Title
Source Characterization and Exposure Modeling of Gas-Phase Polycyclic Aromatic Hydrocarbon (PAH) Concentrations in Southern California.

Authors
Masri, Shahir
Li, Lianfa
Dang, Andy
et al.

Publication Date
2018-03-01

DOI
10.1016/j.atmosenv.2018.01.014

Peer reviewed

HHS Public Access

Author manuscript

Atmos Environ (1994). Author manuscript; available in PMC 2019 March 01.

Published in final edited form as:

Atmos Environ (1994). 2018 March ; 177: 175–186. doi:10.1016/j.atmosenv.2018.01.014.

Source Characterization and Exposure Modeling of Gas-Phase Polycyclic Aromatic Hydrocarbon (PAH) Concentrations in Southern California

Shahir Masri[1,§], **Lianfa Li**[2,3,§], **Andy Dang**[1], **Judith H. Chung**[4], **Jiu-Chiuan Chen**[3], **Zhi-Hua (Tina) Fan**[5], and **Jun Wu**[1,*]

[1]Program in Public Health, College of Health Sciences, University of California, Irvine, CA, 92697, U.S.A

[2]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

[3]Department of Preventive Medicine, University of Southern California, Los Angeles, CA, 90032, U.S.A

[4]University of California, Irvine Medical Center, Orange, 92868, CA, U.S.A

[5]New Jersey Department of Health and Senior Services, Program of Chemical Terrorism, Biomonitoring, and Food Service ECLS/PHILEP, Trenton, NJ, 08625 U.S.A

Abstract

Airborne exposures to polycyclic aromatic hydrocarbons (PAHs) are associated with adverse health outcomes. Because personal air measurements of PAHs are labor intensive and costly, spatial PAH exposure models are useful for epidemiological studies. However, few studies provide adequate spatial coverage to reflect intra-urban variability of ambient PAHs. In this study, we collected 39-40 weekly gas-phase PAH samples in southern California twice in summer and twice in winter, 2009, in order to characterize PAH source contributions and develop spatial models that can estimate gas-phase PAH concentrations at a high resolution. A spatial mixed regression model was constructed, including such variables as roadway, traffic, land-use, vegetation index, commercial cooking facilities, meteorology, and population density. Cross validation of the model resulted in an R^2 of 0.66 for summer and 0.77 for winter. Results showed higher total PAH concentrations in winter. Pyrogenic sources, such as fossil fuels and diesel exhaust, were the most dominant contributors to total PAHs. PAH sources varied by season, with a higher fossil fuel and wood burning contribution in winter. Spatial autocorrelation accounted for a substantial amount of the variance in total PAH concentrations for both winter (56%) and summer (19%). In summer,

*Corresponding author: Program in Public Health, Anteater Instruction & Research Bldg (AIRB) # 2034, University of California, Irvine CA 92697-3957. Tel: 949-824-0548, Fax: 949-824-0529, junwu@uci.edu.

§First authors with equal contribution

other key variables explaining the variance included meteorological factors (9%), population density (15%), and roadway length (21%). In winter, the variance was also explained by traffic density (16%). In this study, source characterization confirmed the dominance of traffic and other fossil fuel sources to total measured gas-phase PAH concentrations while a spatial exposure model identified key predictors of PAH concentrations. Gas-phase PAH source characterization and exposure estimation is of high utility to epidemiologist and policy makers interested in understanding the health impacts of gas-phase PAHs and strategies to reduce emissions.

Keywords

air pollution; polycyclic aromatic hydrocarbon; source; land-use regression; spatial model; Thiessen polygons

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) in ambient air are a common environmental hazard in urban areas. Exposure to PAHs is associated with systemic inflammation in adults and increases the risk of cardiopulmonary and lung cancer mortality (Ahmed et al., 2013; Delfino et al., 2010; Grant, 2009; Lee et al., 2011; Lewtas, 2007). Maternal PAH exposure has also been associated with adverse perinatal outcomes, including low birth weight, intrauterine growth retardation, and *in-utero* fetal death (Dejmek et al., 1999; Dejmek et al., 2000; Perera et al., 1999; Perera et al., 1998; Wu et al., 2010). Emissions of PAHs to the atmosphere are primarily a result of incomplete combustion of fossil fuels and other organic materials (Lewtas, 2007). Major sources include heavy traffic, petroleum spills, commercial/restaurant cooking, smelting operations, coal burning, as well as wood and other biomass burning. Such sources of PAH emissions have been documented in numerous source apportionment studies across different countries (Chen et al., 2009; Khan et al., 2015; Teixeira et al., 2015; Tobiszewski and Namiesnik, 2012).

Ambient concentrations of PAHs can be highly variable spatially (Zhu et al., 2011). Centrally located monitors may therefore underestimate exposure and inter-person variability (Wilson et al., 2005). In recent years, land-use regression (LUR) models have become the state-of-the-art method to estimate ambient air pollutant concentrations (Jerrett et al., 2005). These models estimate local pollutant concentrations using spatially-resolved variables reflecting emission intensity, proximity to sources, and atmospheric dispersion (Hoek et al., 2008). The LUR approach has also been extended to model the spatial variation of air concentrations, with a range of applications including traffic related air pollutants, particulate matter (PM), and volatile organic compounds (Ryan and LeMasters, 2007). In contrast, only few studies have developed LUR models to predict organic compounds (e.g. PAH) (de Hoogh et al., 2013; Jedynska et al., 2015; Saraswat et al., 2013; Vienneau et al., 2013).

Studies aimed at developing spatial models of PAH concentrations have previously been conducted (Jedynska et al., 2014; Noth et al., 2016). Few studies, however, have developed LUR models to estimate ambient concentrations of PAHs (Noth et al., 2013). A study in the eastern U.S. modeled particle-bound benzo-pyrene and coronene based on 50 two-day

samples collected at different times (Polidori et al., 2010). The models utilized four variables (temperature, wind speed, inverse distance to interstate roadways, and atmospheric stability) and explained approximately 66% of the variance in PAH concentrations (4-6 ring PAHs). A study in the western U.S. developed models that explained 80% of the spatial and 18% of the temporal variability in daily PAH concentrations (particle-bound 4-, 5- and 6-ring PAHs) using spatial and temporal variables, and a random effect (Noth et al., 2011). However, characterization of temporal variability in this model relied on continuous ambient measurements at a central site, which may not be available for other study regions. Another U.S. study developed a model using phenanthrene concentrations from pine needle samples across 91 locations (Noth et al., 2013). This model included eight spatial variables and was able to explain 56% of the variability in phenanthrene concentrations. However, instead of measuring ambient PAHs, airborne phenanthrene concentrations were indirectly measured from pine needles with data collected only in the winter season.

Among the studies that model ambient PAH concentrations, all studies modeled particle- as opposed to gas-phase PAHs (Jedynska et al., 2014; Noth et al., 2016; Polidori et al., 2010). The focus of previous studies on particle-bound PAHs was likely due to the concern with carcinogenicity of benzo-a-pyrene, one of the most common particle-bound PAH congeners (ATSDR, 1999). However, gas-phase PAHs are also potentially toxic, and sometimes orders of magnitude higher in terms of their concentration relative to the particle phase. Therefore, the paucity of empirical data on the temporal and spatial distribution of gas-phase PAHs represents a major shortcoming in our understanding of airborne exposures to PAH and their associated health risks (Chen et al., 2009).

Another major shortcoming of previous studies on urban air monitoring of PAHs is the very limited numbers of sampling sites with concurrent measurements, likely constrained by the labor-, equipment-, and time-intensive nature of PAH sampling (Noth et al., 2011; Polidori et al., 2010). This creates challenges in exposure modeling due to the difficulty in separating temporal variability from spatial variability. In addition, all published studies on PAH modeling use linear modeling techniques, without considering potential non-linear relationship between pollutant concentrations and independent variables (Li et al., 2013). Further, previous studies did not appropriately specify the spatial correlations among different sampling locations, although accounting for the spatial structure may improve model performance (Li et al., 2012).

The primary objective of this study was to characterize gas-phase PAH concentrations, examine their emission sources, and model their spatial and temporal distribution in southern California by season at a high spatial resolution. We conducted gas-phase PAH measurements simultaneously over four one-week periods at over 35 locations in the Los Angeles metropolitan area in southern California in both a warm and a cool season. Emission contributions were also examined in terms of their source. LUR models were developed for PAH concentrations by season using Thiessen polygons and roadway and traffic data, land-use patterns, and meteorological data.

2. Methods

2.1 Study Area

This study was conducted in the Los Angeles Metropolitan Area, California. This region contains some of the most polluted places in the country (ALA, 2011). The urban core of the area (Los Angeles-Long Beach-Santa Ana) encompasses the nation's largest marine port complex, housing six major commuter and truck transport freeways (AAPA, 2010). The 15 million registered on-road vehicles in the greater Los Angeles Basin are among the largest contributors to fresh emissions of PAHs (Eiguren-Fernandez et al., 2007). At the time of PAH sampling, the study region had a population density of 2,702 inhabitants per km^2 (U.S.C.B., 2010).

2.2 Sampling Methods

PAH samples were collected using the Fan-Lioy passive PAH sampler (FL-PPS) in south Los Angeles County and Orange County during two separate weeks in summer (July 10-18 and July 24-August 1) and winter (November 13-21 and December 4-12) in 2009 (Fan et al., 2006). The total sampling domain was approximately 3,500 km^2 in area. The sampling sites were residential outdoor locations of a subset of pregnant women who enrolled in our prospective Air Pollution and Birth Outcomes Study as well as the locations of 11 South Coast Air Quality Management District (SCAQMD) stations. Note, samples collected at SCAQMD sites were not collected by the SCAQMD stations, rather these were merely additional locations where our own team collected measurements. Each sampler collected in this study was deployed for a one-week period (maximum variability was 24 h). For summer and winter sampling periods, over 85% of samples represented repeated measures (collected at the same site). Though we were unable to collect repeated measurements at all sites due to issues of willingness and availability of participants, recruitment of additional participants in nearby locations allowed us to nonetheless collect a similar number of measurements for both seasons. Overall, we obtained valid measurements at 40 sites in summer and 39 sites in winter.

The small and light-weight passive PAH sampler provides a cost-effective and convenient means of monitoring ambient PAH concentrations simultaneously at multiple sites. However, using passive samplers comes with limitations, including the ability to measure only gas-phase PAHs and uncertainty in measurements of medium molecular weight PAHs such as fluoranthene and pyrene. Nevertheless, analyses have still shown medium molecular weight PAH congeners to exist mostly in the gas phase (Chen et al., 2009).

2.3 Lab analysis

Eight PAHs were selected for testing, including naphthalene (NAP), acenaphthylene (ACEN), acenaphthene (ACE), anthracene (ANT), fluorene (FLN), phenanthrene (PHE), fluoranthene (FL), and pyrene (PY). Abbreviations are also listed in Table 1, along with physical characteristics of each congener. These compounds are more abundant in the ambient air of urban environments. A significant fraction of these species are distributed in the gas phase for various ambient conditions.

The samples were analyzed directly by thermal desorption. This method is less labor intensive compared to traditional solvent extraction methods and yields lower detection limits since the entire sample is transferred to gas chromatography/mass spectrometry (GC/MS) for analysis. During sample analysis, adsorption units were removed from the FL-PPS holder and then stacked in a 7-in. long with 1/2-in. i.d. stainless steel thermal desorption tube. Known amounts of PAH standards (d_8-naphthalene, d_{10}-phenanthrene, and d_{10}-pyrene) were spiked on the sampler adsorption units before analysis in order to monitor the variation of the analytical system. Thermal desorption was conducted by a Temkar 2000 Thermal Desorber according to the following conditions: 20 min sample tube desorption time, 250 °C desorption temperature, 40 cm^3/min desorption flow rate, 0 °C Cryofocus internal trap temperature, 295 °C desorption preheat temperature, 300 °C injection/desorption temperature, 8 min internal trap desorption time, 40 cm^3/min desorption flow rate, and 300 °C transfer line temperature. The split injection mode was used for FL-PPS analysis because the GC split flow was used to transfer the analyte from the internal trap to the GC column.

A Varian 3900/2000 GC/MS system (Varian,Walnut Creek, CA) equipped with a SPB-5 capillary column (30 m × 0.25 mm i.d. × 0.5 μm film thickness) was used for the PAH analysis. The injection volume was 1 μL, splitless, with an injector temperature of 300 °C and carrier gas flow rate of 1.2 mL/min. The oven temperature program was 45 °C (1 min), ramped at 20 °C/min to 150 °C, ramped at 5 °C/min to 200 °C, then 20 °C/min to 300 °C (1 min). The molecular ion of each PAH was used for quantification. The method detection limits (MDLs) for PAHs were 1-3 ng/m^3 (estimated as 3 times the standard deviation of the response derived from analysis 7 FL-PPS7 “blanks” over a sampling duration of 24 hours).

2.4 Descriptive Data Analysis and Source Characterization

We analyzed data and developed models using R 3.1 (R Foundation for Statistical Computing, Vienna, Austria). Exploratory data analysis was performed, including summary statistics, box plots, identification of outliers, histograms for normality testing, correlation analyses, and scatter plots.

To characterize the source contributions to measured PAH concentrations, we applied molecular diagnostic ratios (DRs), a widely used method for determining the sources of PAH congeners (Khan et al., 2015; Tobiszewski and Namiesnik, 2012). This conventional method is considered less complicated and easier to interpret, in comparison to other methods such as positive matrix factorization and chemical mass balance models. DRs were calculated using the concentrations of specific PAH congeners that are chemically similar and assumed to behave similarly in the environment. The DR of various congeners provides a chemical “fingerprint” of identified distinct emission sources. Previously, DRs have been used to distinguish diesel and gasoline combustion, crude oil processing products, and various biomass burning processes (Tobiszewski and Namiesnik, 2012; Yunker et al., 2002).

In this analysis, two separate DRs were calculated to distinguish between pyrogenic and petrogenic sources. Pyrogenic PAHs are produced from high-temperature combustion of fossil fuels and biomass such as automobile engines, power plants, and wood burning,

whereas petrogenic PAHs originate from petroleum, including crude oil, fuels, lubricants, and their derivatives (Stogiannidis and Laane, 2015).

For source identification using the FL/(FL+PY) ratio, a value < 0.4 was used to indicate the dominance of petrogenic sources, a value between 0.4-0.5 to indicate fossil fuel combustion, and a value > 0.5 to indicated either wood, grass or coal combustion. To identify diesel source emissions, the FLN/(FLN+PY) ratio was used. A value > 0.5 indicated diesel sources, whereas a lower value indicated petrogenic source. DRs and corresponding source interpretations applied in this study represent standard applications that have been commonly used and reported elsewhere (Khan et al., 2015; Tobiszewski and Namiesnik, 2012).

Since PY is used in both of the DR equations applied in this study, a sensitivity analysis was carried out to determine the extent to which PY measurement error could impact each DR, and in turn impact our source characterization results. Specifically, both DRs were calculated using measured PY in addition to measured PY±10% and PY±20%. Results using these percent deviations were then compared to the results that were based off of our true PY measurements.

We did not use the DR based on AN and PHE concentrations to define pyrogenic *versus* petrogenic source characterization, because this alternative ratio has been criticized in terms of its reliability due to the reported differential in photochemical reactions rates of the two congeners, and therefore was not utilized in this analysis (Tobiszewski and Namiesnik, 2012).

Instead of taking averages across seasons, we determined the season-specific DRs for each single site (with 2 winter DRs and 2 summer DRs per site) to better understand the frequency with which PAH samples were influenced by various sources. The resulting DRs were then investigated for seasonal variability. DRs were not calculated for sample sites if the PAH species measurements required to calculate the DRs were below the MDL for one or both species in the DR equations. This did not affect the number of winter samples analyzed, however it resulted in the exclusion of approximately 50% of summertime samples. To assess the sensitivity to source strength characterization, DR analyses for summertime were also carried out using DRs for all samples (zero excluded samples).

Source contribution results using DRs were assessed in three ways. First, DR values for all sample sites were calculated. The percent of samples influenced by each source based on each site's DR value was then calculated. To provide insight about source strength, the mean total PAH concentration across sites within each source category and within each season was calculated. Additionally, total PAH mass associated with each source type was summed and mass fractions for each source type were calculated. The mass fraction for a given source type was calculated relative to the sum of PAH mass across all sites within a season. For instance, if 10 of 40 sites were associated with the fossil fuel source in winter, the fossil fuel contribution for winter would be reported as 25%, the mean would be the mean of total PAH concentrations for the 10 sites, and the mass fraction would be the sum of total PAH mass for the 10 sites divided by the sum of total PAH mass for the 40 sites.

2.5 Spatial Model Variables

Based on prior knowledge, we examined two major categories of variables (emissions-related; meteorological) that potentially contribute to PAH concentrations. Emissions-related variables included data on roadway characterization (proximity to freeways; length of total roadways), traffic densities, land-use, population density, and remote-sensing based vegetation index. The meteorological variables included wind speed, temperature, relative humidity, and atmospheric stability.

2.5.1 Roadway and Traffic-Related Variables—We obtained roadway data for the study region from ESRI StreetMap™ North America 10.0. This dataset was bundled with ArcGIS® software products and included 2003 TeleAtlas® street data. We obtained 2002 annual average daily traffic (AADT) count data from the California Department of Transportation (Caltrans) with continuous coverage of total traffic counts on freeways, highways, and major arterial roads. AADT was produced by Caltrans based on a combination of measurements and modeled values. This is an alternative to the more limited approach of using only traffic counts.

In this study, we calculated AADT-derived traffic density and road length within a buffer, and distance to freeways/highways. Traffic densities were calculated using the density plotting feature of Spatial Analyst in ArcInfo GIS 9.1, which calculates the density of point features around each output raster cell (ESRI, Redlands, CA). Previous studies indicate that ultrafine particles and carbon monoxide drop to near-background levels at 200 m downwind from major roadways during daytime hours (10 AM – 6 PM) and up to 2,000 m downwind during pre-sunrise hours (4:00 – 7:30 AM) (Hu et al., 2009). There is no perfect cut point to define the size of a traffic influenced zone, because changes in roadway pollution dispersion depend greatly on atmospheric stability and wind patterns. Therefore, we used the buffering decaying analysis method to empirically find the optimal buffer distance with the maximum Pearson's correlation with PAH concentrations. Traffic density was defined as the average number of vehicles that occupy one kilometer of road space, expressed in vehicles per kilometer, and was calculated using AADT with an optimal buffering distance of 500 m. Road length within the buffer of 700 m had the maximum correlation with aggregated PAH concentrations and was selected as a covariate.

2.5.2 Land-Use Type—Southern California Association of Governments (SCAG) 2008 land use data at the parcel level were used to identify land use composition around sampling locations. We classified the original 108 land-use types into four major categories: transportation; agriculture, open space and vacant; industrial; and residential. We calculated the percentage of area for each land-use category within different buffer sizes (50 m-15 km) around each sampling location.

2.5.3 Population Density—We obtained block level population data from the U.S. Census 2000. Population kernel density was calculated in ArcGIS (version 10.0; ESRI, Redlands, CA) using a 5,000 m search radius and a 30 m resolution.

2.5.4 Vegetation Index—We obtained 30 × 30 m Enhanced Thematic Mapper Plus (ETM +) data from Landsat's thematic mapper for three to five cloud-free days in each sampling period. From the ETM+ data, we extracted normalized difference vegetation index (NDVI) using Environment for Visualizing Images (ENVI) software (ITT Visual Information Solutions, Boulder, CO). NDVI was calculated using a standard algorithm and data of the near infrared band and red band (ENVI, 2011). NDVI is useful for assessing the density of vegetation, with a scale of 0 to 1; 0 indicating sparse vegetation and 1 indicating dense vegetation. According to NASA, nearly all satellite vegetation indices employ this difference formula to quantify the density of plant growth on the Earth (NASA, 2017). Further details on NDVI can be found on the NASA website (NASA, 2017).

2.5.5 Other Spatial Data—We obtained InfoUSA 2008 business database from InfoUSA from the Southern California Association of Government (SCAG). The InfoUSA data listed detailed commercial data based on specific North American Industry Classification System (NAICS) codes. For analysis, we assessed full-service restaurants, limited-service restaurants, cafeterias, grill buffets, and buffets, snack and nonalcoholic beverage bars, food service, contractors, caterers, and mobile food services.

2.5.6 Meteorological Data—Hourly ambient parameters including temperature, relative humidity, wind speed, and wind direction were obtained from the nearest weather monitoring station operated by National Weather Service and South Coast Air Management District. These meteorological data were aggregated into weekly averages. We additionally obtained atmospheric stability class data every three hours at 40 × 40 km resolution from the nearest (EDAS) modeling grid of the National Oceanic and Atmospheric Administration. We defined EDAS-class E, F and G as a stable category, and A, B, C, and D as unstable and neutral. The percent of time with stable air conditions was calculated for each sampling period.

2.6 Spatial Model Development

In order to identify proper covariates during model development, we conducted stepwise backward elimination. Specifically, we applied the following steps: (1) the variance inflation factor (VIF) for each covariate was generated and the covariates with VIF>10 were removed as predictive covariates; (2) all covariates were used to assess initial model performance (R^2), and then each covariate was removed as a predictive covariate one-by-one in order to assess how their removal impacted the performance metric (R^2); (3) the covariate (when removed) that resulted in the highest model performance (highest increase in R^2) was removed; (4) steps (2) to (3) were repeated until no covariates were removed from the models.

Because of their log-normal distribution, PAH concentrations were log-transformed before model development. Thiessen polygons were constructed around the sampling locations to derive a spatial weight matrix (SWM). This was accomplished using the Analysis Toolbox of ArcGIS 10.1. Thiessen polygons generated for the samples are depicted in Figure 1 (subject locations not shown due to privacy), covering a total area of approximately 3,000 km^2. In the SWM, an entry (i, j) is 1 if Thiessen polygon i is a neighbor of Thiessen polygon

j and 0 if not. The SWM represents the spatial influence of neighbors. A spatial mixed regression model was constructed using spatial covariates as well as a SWM:

$$g(\hat{\mu}_u) = \mu_0 + \sum_{i=1}^{m} f_i(x_u^i, df) + f_{swe}(region(s)) + \varepsilon \quad (1)$$

where μ_0 is the model intercept, $x_u^i \in X$ are the covariates, $f_i(\ldots)$ is the smooth function consisting of series basis functions (representing the non-linear relationship), df is degrees of freedom that controls the smooth degree of the curve fit, and m is the number of covariates. For normal distribution of the log-transformed concentrations, the link function is $g(\mu_u) = \mu_u$. $\varepsilon^c \sim N(0, \Sigma^c)$, SWM, i.e. $\sum^c = [\sigma_{ij}^c]$ represents the spatial autocorrelation that was incorporated into the model. ε represents the normal residual, $\varepsilon \sim N(0, \sigma^2)$. $f_{swe}(region(s))$ is the spatial effect embedded within the model that represents spatial autocorrelation.

Specifically, the spatial effect was modeled using the following formula:

$$f_{swe}(r) | f_{swe}(r'), r \neq r' \sim N\left(\frac{1}{N}\sum_{r' \in \delta_r} f_{swe}(r'), \tau^2/N_r\right) \quad (2)$$

where δ_r denotes the set of neighbors (r') to region r and $N_r = |\delta_i|$ denotes the number of adjacent spatial units. This formula illustrates that any sample point within a certain region r will be affected by their neighboring regions ($r' \in \sigma_r$). τ^2 is *a priori* variance that is weakly informatively determined using inverse gamma distribution (Gelman, 2006). The neighbor topology is determined by the SWM. A spatial random effect was used within the model and embedded as a random additive term within the model (Umlauf et al., 2015).

This model was applied to the aggregate of the six species with the highest molecular weight (univariate and multivariate analysis), which excluded NAP and PHE. Importantly, these were excluded from modeling because these species were present in concentrations that were disproportionately high relative to all other species. Incorporating these species would have therefore produced PAH prediction models that were exclusively driven by these two species (and mostly by NAP), without giving insight as to the distribution of lower level PAH congeners, which was of interest in this study. By excluding these species, we were therefore able to estimate the distribution of lower concentration (less studied) PAH congeners. Models were constructed separately for summer and winter seasons and for annual averages. Both linear and non-linear models were applied and their results were compared.

2.7 Spatial Model Evaluation

Leave-one-out cross validation (LOOCV) was employed to evaluate model performance. LOOCV uses a single observation from the original sample as the validation data and the remaining observations as the training data. This was repeated such that each observation in

the sample was used once as the validation data. To evaluate model performance using LOOCV analysis, we specifically assessed R^2, inter-quartile range of prediction errors, the square root of the mean of the squared prediction errors, and box plots of precision errors.

2.8 Exposure Mapping

To illustrate the spatial pattern of PAH predictions across a subset of our study region, an exposure map was created for summer and winter. PAH estimates were generated by applying the previously described non-linear multivariate model to a subset of data that was collected during a previous study of 140,024 birth certificate locations in 2007 within Los Angeles County (Wu and Kleeman, 2016). The continuous exposure surface of PAH concentrations was generated for Los Angeles County using ordinary krigging in the Spatial Analysis Toolbox of ArcGIS 10.1. For this process, a spherical semivariagram model was specified with an output cell size of 500 m. PAH concentration estimates were not extended into Orange County since no PAH estimates were available for subject locations in Orange County from our previous project.

3. Results

3.1 Descriptive Analysis

Table 1 depicts physical characteristics and summary statistics for all eight PAH congeners as well as total PAH concentrations measured during summer and winter. Winter exhibited the highest PAH concentrations when compared to summer. The average (standard deviation) concentration for summer and winter was 236.5 ng/m^3 (153.6 ng/m^3) and 299.1 ng/m^3 (163.3 ng/m^3), respectively. The ratio of standard deviation to mean concentration was high for both seasons, with the ratio in summer (0.65) being higher than that of winter (0.55) on average.

Figure 2 depicts average concentrations of all eight congeners during summer and winter. Concentrations of gas-phase PAHs were overwhelming dominated by NAP (85-90%) and PHE (3-9%). As shown, the seasonal pattern of PAH concentrations persists even in the absence of NAP and PHE. In general, winter had higher PAH concentrations than summer. For the congeners ACEN, FLN, and PHE, concentrations differed by a factor of two or more. The only exceptions were for the congeners AN and PY, which had higher summertime concentrations.

Table 2 presents the Pearson correlation coefficients for eight PAH congeners, showing the correlation between species varied by season. In summer, the most correlated PAHs were FLN-ACE (0.81), PHE-FL (0.79), and PHE-FLN (0.52). In winter, the species showing the highest correlation were PHE-FLN (0.68) and PHE-FL (0.62), which were the correlated summer species, as well as FLN-FL (0.59) and PY-FL (0.51). Generally, lighter weight congeners tended to correlate better with other light congeners. The same tendency existed for heavier congeners. During summer, the first sampling period showed several more PAH congener concentrations to be highly correlated, including PY-FLN (0.85), PY-ACE (0.82), PY-PHE (0.62), and PHE-ACE (0.52). No other sampling periods showed additional high correlations between congeners.

Table 3 lists descriptive statistics of the spatial variables and the total PAH concentrations of the six heavier congeners (for model use) by season and the corresponding correlations between total PAHs and the spatial variables. There was a roughly two-fold difference between summer (15.6 ng/m^3) and winter (30.2 ng/m^3) average PAH concentrations. Higher spatial variability was observed in the summer than in the winter, with a standard deviation to mean concentration ratio of 0.71 for summer and 0.49 for winter.

3.2 Source Characterization

Figure 3 presents results of the DR analysis using FL and PY congeners in terms of the percent of samples influenced by each source as well as the mean PAH concentrations across different source types. For both summer and winter seasons, all three sources contributed substantially to total PAH concentrations. Seasonal variability in source profiles was noticeable, with fossil fuel sources more predominant in winter samples (28%) than in summer samples (18%). Similarly, the wood, grass, or coal source was more influential across winter samples (60%) than summer samples (38%). Taken together, pyrogenic sources (fossil fuel + wood, grass, coal sources) were predominant in 56% and 88% of samples during summer and winter, respectively. In contrast, petrogenic sources accounted for 44% (summer) and 12% (winter) of samples. DRs for all samples collected during the two sampling periods in summer (N=80) and two periods in winter (N=78) for both DR analyses are provided in table S1.

Results for the DR analysis using FLN and PY congeners indicated a strong influence of diesel emissions (>59% of samples) for both seasons. This suggests a larger fossil fuel contribution than that estimated by the DR that uses FL and PY. A potential explanation for this difference was elucidated by our sensitivity analysis below, which explored the impact of ±10% and ±20% changes in PY congener concentrations.

Sensitivity analyses demonstrated that PY measurement uncertainties of ±10% and ±20% could have a noticeable impact on source characterization results when applying the DR that uses FL and PY congeners, but not when applying the DR that uses FLN and PY congeners. In the former case, the impact was mostly in winter. Using this DR, PY±10% showed the proportion of samples with a predominant fossil fuel contribution in summer to range from 15-18%. This range did not increase at PY±20%. In winter, however, fossil fuels predominantly influenced 18-32% of samples with PY±10% and 18-36% of samples with PY±20%, respectively. In terms of the wood, grass, or coal source, the proportion of samples dominated by this source in summer ranged narrowly (35-40%) with a ±20% deviation in PY. In winter, however, 10% and 20% deviations in PY had a considerable impact, producing ranges of 56-76% and 47-77% of samples dominated by this source, respectively.

Though uncertainties in PY impacted individual source category results in winter, taken together, the proportion of samples influenced by pyrogenic sources (fossil fuel + wood, grass, or coal sources) did not vary substantially during summer (48-55%) or winter (83-98%) even when PY±20% was tested. Similarly, the proportion of samples influenced by the petrogenic source did not vary greatly. In summer, the proportion of samples influenced by the petrogenic source ranged from 45-53% at PY±20%. In winter, PY±10%

and PY±20% resulted in this source influencing between 6-12% and 2-17% of samples, respectively.

The DR that uses FL and PY congeners to identify diesel sources was not sensitive to tested changes in PY. Variations in PY (both PY±10% and PY±20%) had negligible impacts on the proportion of samples showing a predominant diesel source influence. Source contribution results based on mass fraction analyses showed general agreement (similar source distribution) with the results described in this section for both seasons. In summer, petrogenic sources accounted for the majority of total PAH mass (51.3%), followed by the wood, grass, or coal source (37.2%) and fossil fuel source (11.4%). In winter, the wood, grass, or coal source was predominant (67.9.2%), followed by fossil fuel (23.4%) and petrogenic (8.8%) sources. When taken together, pyrogenic sources accounted for 48.7% and 91.2% of PAH mass in summer and winter, respectively.

When looking at mean PAH concentrations across different source types, the same pattern emerges. Such results are also labeled in Figure 3. Petrogenic sources have the highest mean concentration (287.6 ng/m^3) in summer, followed by the wood, grass, or coal source (250.3 ng/m^3) and fossil fuel source (164.4 ng/m^3). In winter, the wood, grass, or coal source exhibited the highest mean concentration (331.5 ng/m^3), followed by the fossil fuel (247.7 ng/m^3) and petrogenic (227.2 ng/m^3) sources.

A sensitivity analysis that calculated DRs for all summertime samples (rather than excluding below MDL species) showed little change in mean PAH concentration compared to the previously described results. Only the wood, grass, or coal source was impacted, showing a slightly reduced (8.4%) mean concentration of 229.2 ng/m^3 after all samples were included.

3.3 PAH Predictors

In the summertime, the proportion of heavy industrial land-use was significantly correlated with PAH levels, while in the winter traffic density and other emission related factors (i.e. high-density residential land-use, agriculture land-use, and park and recreational land-use) were significantly correlated with PAHs. Meteorological factors (relative humidity, wind speed, and atmospheric stability) had no significant association with the total PAH concentrations.

3.4 Univariate Analyses

Table 4 presents the results of univariate analyses, including both linear and non-linear models, each with vs. without spatial autocorrelation, for total PAH concentrations by season. The incorporation of spatial autocorrelation significantly improved model performance, especially during wintertime and for linear models. For example, roadway length was a weak predictor in the models that did not incorporate spatial autocorrelation for winter PAH concentrations ($R^2 \leq 0.01$). However, the R^2 increased to 0.76 when spatial autocorrelation was taken into account.

In univariate models (Table 4), non-linear models performed better than linear models for half of the predictors and worse for one predictor (population density) when no autocorrelation was considered; they performed better for even more predictors when

autocorrelation was included. However, the improvement of the nonlinear model was not remarkable except for relative humidity in winter with no autocorrelation (R^2=0.28 in non-linear model vs. R^2=0.02 in linear model) and temperature in winter with autocorrelation (R^2=0.69 in non-linear model vs. R^2=0.47 in linear model). The highest R^2 values were observed in the winter non-linear PAH models with spatial autocorrelation for traffic-related variables (R^2>0.70) and for the number of cooking facilities. Compared to winter PAH models (R^2 ranged 0.18-0.79 with autocorrelation and 0-0.28 without autocorrelation), much lower R^2 values were observed from summer PAH models (R^2 ranged 0.01-0.34 with autocorrelation and 0-0.18 without autocorrelation).

3.5 Multivariate Analyses

Table 5 shows the non-linear multi-variate models to predict season-specific and annual average PAH concentrations. Concentrations of PAHs in winter were more influenced by traffic emissions. In general, the wintertime model performed better than that of summer, as shown by respective LOOCV results (R^2=0.79 for winter, R^2=0.61 for summer). In the wintertime, traffic density and spatial autocorrelation accounted for approximately 32% and 45% of the variance in PAH concentrations, respectively, while meteorological variables contributed less than 4% to the overall variance explained. In the summertime, emission-related variables including roadway length, meteorological variables (temperature and wind speed), and spatial autocorrelation accounted for 29%, 15%, and 24% of the variance in PAH concentrations, respectively. The annual model had the worst performance among the three models (LOOCV R^2=0.61 in summer, 0.79 in winter, and 0.43 in annual averages). Coefficients for covariates in both the linear univariate and multivariate PAH models can be found in Table S2.

3.6 Exposure Mapping

Figures 4 and 5 illustrate the spatial pattern of PAH predictions during summer and winter in Los Angeles. In summer, the mean and standard deviation of estimated PAH concentrations across our 140,024 subject locations was 18.0 ng/m^3 and 6.2 ng/m^3, respectively, with a range of 6.8-46.6 ng/m^3. Concentration estimates in summer exhibited a clear influence of vehicle traffic as concentrations were visibly highest near major roadways. The map also appears to exhibit a subtle non-linear influence of population density, with higher concentrations in the low and medium population suburban areas and lower concentrations in the more highly populated regions surrounding downtown. In winter, PAH estimates were higher overall, ranging from 17.1-49.8 ng/m^3 with a mean and standard deviation of 28.8 ng/m^3 and 6.6 ng/m^3, respectively. For winter, traffic density still exhibited a strong influence on concentration estimates, although concentrations were more homogenously dispersed. High traffic and high population areas such as Downtown Los Angeles showed the highest predictions. Other concentration hotspots such as West Covina tended to encompass major roadways and roadway intersections.

4. Discussion

This study demonstrated that meteorological and traffic parameters can be used to estimate gas-phase PAH concentrations. Spatial models included optimal spatial covariates, with non-

linear functions, and a better characterization of spatial structure, and performed reasonably well in predicting both summer (LOOCV R^2=0.66) and winter (LOOCV R^2=0.77) PAH concentrations. The ability to estimate PAH concentrations spatially using readily available data is of high utility to epidemiologists investigating the relationship between PAH exposure and diseases. To our knowledge, this is the first study that developed models to estimate gas-phase PAH concentrations by season in a large metropolitan area.

Our results demonstrated that a better characterization of spatial structure of autocorrelation contributed to improvement in model performance, accounting for a significant portion of the variance in PAH concentrations in both summer (19.1%) and winter (55.6%) seasons. Such autocorrelation suggests common influences of PAH concentrations and in turn the utility of local PAH measurements for better understanding relative PAH exposure of nearby locations where additional measurements are otherwise lacking. Our results are consistent with other studies that have demonstrated spatial autocorrelation to improve model performance by accounting for spatial variance that cannot fully be captured by physical or environmental factors in a model (Christakos, 1990; Mercer et al., 2011).

Traffic- and road-related variables were also major predictors of PAHs (15.7-21.4% of variability), including traffic density and roadway length within a 700 m buffer. Meteorological variables (temperature and wind speed) were not major predictors, although still contributed to PAH estimation (~10% of variance explained). The likely reason meteorological variables did not account for a larger fraction of the variance in PAH concentrations is the limited variation in weather that occurred during the specific (within season) sampling windows. Variables related to commercial cooking and population density were least important, and only explained the summertime variance. Additionally, we found that non-linear models did not always improve model predictions in univariate models (especially with spatial correlation incorporated), which may be partly due to the limited number of samples we had.

Due to the small sample size in this study, we modeled spatial autocorrelation using Thiessen polygon rather than spatial covariance methods such as kriging or variograms since a larger sample size is more desirable for the latter approach. In addition, an assumption of spatial homogeneity and high sampling density is required for kriging and semivariograms to have an accurate estimation of the variogram (Burrough, 1986; Mcbratney and Webster, 1986). Furthermore, our spatial random effect modeling approach based on the Markov random field can deal well with the sparse samples without the assumption of regional homogeneity. In fact, Thiessen polygons are often used to determine density of point samples and to build meshes for space-discretized analyses (Seidel, 1995).

Since the samples used in this study are sparse, we took several steps to avoid the potential overfitting issue. We limited the degrees of freedom in our non-linear models to be five in order to prevent overfitting of the non-linear associations; the results of the multivariate models illustrated reasonable associations (Tables 4 & 5). We also visually examined the results of the spatial effects. Lastly, strict LOOCV was applied, producing reasonable results (Table 5).

Our data showed NAP as the predominant congener of total gas-phase PAH concentrations, which is consistent with another study in California by Eiguren-Fernandez et al. (2004). Substantial seasonal variability of PAH concentrations existed, with winter exhibiting much higher average concentrations. This is similar to previous reports in California and elsewhere, and is likely due to reduced atmospheric mixing height, reduced photochemical degradation, and increased burning of wood and fossil fuels for domestic heating during cool months (Eiguren-Fernandez et al., 2004; Manoli et al., 2016). Seasonal variability in PAH concentrations reinforces the importance of taking season into account for the development of spatial PAH models in this study as well as future works, particularly for studies investigating short-term or sub-chronic health effects. Greater pollution heterogeneity in this region has been observed in other studies for NO_x, and is likely due to the decreased stability of the atmosphere in summer, leading to more rapid dispersion of pollutants and consequently greater concentration gradients between sources (Li et al., 2012; Mercer et al., 2011).

Correlation matrices for seasonally averaged PAH congeners showed moderate to high correlation of PHE with FL and FLN for both cool and warm seasons. The highest correlations were mostly between congeners of similar molar mass. This is due to similarities in chemical fate and transport as well as sources. Dissimilar seasonal correlations also existed, with FLN-ACE highly correlated in summer and FLN-FL and PY-FL correlated in winter. Many congeners exhibited weak correlations with one another. In general, weak correlations between congener concentrations, and variable correlations between seasons, reflects the variability in PAH sources (pyrogenic *versus* petrogenic) (Nasher et al., 2013). This is reinforced by source characterization results in this study, which suggested variability in emissions sources.

Results of source characterization showed the dominance of pyrogenic sources during both summer and winter seasons, consistent with previous studies (Chen et al., 2009; Khan et al., 2015; Teixeira et al., 2015). This was the case when assessing source distribution based on the percent of samples associated with each source as well as the PAH mass associated with each source. Since both assessments were in agreement, we can rule out the possibility that a source associated with a low percentage of samples was nonetheless responsible for a high fraction of total PAH pollution (or the converse scenario). This conclusion was further reinforced when assessing mean PAH concentrations across source types, which showed the same source contribution pattern within and across seasons.”

Variability in PAH source contributions existed across seasons, with wintertime showing an enriched influence by fossil fuel sources. This finding has been reported elsewhere and can be attributed to the use of fossil fuels to heat residential and commercial buildings during the cool season (Chen et al., 2009; EIA, 2016).

Winter also showed an increased proportion of samples influenced by wood, grass, or coal sources. Although the DR employed here serves as an indicator of wood, grass or coal sources, in this study it is not reasonable that coal is a contributing source given the lack of coal combustion in the region. Rather, wood and other biomass burning are likely sources. Source apportionment analyses in Los Angeles and other major U.S. cities have similarly

shown wood and biomass burning to contribute substantially to ambient air pollution (Masri et al., 2015; Shirmohammadi et al., 2016). This is largely because wood is still a commonly used resource for domestic heating in the winter (AQMD, 2017; EIA, 2014). Wood burning also takes place in both summer and winter seasons for recreational purposes (e.g. bon fires and camping) in residential areas, State beaches, and campgrounds in the Los Angeles and Orange County regions.

Similar to Teixeira et al. (2015), diesel emissions were found to be a substantial contributor to PAHs during both warm and cool seasons. Of note, however, according to one DR, diesel sources contribute predominantly (nearly two-thirds of samples), whereas, according to the other DR, fossil fuels contributed only modestly (nearly one-third). Since diesel sources constitute a fossil fuel source, this appears to be a conflicting result. An explanation of this discrepancy is elucidated by our sensitivity analysis using the PY congener. The analysis also helps explain why wood burning was reportedly so influential, unexpectedly accounting for the majority of pyrogenic emissions in our study.

Our sensitivity analysis demonstrated that the DR using FLN and PY congeners was a robust indicator of diesel sources. That is, results of the diesel source contribution were not sensitive to tested changes in PY concentration. In contrast, the DR that uses FL and PY congeners exhibited sensitivity to ±10% and ±20% changes in PY, particularly in winter. Since this latter DR did not exhibit much sensitivity in identifying pyrogenic sources as a whole, it appears this DR may be better at identifying petrogenic *versus* pyrogenic sources than at identifying the specific fossil fuel fraction of pyrogenic sources. When examining the results of both DRs, and comparing the proportion of samples influenced by diesel and aggregated pyrogenic sources, the proportion of influenced samples showed much better agreement.

In generating PAH exposure maps of the Los Angeles County region, winter PAH concentration estimates were higher compared to summer. This is consistent with our observational data. Further, PAH estimates averaged by season across 140,024 subject locations showed good agreement with measured PAH concentrations.

For winter, concentrations tended to peak near major roadway features, particularly in downtown Los Angeles and areas in and near the city of West Covina. Downtown is characterized by the convergence of many major roadways such as the Interstate 5, 10, and 710 freeways, as well as State Routes 110, 101, and 60, whereas West Covina is adjacent to the Interstate 605 freeway and State Route 60. Estimates of elevated PAH concentrations in these locations are consistent with regional traffic density patterns as described by the California Environmental Health Tracking Program, which shows the geographic spread of high density traffic areas to be noticeably concentrated around downtown Los Angeles area and West Covina perimeter (CEHTP, 2017). In general, PAHs were more spatially homogenous across the region compared to summer. As described previously, the higher atmospheric stability (less mixing) and reduced atmospheric boundary layer height characteristic of winter in this region helps explain the higher wintertime concentrations and increased spatial homogeneity of PAHs during this season (Li et al., 2012; Mercer et al., 2011). In spite of the overall greater atmospheric stability in winter, PAH estimates also

showed an influence by wind speed during this season. Winds tend to be higher near the coastline compared to adjacent inland areas, which may also help to explain the higher PAH concentration estimates of the inland region (USDE, 2017).

In summer, concentration estimates tended to be less uniform, and exhibited a clear influence of nearby major roadways and roadway intersections. This is consistent with previous studies that have identified narrower impact zones (e.g. 300 m downwind) of primary traffic emissions during the day (less stability and more mixing) and a wider impact zone (e.g. up to 2600 m) prior to sunrise (more stability and less mixing) (Hu et al., 2009; Zhu et al., 2002). Additionally, the summertime pattern of PAH concentrations appeared to exhibit the non-linear influence of population density, with slightly higher concentrations in the low and medium population suburbs and lower concentrations in the more highly populated regions. Lower PAH concentrations in high population regions in summer could be due to higher residential land-use, where PAH sources are less predominant (e.g. major roadways), particularly given the reduced frequency of wood burning for domestic heating during the warm season. This explanation is supported by Figure S1.b, which demonstrates through univariate analysis a negative relationship between PAH concentrations and percent high-density residential land-use in summer.

Further, these exposure maps reflect the seasonal differences in auto-correlation, which was a key predictor in both multi-variate PAH models. That spatial autocorrelation accounted for a large fraction of the variance in wintertime PAH concentrations in this study suggests that a higher sampling density may be required to capture more localized differences in PAH concentrations in this region during the cool season. In summer, by contrast, the majority of PAH variability was explained by roadway, population, and land-use variables, suggesting that the model was less affected by the limited sampling locations in capturing localized differences in PAH concentrations during the warm season.

A major strength of this study is the collection of simultaneous weekly measurements in both summer and winter seasons. This enabled us to understand the spatial variability of PAH concentrations by season, and estimate exposure levels across the region. Air pollutant concentrations are generally influenced by relative strength and distribution of emission sources, metrological factors (e.g. atmospheric stability, wind, precipitation), and local spatial features (e.g. elevation, building height). Samples collected in different times reflect the combined effects of changes in emission sources and meteorology, which makes it difficult to disentangle the individual contributions of each factor. Simultaneous measurements at multiple spatial locations, however, minimizes the impact of meteorology (all the sampling sites share the same or similar meteorology in each sampling period) and thus can effectively reveal the spatial variability of emission sources. The impact of meteorology can be determined by implementing multiple measurement campaigns in different times of year. Yet, multiple field campaigns with simultaneous measurements can be challenging in terms of feasibility and resources, depending on the type and number of sampling devices used, and needed manual labor, which limits its application in air pollution exposure studies. In the literature, simultaneous field measurements are mostly conducted for a small number of pollutants (e.g. NO_2, NO_x, O_3, VOCs), for which portable, easily deployable devices are available and lab analysis is relatively inexpensive (Aguilera et al.,

2008; Saraswat et al., 2013; Su et al., 2010). This may be part of the reason for the lack of simultaneous PAH measurement data in the literature.

An additional strength of this analysis is the focus on gas-phase PAHs. Investigating emission sources and exposure distribution relating to gas-phase PAH concentrations is of particular importance given the toxicity of such species and the current absence of such investigation relative to that of particle-bound PAH species.

A limitation of this study is that sampling sites used for model calibration were home locations (outdoors) of subjects who participated in our air pollution and birth outcome study. Therefore, sites may not have been optimal for capturing pollution hot spots or for identifying the spatial distributions of pollutants throughout the greater Southern California region. Additionally, only a limited number of sample sites were included in this study. To help compensate for these limitations and increase our sample size and distribution, we also included SCAQMD monitoring sites in our analysis. Nonetheless, if more samples had been collected evenly across the study region, the model predictions could have been further improved and some covariates may have explained a higher proportion of the variance.

The use of only gas-phase PAH measurements in this study represents a further study limitation. Since particle-bound PAHs were not accounted for, our modeling results could not be generalized to directly estimate total PAHs in ambient environments where multiple congeners are present in both the gas- and particle-phase. Given the dominance of NAP in our measurements, and because this congener exists mostly in the gas phase, our measurements should provide a good approximation of total PAH levels for the species investigated in this study. Finally, use of DRs to characterize sources is subject to error since DR values can change during the environmental fate and transport of PAHs due to chemical aging and decay. To minimize erroneous DR values and source interpretations, however, we avoided use of less reliable AN and PHE concentrations for DR calculations. Additionally, as noted previously, DR results can be sensitive to measurement error. The importance of specific PAH sources identified in this study, however, were reinforced by both DR source characterization techniques and regression analyses, and were subject to sensitivity analyses.

In future work, the collection of more PAH species, each with a higher number of samples temporally and spatially, would be valuable. This would produce a more robust model and enable better regional extrapolation. Additionally, it would enable techniques such as positive matrix factorization that can better characterize PAH source contributions. Lastly, use of active samplers, though more expensive, would enable a higher collection of PAHs, which would minimize measurement uncertainty and therefore allow the modeling of individual PAH congeners.

5. Conclusions

A spatial mixed regression model was constructed in order to predict summer and winter gas-phase PAH concentrations. Cross validation of the model demonstrated good model performance (R^2 of 0.66 for summer and 0.77 for winter). Meteorological and traffic

parameters were both demonstrated to be useful predictors for PAH concentrations, explaining about one-third of the variability.

Total PAH concentrations exhibited seasonality, with winter having higher concentrations. NAP was overwhelming the greatest contributor to total PAHs during both seasons. Pyrogenic sources including diesel emissions appeared to be the dominant contributor to total PAH concentrations in both seasons. Winter showed an enriched fossil fuel and wood burning contribution.

The ability to estimate PAH concentrations spatially is of high utility to epidemiologists investigating the relationship between PAH exposure and disease. The findings in this study are important to consider for the development of future PAH exposure models. Additionally, results provide insight about PAH sources, which is useful for designing strategies and policies aimed at pollution prevention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of Environmental Health Sciences [NIEHS R21ES016379], Natural Science Foundation of China [41471376], and the Health Effects Institute [HEI 4787-RFA09-4110-3 WU]. We would like to acknowledge Memorial Care Health System for allowing the recruitment of a subset of pregnant women and the placement of samplers at residential subject locations. We would also like to thank Yingjun Ma for help with laboratory analysis of PAH samples, Kimberly Clark for assistance in utilizing SCAG data, and all study subjects and volunteers who participated in the air pollution exposure sampling.

References

AAPA. American Association of Port Authorities, Port Industry Statistics: North American Port Container Traffic (1990 - 2009). 2010

Aguilera I, Sunyer J, Fernandez-Patier R, Hoek G, Aguirre-Alfaro A, Meliefste K, Bomboi-Mingarro MT, Nieuwenhuijsen MJ, Herce-Garraleta D, Brunekreef B. Estimation of outdoor NOx, NO2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. Environmental Science & Technology. 2008; 42:815–821. [PubMed: 18323107]

Ahmed A, Ghosh MK, Choi MC, Choi CH, Kim S. Which hydrogen atom of toluene protonates PAH molecules in (+)-mode APPI MS analysis? J Am Soc Mass Spectrom. 2013; 24:316–319. [PubMed: 23354472]

ALA. State of the AIR American Lung Association. 2011

AQMD. Healthy Hearts. South Coast Air Quality Management District; 2017.

ATSDR. J Toxicol-Cutan Ocul. Vol. 18. US Department of Health & Human Services, public health service, Agency for Toxic Substances and Disease Registry; Washington, DC: 1999. Toxicological profile for polycyclic aromatic hydrocarbons; p. 141-147.August, 1985

Burrough, PA. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press; New York: 1986.

CEHTP. Traffic Density in California, California Environmental Health Tracking Program. 2017

Chen KS, Li HC, Wang HK, Wang WC, Lai CH. Measurement and receptor modeling of atmospheric polycyclic aromatic hydrocarbons in urban Kaohsiung, Taiwan. Journal of Hazardous Materials. 2009; 166:873–879. [PubMed: 19155130]

Christakos G. A Bayesian Maximum-Entropy View to the Spatial Estimation Problem. Math Geol. 1990; 22:763–777.

de Hoogh K, Wang M, Adam M, Badaloni C, Beelen R, Birk M, Cesaroni G, Cirach M, Declercq C, Dedele A, Dons E, de Nazelle A, Eeftens M, Eriksen K, Eriksson C, Fischer P, Grazuleviciene R, Gryparis A, Hoffmann B, Jerrett M, Katsouyanni K, Iakovides M, Lanki T, Lindley S, Madsen C, Molter A, Mosler G, Nador G, Nieuwenhuijsen M, Pershagen G, Peters A, Phuleria H, Probst-Hensch N, Raaschou-Nielsen O, Quass U, Ranzi A, Stephanou E, Sugiri D, Schwarze P, Tsai MY, Yli-Tuomi T, Varro MJ, Vienneau D, Weinmayr G, Brunekreef B, Hoek G. Development of Land Use Regression Models for Particle Composition in Twenty Study Areas in Europe. Environmental Science & Technology. 2013; 47:5778–5786. [PubMed: 23651082]

Dejmek J, Selevan SG, Benes I, Solansky I, Sram RJ. Fetal growth and maternal exposure to particulate matter during pregnancy. Environ Health Perspect. 1999; 107:475–480. [PubMed: 10339448]

Dejmek J, Solansky I, Benes I, Lenicek J, Sram RJ. The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome. Environ Health Persp. 2000; 108:1159–1164.

Delfino RJ, Staimer N, Tjoa T, Arhami M, Polidori A, Gillen DL, Kleinman MT, Schauer JJ, Sioutas C. Association of Biomarkers of Systemic Inflammation with Organic Components and Source Tracers in Quasi-Ultrafine Particles. Environ Health Persp. 2010; 118:756–762.

EIA. Today in Energy. U.S. Energy Information Agency; 2014.

Eiguren-Fernandez A, Miguel Antonio H, Froines John R, Thurairatnam Suresh, Avol EL. Seasonal and spatial variation of polycyclic aromatic hydrocarbons in vapor-phase and pm2.5 in southern california urban and rural communities. Aerosol Sci Tech. 2004; 38:8.

Eiguren-Fernandez A, Avol EL, Thurairatnam S, Hakami M, Froines JR, Miguel AH. Seasonal influence on vapor- and particle-phase polycyclic aromatic hydrocarbon concentrations in school communities located in Southern California. Aerosol Sci Tech. 2007; 41:438–446.

Fan ZH, Jung KH, Lioy PJ. Development of a passive sampler to measure personal exposure to gaseous PAHs in community settings. Environmental Science & Technology. 2006; 40:6051–6057. [PubMed: 17051799]

Gelman A. Prior distributions for variance parameters in hierarchical models(Comment on an Article by Browne and Draper). Bayesian Anal. 2006; 1:515–533.

Grant WB. Air pollution in relation to U.S. cancer mortality rates: an ecological study; likely role of carbonaceous aerosols and polycyclic aromatic hydrocarbons. Anticancer Res. 2009; 29:3537–3545. [PubMed: 19667146]

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

Hu SS, 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. Atmos Environ. 2009; 43:2541–2549.

Jedynska A, Hoek G, Wang M, Eeftens M, Cyrys J, Beelen R, Cirach M, De Nazelle A, Keuken M, Visschedijk A, Nystad W, Akhlaghi HM, Meliefste K, Nieuwenhuijsen M, de Hoogh K, Brunekreef B, Kooter IM. Spatial variations of levoglucosan in four European study areas. Sci Total Environ. 2015; 505:1072–1081. [PubMed: 25461108]

Jedynska A, Hoek G, Wang M, Eeftens M, Cyrys J, Keuken M, Ampe C, Beelen R, Cesaroni G, Forastiere F, Cirach M, de Hoogh K, De Nazelle A, Nystad W, Declercq C, Eriksen KT, Dimakopoulou K, Lanki T, Meliefste K, Nieuwenhuijsen MJ, Yli-Tuomi T, Raaschou-Nielsen O, Brunekreef B, Kooter IM. Development of land use regression models for elemental, organic carbon, PAH, and hopanes/steranes in 10 ESCAPE/TRANSPHORM European study areas. Environ Sci Technol. 2014; 48:14435–14444. [PubMed: 25317817]

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. J Expo Anal Env Epid. 2005; 15:185–204.

Khan MF, Latif MT, Lim CH, Amil N, Jaafar SA, Dominick D, Nadzir MSM, Sahani M, Tahir NM. Seasonal effect and source apportionment of polycyclic aromatic hydrocarbons in PM2.5. Atmos Environ. 2015; 106:178–190.

Lee MS, Magari S, Christiani DC. Cardiac autonomic dysfunction from occupational exposure to polycyclic aromatic hydrocarbons. Occupational and Environmental Medicine. 2011; 68:474–478. [PubMed: 21172795]

Lewtas J. Air pollution combustion emissions: Characterization of causative agents and mechanisms associated with cancer, reproductive, and cardiovascular effects. Mutat Res-Rev Mutat. 2007; 636:95–133.

Li LF, 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; 47:9291–9299. [PubMed: 23859442]

Li LF, Wu J, Wilhelm M, Ritz B. Use of generalized additive models and cokriging of spatial residuals to improve land-use regression estimates of nitrogen oxides in Southern California. Atmos Environ. 2012; 55:220–228.

Manoli E, Kouras A, Karagkiozidou O, Argyropoulos G, Voutsa D, Samara C. Polycyclic aromatic hydrocarbons (PAHs) at traffic and urban background sites of northern Greece: source apportionment of ambient PAH levels and PAH-induced lung cancer risk. Environ Sci Pollut Res Int. 2016; 23:3556–3568. [PubMed: 26490935]

Masri S, Kang CM, Koutrakis P. Composition and sources of fine and coarse particles collected during 2002-2010 in Boston, MA. J Air Waste Manage. 2015; 65:287–297.

Mcbratney AB, Webster R. Choosing Functions for Semi-Variograms of Soil Properties and Fitting Them to Sampling Estimates. J Soil Sci. 1986; 37:617–639.

Mercer LD, Szpiro AA, Sheppard L, Lindstrom J, Adar SD, Allen RW, Avol EL, Oron AP, Larson T, Liu LJS, Kaufman JD. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NOx) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). Atmos Environ. 2011; 45:4412–4420.

NASA. MODIS Data Product Non-Technical Description - MOD 13. National Aeronautics and Space Administration; 2017.

Nasher E, Heng LY, Zakaria Z, Surif S. Concentrations and Sources of Polycyclic Aromatic Hydrocarbons in the Seawater around Langkawi Island, Malaysia. J Chem-Ny. 2013

Noth EM, Hammond SK, Biging GS, Tager IB. A spatial-temporal regression model to predict daily outdoor residential PAH concentrations in an epidemiologic study in Fresno, CA. Atmos Environ. 2011; 45:2394–2403.

Noth EM, Hammond SK, Biging GS, Tager IB. Mapping and modeling airborne urban phenanthrene distribution using vegetation biomonitoring. Atmos Environ. 2013; 77:518–524.

Noth EM, Lurmann F, Northcross A, Perrino C, Vaughn D, Hammond SK. Spatial and temporal distribution of polycyclic aromatic hydrocarbons and elemental carbon in Bakersfield, California. Air Qual Atmos Health. 2016; 9:9.

Perera FP, Jedrychowski W, Rauh V, Whyatt RM. Molecular epidemiologic research on the effects of environmental pollutants on the fetus. Environ Health Persp. 1999; 107:451–460.

Perera FP, Whyatt RM, Jedrychowski W, Rauh V, Manchester D, Santella RM, Ottman R. Recent developments in molecular epidemiology: A study of the effects of environmental polycyclic aromatic hydrocarbons on birth outcomes in Poland. Am J Epidemiol. 1998; 147:309–314. [PubMed: 9482506]

Polidori A, Kwon J, Turpin BJ, Weisel C. Source proximity and residential outdoor concentrations of PM(2.5), OC, EC, and PAHs. J Expo Sci Environ Epidemiol. 2010; 20:457–468. [PubMed: 19623217]

Ryan PH, LeMasters GK. A review of land-use regression for characterizing intraurban air models pollution exposure. Inhal Toxicol. 2007; 19:127–133. [PubMed: 17886060]

Saraswat A, Apte JS, Kandlikar M, Brauer M, Henderson SB, Marshall JD. Spatiotemporal Land Use Regression Models of Fine, Ultrafine, and Black Carbon Particulate Matter in New Delhi, India. Environmental Science & Technology. 2013; 47:12903–12911. [PubMed: 24087939]

Seidel R. The Upper Bound Theorem for Polytopes - an Easy Proof of Its Asymptotic Version. Comp Geom-Theor Appl. 1995; 5:115–116.

Shirmohammadi F, Hasheminassab S, Saffari A, Schauer JJ, Delfino RJ, Sioutas C. Fine and ultrafine particulate organic carbon in the Los Angeles basin: Trends in sources and composition. Sci Total Environ. 2016; 541:1083–1096. [PubMed: 26473710]

Stogiannidis E, Laane R. Source Characterization of Polycyclic Aromatic Hydrocarbons by Using Their Molecular Indices: An Overview of Possibilities. Rev Environ Contam T. 2015; 234:49–133.

Su JG, Jerrett M, Beckerman B, Verma D, Arain MA, Kanaroglou P, Stieb D, Finkelstein M, Brook J. A land use regression model for predicting ambient volatile organic compound concentrations in Toronto, Canada. Atmos Environ. 2010; 44:3529–3537.

Teixeira EC, Agudelo-Castaneda DM, Mattiuzi CD. Contribution of polycyclic aromatic hydrocarbon (PAH) sources to the urban environment: A comparison of receptor models. Sci Total Environ. 2015; 538:212–219. [PubMed: 26298853]

Tobiszewski M, Namiesnik J. PAH diagnostic ratios for the identification of pollution emission sources. Environ Pollut. 2012; 162:110–119. [PubMed: 22243855]

U.S.C.B. US Census Bureau Population and Housing Occupancy Status: 2010 - United States -- Metropolitan Statistical Area. U.S. Census Bureau; Washing, DC: 2010.

Umlauf N, Adler D, Kneib T, Lang S, Zeileis A. Structured Additive Regression Models: An R Interface to BayesX. J Stat Softw. 2015; 63:1–46.

USDE. California 30-Meter Residential-Scale Wind Resource Map. U.S. Department of Energy: Wind Exchange; 2017.

Vienneau D, de Hoogh K, Bechle MJ, Beelen R, van Donkelaar A, Martin RV, Millet DB, Hoek G, Marshall JD. Western European Land Use Regression Incorporating Satellite- and Ground-Based Measurements of NO2 and PM10. Environmental Science & Technology. 2013; 47:13555–13564. [PubMed: 24156783]

Wilson JG, Kingham S, Pearce J, Sturman AP. A review of intraurban variations in particulate air pollution: Implications for epidemiological research. Atmos Environ. 2005; 39:6444–6462.

Wu J, Hou H, Ritz B, Chen Y. Exposure to polycyclic aromatic hydrocarbons and missed abortion in early pregnancy in a Chinese population. Sci Total Environ. 2010; 408:2312–2318. [PubMed: 20219237]

Wu, J., L, O., L, L., H, J., Kleeman, M. Adverse reproductive health outcomes and exposure to gaseous and particulatematter air pollution in pregnant women. Health Effects Institute; 2016.

Yunker MB, Macdonald RW, Vingarzan R, Mitchell RH, Goyette D, Sylvestre S. PAHs in the Fraser River basin: a critical appraisal of PAH ratios as indicators of PAH source and composition. Org Geochem. 2002; 33:489–515.

Zhu X, Fan ZT, Wu X, Jung KH, Ohman-Strickland P, Bonanno LJ, Lioy PJ. Ambient concentrations and personal exposure to polycyclic aromatic hydrocarbons (PAH) in an urban community with mixed sources of air pollution. J Expo Sci Environ Epidemiol. 2011; 21:437–449. [PubMed: 21364704]

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

Highlights

- Developed a spatiotemporal regression model to estimate gas-phase PAH concentrations
- Applied Diagnostic Ratios to identify potential PAH sources
- Roadway, population, and meteorological variables were key predictors of PAHs
- PAHs were highest in winter, with a high contribution from fossil fuels and wood burning
- A map was generated to estimate PAH exposure in Los Angeles, CA

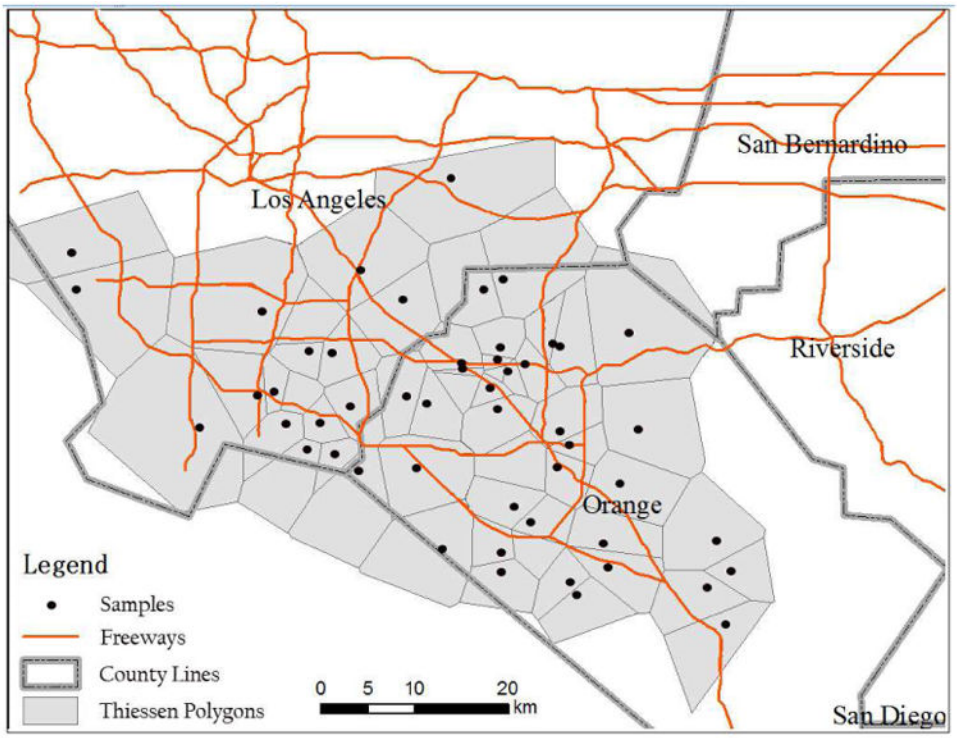

Figure 1.
Thiessen polygons for construction of spatial weight matrix in Los Angeles and Orange counties.

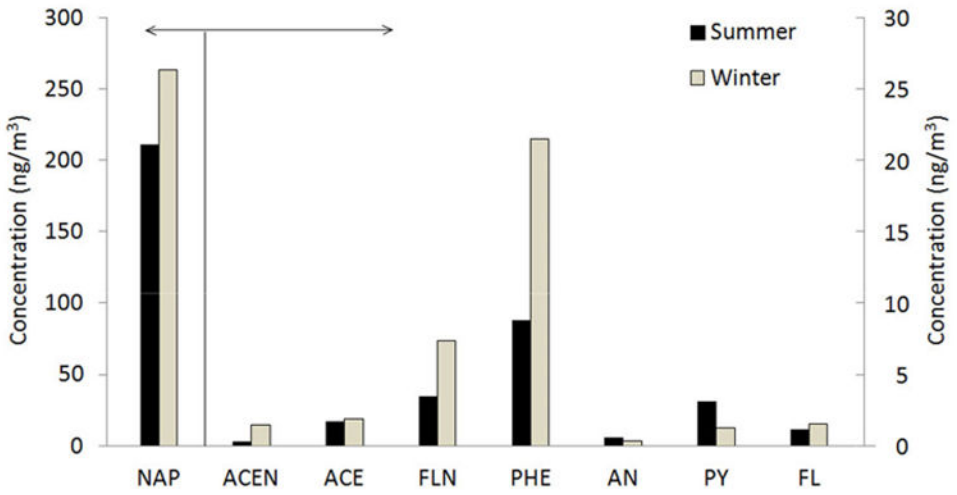

Figure 2.
Average concentration of eight PAH congeners in order of increasing molar mass.

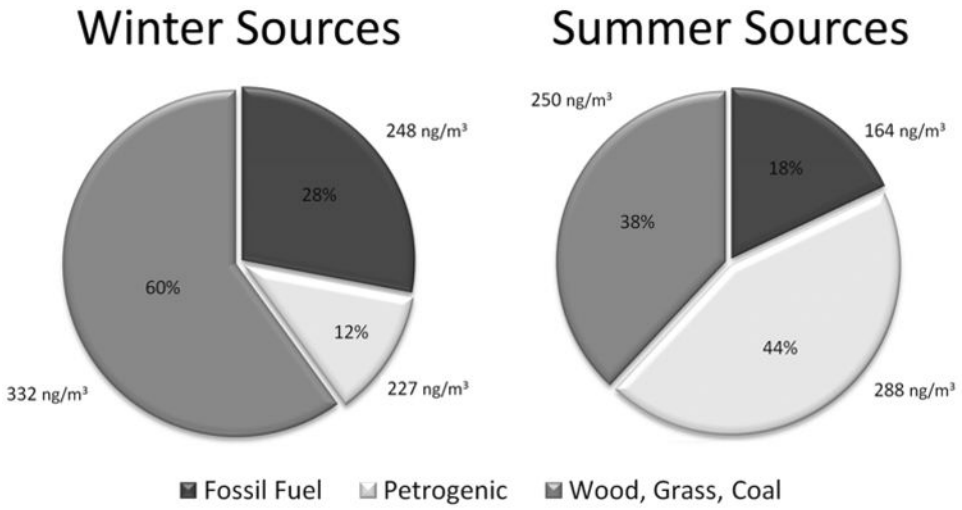

Figure 3.
Source characterization for total PAHs in winter and summer.

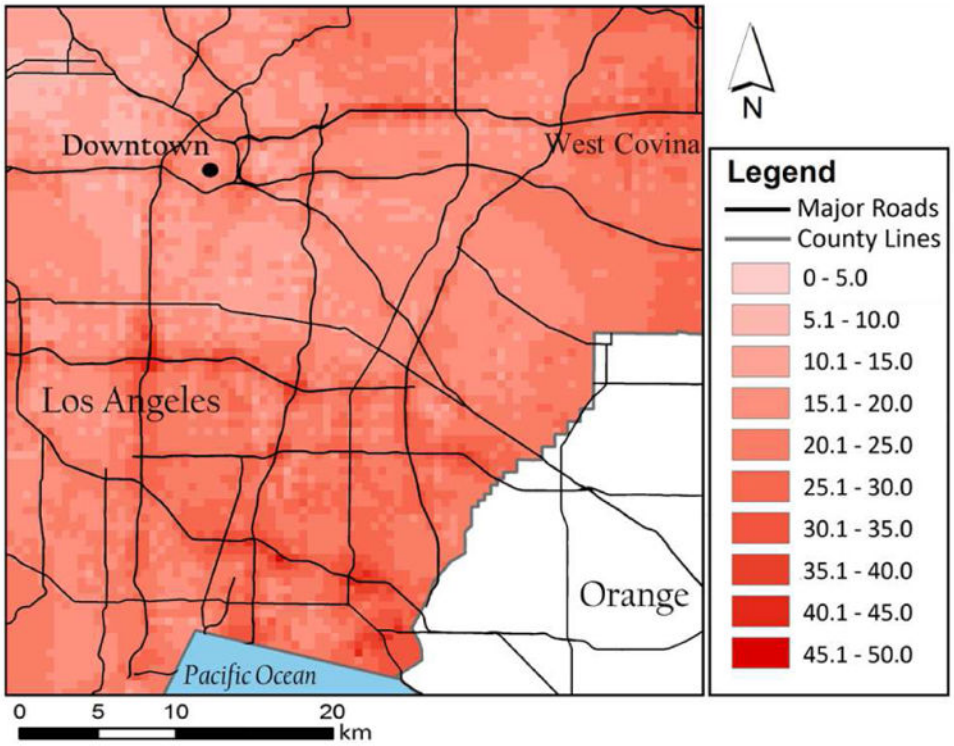

Figure 4.
PAH concentration estimates (ng/m^3) in Los Angeles County during summer.

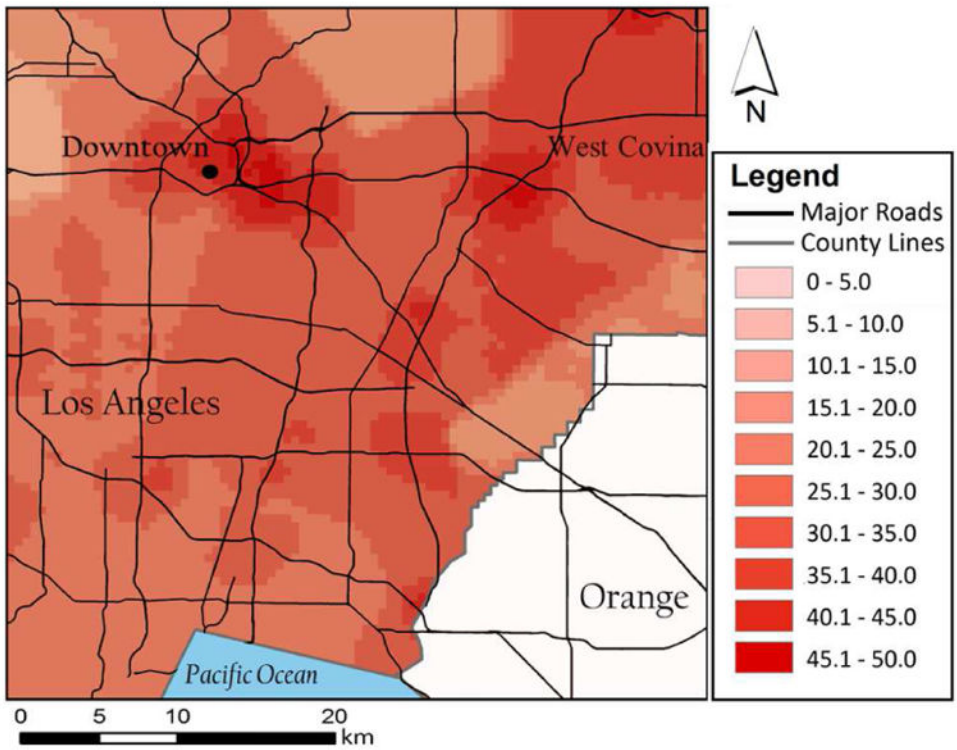

Figure 5.
PAH concentration estimates (ng/m^3) in Los Angeles County during winter.

Table 1

Physical characteristics, summary statistics, and abbreviations for PAHs.

				Summer Concentration (ng/m^3)		Winter Concentration (ng/m^3)	
PAH	**Abbreviation**	**Molar Mass [g]**[a]	**# Rings**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
Naphthalene	NAP	128.2	2	217.5	146.5	263.9	153.0
Acenaphthylene	ACEN	152.2	3	0.3	0.3	1.4	3.4
Acenaphthene	ACE	154.2	3	1.7	2.0	1.9	4.4
Fluorene	FLN	166.2	3	3.5	4.1	7.3	3.5
Phenanthrene	PHE	178.2	3	8.8	10.0	21.5	15.1
Anthracene	AN	178.2	3	0.5	1.2	0.3	0.2
Pyrene	PY	202.3	4	3.1	6.5	1.2	0.7
Fluoranthene	FL	202.3	4	1.1	1.3	1.5	0.8
Total PAHs	-	-	-	236.5	153.6	299.1	163.3

[a] From (Tobiszewski et al. 2011)

Table 2

Pearson correlation coefficients for eight PAH congener concentrations during summer and winter in order of increasing molar mass.

	NAP	ACEN	ACE	FLN	PHE	AN	PY	FL
Summer								
NAP	1.00	0.12	**0.24**	**0.35**	**0.27**	0.03	0.20	**0.29**
ACEN		1.00	-0.05	0.03	0.09	0.03	-0.04	0.12
ACE			1.00	**0.81**	**0.49**	0.10	0.14	0.14
FLN				1.00	**0.52**	0.05	0.16	**0.23**
PHE					1.00	0.01	-0.06	**0.79**
AN						1.00	-0.07	0.00
PY							1.00	-0.06
FL								1.00
Winter								
NAP	1.00	0.17	**0.27**	**0.43**	**0.37**	0.11	0.13	**0.35**
ACEN		1.00	0.03	**0.30**	0.13	-0.01	0.12	0.22
ACE			1.00	**0.28**	**0.26**	-0.07	0.19	0.17
FLN				1.00	**0.68**	-0.02	**0.24**	**0.59**
PHE					1.00	0.21	0.16	**0.62**
AN						1.00	0.07	0.10
PY							1.00	**0.51**
FL								1.00

[a] Values in bold and underlined indicate statistical significance at the p=0.05 level.

Table 3

Descriptive statistics of seasonal total concentrations of six PAH species[a] and their correlation with predictive variables

		Summer			Winter		
Variables (optimal buffer size if available)		**Mean**	**St. Dev.**	**Pearson's Corr.**	**Mean**	**St. Dev.**	**Pearson's Corr.**
PAH (ng/m^3)		**15.6**	**11.0**	**NA**	**30.2**	**14.9**	**NA**
Emission-related							
Land-use	Proportion of heavy industrial land-use (4.5 km)	1%	3%	0.43*	1%	1%	0.04
	Proportion of high-density residential land-use (150 m)	21%	27%	-0.20	20%	25%	-0.30*
	Proportion of agricultural land-use (5 km)	1%	2%	-0.10	1%	1%	-0.30*
	Proportion of park and recreational land-use (3.5 km)	5%	6%	-0.10	7%	7%	-0.43*
Traffic	Shortest distance to freeway (m)	3458	3124	-0.12	3792	2874	0.00
	Roadway length (700 m buffer, m)	7938	3766	0.37	5994	2818	-0.07
	Traffic density (vehicles per km)	122	68	0.25	106	66	0.27*
	Population density	0.002	0.001	0.03	0.003	0.001	0.14
Others	Shortest distance to commercial cooking facilities	91	56	0.23	90	55	-0.16
	NDVI (150 m)	0.24	0.11	0.00	0.25	0.09	-0.37
Meteorology							
	Temperature (°C)	19.9	2.7	0.09	16.1	2.6	0.23
	Relative humidity (%)	62%	6.7%	-0.10	60%	4.8%	-0.16
	Wind speed (m/s)	2.1	0.5	-0.14	1.9	0.5	0.08
	Proportion of stable atmosphere[b]	51%	9%	0.18	59%	11%	0.08

[a] Total PAH included fluorene, phenanthrene, anthracene, acenaphthene, fluoranthene, and pyrene.

[b] Defined as proportion of time with an estimated atmospheric stability of Class E (slightly stable), F (moderately stable), or G (extremely stable).

* indicates statistical significance.

Table 4

Non-linear (linear) univariate models with and without spatial autocorrelation for estimating seasonal total concentrations of six PAH species[a]

		Summer		Winter	
	Variables	**Non-linear (linear) R^2**	**Non-linear (linear) R^2 with spatial corr.**	**Non-linear (linear) R^2**	**Non-linear (linear) R^2 with spatial corr.**
Emission related					
Land-use	Proportion of heavy industrial land-use (4.5 km)	0.10 (0.10)	0.32 (0.32)	0.02 (0.03)	0.54(0.54)
	Proportion of high-density residential land-use (150 m)	0.02 (0.01)	0.10(0.10)	0.08 (0.08)	0.54 (0.54)
	Proportion of agricultural land-use (5 km)	0.01 (0.01)	0.14(0.03)	0.03 (0.02)	0.51 (0.51)
	Proportion of park and recreational land-use (3.5 km)	0.01 (0.01)	0.06 (0.05)	0.25 (0.20)	0.25 (0.21)
Traffic	Shortest distance to freeway (m)	0.06 (0.02)	0.20 (0.20)	0.03 (0.02)	0.70 (0.57)
	Roadway length (700 m buffer; m)	0.11(0.10)	0.03(0.03)	0.02 (0.02)	0.56 (0.56)
	Traffic density (vehicles per km)	0.03 (0.02)	0.25 (0.25)	0.10 (0.07)	0.73 (0.56)
	Population density	0.02 (0.01)	0.12 (0.12)	0.01 (0.01)	0.51 (0.51)
Others	Shortest distance to commercial cooking facilities	0.22 (0.12)	0.28 (0.2)	0.01 (0.01)	0.53 (0.53)
	NDVI (150 m)	0.01 (0.00)	0.13 (0.12)	0.07 (0.06)	0.52 (0.52)
Meteorology					
	Temperature (°C)	0.04 (0.01)	0.22 (0.16)	0.08 (0.07)	0.63 (0.63)
	Relative humidity	0.04 (0.04)	0.11 (0.12)	0.26 (0.01)	0.45 (0.46)
	Wind speed (m/s)	0.02 (0.02)	0.27 (0.24)	0.25 (0.01)	0.48 (0.48)

[a]Total PAH included fluorene, phenanthrene, anthracene, acenaphthene, fluoranthene, and pyrene.

Table 5

Non-linear multi-variable regression models with and without spatial autocorrelation for total concentrations of six PAH species[a]

Season	Predictors (variance explained)	R^2 : RMSE with (without) spatial autocorrelation	LOOCV R^2 : RMSE with (without) spatial autocorrelation
Summer	Spatial autocorrelation effect (19.1%); proportion of commercial cooking facilities (2.1%); ambient temperature (7.1%); ambient wind speed (2.1%); population density (15.2%); roadway length within a 700-m buffer (21.4%)	0.67 : 6.95 ng/m^3 (0.57:7.90 ng/m^3)	0.66: 7.00 ng/m^3 (0.57: 7.90 ng/m^3)
Winter	Spatial autocorrelation effect (55.6%); traffic density (15.7%); ambient temperature (2.4%); ambient wind speed (9.3%)	0.83:6.61 ng/m^3 (0.67: 9.14 ng/m^3)	0.77: 7.76 ng/m^3 (0.59: 10.16 ng/m^3)
Annual average	Spatial autocorrelation effect (11.6%); season (38.5%); proportion of park and recreational land-use (1.2%); traffic density (2.8%)	0.54: 10.81 ng/m^3 (0.33: 12.99 ng/m^3)	0.53: 10.90 ng/m^3 (0.33: 13.00 ng/m^3)

[a] Total PAH included fluorene, phenanthrene, anthracene, acenaphthene, fluoranthene, and pyrene