pollution (NRP) in ambient air has adverse health effects. However, specific components of the NRP mixture responsible for these effects have not been established. A major limitation for health studies is the lack of exposure models that estimate NRP components observed in epidemiological studies over fine spatial scale of tens to hundreds of meters. In this study, exposure models were developed for fine-scale variation in biologically relevant elemental carbon (EC). Measurements of particulate matter (PM) and EC less than 2.5 µm in aerodynamic diameter ($EC_{2.5}$) and of PM and EC of nanoscale size less than 0.2 µm were made at up to 29 locations in each of eight Southern California Children's Health Study communities. Regression-based prediction models were developed using a guided forward selection process to identify traffic variables and other pollutant sources, community physical characteristics and land use as predictors of PM and EC variation in each community. A combined eight-community model including only CALINE4 near-roadway dispersion-estimated vehicular emissions accounting for distance, distance-weighted traffic volume, and meteorology, explained 51% of the $EC_{0.2}$ variability. Community-specific models identified additional predictors in some communities; however, in most communities the correlation between predicted concentrations from the eight-community model and observed concentrations stratified by community were similar to those for the community-specific models. $EC_{2.5}$ could be predicted as well as $EC_{0.2}$.

Corresponding author: Rob McConnell, Division of Environmental Health, Keck School of Medicine, University of Southern California, 2001 N. Soto Street, MC 9237, Los Angeles, CA 90089, USA, Phone: 323, 442 1593, Fax: 323 442 3272, rmcconne@usc.edu.

$EC_{2.5}$ estimated from CALINE4 and population density explained 53% of the within-community variation. Exposure prediction was further improved after accounting for between-community heterogeneity of CALINE4 effects associated with average distance to Pacific Ocean shoreline (to 61% for $EC_{0.2}$) and for regional NO_x pollution (to 57% for $EC_{2.5}$). PM fine spatial scale variation was poorly predicted in both size fractions. In conclusion, models of exposure that include traffic measures such as CALINE4 can provide useful estimates for $EC_{0.2}$ and $EC_{2.5}$ on a spatial scale appropriate for health studies of NRP in selected Southern California communities.

1. Introduction

Emerging evidence suggests that near-roadway air pollution is associated with chronic respiratory, cardiovascular, and neurological diseases (Guxens and Sunyer, 2012; HEI, 2010). Considerable uncertainty exists as to the components of the near-roadway pollutant mixture responsible for chronic health effects. Oxides of nitrogen have been commonly measured to develop near-roadway prediction models because of the close association between NO_x and vehicular emissions and the existence of inexpensive passive NO_x monitors (HEI, 2010). Although acute effects of NO_2 have been observed at ambient concentrations, toxicological studies have identified components of ambient particulate matter as more likely to be responsible for the chronic effects of near-roadway exposures. Recent epidemiological studies have reported health associations with estimated exposure to particulate elemental carbon (EC), employing models based on traffic metrics and other land use (Brauer et al., 2007; Morgenstern et al., 2007; Ryan et al., 2007). Particles with EC may also contain transition metals and organic compounds that cause oxidative stress and inflammation known to be involved in the pathogenesis of asthma and other respiratory diseases (Ghio et al., 2012; Riedl and Diaz-Sanchez, 2005). EC, especially smaller particles, carries these toxicologically relevant particle components deep into the lung. However, there have been few exposure models estimating components of particulate matter on a fine spatial scale of tens to a few hundred meters that is relevant for epidemiological studies examining near roadway effects.

In Southern California, EC is a useful marker for vehicular combustion products, especially from diesel powered vehicles, which are the primary EC source (Schauer, 2003). Smaller contributions to ambient EC are made by wood smoke (little used in our study communities), ship emissions, railways, and off-road vehicles (EPA, 2012). For this study, we measured and modeled Southern California EC concentrations in the fine respirable fraction less than 2.5 μm in aerodynamic diameter ($EC_{2.5}$) and in a nanoscale size fraction less than 0.2 μm ($EC_{0.2}$) that we anticipated might better reflect the near-roadway gradient in ultrafine particles in communities participating in the Children's Health Study (CHS), a large prospective study of cardio-respiratory health (McConnell et al., 2010). The study is notable for the fine spatial scale at which these measurements were made in order to assess small-scale intra-community variation. Information on traffic, land use and other community features were used to develop models of within-community exposure, based on measurements made at informatively selected locations in each study community. We also measured and modeled intra-community variation in particulate matter (PM) mass in the 2.5

and 0.2 µm size fractions ($PM_{2.5}$ and $PM_{0.2}$). Additionally, we assessed both within- and between-community variation of these pollutants.

2. Methods

2.1 Study locations and air sampling

Air pollution samplers for size-resolved PM mass and components were deployed from November 2008 until December 2009 in up to 29 informatively selected locations within each of eight Southern California communities (see Figure 1) in which CHS participants are currently being studied. Sample locations were selected from among participants' homes based on high or low impacts of freeway, non-freeway, and other non-traffic sources. All samplers were deployed at the same time in each community for two consecutive two-week periods during warm and cool times of year, for a total of four two-week sampling periods per community. Size-resolved PM less than 0.2 ($PM_{0.2}$) and 0.2 to 2.5 µm in aerodynamic diameter ($PM_{0.2\text{-}2.5}$) were collected on modified Harvard cascade impactors (Lee et al., 2006). $PM_{2.5}$ was estimated by summing the $PM_{0.2\text{-}2.5}$ and $PM_{0.2}$ stage data. EC was collected from different sampling lines and measured using a thermal-optical transmittance method. Additional information on the selection of sampling locations and on air monitoring is available in the Online Supplement and in a previous report (Fruin et al., In Press).

2.2 Predictors of EC and PM mass

Potential predictors of EC included distance (and inverse distance) to roadways and other sources, traffic density in distance buffers around sampling locations, dispersion modeled traffic pollutant exposure, length of road and amount of green space in buffers around sampling locations, population density and elevation. Predictors were linked to GPS measurements made at the sampling locations using GIS software (ArcGIS). Details are provided in the Online Supplement.

Annual average daily traffic (AADT) volumes on roadways and truck percentage were obtained from the California Department of Transportation (Caltrans) milepost data for freeways and numbered state highways for 2009 (CALTRANS, 2010) and Dynamap Traffic Count (Version 10.2) datasets produced by TeleAtlas (Boston, Massachusetts) for other roads. Roadway classification was based on the Functional Class Code (FCC) as found in the Dynamap dataset. Density plots were generated within the GIS using a linear decay function that approximated the decrease in ambient concentrations with increasing distance away from roadways, i.e., decays to background between 150 and 300 meters (Zhu et al., 2002).

Estimates of the contributions of local on-road motor vehicle emissions to air quality were obtained using the CALINE4 Gaussian line-source dispersion model (Benson, 1989). The CALINE4 dispersion model uses distance to roadways, vehicle counts, vehicle emission rates, and meteorological conditions as inputs. Although the CALINE4 model provides estimates of the near-roadway contribution to EC and $PM_{2.5}$ (and of multiple other near-roadway pollutants), these estimates are all highly correlated and should be regarded as markers for the primary near-roadway mixture. Separate estimates were made for the contribution of local traffic on freeways and on all other roadways (non-freeway roads) to concentrations of EC and $PM_{2.5}$. Total CALINE4 was computed as the sum of the contributions from both freeway and non-freeway roads. Total CALINE4 was highly correlated with the freeway component of CALINE but not with the non-freeway component of CALINE. See Online Supplement for more details.

Distances to freeways, active railways, combustion point sources (eg. a port or a refinery), intermodal transportation facilities (for example, where train to truck transfer of cargo occurs), and to the nearest Pacific Ocean shoreline were computed by GIS. To provide another indicator of emissions proximity, roadway lengths within various buffer distances (50, 100, 150, 200, 250, and 300 m radius) were computed for each FCC road class and summed together to provide the total length in each buffer (Eckel et al., 2011).

Population density data at the block group level were obtained from US Census Bureau (2000 data projected to year 2010) via ESRI's data repository. The population density within 300m radius buffers of each sampling location was computed as an aerial extent-weighted average of each block group's density in the buffer.

Elevation data with ~10m resolution were obtained from the US Geological Survey (USGS) website (http://seamless.usgs.gov). For each sampling location, we computed a mean elevation at a neighborhood level based on 10m-grid elevation values within a 300m buffer.

The Normalized Difference Vegetation Index (NDVI) is an indicator of live green vegetation derived from satellite remote sensing data (Pettorelli et al., 2005) and was included as a predictive variable as a metric of the absence of traffic and other pollution sources.

Additional description of predictor covariates is provided in the Online Supplement.

2.3 Regional pollutants

Additional measures of regional air pollutants were continuously collected at regulatory agency regional air monitoring stations in each of the study communities. Pollutants of interest from these measurements included $PM_{2.5}$ and NO_x as described previously (Gauderman et al., 2004). Daily measures of these regional pollutants were integrated over the multi-week study periods and were used in the modeling as modifiers of the association between predictors and outcome.

2.4 Statistical methods

EC and PM samples collected across seasons were averaged to derive a single eight-week average concentration for each sampling location. Sampling locations that were not the same across seasons were not included in the primary analyses. At schools and central sites where duplicate measurements were made, concentrations were averaged. All eight-week averaged concentrations were natural log transformed to help satisfy the normality and homoscedasticity assumptions of linear regression and to ensure model predictions would be positive.

Because the focus of this study was to examine the factors that affected within-community variation of fine and nanoscale EC ($EC_{2.5}$ and $EC_{0.2}$) and particulate mass ($PM_{2.5}$ and $PM_{0.2}$), we implemented a strategy similar to the one used in Franklin et al. (2012) for NO_2, NO, and NO_x. To parse out the within-community variation from the cross-community variation, we subtracted (or "deviated") the mean concentration of pollutant *Y* in community ***c*** from the pollutant measurement *Y* at location *i* in community *c* ($dY_{ci} = Y_{ci} - Y_c$). We similarly performed a transformation of each predictor (*X*) previously listed (i.e. $dX_{ci} = X_{ci} - X_c$) to estimate a within-community distribution of the predictors. Some predictor variables (CALINE4 estimates, traffic density) were positively skewed and were log transformed (prior to deviating) to minimize the potential influence of very high values. We also evaluated the components of within-community and between-community variation for each pollutant using the VARCOMP procedure in SAS version 9.3 (SAS Institute Inc., Cary, NC).

We developed both community-specific models and a combined (eight-community) model for each pollutant. For community-specific analyses, point sources farther than 10km were excluded from consideration. In the combined model, we weighted the distance to intermodal facilities by dividing by the mean distance (i.e. dX_{ci}/Xc) , thus giving less weight to communities without intermodal facilities within 10km. We excluded from the combined models combustion point source locations with NO_x emission rates that were greater than 50 tons per year as predictors because most communities had none and distance to a shoreline was excluded because variation at the within-community scale was not meaningful for most communities that were many kilometers inland.

We calculated Pearson correlations between each of the deviated pollutants (dY_{ci}) and predictors (dX_{ci}) to understand how they varied together within communities. Supervised forward selection, similar to the one used in the European ESCAPE study (Eeftens et al., 2012a), was used to develop combined models as well as community-specific models. Model selection began with the predictor that produced the highest adjusted R^2 and that had a beta coefficient in the expected direction. Remaining predictors were then added one at a time until the addition did not result in at least a 1% improvement in the adjusted R^2. The direction of all beta coefficients was checked during each step of this model selection process. From this group of predictors, those that were not significant at the 0.10 level were dropped one at a time starting with the least significant predictor. Variables that had a variance inflation factor (VIF) greater than 3 were also dropped from the model.

Leave-one-out cross-validated (LOOCV) and (for the combined models) leave-one-community-out cross-validated (LOCOCV) R^2 were calculated to assess how well the models performed across communities and how transferable they might be to other communities in Southern California. To examine the performance of the combined model in each community, we took the predicted dY_{ci} from this model (using the LOOCV approach) and calculated the correlation with the observed values by community. The correlation was then squared to estimate the proportion of variation in each community that was explained by the combined model.

Finally, we fitted a mixed-effects model to consider the possibility that community-level variables might modify intra-community prediction models. The community-level variables considered included measured concentrations of PM_{10} and NO_x from the regulatory agency regional air stations as well as the community average of measured $EC_{2.5}$ and of selected predictors (population density, which might be an indicator of additional combustion sources in densely populated areas, and community-average distance to the shoreline as a proxy for meteorological characteristics that might affect the models). This mixed model took the form $dY_{ci} = \alpha + \beta_1 * dX_{ci} + \beta_2 * dX_{ci} * C + \beta_4 * d\boldsymbol{Z}_{ci} + f_c * dX_{ci} + e_c + e_{ci}$, where *Y* was the outcome ($EC_{0.2}$ or $EC_{2.5}$), *X* was the predictor of interest, C was a possible community-level modifier, ***Z*** were adjustment covariates from the combined model, and the e_c and f_c were community-level random effects, assumed to be bivariate normally distributed and independent of the subject-specific random effect e_{ci}. The parameter β_2 and its corresponding level of statistical significance was used to determine whether the intra-community relationship between X and Y varied by C. All analyses were conducted using SAS version 9.3.

3. Results

In this section, the sample size and distribution of each exposure outcome is described. The distribution of key covariates and their univariate association with PM and EC by community and size fraction is illustrated. The predictors from a unified model across all communities and the associated heterogeneity in the cross-validated predictions from this model in individual communities were examined, and the results from this approach were compared with a more traditional community-specific modeling approach. In sensitivity analyses, we examined the influence of regional pollution and other community characteristics on the heterogeneity of effects of near-roadway traffic metrics. Because these models will be applied to health outcomes in the CHS, a combined model of $EC_{2.5}$ exposure was developed restricted to communities in which within-community variability was well-predicted by near-roadway traffic metrics.

3.1 Characterization of sampling sites and key predictors

Samples were collected from 228 locations across the eight communities. Of these locations, 177 remained unchanged across the entire study design and were eligible for the analysis of eight-week average concentrations across seasons, as described in the Online Supplement. After eliminating locations with invalid data, mostly due to equipment failures or power interruptions, we had 148, 152, 130, and 137 locations with valid eight-week $EC_{2.5}$, $EC_{0.2}$, $PM_{2.5}$, and $PM_{0.2}$ data, respectively.

The community-specific distribution of the average of the eight-week measurements is shown in Figure 2 and the corresponding geometric means and coefficients of variation are shown in Table 1. The smaller nanoscale fraction of EC (0.2μm) had a similar pattern of within-community variability to $EC_{2.5}$, based on the coefficients of variation in Table 1. The within-community variation of $EC_{2.5}$ and $EC_{0.2}$ was about half that of the between-community variation (Table 2). In contrast, the within-community variance of $PM_{0.2}$ was greater than its between-community variance. The between-community variance of $PM_{2.5}$ was about ten times as large as its generally small within-community variance. However, one community (Mira Loma) contributed most of the between-community variability in $PM_{2.5}$ (Figure 2). There was strong correlation between $EC_{2.5}$ and $EC_{0.2}$ across all locations (0.83; Supplement Table 1). The community adjusted (within-community) correlation was almost as large (0.76). Both size fractions of particulate mass were weakly correlated with one another and with each EC size fraction.

There was substantial variability in the distribution of the predictor variables at sampling locations in different communities, for example for CALINE4 EC estimates and population density, which were key explanatory variables in combined models of exposure (as described below). The CALINE4-modeled freeway concentration varied by almost 7-fold, considerably more than the CALINE4-modeled concentration from all other roads. Mean population density varied by approximately 3-fold (See Figure 3).

The strongest correlations of measured EC pollutant concentrations in both size fractions were with traffic metrics (Table 3). Correlations with freeway and with the sum of freeway and non-freeway CALINE4 were approximately 0.7. Weaker correlations were observed with other predictors. Correlations of traffic and other predictor variables with $PM_{2.5}$ and $PM_{0.2}$ were much weaker than with $EC_{2.5}$ and $EC_{0.2}$ with few exceptions (e.g. NO_x point sources with $PM_{0.2}$). We also examined the community-specific correlations of $EC_{2.5}$ and $EC_{0.2}$ with potential predictor variables and found that there was considerable heterogeneity across communities for each pollutant (Supplement Tables 2 and 3). For $EC_{2.5}$, there were consistently strong and common traffic associations in five of the eight communities. A different traffic metric, truck count on the nearest freeway, was strongly correlated with $EC_{2.5}$ in Long Beach. However, in Mira Loma and San Dimas, traffic was poorly correlated with $EC_{2.5}$. $EC_{0.2}$ showed strong traffic associations in all but one community (San Dimas).

3.2 Combined eight-community model

CALINE4 was included in the best combined model for EC in each size fraction, and some form of traffic exposure was included in the best model for every pollutant studied (Table 4). After the traffic metrics, population density had the next largest effect estimates. Cross validation R^2 for both EC sizes were about 0.5, while the cross validation R^2 for PM mass were much smaller.

Although the LOOCV R^2 was 51% for $EC_{2.5}$ from the combined model, the performance varied substantially when applied to each community separately. For example, the R^2 of predicted with measured $EC_{2.5}$ in Santa Barbara was 82% but the model explained none of the intra-community variation in Long Beach (Table 5). Concentrations of $EC_{2.5}$ were also poorly predicted in Mira Loma and San Dimas. The $EC_{0.2}$ model predictions explained at least 30% of the measured variation in seven of the eight communities. In contrast, the model for $PM_{2.5}$ explained 30% or more of the measured variation in only two communities and the model for $PM_{0.2}$ in no community.

3.3 Community-specific models

Models identifying community-specific predictors were fitted for $EC_{2.5}$ (Table 6). These models explained the variation in some communities, particularly San Dimas (in which traffic metrics did not contribute to the model), Long Beach and Riverside, considerably better than the combined models, but the R^2 was still relatively low in San Dimas. Community-specific models did not substantially improve the R^2's for $EC_{0.2}$, except in San Dimas, in which traffic metrics did not contribute to the model and community-specific R^2 increased only to 0.21 (Table 7). Either freeway or (correlated) total CALINE was selected in most community-specific models.

3.4 Sensitivity analyses

As a post-hoc analysis for EC, we developed models in just those communities in which the combined model predicted at least 30% of community-specific variability (Table 5). For $EC_{2.5}$, a five-community model (excluding Long Beach, Mira Loma, and San Dimas) was able to explain 66% of the measured variation (Supplement Table 4), compared with 51% using data from all communities (from Table 4). The five-community model included total CALINE4, population density, and NDVI as predictors. Only San Dimas was excluded from the sensitivity model for $EC_{0.2}$, which like the model for all eight communities included only total CALINE4. The LOOCV R^2 was 53%, compared with 49% in all communities from Table 4. The LOCOCV R^2's were similar to those for LOOCV.

In previous analyses examining within-community NO_x variability, we observed larger CALINE4 effects in communities with lower average concentrations (Franklin et al., 2012). Therefore, we investigated whether the heterogeneity in traffic effect estimates in different communities might be explained by the average of community exposures and by the average of the continuous regional pollutant measurements made at the central site monitors during the time of sampling in each community. We focused on the variability in effects of total CALINE4 as this was a strong predictor in the eight-community models for both $EC_{2.5}$ and $EC_{0.2}$ and was selected in a majority of the community-specific models. In some models that included an interaction between total CALINE4 and these community-level modifiers, there was substantial improvement in the LOOCV R^2 (Table 8). In the $EC_{2.5}$ model, we found that the strongest association between total CALINE4 and $EC_{2.5}$ were in communities with low levels of regional NO_x (Supplement Figure 1), while associations between $EC_{0.2}$ and total CALINE4 were strongest in communities nearest to the shoreline (Supplement Figure 2).

4. Discussion

Notable features of this analysis included (1) the heterogeneity of the strength of EC-traffic associations across communities and the potential to partially explain this variability by community characteristics, a finding with potentially broad implications for spatial exposure modeling; (2) a comparison of model performance across multiple communities using two complementary approaches (a combined model and more traditional community-specific models); (3) models that were able to predict EC on a fine spatial scale; and (4) model development for a novel size fraction ($PM_{0.2}$ and $EC_{0.2}$).

Combined prediction models that captured the fine spatial scale of EC across eight communities were developed. The combined model R^2's were substantially improved by accounting for community characteristics that modified the effects of CALINE4 (from 51% to 57% for $EC_{2.5}$ by accounting for regional NO_x and 49% to 61% for $EC_{0.2}$ by accounting for average shoreline distance; Table 8 and Supplement Figures 1 and 2). It is possible that in the setting of a noisy and more complex high regional pollution background that a small local traffic effect on $EC_{2.5}$ was not identified, whereas in a community with little transported pollution, the effect of small primary traffic sources was apparent. This finding is consistent with our previous report with a larger number of communities in which we measured the within-community variation in NO, NO_2 and NO_x (Franklin et al., 2012). In that study, CALINE4 predicted variation better in less polluted communities outside of the Los Angeles air basin, where regional pollution is lower, than within the basin. For $EC_{0.2}$, it is not entirely clear why the effect of CALINE4 is stronger closer to the shoreline, but we speculate that it might be due to onshore winds creating clearer gradients of $EC_{0.2}$ concentrations and producing larger contrasts in communities closer to the shoreline with cleaner background concentrations. These findings, especially the variability by regional pollution levels, have potentially broad relevance to modeling of near-roadway pollution and merit further study in other geographic regions.

In community-specific models, the inclusion of truck counts on the nearest freeway in Long Beach improved the $EC_{2.5}$ R^2 substantially compared to the combined model (0.54 in Table 6 and 0 in Table 5). There were three freeways in Long Beach with markedly different truck counts. However, there was little variation in truck counts within each freeway and there was weak association of $EC_{2.5}$ with distance to nearest freeway (Supplement Table 2). Therefore, truck counts may have reflected background levels associated with the areas of the city corresponding to the three freeways rather than a near-roadway effect of truck exhaust. Long Beach is a coastal community with a major shipping port, refineries, and rail activity. It has complex air flows due to convergence of westerly and southerly onshore flows during the day. The CALINE4 estimates of near-roadway traffic impacts may be less accurate than in other communities because (1) the modeling relied on a single meteorological monitoring site which did not represent the complex flows and (2) the on-road emission estimates probably underestimated the heavy truck traffic on arterial corridors due to traffic from the port. Riverside is another relatively large community with heavy truck traffic en route from the port to large local warehouse transfer facilities. In this community, truck count on the nearest freeway and distance to these intermodal transfer facilities improved $EC_{2.5}$ prediction. $EC_{2.5}$ variability was also explained by predictors other than near-roadway metrics in San Dimas. Higher elevation predicted lower $EC_{2.5}$ (Table 6) and, along with vegetation, $EC_{0.2}$ (Table 7). While the CALINE4 predictions capture some aspects of meteorology, the absence of local meteorological measurements may have contributed to the poor predictability of EC in San Dimas. This community extends into the foothills of the San Gabriel Mountains, which may have strong and local influence on wind speed and direction that is not identified by the two closest monitoring stations (Azusa and Pomona, both far away from the major terrain features). A Mira Loma specific model for $EC_{2.5}$ was not reported due to a technical problem with samplers during one particle collection wave, leaving only 13 locations for analysis. The small sample size might explain the poor model fit of the combined model in this community.

The leave-one-community-out cross-validated R^2 was approximately 50% for both $EC_{2.5}$ and $EC_{0.2}$, based on the combined models. However, the poor performance of these models in some communities indicates that further study is warranted to determine how transferable the models could be to other Southern California communities, and whether communities to which the combined model would not be transferable could be identified a priori based on complexity of geographic topology, meteorology and other pollution sources (eg. San Dimas and Long Beach). A few other studies in Europe and North America have found that exposure models developed from land use in one city had reduced R^2 when used to predict measurements in other cities (Allen et al., 2011; Poplawski et al., 2009; Vienneau et al., 2010).

The combined $EC_{2.5}$ model predicted poorly in some communities because the near-roadway exposure metrics that determined variability overall did not explain variability in some communities. European studies that have examined the variability of $EC_{2.5}$ through land-use regression modeling also found measures of traffic to be important predictors but to vary between regions (Beelen et al., 2007; Brauer et al., 2003; Carr et al., 2002; Eeftens et al., 2012a; Hochadel et al., 2006; Morgenstern et al., 2007). In the ESCAPE study, separate models were developed using information on local land use for each of 20 large European cities, and cross-validated R^2 for the separate models ranged from 40% to 95% (Eeftens et al., 2012a). The R^2's of 36% to 77% in the community-specific models for $EC_{2.5}$ in our communities (Table 6) were somewhat lower but, as in the European studies, were heterogeneous across communities.

Possible reasons for the higher model R^2 in ESCAPE include a wider range of measured $EC_{2.5}$ concentrations to be explained by traffic and other land uses across large metropolitan regions in Europe, compared with the range in the generally smaller communities in the CHS as the focus of our modeling was to predict fine spatial scales of near-roadway mixtures (e.g. 50-150m). Comparing the exposures across the two studies is not straight forward, because $EC_{2.5}$ was assessed by light absorbance of $PM_{2.5}$ in ESCAPE. Although absorbance is highly correlated with measured $EC_{2.5}$, the relationship between the two measurements of EC can vary depending on location (Cyrys et al., 2003). The ratio of the range to mean of $PM_{2.5}$ absorbance was provided in the European study (Eeftens et al., 2012b), and we have calculated this index in each CHS study community in order to compare the variability across studies (Supplement Table 5). In ESCAPE, this measure of variability ranged from 68% in Gyor (Hungary) to 235% in London/Oxford (United Kingdom) and about half of the 20 study areas in ESCAPE had values that were greater than 100%. In contrast, only one among our eight study communities had a value greater than 100% for $EC_{2.5}$ (101% in Anaheim). Levels of residential $EC_{2.5}$ in European cities can be high (Putaud et al., 2004; Putaud et al., 2010), because unlike Southern California there is a high proportion of diesel powered passenger vehicles that travel on secondary roads in close proximity to residences.

We observed stronger correlations between $EC_{2.5}$ with freeway sources, compared with non-freeway sources, of CALINE4-predicted concentrations. A large proportion of EC is attributable to diesel exhaust from trucks, which are found largely on freeways in Southern California (Kam et al., 2012; Schauer, 2003) and elsewhere (Kinney et al., 2000). A Cincinnati study also found diesel sources, including length of bus routes and truck intensity within 300 meters of monitoring locations, to be strong predictors of $EC_{2.5}$ concentrations

(Ryan et al., 2008). In a Boston study, the strongest traffic predictor of $EC_{2.5}$ (measured via absorbance) was length of roadway in a 200m buffer (Clougherty et al., 2008).

For $EC_{2.5}$ population density contributed to the eight-community model, suggesting either that there were other anthropogenic sources of $EC_{2.5}$ or that population density provided additional information on traffic emissions. For example, residential population might be an indicator of "cold starts" that produce more EC after a prolonged period with the engine off. Other studies have also found population density to predict near-roadway air pollutants (Beelen et al., 2007; Brauer et al., 2003; Eeftens et al., 2012a). However, in our study this variable added little to the R^2 (~1%) in models also containing total CALINE4.

To our knowledge, few previous studies have examined the within-community spatial distribution and predictors of EC in size fractions that are smaller than 2.5μm. We hypothesized that $EC_{0.2}$ (and to a lesser extent $PM_{0.2}$) would be better markers for fresh near-roadway combustion than $EC_{2.5}$, which might contain a larger proportion of regionally transported EC. A recent study showed that a larger proportion of EC along busy roads was found in the smaller 0.25 μm size fraction compared to the 2.5-0.25 μm size fraction and the concentration of these fractions were highest on a stretch of freeway containing a large number of diesel trucks (Kam et al., 2012). Contrary to our hypothesis and these previous results, in our study the ratio of within-to between-community variance was similar for both size fractions (Table 2), suggesting that the accumulation mode (i.e, 0.1 to 1 μm), which should account for most of the $EC_{0.2}$ mass, has a substantial transported regional component in Southern California. $EC_{0.2}$ was highly correlated with $EC_{2.5}$ (Supplement Table 1), both within communities (R=0.76) and across all measurements (R=0.83). In addition, only total CALINE4 was a predictor of the within-community variation in $EC_{0.2}$ in the combined model. Although $EC_{0.2}$ is likely to penetrate more deeply into lungs and therefore may be a more biologically relevant exposure, these results suggest that modeling $EC_{0.2}$ exposure may provide little information for assessing health effects of within-community exposure to primary traffic source beyond what is provided by $EC_{2.5}$ (or by the CALINE4 estimate).

The within-community variability of $PM_{0.2}$ was almost as large as for EC (based on coefficients of variation in Table 1), but the cross-validation R^2 for within-community variability in $PM_{0.2}$ (0.12 from Table 4) was poor. Although $PM_{0.2}$ is enriched with ultrafine particles less than 0.1 μm in diameter, which have large spatial gradients downwind from major roadways (Beckerman et al., 2008; Zhu et al., 2002), ultrafine particles have little mass and most of the $PM_{0.2}$ mass is likely to be greater than 0.1 μm in diameter. Determinants of $PM_{0.2}$ variability merit further investigation.

The model for $PM_{2.5}$, which included length of road in a 100m buffer, nearest freeway truck count, population density, elevation and distance to the nearest point source of NOx (from Table 4), nevertheless poorly predicted the within-community variation in our study (cross-validated R^2 0.17). This is consistent with the regional character of $PM_{2.5}$ mass and with other studies that have shown little variation with local traffic predictors (Clougherty et al., 2008). Given the relatively smaller within- to between-community ratio in variance for $PM_{2.5}$ (Table 2), it is unlikely that predicted exposures from within-community models would contribute substantially more to understanding health effects than measurements from a single central site monitoring station. Other studies have reported better R^2 for $PM_{2.5}$ based on traffic and land use, perhaps because they have been conducted across generally larger metropolitan regions with both regional and local variation in $PM_{2.5}$ (Brauer et al., 2003; Eeftens et al., 2012a; Moore et al., 2007).

5. Conclusion

In Southern California, a combined model for land use effects on EC on a fine spatial scale within multiple communities was generally robust, although there was marked heterogeneity in effect estimates for the CALINE4 near-roadway traffic metric that could partially be explained by regional pollutant concentrations and distance to shoreline. Predictors other than near-roadway traffic metrics substantially improved model fit in some communities. In addition, traffic prediction models for a novel 0.2 size fraction we had hypothesized would be a better marker for near-roadway pollution were not substantially better than for $EC_{2.5}$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Support was provided by the National Institutes of Health (grants P01ES011627, P01ES022845, P01ES009581, R01ES016535, R01HL076647, P30ES007048 and K25ES019224), U.S. Environmental Protection Agency (grants RD83544101, R826708, RD831861) and the Hastings Foundation. The authors thank Meng Wang for guidance with the analysis, Maryam Taher and John Wilson for preparation of covariates, and Ed Rappaport and Lisa Grossman for assistance with data management.

References

Allen RW, Amram O, Wheeler AJ, Brauer M. The transferability of NO and NO2 land use regression models between cities and pollutants. Atmos Environ. 2011; 45:369–378.

Beckerman B, Jerrett M, Brook JR, Verma DK, Arain MA, Finkelstein MM. Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. Atmos Environ. 2008; 42:275–290.

Beelen R, Hoek G, Fischer P, van den Brandt PA, Brunekreef B. Estimated long-term outdoor air pollution concentrations in a cohort study. Atmos Environ. 2007; 41:1343–1358.

Benson P. CALINE4 - A dispersion model for predicting air pollution concentrations near roadways. 1989 Created November 1984, updated June and November 1989.

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; 29:879–888. [PubMed: 17251230]

Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U, Heinrich J, Cyrys J, Bellander T, Lewne M, Brunekreef B. Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems. Epidemiology. 2003; 14:228–239. [PubMed: 12606891]

CALTRANS. 2009 Annual Average Daily Traffic (AADT) for all vehicles on California State Highways. California Department of Transportation; Sacramento, CA: 2010.

Carr D, von Ehrenstein O, Weiland S, Wagner C, Wellie O, Nicolai T, von Mutius E. Modeling annual benzene, toluene, NO2, and soot concentrations on the basis of road traffic characteristics. Environmental Research. 2002; 90:111–118. [PubMed: 12483801]

Clougherty JE, Wright RJ, Baxter LK, Levy JI. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. Environ Health-Glob. 2008; 7

Cyrys J, Heinrich J, Hoek G, Meliefste K, Lewne M, Gehring U, Bellander T, Fischer P, van Vliet P, Brauer M, Wichmann HE, Brunekreef B. Comparison between different traffic-related particle indicators: elemental carbon (EC), PM2.5 mass, and absorbance. J Expo Anal Environ Epidemiol. 2003; 13:134–143. [PubMed: 12679793]

Eckel SP, Berhane K, Salam MT, Rappaport EB, Linn WS, Bastain TM, Zhang Y, Lurmann F, Avol EL, Gilliland FD. Residential traffic-related pollution exposures and exhaled nitric oxide in the children's health study. Environ Health Perspect. 2011; 119:1472–1477. [PubMed: 21708511]

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, Declercq C, Dedele A, Dons E, de Nazelle A, Dimakopoulou K, Eriksen KT, Falq G, Fischer P, Galassi C, Grazuleviciene R, Heinrich J, Hoffmann B, Jerrett M, Keidel D, Korek M, Lankki T, Lindley S, Madsen C, Molter A, Nador G, Nieuwenhuijsen MJ, Nonnemacher M, Pedeli X, Raaschou Nielsen O, Patelarou E, Quass U, Ranzi A, Schindler C, Stempfelet M, Stephanou EG, Sugiri D, Tsai M, Yli-Tuomi T, Varro MJ, Vienneau D, von Klot S, Wolf K, Brunekreef B, Hoek G. Development of land use regression models for PM2.5, PM2.5 absorbance, PM10 and PMcoarse in 20 European study areas; results of the ESCAPE project. Environmental Science & Technology. 2012a

Eeftens M, Tsai MY, Ampe C, Anwander B, Beelen R, Bellander T, Cesaroni G, Cirach M, Cyrys J, de Hoogh K, De Nazelle A, de Vocht F, Declercq C, Dedele A, Eriksen K, Galassi C, Grazuleviciene R, Grivas G, Heinrich J, Hoffmann B, Iakovides M, Ineichen A, Katsouyanni K, Korek M, Kramer U, Kuhlbusch T, Lanki T, Madsen C, Meliefste K, Molter A, Mosler G, Nieuwenhuijsen M, Oldenwening M, Pennanen A, Probst-Hensch N, Quass U, Raaschou-Nielsen O, Ranzi A, Stephanou E, Sugiri D, Udvardy O, Vaskoevi E, Weinmayr G, Brunekreef B, Hoek G. Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2 - Results of the ESCAPE project. Atmos Environ. 2012b; 62:303–317.

EPA. Report to Congress on Black Carbon. United States Environmental Protection Agency; 2012. EPA-450/R-12-001

Franklin M, Vora H, Avol E, McConnell R, Lurmann F, Liu F, Penfold B, Berhane K, Gilliland F, Gauderman WJ. Predictors of intra-community variation in air quality. J Expo Sci Environ Epidemiol. 2012; 22:135–147. [PubMed: 22252279]

Fruin S, Urman R, Lurmann F, McConnell R, Gauderman WJ, Rappaport E, Franklin M, Gilliland F, Shafer M, Gorski P, Avol E. Spatial Variation in Particulate Matter Components over a Large Urban Area. Atmos Environ. In Press. 10.1016/j.atmosenv.2013.10.063

Gauderman WJ, Avol E, Gilliland F, Vora H, Thomas D, Berhane K, McConnell R, Kuenzli N, Lurmann F, Rappaport E, Margolis H, Bates D, Peters J. The effect of air pollution on lung development from 10 to 18 years of age. N Engl J Med. 2004; 351:1057–1067. [PubMed: 15356303]

Ghio AJ, Smith CB, Madden MC. Diesel exhaust particles and airway inflammation. Curr Opin Pulm Med. 2012; 18:144–150. [PubMed: 22234273]

Guxens M, Sunyer J. A review of epidemiological studies on neuropsychological effects of air pollution. Swiss Med Wkly. 2012; 141:w13322. [PubMed: 22252905]

HEI. Special Report: Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects. Health Effects Institute; Boston, MA: 2010.

Hochadel M, Heinrich J, Gehring U, Morgenstern V, Kuhlbusch T, Link E, Wichmann HE, Kramer U. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. Atmos Environ. 2006; 40:542–553.

Kam W, Liacos JW, Schauer JJ, Delfino RJ, Sioutas C. Size-segregated composition of particulate matter (PM) in major roadways and surface streets. Atmos Environ. 2012; 55:90–97.

Kinney PL, Aggarwal M, Northridge ME, Janssen NAH, Shepard P. Airborne concentrations of PM2.5 and diesel exhaust particles on Harlem sidewalks: A community-based pilot study. Environmental Health Perspectives. 2000; 108:213–218. [PubMed: 10706526]

Lee SJ, Demokritou P, Koutrakis P, Delgado-Saborit JM. Development and evaluation of personal respirable particulate sampler (PRPS). Atmos Environ. 2006; 40:212–224.

McConnell R, Islam T, Shankardass K, Jerrett M, Lurmann F, Gilliland F, Gauderman J, Avol E, Kunzli N, Yao L, Peters J, Berhane K. Childhood incident asthma and traffic-related air pollution at home and school. Environ Health Perspect. 2010; 118:1021–1026. [PubMed: 20371422]

Moore DK, Jerrett M, Mack WJ, Kunzli N. A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. J Environ Monitor. 2007; 9:246–252.

Morgenstern V, Zutavern A, Cyrys J, Brockow I, Gehring U, Koletzko S, Bauer CP, Reinhardt D, Wichmann HE, Heinrich J. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. Occup Environ Med. 2007; 64:8–16. [PubMed: 16912084]

Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ, Stenseth NC. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol Evol. 2005; 20:503–510. [PubMed: 16701427]

Poplawski K, Gould T, Setton E, Allen R, Su J, Larson T, Henderson S, Brauer M, Hystad P, Lightowlers C, Keller P, Cohen M, Silva C, Buzzelli M. Intercity transferability of land use regression models for estimating ambient concentrations of nitrogen dioxide. J Expo Sci Environ Epidemiol. 2009; 19:107–117. [PubMed: 18398445]

Putaud JP, Raes F, Van Dingenen R, Bruggemann E, Facchini MC, Decesari S, Fuzzi S, Gehrig R, Huglin C, Laj P, Lorbeer G, Maenhaut W, Mihalopoulos N, Mulller K, Querol X, Rodriguez S, Schneider J, Spindler G, ten Brink H, Torseth K, Wiedensohler A. European aerosol phenomenology-2: chemical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe. Atmos Environ. 2004; 38:2579–2595.

Putaud JP, Van Dingenen R, Alastuey A, Bauer H, Birmili W, Cyrys J, Flentje H, Fuzzi S, Gehrig R, Hansson HC, Harrison RM, Herrmann H, Hitzenberger R, Huglin C, Jones AM, Kasper-Giebl A, Kiss G, Kousa A, Kuhlbusch TAJ, Loschau G, Maenhaut W, Molnar A, Moreno T, Pekkanen J, Perrino C, Pitz M, Puxbaum H, Querol X, Rodriguez S, Salma I, Schwarz J, Smolik J, Schneider J, Spindler G, ten Brink H, Tursic J, Viana M, Wiedensohler A, Raes F. A European aerosol phenomenology-3: Physical and chemical characteristics of particulate matter from 60 rural, urban, and kerbside sites across Europe. Atmos Environ. 2010; 44:1308–1320.

Riedl M, Diaz-Sanchez D. Biology of diesel exhaust effects on respiratory function. J Allergy Clin Immunol. 2005; 115:221–228. quiz 229. [PubMed: 15696072]

Ryan PH, Lemasters GK, Biswas P, Levin L, Hu S, Lindsey M, Bernstein DI, Lockey J, Villareal M, Khurana Hershey GK, Grinshpun SA. A comparison of proximity and land use regression traffic exposure models and wheezing in infants. Environ Health Perspect. 2007; 115:278–284. [PubMed: 17384778]

Ryan PH, LeMasters GK, Levin L, Burkle J, Biswas P, Hu SH, Grinshpun S, Reponen T. A land-use regression model for estimating microenvironmental diesel exposure given multiple addresses from birth through childhood. Sci Total Environ. 2008; 404:139–147. [PubMed: 18625514]

Schauer JJ. Evaluation of elemental carbon as a marker for diesel particulate matter. J Expo Anal Environ Epidemiol. 2003; 13:443–453. [PubMed: 14603345]

Vienneau D, de Hoogh K, Beelen R, Fischer P, Hoek G, Briggs D. Comparison of land-use regression models between Great Britain and the Netherlands. Atmos Environ. 2010; 44:688–696.

Zhu YF, Hinds WC, Kim S, Sioutas C. Concentration and size distribution of ultrafine particles near a major highway. Journal of the Air & Waste Management Association. 2002; 52:1032–1042. [PubMed: 12269664]

Highlights

- Fine spatial scale EC and PM were measured in Southern Californian communities.
- Combined multi-community prediction models were generally robust for EC.
- There was substantial heterogeneity in effects of near-roadway traffic metrics on EC.
- This heterogeneity varied by regional pollution and distance to shoreline.
- Model R^2 of a novel 0.2 μm EC fraction was not larger than for $EC_{2.5}$.

Figure 1. Map of communities

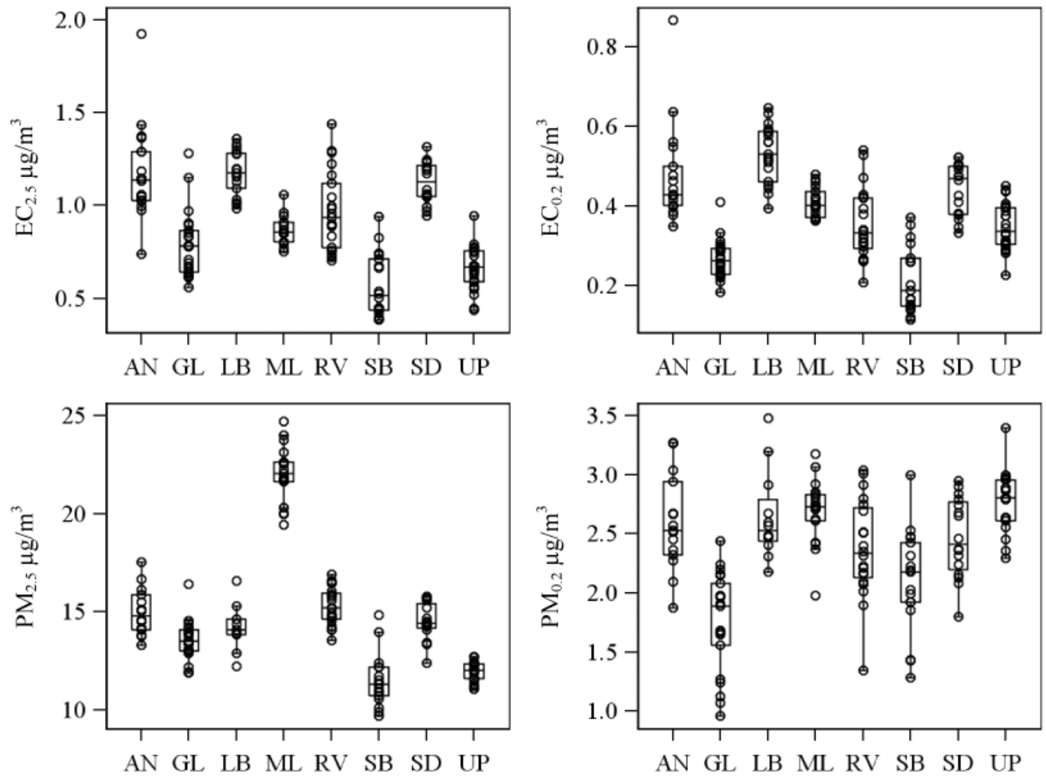
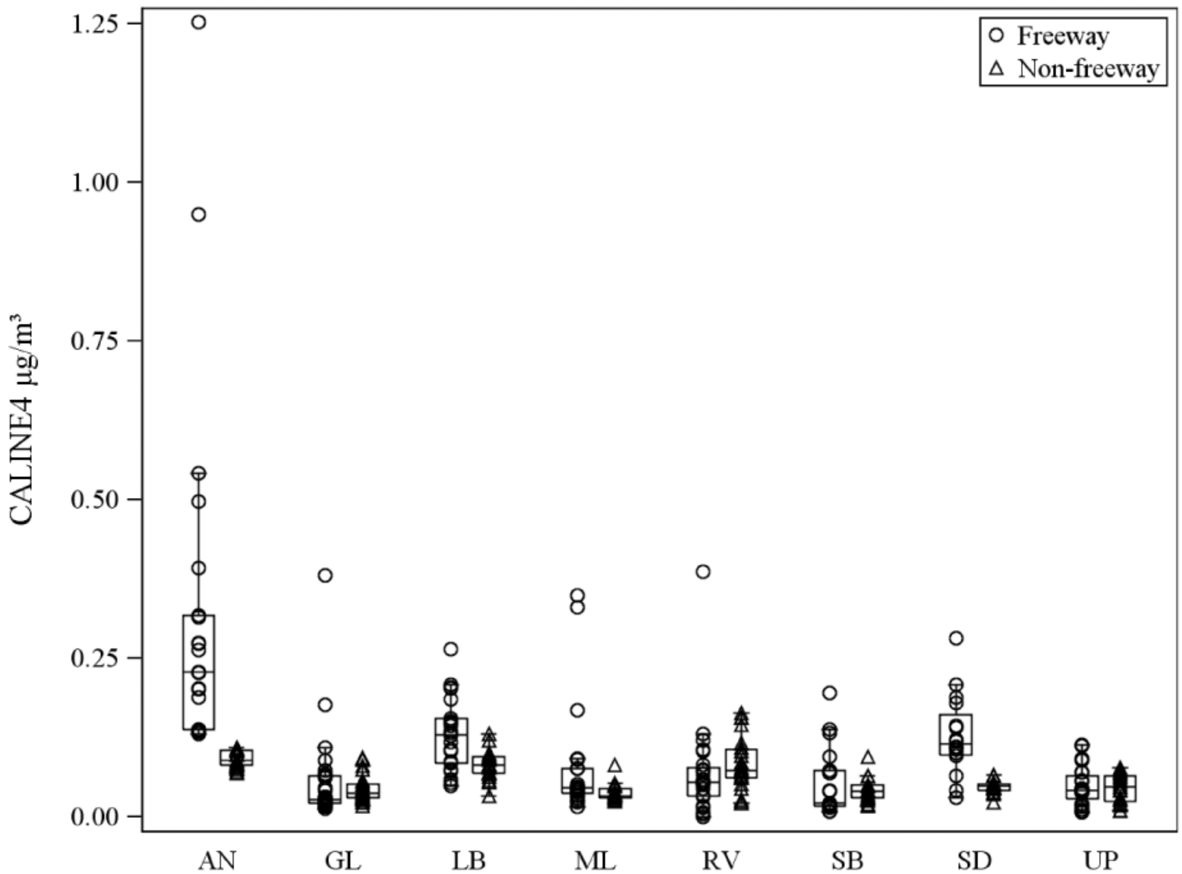


Figure 2.

Distribution of eight-week averaged concentrations of EC and PM in 2.5 and 0.2 µm size fractions.


Freeway
Non-freeway
CALINE4 μg/m3
1.25
1.00
0.75
0.50
0.25
0.00
AN
GL
LB
ML
RV
SB
SD
UP

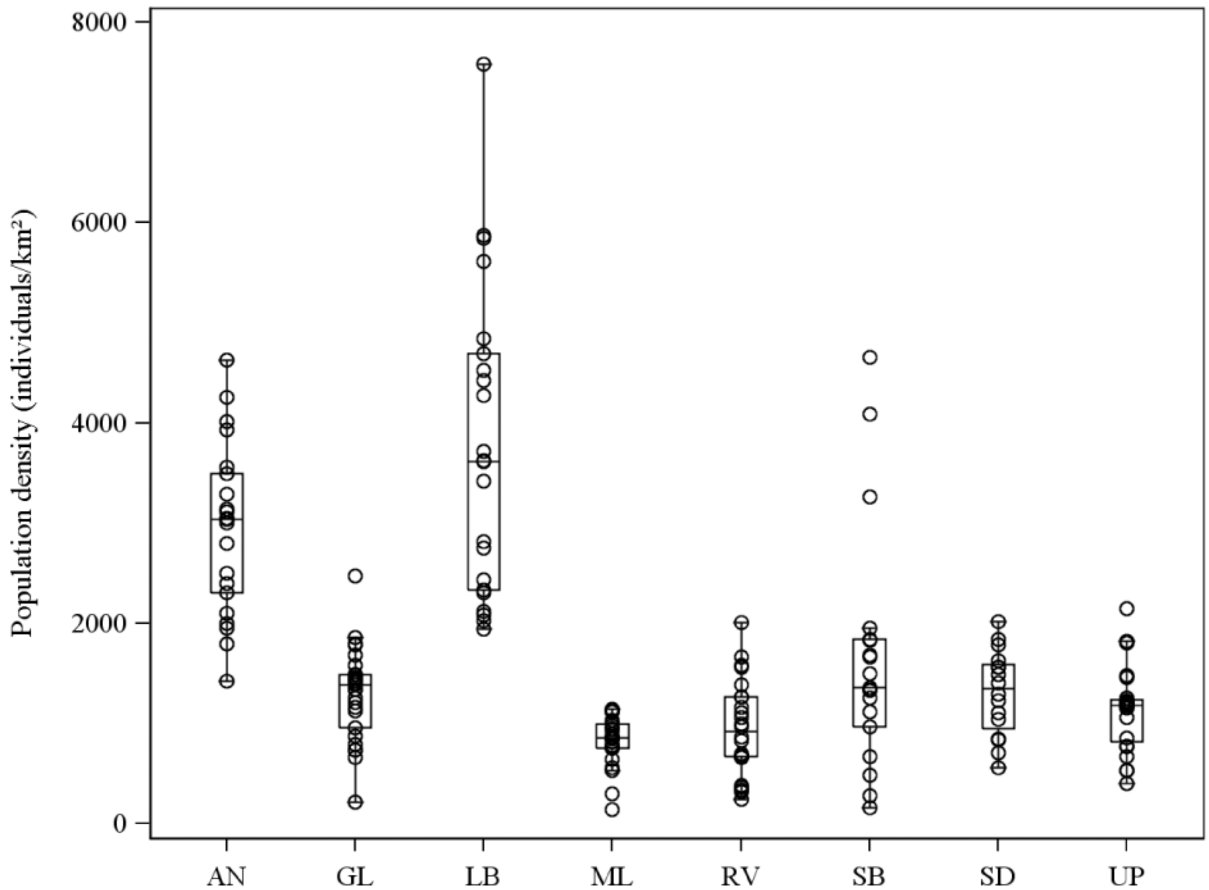


Figure 3.
Distribution of selected predictor variables by community.

Table 1

Eight-week geometric mean concentration of measured pollutants (in μg/m^3) and coefficient of variation (CV in %) in each community.

	$EC_{2.5}$		$EC_{0.2}$		$PM_{2.5}$		$PM_{0.2}$	
Community	**Geo Mean**	**Geo CV**	**Geo Mean**	**Geo CV**	**Geo Mean**	**Geo CV**	**Geo Mean**	**Geo CV**
Anaheim (AN)	1.15	20.5	0.46	22.6	14.97	8.0	2.56	15.7
Glendora (GL)	0.77	21.5	0.26	17.7	13.52	7.4	1.70	27.1
Long Beach (LB)	1.17	10.6	0.53	14.0	14.15	8.9	2.62	13.5
Mira Loma (ML)	0.87	9.6	0.41	8.9	21.98	6.7	2.68	10.7
Riverside (RV)	0.95	21.7	0.35	25.3	15.26	6.2	2.34	19.1
Santa Barbara (SB)	0.55	29.3	0.20	39.8	11.42	11.3	2.05	22.8
San Dimas (SD)	1.12	10.1	0.44	15.5	14.48	6.7	2.45	14.2
Upland (UP)	0.65	18.6	0.34	17.4	11.94	4.3	2.76	9.6
All measurements[a]	0.86	33.2	0.36	36.6	14.45	20.7	2.34	24.0
8 community mean[b]	0.90	17.7	0.37	20.1	14.71	7.4	2.40	16.6

[a]Represents the geometric mean and CV using all measurements across all communities. The geometric CV is a mixture of between and within-community variation.

[b]Unweighted average of the reported community specific geometric means and CVs. The geometric CV represents the within-community variation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Components of within- and between-community variance (and ratio of within- to between-community variance).

	Within Variance	Between Variance	Ratio
$EC_{2.5}$	0.03	0.05	0.53
$EC_{0.2}$	0.006	0.01	0.54
$PM_{2.5}$	1.23	11.0	0.11
$PM_{0.2}$	0.14	0.12	1.14

Table 3

Pairwise correlations between deviated (community-centered) pollutants levels and potential predictors.

Predictors	$EC_{2.5}$[a]	$EC_{0.2}$[a]	$PM_{2.5}$[a]	$PM_{0.2}$[a]
CALINE4[a]				
Freeway	0.69 **	0.65 **	0.39 **	0.29 **
Non-freeway	0.35 **	0.39 **	0.16	0.16
Total	0.72 **	0.71 **	0.41 **	0.28 **
Distance				
Freeway	-0.49 **	-0.48 **	-0.23 **	-0.21 *
Large arterial roads	-0.04	-0.16	-0.10	-0.01
Traffic density[a]				
150m buffer	0.49 **	0.57 **	0.34 **	0.14
300m buffer	0.53 **	0.54 **	0.30 **	0.15
Freeway truck count	0.17 *	0.17 *	0.26 **	0.17
Road buffers (all roads)				
50m	0.23 **	0.26 **	0.22 *	0.26 **
100m	0.33 **	0.35 **	0.28 **	0.14
150m	0.33 **	0.37 **	0.22 *	0.08
200m	0.33 **	0.37 **	0.22 *	0.08
250m	0.31 **	0.35 **	0.20 *	0.07
300m	0.33 **	0.34 **	0.17	0.07
Elevation	-0.45 **	-0.47 **	-0.36 **	-0.22 **
Population density (300m buffer)	0.33 **	0.27 **	0.30 **	0.26 **
Normalized difference vegetation index (NDVI)	-0.29 **	-0.24 **	-0.19 *	-0.05
Distance to railway	-0.40 **	-0.44 **	-0.18 *	-0.17
Distance to intermodal facility (weighted)	-0.21 *	-0.26 **	-0.15	-0.14
Distance to point source of NO_x (10-50 tons/yr)	-0.11	-0.03	-0.04	-0.18 *

[a]On log scale (and in all following tables).

[*]p-value<0.05;

[**]p-value<0.01

Table 4

Prediction models across all eight communities[a]

Predictors[a]	$EC_{2.5}$	$EC_{0.2}$	$PM_{2.5}$	$PM_{0.2}$
Logged freeway CALINE4				0.059
Logged total CALINE4	0.255	0.291		
Total length of roads in 50m buffer				0.074
Total length of roads in 100m buffer			0.039	
Freeway truck count			0.041	
Elevation			-0.039	
Population density (300 m buffer)	0.050		0.037	0.071
Distance to point source of NO_x (10-50 tons)				-0.048
Adjusted R^2	0.53	0.51	0.27	0.16
Leave-one-out cross-validated (LOOCV) R^2	0.51	0.49	0.21	0.12
Leave-one-community-out cross-validated (LOCOCV) R^2	0.48	0.47	0.20	0.14

[a]Reported betas are scaled to two standard deviations of deviated predictors across all eight communities as follows: 1.6 units for logged freeway CALINE4, 1.1 units for logged total CALINE4, 116.7 meters for total length of roads in 50m buffer, 340.9 meters for total length of roads in 100m buffer, 6250 trucks for freeway truck count, 84.5 meters for elevation, 1574 individuals/km^2 for population density, and 3210 meters for distance to point source of NO_X (10-50 tons).

Table 5

Leave-one-out cross-validated (LOOCV) R^2 for prediction models in Table 4 applied to each community.

Towns	**$EC_{2.5}$**	**$EC_{0.2}$**	**$PM_{2.5}$**	**$PM_{0.2}$**
Anaheim	0.64	0.91	0.26	0.03
Glendora	0.74	0.42	0.24	0.14
Long Beach	0.00	0.54	0.27	0.14
Mira Loma	0.02	0.37	0.11	0.04
Riverside	0.50	0.50	0.03	0.16
Santa Barbara	0.82	0.80	0.39	0.23
San Dimas	0.08	0.09	0.43	0.12
Upland	0.51	0.54	0.23	0.11

Table 6

Community specific $EC_{2.5}$ models

Predictors[a]	AN	GL	LB	ML[b]	RV	SB	SD	UP
Logged freeway CALINE4	0.395			-				
Logged total CALINE4		0.324		-	0.304	0.338		0.222
Elevation				-			-0.214	
Population density (300m buffer)				-		0.097		
Distance to shoreline				-				
NDVI				-	-0.252			
Freeway truck count			0.120		0.204			
Distance to intermodal facility					-0.048			
Adjusted R^2	0.67	0.73	0.61		0.79	0.82	0.44	0.50
LOOCV R^2	0.58	0.70	0.54	-	0.75	0.77	0.36	0.44
N	18	22	17	13	21	18	16	23

[a] Reported betas are scaled to two standard deviations of the deviated predictors across all eight communities as follows: 1.6 units for logged freeway CALINE4, 1.1 units for logged total CALINE4, 84.5 meters for elevation, 1574 individuals/km^2 for population, 5260 meters for distance to shoreline, 0.1 for NDVI, 6250 trucks for freeway truck count, and 4300 meters for distance to intermodal facility.

[b] A stable model could not be fit for Mira Loma (ML), which was likely attributable to the small number of samples available for this analysis.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 7

Community specific $EC_{0.2}$ models (reported betas in each column followed by R^2 for each community)[a]

Predictors[a]	AN	GL	LB	ML	RV	SB	SD	UP
Logged freeway CALINE4			0.351					
Logged non-freeway CALINE4								0.164
Logged total CALINE4	0.446	0.170			0.250	0.508		
Distance to FCC3						-0.184		-0.141
Logged traffic density (150m radius)				0.142				
Elevation		-0.145					-0.202	
NDVI							-0.107	-0.113
Adjusted R^2	0.91	0.48	0.56	0.57	0.49	0.82	0.31	0.70
LOOCV R^2	0.86	0.38	0.49	0.52	0.40	0.74	0.21	0.57
N	17	22	19	19	20	17	16	22

[a]Reported betas are scaled to two standard deviations of the deviated predictors across all eight communities as follows: 1.6 units for logged freeway CALINE4, 0.8 units for logged non-freeway CALINE4, 1.1 units for logged total CALINE4, 393 meters for distance to FCC3, 1.9 units for logged traffic density, 84.5 meters for elevation, and 0.1 for NDVI .

Table 8

Leave-one-out cross-validated (LOOCV) R^2 of various hierarchical combined models[a].

Community level modifier of total CALINE4	$EC_{2.5}$	$EC_{0.2}$
*No modifier (from*Table 4)	0.51	0.49
NO_x central site	0.57	0.52
$PM_{2.5}$ central site	0.54	0.54
Average of community $EC_{2.5}$		
measurements	0.52	0.50
Averaged population density	0.52	0.54
Averaged shoreline distance	0.54	0.61

[a] NO_x and $PM_{2.5}$ measurements came from fixed monitoring sites, while $EC_{2.5}$, population density, and distance to shoreline were averaged across pollution measurement sites by community.