Title
Prediction of real time particulate matter concentrations on highways using traffic information and emission model

Authors
Wu, Guoyuan
Hao, Peng
Pham, Liem
et al.

Publication Date
2017

Peer reviewed

Prediction of real time particulate matter concentrations on highways using traffic information and emission...

Guoyuan Wu
University of California, Riverside

Kanok Boriboonsomsin
University of California, Riverside

Prediction of real time particulate matter concentrations on highways using traffic information and emission model

Guoyuan Wu, Ph. D
Center for Environmental Research and Technology
University of California at Riverside
1084 Columbia Ave, Riverside, CA 92507, USA
Phone: (951) 781-5630, Fax: (951) 781-5790
E-mail: gywu@cert.ucr.edu

Liem Pham, Ph. D Candidate
Center for Environmental Research and Technology
University of California at Riverside
1084 Columbia Ave, Riverside, CA 92507, USA
E-mail: lpham@ucr.edu, Fax: (951) 781-5790

Peng Hao, Ph. D
Center for Environmental Research and Technology
University of California at Riverside
1084 Columbia Ave, Riverside, CA 92507, USA
Phone: (951) 781-5777, Fax: (951) 781-5790
E-mail: haop@cert.ucr.edu

Heejung Jung, Ph. D
Center for Environmental Research and Technology
University of California at Riverside
1084 Columbia Ave, Riverside, CA 92507, USA
Phone: (951) 781-5742, Fax: (951) 781-5790
E-mail: heejung@cert.ucr.edu

Kanok Boriboonsomsin, Ph.D
Center for Environmental Research and Technology
University of California, Riverside
1084 Columbia Avenue, Riverside, CA 92507, USA
Phone: (951) 781-5792, Fax: (951) 781-5790
Email: kanok@cert.ucr.edu

[4 Tables and 6 Figures: 2,500 words]
[Text 4,936 words]

Word count: 7,436 words

Paper for the 96th Annual Meeting of
Transportation Research Board
Washington, D.C. January 2017

ABSTRACT

The public raises concerns about the exposure to particulate matter (PM) which has been strongly associated with illness and mortality. However, most of the studies rely on the measurements from stationary monitoring sites which cannot capture the actual PM exposure for those people in or near the source. In this study, we first set up a comprehensive mobile monitoring platform to measure both PM concentration and traffic conditions on some major highways in Southern California. Then, we developed an integrated database to fuse different data sources and to facilitate the investigation of relationship between traffic conditions and highway PM concentration. Using the fused datasets and combining with Emission FACtor (EMFAC) model, contour plots based on estimated PM emissions were generated with the overlay of particle concentration measurements. Analyses of the results indicate that there are numerous particle concentration peaks cause by traffic congestions and vehicle acceleration. PM concentrations may be affected by traffic conditions on the other side of the highway as shown in both measurement and emission models. In view of the complicated physical nature of PM concentration on highways, we applied the *Multivariate Adaptive Regression Splines (MARS)* model to the integrated database, and identified the eleven traffic-related variables that have the most impacts on in-source PM concentration prediction. The high coefficient of determination (i.e., $R^2 = 0.72$) indicates the capability of the model to address the variance in PM concentration.

Keywords:
Particulate matter (PM); PM concentration; mobile monitoring; emission model; pollutant emissions

1. INTRODUCTION

Traffic congestion has been the daily norm in many metropolitan areas. The associated socio-economic issues, such as the waste in energy consumption and air pollution, have received increasingattentions from the public. The major pollutants emitted by vehicles include carbon monoxide (CO), volatile organic compounds (VOCs), nitrogen oxides (NO_x), particulate matter (PM), and polycyclic aromatic hydrocarbons (PAHs) (*1*). It is estimated by the U.S. Environmental Protection Agency (USEPA) that the nationwide CO, NOx, PM (including $PM_{2.5}$ and PM_{10}), and sulfur dioxide (SO_2) emissions due to transportation activities were about 36.30, 7.16, 0.49, 0.34 and 0.10 million metric tons, respectively, in Year 2014 (*2*). Of all these commonly-seen air pollutants, PM has been strongly associated with illness and mortality, such as respiratory inflammation, allergy, and asthma attacks, as indicated in many studies (3, 4, 5, 6). For example, California Air Resources Board (CARB) estimated that annually about 9,200 people in California die prematurely as a result of exposure to $PM_{2.5}$ (*7*). Other detrimental health effects caused by the exposure to PM may include respiratory and cardiovascular morbidity (*8*, *9*). In addition, the ultrafine particles (less than 100 nanometers in diameter) whose dominant sources are diesel engine powered vehicles (*10*), have been considered to be more toxic by many researchers due to their unique physical properties, interactions with tissues and cells, and the potential for translocation beyond the lung (*11*). The National Ambient Air Quality Standards (NAAQS) set by USEPA suggest that the annual mean for primary $PM_{2.5}$ should not exceed 12 $\mu g/m^3$ and the temporal average of PM_{10} within any 24-hour period should not exceed 150 $\mu g/m^3$ (*12*).

Although a substantial body of research has been focused on assessment of public exposure to PM and the associated health effects, most of the measurement data were obtained from stationary monitoring sites which are not close enough to the sources, such as highways. This may lead to discrepancy from the actual PM exposure for those people who are in or near the sources, e.g., travelers in the traffic flow. It was reported that the average time an American spent traveling in car is nearly 1 hour everyday (*13*). Furthermore, previous studies estimated that in-cabin exposures to ultrafine particles (UFPs) might be 10 times higher than ambient levels and were responsible for 10 – 50% of total daily UFP exposure for Los Angeles commuters (*14*). In consideration of all these concerns, USEPA's new air pollution rules require near-road monitoring starting from January 2014. South Coast Air Quality Management District (SCAQMD) has also set up 4 air pollution surveillance stations in the proximity of major highways in South Coast Air Basin to monitor NOx, fine particulate matter (e.g., soot) and CO (cite). Such effort significantly improves the accuracy in measuring the PM concentration near the mobile sources. However, the measurements are highly restricted by the locations and sparsity of surveillance stations, resulting in difficulty to capture spatial variations of in-/near-source PM concentration.

To address the aforementioned issues, there are two promising approaches: 1) *measurement-based* approach; and 2) *model-based* approach. For the *measurement-based* approach, mobile PM monitoring or Lagrangian PM monitoring has become an attractive strategy, which is able to cover long spatial range with high temporal resolution and to conduct real-time assessment on people's exposure to in-/near-source PM concentration. For example, Fruin et al. (*15*) used a mobile monitoring platform including a scanning mobility particle sizer (SMPS), to measure particle counts and size distributions. There are other commercially available instruments for ambient particle monitoring in real time. A condensational particle counter (CPC) can effectively measure number-based particle concentration but not detailed information on particle size. An electrical aerosol detector (EAD) can measure aerosol diameter concentration and can implicitly estimate the effective surface area of particles. Most of these studies have been only focused on very limit-scale measurements of PM characteristics due to the significant cost for real-world experimentation. Very few studies have investigated the relationship between traffic conditions and in-/near-source PM concentration (e.g., on highways).

On the other hand, the *model-based* approach heavily relies on detailed traffic conditions and emissions models (*16*). Based on the resolution of available traffic information, microscopic, mesoscopic or macroscopic motor emission models such as MOVES (*17*), EMFAC (*18*) and PHEM (*19*), can be applied to estimate the tailpipe PM emissions. For example, Reynolds et al. applied a self-developed emission model for mobile sources to the traffic data at a test intersection to assess the traffic related PM emissions (*20*). Abou-Senna et al. used VISSIM to simulate real-world traffic condition and predict the mobile source emissions using MOVES model (*21*). In Hao et al. 2015, the authors developed a statistical model to estimate the vehicle speed trajectory based on sparse mobile sensor data from the probe vehicle, and estimated the PM emissions by applying a microscopic emission model (*22*). Compared to the *measurement-based* approach, the *model-based* one can be applied to the in-/near-source PM emissions assessment at a much larger scale in a much more economical manner. However, the model accuracy and reliability for on-road traffic is still questionable, since most models were developed using dynamometer tests from standard drive cycles, which may not necessarily apply well to real-world driving due to the effects of road grades, driving behavior, fleet composition, and traffic conditions (*23*).

The objective of this study is 1) to explore the connection between the *measurement-based* approach for in-/near-source PM concentration assessment and *model-based* approach for on-road PM emissions assessment; and 2) to identify the key traffic-related factors and their impacts on in-/near-source PM concentrations. In this work, we built a mobile monitoring platform (on a probe vehicle) to collect on-road PM concentration data, and developed a comprehensive database to fuse information from various sources (including probe vehicle activity, traffic conditions, PM concentration measurement and PM emissions inventory) for modeling and analysis purpose. The rest of this paper is organized as follows: Section 2 describes the data collection effort in detail, followed by the presentation of methodologies for data processing and database construction in Section 3. Based on the database, statistical models are developed to predict on-road PM concentration with traffic information and the analysis results are presented in Section 4. The last section concludes this paper with further discussion on potential research topics.

2. DATA COLLECTION AND DESCRIPTION

2.1. Experiment Setup

As aforementioned, we set up a mobile monitoring platform on a testing vehicle to measure the in-/near-source PM concentration along highways. In this section, we will detail the experiment setup for data collection.

The testing vehicle was a 2011 NISSAN Infinity M37 (powered by gasoline only), equipped with the specifically designed data acquisition system for NISSAN vehicles – CONSULT III plus kit (including a PC, called Toughbook and a vehicle interface), and a Trimble R8 GPS receiver with RTK (real-time kinetic) positioning under the synchronized mode. The CONSULT III plus kit was capable of accessing high-resolution (every 0.01 second) on-board diagnostics (OBD) data from the testing vehicle, such as engine speed, vehicle speed and radar detection information (e.g., relative distance and relative speed with respect to the preceding vehicle along the same lane). The Trimble GPS receiver could report the vehicle's location (in terms of latitude, longitude and altitude) at the centimeter-accuracy level (*24*). A forward-facing camera was also installed on the front panel to capture the preceding traffic conditions (e.g., vehicle type, congestion level) for the verification purpose, which was also done in other studies (*25*).

The sampling port was facing toward the front of the vehicle through the front passenger window (see Figure 1). A CPC (TSI, 3022A) was used to measure particle number concentration with a cut-off diameter of 7 nm. An EAD (TSI, 3070A) was also equipped to measure particle surface area. Both CPC and EAD can measure down to per second time resolution, which is important to capture transient emissions during traffic congestion.

Besides, a NanoScan SMPS (TSI, NanoScan SMPS 3910) was used for particle size distribution measurement with a sizing range from 10 nm to 420 nm. CO_2 concentrations were measured by using PP Systems CIRAS-SC. All instruments in the trunk were powered by two deep cycle marine batteries (U.S. Battery, US 2200 XC2) with a 12 VDC to 120 VAC inverter. Figure 1 presents a detailed layout of the mobile monitoring platform.

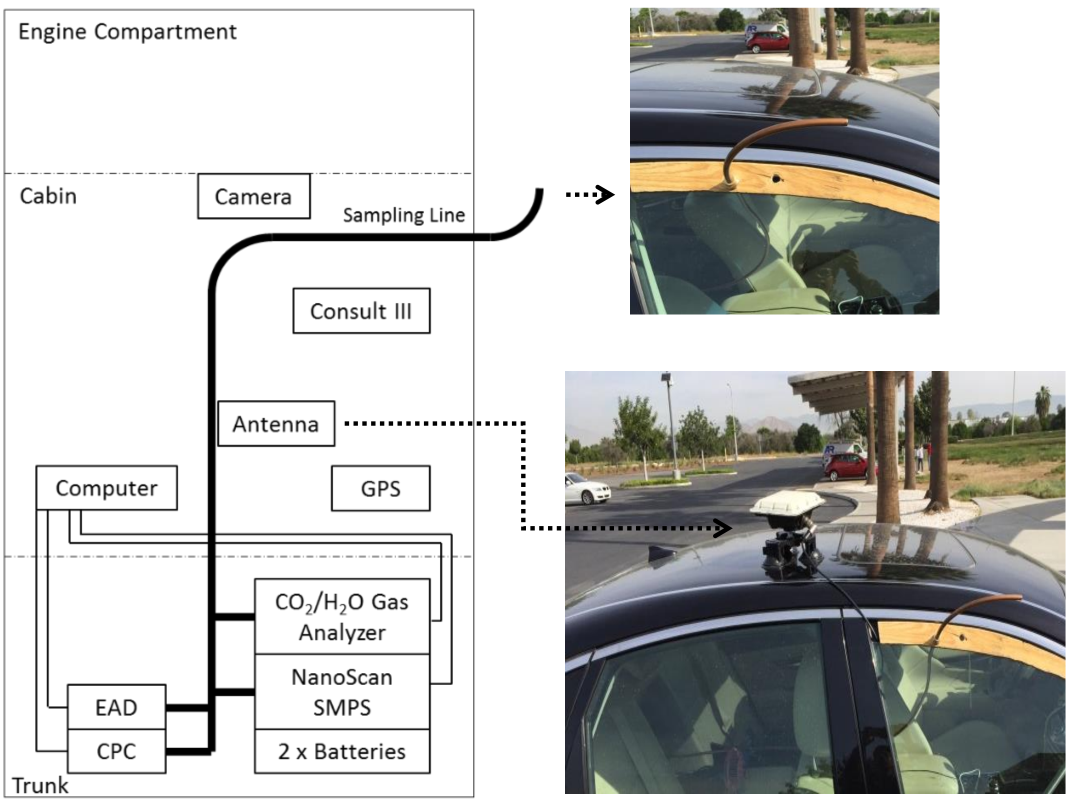


Figure 1. Layout of the mobile monitoring platform.

2.2. Study Routes and Dates

In this study, we chose two major routes in Southern California (see Figure 2) for testing:

1) *California State Route 91 (SR-91)*: a segment between California State Route 55 (SR-55) and California State Route (SR-60) which includes a major recurrent bottleneck (in Yorba Linda, CA). The typical traffic mix along this route is dominated by light duty vehicles (LDVs).
2) *Interstate 710 (I-710)*: a segment between SR-60 and the Port of Long Beach where heavy-duty trucks (HDTs) account for a significant portion of highway traffic. Since the majority

of these HDTs are diesel engine powered, it is expected to observe distinctive particle size distributions from diesel engines' emissions.

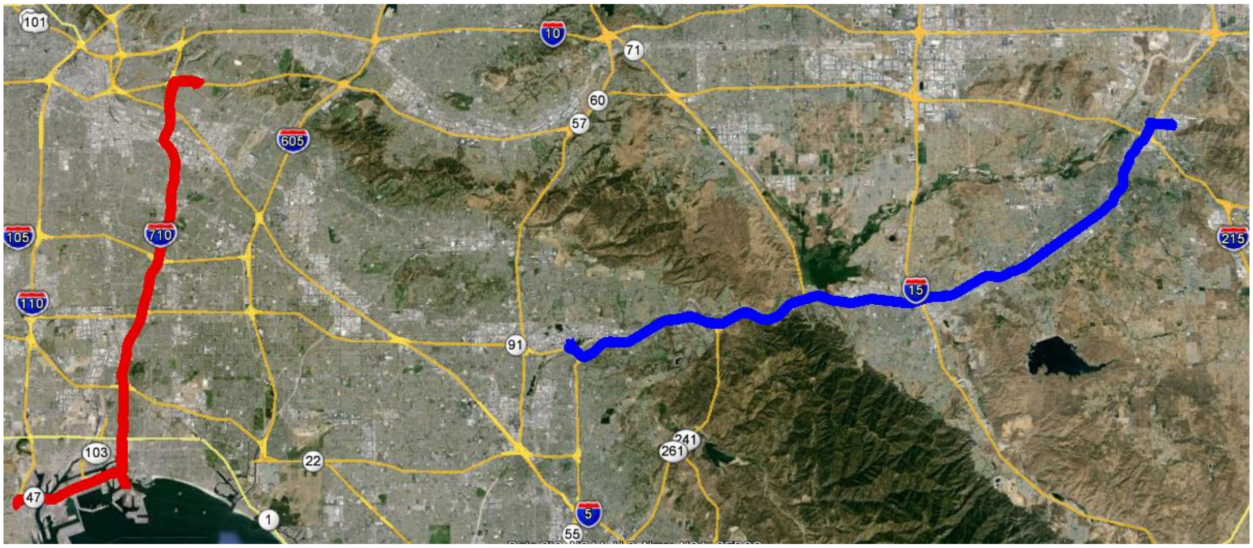


Figure 2. Illustration of study routes in Google Earth: SR-91 – blue; I-710 – red.

Table 1 summarizes the dates and time periods for field data collection. It is noteworthy that some of the testing periods were selected in order to capture the peak hours of traffic along the designated route and direction. In addition, the testing vehicle was consistently driven along the second left-most lane by following the typical speed of mainline traffic except upon exiting the highway.

Table 1. Summary of study routes and period

Date	Time	Route	Direction
03/26/2015	6 ~ 7 p.m.	SR-91	West
	7 ~ 8 p.m.*		East
03/27/2015	7 ~ 8 a.m.*		West
	8 ~ 9 a.m.		East
03/30/2015	11 a.m. ~ 12 p.m.	I-710	South
	11 a.m. ~ 12 p.m.		North

* indicate the peak hour for the specified route and direction

2.3. Other Data Sources

Traffic Data

Since the measurement is the PM concentration in/near the source (i.e., the mainline traffic flows along highways), it is critical to obtain the traffic conditions (of both directions) around the probe. In this study, the major traffic data source was the California Performance Measurement System or PeMS (*26*) which receives real-time 30-second raw measurements of traffic count and lane occupancy from each inductive loop detector throughout the California freeway system, detects the invalid or missing data, and rectifies them or fills the "holes". Based on the rectified traffic flow and lane occupancy data for each lane, aggregate traffic speed at each single loop detector can be estimated using the *g-factor* algorithm (*27*). PeMS also estimates the truck volume based on the algorithm proposed by Kwon et al. (*28*). In addition, all these raw data have been aggregated at various temporal levels, e.g., 5 minutes, for different purposes of analyses. It is noted that PeMS also archives some geographic information, such as the latitude and longitude of each vehicle detection station (VDS) and the associated post-mile. With such information, we can identify the closest VDS with respect to the testing vehicle's location at each time step. The data association effort will be detailed in the next section.

Meteorological Data

It is well understood that the meteorology conditions (e.g., wind direction and speed) may affect the PM data collection and results interpretation. Therefore, we also acquired the meteorological data from the California Air Resources Board (CARB) database (*29*). It turns out that during the testing period listed in Table 1, the observed magnitudes of wind from all the nearby stations were no more than 5 mph. Therefore, we ignore the wind effects in this study. But in a general situation, the concept of *apparent wind* which considers information (e.g., direction. velocity) from both *true wind* and traffic, should be used to account for the wind impacts.

3. DATA FUSION

As mentioned in the Section 2, there are multiple data sources. To facilitate our data analyses and statistical modeling of the relationship between traffic conditions and highway PM concentration, we fused all data sources, developed an integrated database and conducted more in-depth data processing/cleaning and. Nevertheless, before any data fusion or data processing effort, we conducted data cleaning by detecting and removing the outliers. Missing data were also imputed using the linear interpolation technique.

3.1. Time Synchronization

One of the key steps in data fusion is to synchronize the time of all data sources. For PM measurements, all the instruments were connected to a PC whose clock had been already synchronized with an internet time server right before the experiment. The Trimble GPS receiver had also synchronized itself to the highly accurate atomic clocks. However, time stamps in the output files from CONSULT III plus kit do not include the computer clock time. One possible way is to estimate the starting time based on the file modification time if the clock of Toughbook is synchronized. But in this study, we applied the cross-correlation technique to the vehicle speed data to synchronize the output from CONSULT III plus kit with that from the Trimble GPS receiver. Compared with the instantaneous speed estimated from the Trimble GPS receiver, the information reported by CONSULT III plus kit is more reliable because it is directly accessed through the CAN bus of the vehicle. Therefore, in the final database, we used the outputs from CONSULT III plus kit as the ground truth of vehicle dynamics.

3.2. Ambient Traffic State Association

In this study, we followed a 3-step procedure to associate the mobile platform data with PeMS database in order to identify the surrogate ambient traffic conditions (for both bounds):

Step 1 *Map-matching*. For each GPS data point (with latitude and longitude), we projected it to the specified study route and determined the location (or route distance) with respect to a referenced starting point, based on a list of survey nodes – "Postmile to Latitude & Longitude" in PeMS – that maps node (along the route) coordinates onto postmile highway location markers with a separation of 0.1 mile.

Step 2 *Vehicle Detection Station* (*VDS*) *association*. With the identified route location (in terms of postmile) for each GPS record, we searched for the PeMS "Station Metadata" database, found the nearest upstream and downstream VDSs along both directions and extracted the associated 5-min traffic data, including traffic flow, speed and truck flow.

Step 3 *Ambient traffic state estimation*. Based on the time *t* and location *x* of the probe vehicle, the ambient traffic state (e.g., speed) along the traveling direction was estimated by applying a 2-D interpolation technique (*29*) as illustrated in Figure 4:

$$v(x,t) = \frac{A_{11}\cdot v_{11} + A_{12}\cdot v_{12} + A_{21}\cdot v_{21} + A_{22}\cdot v_{22}}{A_{11} + A_{12} + A_{21} + A_{22}}$$

$$\forall x \in (x_1, x_2) \text{ and } t \in (t_1, t_2) \qquad (1)$$

where v_{ij} denotes the measurement from VDS located at x_i during the time interval between $[t_{j-1}, t_j]$. A_{ij} represents the "area" (or weight) in the time-space diagram (see Figure 3) associated with the measurement at (x_i, t_j) for the calculation purpose.

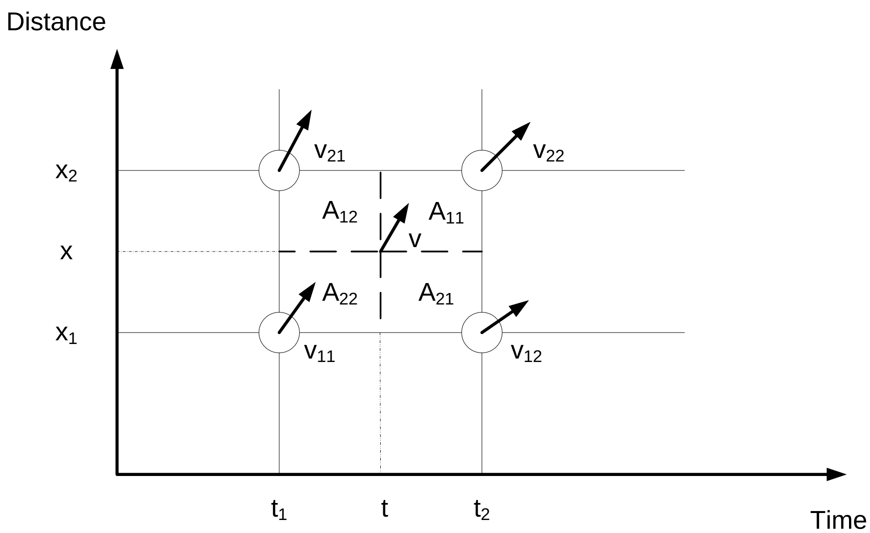


Figure 3. Illustration of ambient traffic state (e.g., speed) estimation using 2-D interpolation method.

3.3. Traffic Related PM Emissions Estimation

With the associated traffic data as described in Section 3.2, we used the Emission FACtors (EMFAC) 2014 model (*30*) developed by CARB to estimate the tailpipe PM emissions from traffic. Basically speaking, EMFAC is a speed bin based emission factor model with correction/adjustment customized for all motor vehicles operating in California. According to the vehicle category description by CARB, we chose gasoline "LDA" (i.e., passenger cars) of "aggregated model year" to represent general traffic, while diesel "T7 POLA" (i.e., Heavy-Heavy Duty Diesel Truck near South Coast) of "aggregated model year" for truck traffic. Since the model only provides emission factor at each discrete speed level (usually ranging from 5 mph to 70 mph with 5 mph interval), linear interpolation was applied to the emission factors of bracketed speed levels to estimate the emission factor of certain speed in between.

3.4. Integrated Database Development

To summarize the data fusion effort, Figure 4 illustrates the flow chart for integrated database development. In case more experiments will be conducted in the future and/or other pollutants are of interest, the database can be easily expanded by following the similar procedures as in this flow diagram.

The data sources include: 1) dynamics of testing vehicle and its preceding traffic (from CONSULT III, Trimble unit, Map database, and on-board radar); 2) in-/near-source PM concentration measurements (from CPC, EAD, NanoScan SMPS, and CO_2/H_2O Gas Analyzer); 3) traffic data (from PeMS); 4) meteorological data (from CARB website); 5) mobile source PM emissions (from EMFAC 2014). Table 2 provides more detailed description of information stored in the integrated database. With such database, we conducted our analysis as presented in the following section.

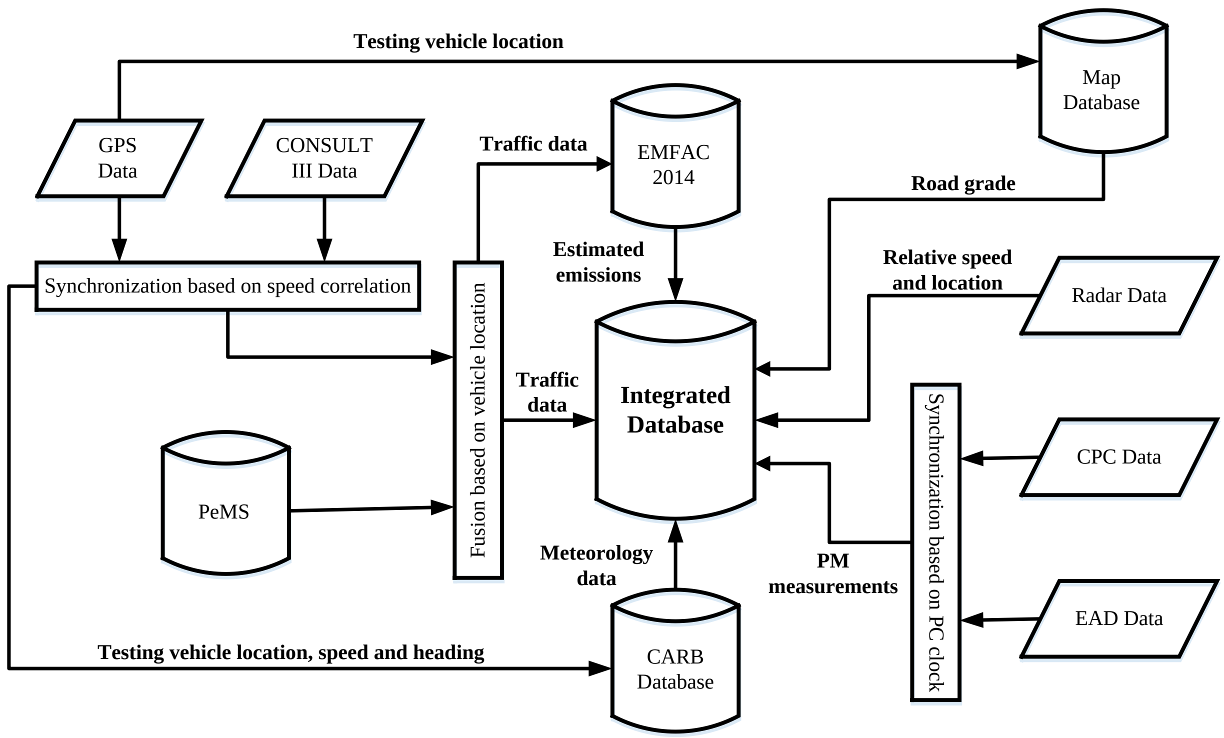


Figure 4. Flow Diagram for Integrated Database Development.

4. RESULTS AND ANALYSES

As can be expected, it is quite challenging to link the traffic-related factors with PM measurements (both concentration and size distribution). In the following, we explore this relationship from four perspectives: 1) to analyze basic statistics for PM measurements across different routes; 2) to correlate probe vehicle speed with PM measurements; 3) to investigate the estimated tailpipe PM emissions from EMFAC 2014; and 4) to develop a non-parametric model to address the influential factors for highway PM concentration.

Table 2. List of variables in the database

Category	***Data Entry***
General information	Time Stamp Test route name, direction
Probe vehicle information	Location (latitude, longitude, altitude and postmile with respect to the test route) Road grade Instantaneous speed and acceleration Preceding vehicle information (gap, relative speed)
VDS information	Features of downstream VDS on the same direction (postmile, number of lanes) Features of upstream VDS on the same direction (postmile, number of lanes) Features of downstream VDS on the other direction (postmile, number of lanes) Features of upstream VDS on the other direction (postmile, number of lanes)
Traffic information	Traffic and truck volume, speed (from downstream VDS on the same direction) Traffic and truck volume, speed (from upstream VDS on the same direction) Traffic and truck volume, speed (from downstream VDS on the other direction) Traffic and truck volume, speed (from upstream VDS on the other direction) Traffic and truck volume, speed (surrounding of the probe on the same direction)[a] Traffic and truck volume, speed (surrounding of the probe on the other direction)[a]
Pollutant information	PM measurements (concentration, size distribution) Tailpipe PM estimation (mass rate)

[a] *Estimated traffic conditions based on field measurements from the nearest upstream/downstream VDS*

4.1. Basic Statistics for PM Measurements

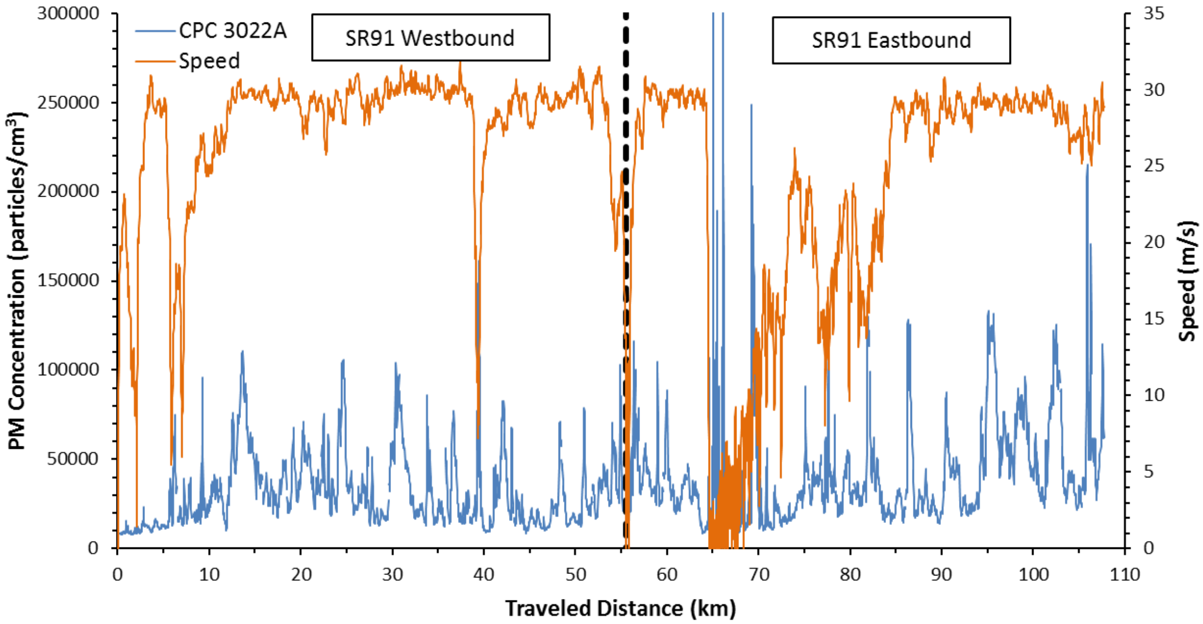
In terms of health effects, the PM measurements captured the exposure of both people driving on the highway and living in the nearby communities. For comparison purpose, we measured the ambient (outdoor) PM concentration before entering the study route on each testing day. The measurements were approximately 8100, 11000, and 12400 particles/cm^3, respectively, for SR-91 (on 03/26/15), SR-91 (on 03/27/15) and I-710 (on 03/30/15). These values were much (4 ~ 15 times) smaller than the average PM concentration (i.e., 31600, 55000, and 164000 particles/cm^3, respectively.) measured in the traffic flow (see Table 3). As can be observed from the table, PM concentrations have peaks that are greater than 350000 particles/cm^3 in some events. An interesting finding is that the PM concentration (both average and peak value) in I-710 outnumbers that in SR-91 (either day), which may be contributed by the much higher flow and density of heavy-duty trucks (diesel engine powered) along I-710.

Table 3. Basic statistics for PM Concentration along the study routes

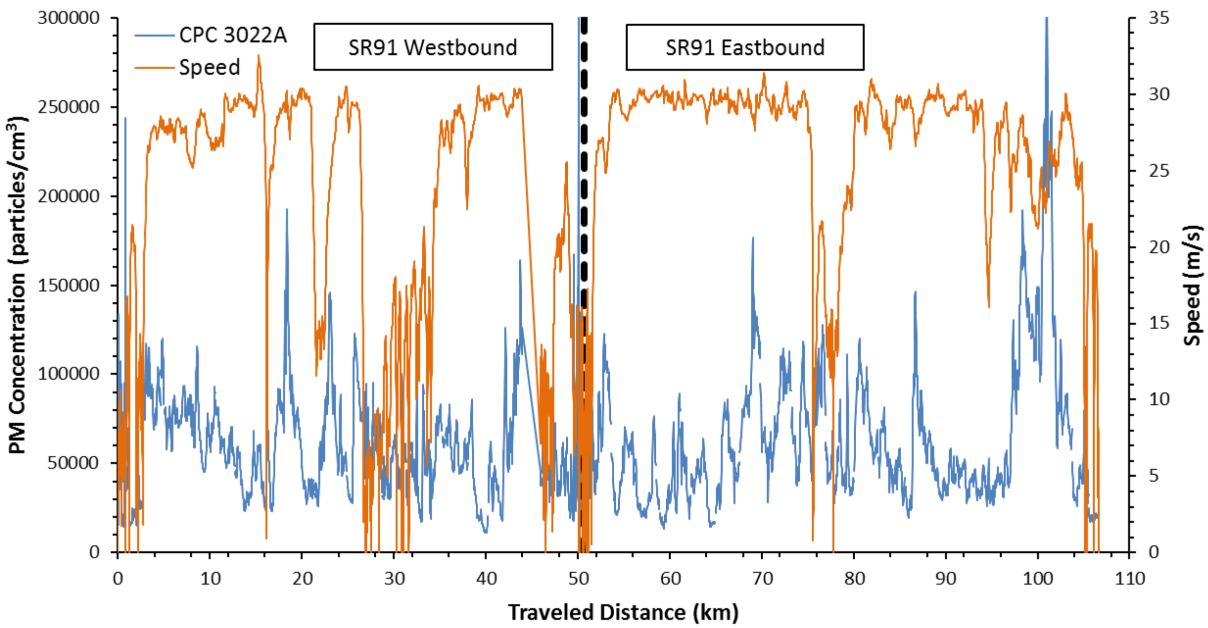
Route	**Particle Concentration (particles/cm^3)**		**Active Surface Area Concentration (mm/cm^3)**	
	Average	**Peak**	**Average**	**Peak**
SR-91 (03/26)	31600	392500	1.00	15.7
SR-91 (03/27)	55000	371500	2.24	33.5
I-710 (03/30)	164000	2065000	4.60	38.1

4.2. Probe Vehicle Speed vs. PM Measurements

In this subsection, we investigate the relationship between the probe vehicle speed and en-route PM measurements. By intuition, the probe vehicle speed may provide indication of the immediate surrounding traffic conditions. For example, if the speed is high, then a light traffic condition around may be inferred with certain confidence. It is noteworthy to mention that in the analysis of PM concentration, the peak pattern (representing the particle burst) is of our major interest. In the following, we use SR-91 test (peak periods indicated in Table 1) as an example.



(a) Results on 03/26/15



(b) Results on 03/27/15

Figure 5. Particle concentration versus speed measurements with respect to traveled distance along SR-91

Figure 5(a) presents the particle concentration for SR-91 test in the afternoon on 03/26/15 when the eastbound traffic experienced much more severe congestions (as indicated by the significant drop of the orange line in the figure), especially within the major recurrent bottleneck in Yorba Linda area (traveled distance between 64 and 70 km). In the same location, there was a prominent peak in PM concentration which may result from the heavy traffic congestion. Besides this major peak, there were many other concentration spikes that were correlated with a change in vehicle speed, which were caused by the engine loads during acceleration. The diameter concentration which is more related to the effective surface areas of particles, followed the similar pattern with particle (in number) concentration.

Results for SR-91 test in the morning (peak period for westbound traffic) on 03/27/15 are illustrated in Figure 5(b). Similar to Figure 5(a), there are also many concentration spikes that are correlated with the speed change. However, some major peaks (traveled distance at about 1 and 102 km) do not occur at the place where the speed drops significantly (e.g., traveled distance at around 30 km). Further investigation reveals that these interesting locations are exactly where two or more major freeways intersect. For example, at traveled distances of 1 and 102 km (symmetric due to round trip), is the interchange of SR-91, SR-60 and I-215 near UC Riverside. The interchange of SR-91 and I-15 is located at the traveled distance of 30 km. Therefore, it is very likely the correlation between the PM concentration and vehicle speed at the interchange was distorted by the traffic conditions on I-15. Another interesting finding is that there seems to be a mirror-symmetric pattern (to some degree) in concentration spike along the traveled distance. A hypothesis is that PM emissions from the traffic in other direction of the same highway segment would considerably affect the PM concentration measurements along the traveling direction.

4.3. Estimated PM Emissions Contour Plot vs. PM Concentration Measurements

In the integrated database, we applied the EMFAC 2014 model to the traffic data to estimate the "background" traffic-related PM (e.g., PM2.5) emissions from tailpipe. For the visualization purpose, we developed a so-called PM Emissions Contour plot in this study which may be used to understand the evolution of pollutant emissions from mobile sources over time and space domain. Figure 6 gives an example of such contour plots for PM2.5 emissions along SR-91E within the range of interest on March 26th, 2015, and the PM concentration measurements (width of the line is proportional to the value of PM concentration) by the mobile platform is overlaid as red line for reference. As can be observed from the figure, the estimated contour plot matches the peak of the measured PM concentration well. For example, this can be seen at postmile 40, 42, 43, 46, and 54. However, note that the contour plot only indicates for one direction rather than total emissions estimation from both directions. Therefore, the particle measurements may not be perfectly aligned with the color of the contour.

4.4. Statistical Analysis on Influential Factors for PM Concentration

Preliminary analysis indicates that the relationship between in-source PM concentration and traffic-related parameters is very complicated and could be highly nonlinear. In this study, we applied a nonparametric regression technique, called *Multivariate Adaptive Regression Splines* (MARS) model (*31*) to the integrated database for further exploring the influential factors for highway PM concentration. Although the statistical properties of the resulting estimators are more difficult to determine, non-parametric regression techniques require fewer assumptions and provide better fit than parametric techniques (*32*). In addition, the MARS model can be regarded as an extension of the linear models that automatically captures nonlinearities and interactions using the form

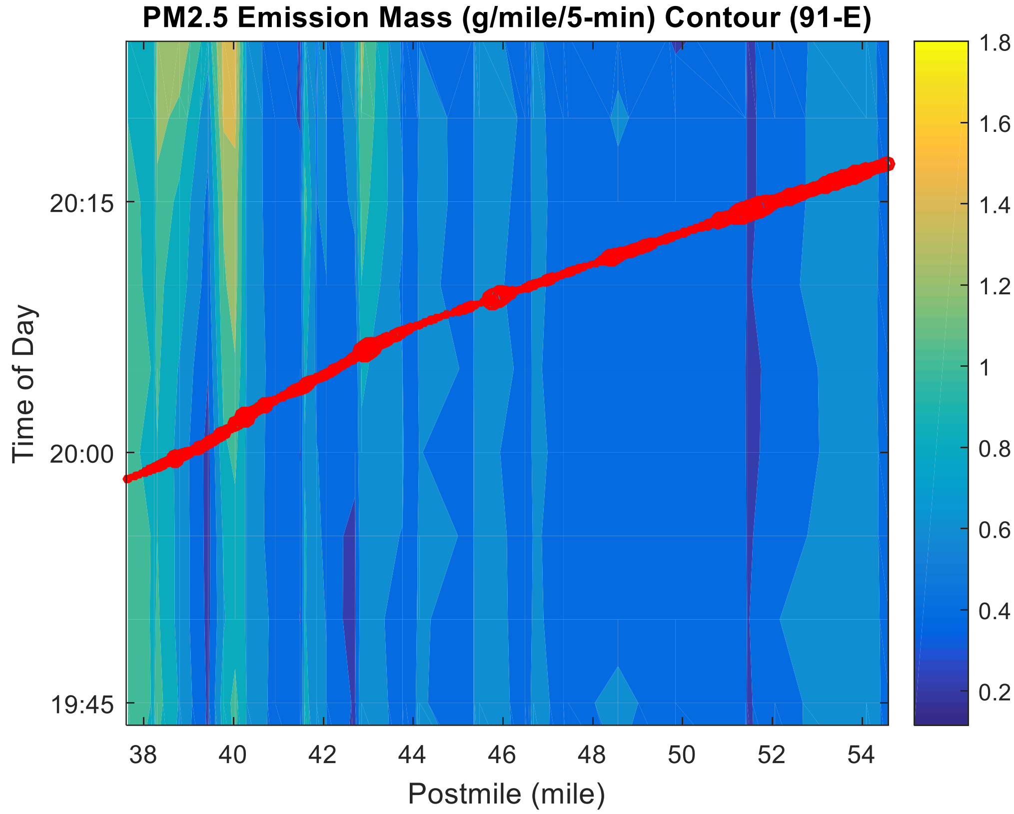


Figure 6. An example contour plot for estimated PM2.5 emissions (overlaid with PM concentration measurements) for SR-91E on March 26th, 2015

$$\hat{f}(x) = \sum_i c_i \cdot B_i(x) \tag{2}$$

where $\hat{f}(x)$ is the estimated model output; $B_i(x)$ is the *i*-th basis function which can be a constant 1, a hinge function, or a product of two or more hinge functions. The hinge functions can take the form

$$max(0, x - const.) \tag{3}$$

or,

$$max(0, const. - x) \tag{4}$$

and automatically partition the input data, so the effect of outliers can be attenuated. The MARS model can handle both numeric and categorical data and tends to have a good bias-variance tradeoff due to its flexible but sufficiently constrained form of basis functions to model nonlinearity with fairly low bias and fairly low variance.

In this work, the state-of-the-art statistical software, *R* (*33*), is applied to the entire database (i.e., all the "cleaned" data for SR-91 and I-710) whose sample size is 6848 and the number of predictor is 23 (including the intercept). The basis functions and associated coefficients (totally 32 terms) of the MARS model (see Eq. (2)) for PM concentration (count/cm^3) are listed in Table 4.

Table 4. List of Basis functions and the associated coefficients of for MARS model for PM Concentration

i	c_i	$B_i(\cdot)$	i	c_i	$B_i(\cdot)$	i	c_i	$B_i(\cdot)$
1	508477	Intercept	12	17584	$max(0, x_4 - 19.9)$	23	-30935	$max(0, 9.9 - x_8)$
2	4804	$max(0, x_1 - 25.7)$	13	615	$max(0, 70.1 - x_5)$	24	-261318	$max(0, x_8 - 9.9)$
3	-189	$max(0, 87.3 - x_2)$	14	153880	$max(0, x_5 - 70.1)$	25	269294	$max(0, x_8 - 11.9)$
4	-458	$max(0, x_2 - 87.3)$	15	-33	$max(0, 876 - x_6)$	26	560	$max(0, 67.4 - x_8)$
5	-72170	$max(0, x_3 - 3.9)$	16	25421	$max(0, x_6 - 876)$	27	-9814	$max(0, x_9 - 67.4)$
6	275269	$max(0, x_3 - 9.8)$	17	-190085	$max(0, x_6 - 881)$	28	-18749	$max(0, 19.9 - x_{10})$
7	2868073	$max(0, x_3 - 11.2)$	18	-511226	$max(0, x_7 - 53.3)$	29	-13347	$max(0, x_{10} - 19.9)$
8	-28511	$max(0, 11.5 - x_3)$	19	681	$max(0, 53.7 - x_7)$	30	3350	$max(0, x_{11} - 33.5)$
9	-4131297	$max(0, x_3 - 11.5)$	20	678568	$max(0, x_7 - 53.7)$	31	-7088	$max(0, x_{11} - 52.6)$
10	1044194	$max(0, x_3 - 12)$	21	-157765	$max(0, x_7 - 55)$	32	4076	$max(0, x_{11} - 61.8)$
11	21989	$max(0, 19.9 - x_4)$	22	-11336	$max(0, x_7 - 60.8)$			

where

x_1 = probe vehicle's speed, m/s;
x_2 = gap with preceding vehicle, m;
x_3 = probe vehicle's route location, mile;
x_4 = upstream VDS (same direction as probe vehicle) route location, mile;
x_5 = traffic speed measured by upstream VDS (same direction as probe vehicle), mph;
x_6 = traffic volume measured by downstream VDS (same direction as probe vehicle), veh/5-min;
x_7 = traffic speed measured by downstream VDS (same direction as probe vehicle), mph;
x_8 = upstream VDS (other direction as probe vehicle) route location, mile;
x_9 = traffic speed measured by upstream VDS (other direction as probe vehicle), mph;
x_{10} = downstream VDS (other direction as probe vehicle) route location, mile; and
x_{11} = traffic speed measured by downstream VDS (other direction as probe vehicle), mph;

According to Table 4, the variables of importance (i.e., the ones used in the MARS models) are x_1 through x_{11} as listed above, and the values in those basis functions represents the associated "knots" for different predictors which are critical to the range partitioning for a certain set of numerical explanatory variables. For example, 25.7 (m/s) is a critical partitioning point for the probe vehicle's speed (a surrogate of surrounding traffic speed as mentioned in Section 4.2) Please note that the traffic speeds of both downstream and upstream (with respect to the location of probe vehicle) for both directions play a statistically significant role in estimating/predicting highway PM concentration (count/cm^3). The R^2 values for the MARS model is 0.72, which is satisfactory considering the highly complicated process to model from traffic condition to highway PM concentration.

5. CONCLUSION AND FUTURE WORK

In this study, we took non-trivial effort to set up the mobile monitoring platform for real-time PM concentration measurement and traffic data collection along some major highways in Southern California. We developed an integrated database by fusing a variety of data sources. Based on the archived data, we investigated the relationship between traffic conditions and highway PM concentration. We proposed an innovative tool, so-called PM Emissions Contour plot which can provide more in-depth insight for assessing in-source PM emissions (e.g., on highways). In addition, we applied MARS model to the integrated database to evaluate the impacts of traffic-related parameters on PM concentration. Some major findings include:

- The measured PM concentration along highways is significantly higher than the ambient outdoor measurement, which indicates people driving on the highway and living in the nearby communities have much higher exposure to PM.
- Many prominent peaks in PM concentration and surface area measurements were due to the heavy traffic congestion and the change in vehicle speed from vehicle acceleration and deceleration.
- The estimated PM Emissions Contour plot shows good agreement with PM number concentration.
- In spite of the highly complex physical nature of the emission source on highways, the MARS model can provide a satisfactory prediction results where R^2 value is as high as 0.72.

As one of the major future work, the integrated database will be improved by: 1) feeding more experiment data, 2) incorporating other potential data sources (e.g., NanoScan SMPS data for particle mass estimation); 3) enhancing the surrogate ambient traffic condition estimation; and 4) adding an appropriate dispersion model.

ACKNOWLEDGMENTS

This research is jointly supported by the National Center for Sustainable Transportation (NCST) and National Science Foundation. We would like to thank TSI for lending us NanoScan SMPS for the test. We are also grateful to Mike Todd and Daniel Sandez for their kind support in setting up the mobile monitoring platform.

The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of or policy of the sponsor. This paper does not constitute a standard, specification or regulation.

REFERENCES

1. California Air Resources Board (CARB). The California Almanac of Emissions and Air Quality – 2013 Edition. http://www.arb.ca.gov/aqd/almanac/almanac.htm. Accessed on July 6th, 2016
2. U.S. Environmental Protection Agency. *National Emissions Inventory (NEI) Air Pollutant Emissions Trends Data*. http://www.epa.gov/ttn/chief/trends/index.html. Accessed on July 6th, 2016
3. Health Effects Institute (HEI). *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*. January 2010
4. Atkinson, R.; Mills, I.; Walton, H.; Anderson, H. R. Fine particle components and health-a systematic review and meta-analysis of epidemiological time series studies of daily mortality and hospital admission. J. of Exposure Science and Env. Epidemiology. **2014**, 1-7

5. Li, N.; Hao,M.Q.; Phalen, R. F.; Hinds,W.C.; Nel, A. E. Particulate air pollutants and asthmasA paradigm for the role of oxidative stress in PM-induced adverse health effects. Clin. Immunol. **2003**, 109, 250-265.
6. McCormack, M. C.; Breysse, P. N.; Matsui, E. C.; Hansel, N. N.; Peng, R. D.; Curtin-Brosnan, J.; Williams, D. L.; Wills-Karp, M.; Diette, G. B. Indoor particulate matter increases asthma morbidity in children with non-atopic and atopic asthma. Ann Allergy Asthma Immunol. **2011**;106:308-315.
7. California Air Resources Board (CARB). *Estimate of Premature Deaths Associated with Fine Particle Pollution (PM2.5) in California Using a U.S. Environmental Protection Agency Methodology*. August 2010
8. Dockery, D. W., C. A. Pope, X. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. Ferris, Jr. and F. E. Speizer. *An Association between Air Pollution and Mortality in Six US Cities.* New England Journal of Medication,1993, 329: 1753 – 1759
9. World Health Organization (WHO). *Health Effects of Particulate Matter*. 2013, Final Report, 20 pages
10. Health Effects Institute (HEI). *Understanding the Health Effects of Ambient Ultrafine Particles*. January 2013
11. Shah, A. P. et al.. *Effect of Inhaled Carbon Ultrafine Particles on Reactive Hyperemia in Healthy Human Subjects*. Environmental Health Perspectives, 2008, 116: 375 – 380
12. U.S. Environmental Protection Agency. *National Ambient Air Quality Standards* (*NAAQS*): http://www.epa.gov/air/criteria.html. Accessed on July 20th, 2015
13. U. S. Department of Transportation. Summary of Travel Trends – 2009 National Household Travel Survey. Final Report FHWA-PL-ll-022, June 2011, 83 pages
14. Zhu, Y., A. Eiguren-Fernandez, W. C. Hinds, and A. H. Miguel. *In-cabin commuter exposure to ultrafine particles on Los Angeles freeways*. Environmental Science Technology, 2007, 41: 2138 – 2145
15. Fruin, S., D. Westerdahl, T. Sax, C. Sioutas, P. M. Fine. *Measurements and Predictors of On-Road Ultrafine Particle Concentrations and Associated Pollutants in Los Angeles*. Atmospheric Environment, 2008, 42: 207 – 219
16. Pallavi Pant, Roy M. Harrison, (2013) "Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review", Atmospheric Environment, 77, pp. 78 – 97
17. Reynolds, A.W., B. M. Broderick. *Development of an emissions inventory model for mobile sources.* Transportation Research Part D, 5 (2000) 77-101
18. Environmental Protection Agency, *MOVES 2014a User Guide*, 201, EPA-420-B-15-095
19. California Air Resources Board. *EMFAC2014 User's Guide*. 2014, v1.0.7
20. Hausberger, S., M. Rexeis, M. Zallinger, R. Luz. *Emission Factors from the Model PHEM for the HBEFA Version 3*. Graz University of Technology. Institute for Internal Combustion Engines and Thermodynamic (2009) Report Nr. I-20a/2009 Haus-Em 33a/08/679
21. Abou-Senna, H., Essam Radwan, Kurt Westerlund, C. David Cooper. *Using a traffic simulation model (VISSIM) with an emissions model (MOVES) to predict emissions from vehicles on a limited-access highway.* Journal of the Air & Waste Management Association, 63(7):819-831, 2013
22. P. Hao, G. Wu, K. Boriboonsomsin, and M. Barth. *Modal Activity-Based Vehicle Energy/Emissions Estimation Using Sparse Mobile Sensor Data*. Transportation Research Board Annual Meeting, Washington D.C., January 10-14, 2016Jamriska, L., L. Morawska. *A*

model for determination of motor vehicle emission factors from on-road measurements with a focus on submicrometer particles. The Science of the Total Environment, 264 (2001, 241-255
23. Trimble. *Trimble® R7/R8 GPS Receiver User Guide*. Version 1.00, September 2003
24. Westerdahl, D., S. Fruin, T. Sax, F. Philip, C. Sioutas. *Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles*. Atmospheric Environment, 2005, 39: 3597-3610.
25. California Performance Measurement System (PeMS). Accessed on July 10th, 2016 http://pems.dot.ca.gov/
26. Jia, Z., C. Chen, B. Coifman, P. Varaiya. *The PeMS Algorithms for Accurate, Real-time Estimates of g-factors and Speeds from Single-loop Detectors*. Proceedings of IEEE Intelligent Transportation Systems Conference, Oakland, CA, August, 2001, pp. 536 – 541
27. Kwon, J., P. Varaiya, A. Skabardonis. *Estimation of Truck Traffic Volume from Single-loop Detectors with Lane-to-lane Speed Correlation*. Transportation Research Record, No. 1856, 2003, pp. 106 – 117
28. California Air Resource Board (CARB). *Meteorology data query tool*. Updated in July 2016
29. Wu, G., K. Boriboonsomsin, M. Barth. *Development and Evaluation of an Intelligent Energy-Management Strategy for Plug-in Hybrid Electric Vehicles*. IEEE Transactions on Intelligent Transportation Systems, 15(3), pp. 1091 – 1100
30. California Air Resource Board (CARB). *EMFAC2014 Web Database*. http://www.arb.ca.gov/emfac/2014/ Accessed on July 8th, 2016
31. Friedman, J. H. (1991) Multivariate adaptive regression splines. The Annals of Statistics, Vol. 19, No. 1, pp. 1 – 141
32. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984) Classification and regression trees. Chapman & Hall, USA
33. R Development Core Team (2016) R: a language and environment for statistical computing. Available from: <http://www.R-project.org accessed November, 2016>