COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING

Computer-Aided Civil and Infrastructure Engineering **29** (2014) 546–558

INDUSTRIAL APPLICATION

Real-Time Estimation of Pollution Emissions and Dispersion from Highway Traffic

Samitha Samaranayake* & Steven Glaser

Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

David Holstius

Department of Environmental Health Sciences, University of California, Berkeley, CA, USA

Julien Monteil

French Authority for the Quality of Transportation Services (AQST) Ministry of Ecology, Sustainable Development and Energy, Paris, France

Ken Tracton

Palo Alto, CA, USA

Edmund Seto

Department of Environmental and Occupational Health Sciences, University of Washington, WA, USA

&

Alexandre Bayen

Department of Electrical Engineering and Computer Science and Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

Abstract: *Traffic-related air pollution is a serious problem with significant health impacts in both urban and suburban environments. Despite an increased realization of the negative impacts of air pollution, assessing individuals' exposure to traffic-related air pollution remains a challenge. Obtaining high-resolution estimates are difficult due to the spatial and temporal variability of emissions, the dependence on local atmospheric conditions, and the lack of monitoring infrastructure. This presents a significant hurdle to identifying pollution concentration hot spots and understanding the emission sources responsible for these hot spots, which in turn makes it difficult to reduce the uncertainty of health risk estimates for communities and to develop policies that mitigate these risks. We present a novel air pollution estimation method that models the highway traffic state, highway traffic-induced air pollution emissions, and pollution dispersion, and describe a prototype implementation for the San Francisco Bay Area. Our model is based on the availability of real-time traffic estimates on highways, which we obtain using a traffic dynamics model and an estimation algorithm that augments real-time data from both fixed sensors and probe vehicles. These traffic*

*To whom correspondence should be addressed. E-mail: *samitha@alum.mit.edu*.

DOI: 10.1111/mice.12078

estimates combined with local weather conditions are used as inputs to an emission model that estimates pollutant levels for multiple gases and particulates in real-time. Finally, a dispersion model is used to assess the spread of these pollutants away from the highway source. Maps generated using the output of the dispersion model allow users to easily analyze the evolution of individual pollutants over time, and provides transportation engineers and public health officials with valuable information that can be used to minimize health risks.

1 INTRODUCTION

Air pollution from automotive sources is typically the single largest source of regional air pollution in urban areas. Air pollutant concentrations adjacent to and downwind of major traffic routes are significantly higher than regional background levels (Zhu et al., 2002), which endangers near-by populations and disproportionately exposes them to traffic-related air pollution. Among the potential negative health impacts associated with living close to traffic sources are premature mortality, exacerbation of preexisting respiratory health conditions such as asthma, and poor cardiovascular health (Health Effects Institute, 2010). Impacted communities would benefit from accurate and timely localized knowledge of air pollution levels for immediate protection (asthmatics would stay indoors when air quality is poor), as well as for longer term mitigation of high pollution areas and traffic planning. Such estimates would be an invaluable tool for tackling environmental justice issues, for example the degree to which certain communities bear more of the burden of air pollution. There is a strong need, specifically, to estimate air pollutant concentrations near major roadways in urban areas, where land resources are scarce, and most planning for new land use development and redevelopment is at sites near major roadways. The focus of our current work is to support the need for planning tools to better assess near-roadway pollution levels for both land use and transportation planning agencies.

At present, fine spatial and temporal resolution air pollution concentration data do not exist at a neighborhood scale. Regulatory mandated fixed-site monitoring is spatially sparse since it is aimed at characterizing regional air pollution concentrations rather than identifying local hot spots. Moreover, regional air monitoring captures the pollution from both mobile and stationary sources, which makes it difficult for transportation planners to quantify the roadway attributable fraction. Thus, fine resolution and roadway-specific air pollution data would serve as a powerful resource for understanding the traffic-related pollution exposure levels and strategies to deal with them.

Despite the sparsity of air quality measurements, air pollution can still be estimated using appropriate models. As vehicle emissions are a major source of pollution in urban environments, emission models that use traffic state estimates can provide valuable information (Aziz and Ukkusuri, 2012). Models of roadway emissions and their dispersion are important tools that can be used both to study the impacts of vehicle emissions, and as an input to more sophisticated air pollution models that account for other sources of pollutants. Static maps of air pollutant concentrations can be easily produced by using estimates relying on average traffic and weather conditions, but these maps are rough estimates which do not account for the temporal variance in the pollution levels. Existing research in this area uses either fixed origin–destination matrices (Gualtieri and Tartaglia, 1998), traffic assignment and land use models (ENV4-CT96-0201, 1998; Karppinen et al., 2000), or simulations (Namdeo et al., 2002; Zegeye et al., 2011) for generating traffic state data. However, real-time estimates of traffic and weather conditions are required for an accurate dynamic environmental monitoring and modeling system.

Advances in traffic estimation and sensing technology make it possible to generate real-time pollution estimates that are accurate and rich enough for such a system. Macroscopic traffic models (Lightwill and Whitham, 1955; Work et al., 2010) based on the physical dynamics of traffic combined with statistical data assimilation techniques can provide reliable traffic estimates in a computationally tractable manner. However, these algorithms still require large amounts of traffic data, either in the form of vehicle counts that are usually measured using magnetic loop detectors or radars, or instantaneous speed estimates obtained from GPS tracking of vehicles. The increasing deployment of smartphones, which now provide mobile monitoring capabilities (GPS, accelerometers) and wireless connectivity (GPRS, Wi-Fi, bluetooth), enables such large-scale collection of vehicle speed data.

This article presents the theory, models, and tools needed to implement a real-time highway pollution estimation system and describes a prototype implementation within the Mobile Millennium (Herrera et al., 2010) traffic information system at the University of California, Berkeley. This multi-layer system collects real-time traffic, weather, and air quality information from multiple live feeds and aggregates the information in a database so that it can be efficiently accessed by the models and estimation algorithms. Data feeds and maps produced from the output of the system allow users to query and monitor pollution metrics in a number of different scales and dimensions.

The goal of this work is to create a framework that can provide real-time traffic-induced emissions estimates at a neighborhood scale in a computationally tractable manner, using components that have been technically validated and accepted by practitioners. Our work extends the state of the art in terms of traffic-related emissions estimation and dispersion systems, by presenting a methodology for obtaining real-time emissions estimates and dispersion maps using a readily available data source and no additional infrastructure deployments, providing a valuable tool for practitioners in the areas of transportation planning, public health, and urban planning.

The rest of the article is organized as follows. In Section 2, we present an overview of the system used for this work, that is the inputs, models, and outputs of our system. In Section 3, we present our traffic model and estimation algorithms, which enable real-time estimation of traffic conditions. The emissions model we use is presented in Section 4. The outputs of the emissions model are then used as inputs to the dispersion model that is presented in Section 5. In Section 6, we describe the data delivery service, which consists of visualization and analysis tools used to assess the health impacts of emission and dispersion. An example of the emissions and dispersion output generated by the model for the San Francisco Bay Area and potential applications of these results are given in Section 7. Finally, Section 8 provides some conclusions and directions for future work.

2 SYSTEM OVERVIEW

2.1 Project background

The effort presented in this article is an extension of the Mobile Millennium project, a partnership between Nokia, NAVTEQ, the U.S. and California Departments of Transportation (CalTrans) and the University of California, Berkeley. Mobile Millennium is a large-scale system that provides a number of services such as real-time vehicle speed and flow estimates, travel time information, route choice advice, incident detection, and traffic forecasting in the San Francisco Bay Area. This information is mapped in real-time through a visualization interface available on the web. As exposure to air pollution is an increasing concern, especially in urban areas, real-time pollution estimation is a natural extension to the project.

The goal of this work is to model the air pollution contribution of highway traffic within an urban environment in real-time at a high granularity. The estimation models create a database of location specific concentrations of select pollutants that can then be queried by various air quality management and transportation agencies to incorporate into their systems, and which can also be visualized in the form of a "heat map" (i.e., a color map of pollution levels). A prototype implementation of the system is demonstrated for the San Francisco Bay Area, but the system can be extended to any geographical area where traffic data, emissions factors, and local weather data are available. This article presents the modeling process and system design for estimating real-time highway emissions and dispersion levels.

The Mobile Millennium system consists of a collection of data feeds, databases, traffic models, estimation and filtering algorithms, and visualization tools that enable the real-time conversion of raw traffic data to system-wide traffic estimates. We first describe the traffic system architecture before explaining the air pollution emissions and dispersion extensions that were implemented as part of this project.

The highway traffic estimation module combines traffic data with a physical model of traffic dynamics to generate traffic estimates in real time. The system uses fixed sensor data from the *Caltrans Performance Measurement System* (PeMS), probe data from GPS-enabled mobile phones and taxi fleet data from 500 equipped vehicles in San Francisco, which provide real-time measurements of traffic conditions on the road network. PeMS data originate from inductive loop detectors that are stationed at fixed highway locations. Mobile probe data are obtained through a partnership with Nokia using *virtual trip lines* (VTL) (Hoh et al., 2008). VTLs are geographic markers that indicate locations where GPS data can be sent from smartphones in a privacy preserving manner. Nokia retrieves VTLs from cell towers through a transmission and cloaking process and sends this data back to the Berkeley data server system.

Data that are obtained from the feeds are stored in our system and refined as needed for downstream modeling tasks. For example, traffic data feeds have to be filtered for faulty sensors before they are used as inputs to the traffic estimation model. The traffic models and estimation algorithms process these data and output the traffic state for each road segment. These results are then shared with partners and other users through a web interface. In summary, the data server at Berkeley provides database services to other machines and computers through a global database management system.

2.2 Pollution emissions and dispersion system

The roadway emissions module is a new addition to the Mobile Millennium system that enables the real-time estimation of vehicle emissions. The emissions estimation is based on the traffic state outputs from the highway traffic estimation module and the Wunderground weather feed provided by Weather Underground, which provides weather conditions at discrete locations in the Bay Area. The highway traffic estimation module processes the filtered traffic data feeds and generates a traffic speed map (averaged over 5 minute periods) for the entire Bay Area highway system. A historical traffic patterns feed from NAVTEQ provides historical traffic speed data by road segment in 15 minute intervals for each day of the week. These historical data are used when the speed estimate of the highway model is not reliable due to the unavailability of data or errors in the data that are detected by the individual data feed filters. Details are provided in Section 4.4.

The traffic estimates are combined with temperature and relative humidity data from Weather Underground and emissions factors for different gas and particulate pollutants obtained from the *California Air Resources Board* (CARB) (California Air Resources Board, 2006) to compute emissions levels for each road segment. These emissions levels are calculated by multiplying the estimated vehicle flows from the highway model with the estimated per vehicle emission strengths for each road segment given the weather conditions and vehicles speeds. The emissions values are then stored in a database as outputs of the roadway emission system. A detailed description of the emissions model will be provided in Section 4.1. Emissions levels are visualized in the form of an emission color map on the highway links.

The dispersion model uses the outputs of the road emissions module and the wind strength and direction information from the weather feed to estimate the concentration of a specified pollutant for a dense set of receptor coordinates. The dispersion computation requires significant computing power and therefore is run on a dedicated server.

The concentration values computed by the dispersion system only include pollutants that are attributable to the road traffic. Optionally, regional background concentrations can be added to the roadway-attributable concentrations. These background concentrations are available for limited regions as a feed from the CARB. In areas that are not covered by CARB, concentrations can be better informed using a probe sensing module that can be installed in a set of fleet vehicles such as garbage trucks or postal vehicles that cover a large area of the geography. We are currently in the process of equipping such a set of fleet vehicles with pollution sensors to enable and test this idea. This is the subject of future work.

The final emissions and dispersion values are stored in a database and then pushed to our partners, who can then analyze the data to suit their planning questions as well as potentially use their own web interfaces to display the results in ways that best communicate to their agency's audience the degree of air pollution risk.

3 TRAFFIC ESTIMATION

This section describes the estimation model used to produce traffic velocity and density estimates that are required by the emissions model. We use a velocity evolution model (Work et al., 2010) augmented with a data assimilation algorithm that is implemented in the Mobile Millennium system and provides robust and reliable velocity estimates for the entire California Bay Area. Traffic densities can be approximated using the velocity information and road segment specific calibrated macroscopic parameters.

3.1 Velocity evolution model

In traffic theory, the evolution of vehicles along a roadway is typically modeled using the mass conservation law prescribed by the *Lighthill-Whitham-Richards* (LWR) *partial differential equation* (PDE) (Lightwill and Whitham, 1955). This is a macroscopic traffic model that models the evolution of vehicle density for a roadway section of length L over a time T using the following relation:

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial Q(\rho(x,t))}{\partial x} = 0 \quad (x,t) \in (0,L) \times (0,T) \tag{1}$$

where x and t are the position and the time, ρ is the vehicle density, and Q is the vehicle flow. This provides a density evolution model for traffic. A major drawback of this model is that it is unable to incorporate velocity data obtained from probe vehicles, which would result in ignoring a large percentage of the available traffic data and thus degrade the traffic state estimation. Therefore, we use a corresponding velocity evolution model to accurately estimate the traffic velocity and incorporate probe-based speed measurements into the estimation model. The Mobile Millennium system implements the velocity evolution model described in (Work et al., 2010), which we use use for our velocity and flow estimation purposes. A very brief high-level description is provided in the rest of this section. Please see Work et al. (2010) for a detailed description of the system.

The velocity V is related to the flow and density through the following relation:

$$Q(\rho(x,t)) = \rho(x,t)V(\rho(x,t)) \tag{2}$$

Equation (2) is inverted to obtain an evolution equation for velocity and to obtain the density as a function of velocity. The relationship between density ρ and velocity V is given by the so-called fundamental diagram (Daganzo, 1995). The standard Daganzo-Newell triangular fundamental diagram is however noninvertible. Therefore, it is approximated using a hyperbolic-linear velocity function to obtain the following:

$$\rho = \begin{cases} \rho_{\max}\left(1-\dfrac{v}{v_{\max}}\right) & \text{if} \quad v \geq v_c \\ \rho_{\max}\left(\dfrac{1}{1+\dfrac{v}{\omega_f}}\right) & \text{otherwise} \end{cases} \tag{3}$$

where $\rho_{\max}$ is the jam density, ρ_c the critical density, ω_f is the shockwave speed, $v_{\max}$ is the free-flow speed, and v_c is the critical velocity defined as $v_c = V(\rho_c)$.

To model the evolution of traffic in time and space, we use the discretized Godunov scheme (Godunov, 1959) which computes an approximate solution to the LWR-PDE (Equation (1)). For the discretization, we introduce a time step ΔT indexed by n and a space step Δx indexed by i. This discretization needs to satisfy the Courant-Friedrichs-Lewy (CFL) condition $\alpha_{\max}\frac{\Delta T}{\Delta x}\leq 1$, where $\alpha_{\max}$ is the maximum characteristic vehicle speed, to ensure the numerical stability of the solution (Leveque, 1992).

3.2 State estimation via the ensemble Kalman filter

Velocity measurements, obtained from filtered PeMS data, probe vehicles (VTL), and taxi trajectories, are integrated into the flow velocity evolution model using a data assimilation process. A data assimilation process is required because the spatiotemporal granularity and quality of the measurements is not sufficient for obtaining the real-time traffic state of the entire network. Due to the nonlinear nature of the velocity evolution model, exact methods such as the Kalman filter can not be used for the state estimation. We use an inverse modeling method known as the ensemble Kalman filter (Evensen, 2007), which is preferable to the alternatives such as the extended Kalman filter due to its lower computation cost and ability to handle the nondifferentiable velocity flux equation (Work et al., 2010).

3.3 Model output

The resulting speed estimates are used as inputs to the roadway emission model. We define an output time step $j\cdot\Delta T$ (with $j\in\mathbb{N}$, this time step is therefore a multiple of the previously defined time step) for velocity. We choose to keep the same index scale for the rest of the article so that the time index is now written N, with $N=j\cdot n$. A corresponding confidence value $C_{N,e}$ is computed by setting it equal to 1 if there has been a measurement in the neighboring edges in the previous time step and setting it equal to 0 otherwise. The highway model provides an average speed output V_e^N and a confidence value $C_{N,e}$ on each highway edge at each output time step index N. See Section 4.4 for details on the use of $C_{N,e}$.

4 ROADWAY EMISSIONS ESTIMATION

4.1 Emission model

The emission model maps traffic outputs into emissions strengths for each roadway segment. Based on discussions with the local Bay Area Air Quality Management District, we chose to implement an emissions model that has been accepted by their land use planning processes to maximize the practical use of our system. We first generate an emission inventory by using the EMFAC2007 (EMissions FACtors) model (California Air Resources Board, 2006) developed by CARB as an onroad motor vehicle emissions inventory for transportation and land use planning. The emissions factors are based on empirical data from the U.S. Environmental Protection Agency's (U.S. EPA) federal testing of emissions for different classes of vehicles under different operating conditions, tailored specifically to California's light duty vehicle standards, and vehicle fleet composition data from the California Department of Motor Vehicles' vehicle registration system. This model provides emissions factors E_{type} in (grams/vehicle mile) for various tail pipe air pollutants (e.g., CO, CO_2, particulate matter, NO_x, SO_2) as a function of year and location within California, and atmospheric temperature and humidity. In the case of our system, the emissions factors obtained from the emission model are average emissions for the San Francisco Area vehicle fleet in the year 2010. The model considers all vehicle model years from 1965 onward, and it assumes that vehicles from before 1965 have the same emissions rates as 1965 models. This emissions factors inventory is stored in our systems database, so that we can query it in real time.

Based on the stored emissions factors for the 2010 vehicle fleet, the emissions factor E_{type} for a given type of pollutant as a function of vehicle speed V, location temperature T, and humidity H for the roadway location consists of a database query:

$$E_{\text{type}} = f_{\text{type}}(V, T, H) \tag{4}$$

The effects of acceleration/deceleration are not incorporated in the current emissions model due to the difficulty of measuring these values in practice. The EMFAC model has this capability though, and the emissions model can be extended in the future if such information is available. By multiplying the emissions factor by the average flow Q on a stretch of highway, we get an emissions strength per roadway length, E in (grams/mile second). At a time step index N and on a highway edge e, the estimated flow value is Q_e^N. For a given pollutant, the emissions strength per roadway length is then computed as follows:

$$E_e^N = Q_e^N \cdot E_{\text{type}} \quad (5)$$

For notational simplicity, we drop the *type* subscript from the emissions strength and consider a single type of pollutant in the rest of the discussion. Multiplying the emissions strength by the specific roadway length $E_e^N \cdot L_e$, results in the mass per time (g/s) emissions strengths for each roadway segment, which is the commonly encountered metric in roadway air pollution modeling. These roadway segment-specific emissions strengths are also used as inputs to the dispersion model, as described in Section 5.3.

4.2 Estimating the truck penetration rate

It is important to note that the vehicle emissions estimate computed in this manner assumes that the truck penetration rate on each roadway corresponds to the 2010 average truck penetration rate. This can be an issue when certain road segments have a high concentration of large vehicles such as trucks that have higher emissions values. To address this issue, we are in the process of adding a module to the system that estimates the percentage of trucks on each road segment in real time and adjusts the emissions estimate accordingly. The truck percentage is estimated using vehicle occupancy and flow information obtained from the PeMS and Mobile Millennium traffic systems. Preliminary results show that the truck percentage can be significant on specific highways at certain times of the day. Accounting for the percentage of truck can significantly improve the accuracy of emissions estimates in these cases.

4.3 Weather data

The weather data needed for emissions estimation, the temperature (T), and humidity (H), are retrieved from the Wunderground weather feed via an XML stream. Wunderground is a weather feed provided by Weather Underground, a weather service that publishes real-time weather data on the Internet. Most Weather Underground data come from the National Weather Service, and part of it from registered personal weather stations. Weather data are collected at nine locations in the Bay Area and include temperature, humidity, wind direction, and wind intensity measurements. After matching the road segments of the highway network to the nearest weather location, we assign to each segment of the network the most recent available weather data. Temperature and humidity values are hence defined for a given time step index N and a given edge e as T_e^N and H_e^N.

4.4 Emissions estimation algorithm

The first step in estimating real-time emissions values is to obtain the corresponding velocity outputs from the highway model at the given time step N for road segment e. If the confidence value ($C_{N,e}$) of the velocity output is equal to 0, we will look for valid estimates with $C_{N,e} = 1$ in the following order: (1) velocity values at this location in previous time steps, (2) velocity values close to this location at approximately the same time window in recent weeks, and (3) historical velocity patterns. A confidence value of 0 corresponds to road segments where there is insufficient VTL, taxi, and PeMS data. The localized variance of the estimate given by the ensemble Kalman filter for these locations can be quite high and therefore the estimates can not be trusted. We have observed that on average 70% of the network has a valid velocity estimate, but expect this number to increase as we incorporate more VTL and other probe data into the system.

Next, we match the weather data with the road segments, and compute the emissions factor values for each pollutant (4). The vehicle flow is obtained by using the relationship in Equation (2) that gives flow as a function of density and velocity, and Equation (3) that gives density as a function of velocity. The relationship in Equation (3) can be highly sensitive to its inputs when the road segment is in free flow. Therefore, we use direct flow estimates from the PeMS system in locations were flow data are available. We are also currently in the process of implementing a direct density estimation model using an approach similar to the work of Gomes et al. (2001) to improve the accuracy of the flow estimate. Since the highway traffic estimation model provides velocity estimates at 30 second intervals we choose to compute the flow and emissions values for an aggregated period. Note that the traffic model provides a single velocity estimate for all the vehicles in a given road segment at a given time. We use a default aggregation that gives emissions estimates in 5 minute intervals.

The following algorithm describes how the real-time emissions estimates are computed:

1. *Compute emissions values*: For each edge e and time step N,
   - If $C_{N,e} = 1$:
     Obtain the velocity V_e^N and compute density ρ_e^N, flow Q_e^N, and emissions values E_e^N according to (2) and to the following relations:

$$Q_e^N = \rho_e^N \cdot V_e^N \quad (6)$$

$$E_e^N = Q_e^N \cdot E_{\text{type}} = Q_e^N \cdot f_{\text{type}}(V_e^N, T_e^N, H_e^N) \quad (7)$$

- If $C_{N,e} = 0$:
  - Iterate through the N_{pre} previous velocity outputs $V_e^{N'}$, with $N' \in [N - N_{\text{pre}}, \cdots, N - 1]$. If $C_{N',e} = 1$ for any $N' \in [N - N_{\text{pre}}, \cdots, N - 1]$, let $V_e^N = V_e^{N'}$ for the first N' such that $C_{N',e} = 1$ and update the flow and emissions values.
  - If not, look for historical data $V_e^{N''}$ from previous weeks at the same time period. We look at the last 4 weeks of data and look at N_{pre} time steps around the time of interest. If $C_{N'',e} = 1$ for any of these $V_e^{N''}$, then update the flow and emissions values using the velocity $V_e^{N''}$ from the most recent week and closest time period such that $C_{N'',e} = 1$.
  - If not, use the historical average speed for the the road segment based on daily temporal patterns and update flow and emissions values.

2. *Aggregate*: Write the final flow and emissions values into the database at each aggregation time step. A default value of $k = 10$ with a 30 second time discretization results in 5 minute aggregates.

$$Q_e^N = \frac{1}{k} \sum_{j=N-k}^{N} Q_e^j \tag{8}$$

$$E_e^N = \frac{1}{k} \sum_{j=N-k}^{N} E_e^j \tag{9}$$

5 DISPERSION MODEL

5.1 Model selection

Numerous air quality models exist within routine air quality management and planning practice as well as within academic research for near-roadway applications as reviewed by the U.S. EPA (Pierce et al., 2008), which include box, computational fluid dynamic, Gaussian plume, geographic information system-based, puff, and statistical models. Other reviews include those conducted by the following (Holmes and Morawska, 2006; Sharma and Rao, 2004; Benson, 1992; Vardoulakis et al., 2003). Considering these reviews as background, our selection of air quality model was governed by these criteria:

1. Model is appropriate for estimating near-roadway micro-scale concentrations.
2. Model has been previously validated.
3. Model can run in real time (update hourly) on a single server for the entire Bay Area with a dense receptor grid.
4. Model is accepted by the Bay Area Air Quality Management District (BAAQMD) for land use and transportation planning.

Based on the criteria, we selected the CALINE3 model (Benson, 1979), a finite line source Gaussian dispersion model. Designed specifically as a screening tool for transportation planning applications, it takes roadway edge emissions strengths as inputs and estimates roadway attributable concentrations as outputs, making it a natural fit to the goals for our system. The CALINE model is recommended for use in limited scenarios, such as those in which the U.S. EPA has authorized its use, including for near highway studies of nonreactive pollutants, and under assumptions of horizontally homogeneous nonzero wind flow and stable meteorology, and relatively simple terrain to avoid wind channeling that can affect wind speeds and directions (Benson, 1992). However, in complex terrain conditions, finite line source models may be useful for hot-spot analyses of small areas (e.g., intersections), where the wind and meteorological conditions of the modeling domain are somewhat homogeneous (Benson, 1986). As such, we have employed CALINE in our system to only model highways and to estimate near-roadway concentrations.

CALINE has been validated in previous studies in various settings, originally with tracer studies (Cadle et al., 1977; Benson, 1984b) and freeway studies (Claggett and Miller, 1979; Rodes and Holland, 1979; Benson, 1984a), with general findings indicating CALINE's ability to model spatial patterns of concentrations that decay with downwind distance, and that generally predicted concentrations fall within a factor of two of measured concentrations (Benson, 1992). Because of the relatively good measured versus predicted spatial correlation, and because most errors are overpredictions (Benson, 1992), the model has been accepted for screening analyses for transportation planning purposes. More recent evaluations of the CALINE model, with data collected in Northern California have had similar general findings and consistency with Benson's original recommended use scenarios (Yura et al., 1984). While a detailed air pollution model validation study is outside of the scope of this current article, recent data provided by the Bay Area Air Quality Management District for near-roadway air pollutant concentrations suggest similar near-roadway downwind concentration decay patterns as those summarized by other researchers (Karner and Niemeier, 2010), which suggest the applicability of CALINE models to our Bay Area-based system, with the aforementioned understanding of CALINE model use recommendations and validation findings.

Moreover, CALINE3 models are familiar to air quality modelers within state and regional agencies: BAAQMD, CARB, and Caltrans. In fact, the CALINE family of models is an established U.S. EPA regulatory model (Pierce et al., 2008). Although the model has its limitations, including poor performance in low wind speed conditions, lack of consideration of complex terrain, short distance effects, and applicability only to inert gases and particulate matter, these limitations are well understood by the above agencies that routinely use the model. The one major limitation to the CALINE model is that it was originally designed for project-specific analyses, and changes to the model were needed to make it sufficiently computationally efficient to be able to model the entire Bay Area domain. Next, we describe the model, and the changes we made to integrate it into our system and improve its performance for our application.

5.2 The CALINE3 model

Under the CALINE3 model, emissions attributable to a stretch of highway can be inferred at downwind locations using a multilayered approximation to a steady-state Gaussian diffusion model. The approximations and fundamental algorithm are detailed in (Benson, 1992). In this model, each direction of the highway is represented as a set of planar rectangles ("segments"), and each 2-D rectangle is replaced by a limited number of 1-D line sources oriented perpendicular to the prevailing wind (finite line sources). It is assumed that emissions are uniform for the segment, which in our system are assigned as averages over a predefined aggregation time (Equation 9)). Concentrations are then estimated over a regularly spaced grid of receptor locations, chosen at a resolution likely to reveal interesting patterns in spatial variability (e.g., at 100 m resolution). For each receptor location the coordinate system is transformed so that the new (x, y) axis is aligned to and centered on the wind vector (x-axis) crossing the midpoint of the finite line source (y-axis). Following (Benson, 1992) the concentration at a given (x, y) receptor location is computed for each line source of length dy according to the Gaussian diffusion equation:

$$C_{x,y} = \frac{q_e}{\pi \sigma_z u} \int_{y_1}^{y_2} \exp\left(\frac{-y^2}{2\sigma_y^2}\right) dy \tag{10}$$

where $C_{x,y}$ is the concentration of the pollutant at (x, y), u is the wind speed, and q_e is the linear source strength (emissions in grams per unit time) for a given highway link e. The parameters σ_y and σ_z account for horizontal and vertical dispersion and are dependent upon the standard deviation of the wind angle, receptor coordinate along the x-axis, and an assumption of the presence of an intense mixing zone directly above the roadway that characterizes vertical residence times. Adjustments are then made for different road types (bridge, fill, etc.), relative height and number of lanes, as well as roughness of the intermediate terrain and prevailing atmospheric stability.

5.3 Implementation details

In order to model the entire Bay Area efficiently in real time (new update every hour) our system implements a parallelized version of the CALINE3 model that was written specifically to integrate inputs from the highway traffic model, meteorological data, and emissions models described earlier. Without parallelization it is not possible to compute concentrations at the finely spaced receptor locations (described below) for all the primary roadway links in our modeling domain on a single CPU computer within one hour. We use a single server with 12 Intel X5650 2.66GHz cores and 16GB of RAM. As a steady-state model, CALINE3 is well suited to parallelization. The model can be evaluated independently at each receptor location and for a set of hourly meteorological conditions. Thus, we parallelized the model runs across receptors and highway links, allowing it to scale to a large amount of real-time inputs, while simultaneously computing estimates at high resolution. With this parallelized system, it was possible to compute concentrations at all receptor locations within 45 minutes, leaving ample time for remaining data storage, generation of data feeds, and visualization tasks.

At each hour, we perform the following steps: first, we compute the emissions strengths for each roadway edge, as described before $q_e = E_{Ne} \cdot L_e$, in units of (g/s), scaling it to a 1-hour averaging time. Windspeed u is obtained from the meteorology feed, using the first observation during the hour. Based on the requirements of the CALINE3 model, wind speeds are set to a minimum of 1 m/s. Roadway geometry parameters are obtained directly from the Navteq roadway data.

Second, we define a receptor grid of x, y coordinates based on varying buffer distances from the roadway geometry. The buffering approach is helpful so as to not place receptors on top of the roadways. Buffers closer to the road have receptors spaced at a finer resolution than those buffers that are farther away, which allows for more computation time to be spent close to the roadways where concentrations are likely to vary more than farther from the roadway. The buffering and grid spacings were determined through trial and error, so as to roughly approximate a uniform 100 m resolution grid when spatially interpolated. In the future, working more closely with land use and transportation agencies, we may dynamically define the receptors to fit their specific planning questions and areas of concern.

Third, we divide the CALINE3 runs for each of the individual receptor cells across all the available cores of the server. Jobs are assigned evenly across all the cores. On each core, nested loops of the CALINE3 model are run to estimate the air pollution concentration from each of the Bay Area road segments to each of the receptor locations within the region of influence for that road segment. As each run of the model finishes, its results are saved within a system database table. This table is used by the data delivery service described next.

6 DATA DELIVERY SERVICE

The emissions and dispersion data are made available to Nokia and other partners through a real-time data feed. Archived historical data are also made available as needed.

6.1 Emissions estimates

Emissions values are computed over 5 minute intervals for the entire network and available both via the emissions data feed and an emissions visualizer in real time. The visualization tool is implemented as a web application. Figure 1 shows an example output of the visualizer. This application is an extension of the mapping engine used to display traffic state estimates in the Mobile Millennium system. A graphical interface allows users to select different geographic regions and the pollutant of interest. User queries that originate from the visualizer are sent to a data server that relays the necessary data back to the client. The query includes information such as the name of the layer, bounds of the map view, current zoom level, and time of the day. The server then processes the request and returns a vector layer object which contains all the features (geometry and attribute) for that request. Finally, the server sends back the layer object to the client, which translates it and displays it on the map. Detailed attributes of the vector objects can be accessed by clicking on the map.

6.2 Dispersion estimates

Dispersion estimates are currently only computed on an hourly basis over the aggregate values of flow, wind speed, and wind direction per highway link. This is to balance the computational needs of our preliminary implementation with the physical meaning of the time resolution. The current model also does not estimate ambient concentrations of pollutants, so as a first step, it is possible to approximate these values by interpolating air quality measurements from the CARB feed for the Bay Area. This is a very crude estimate given the sparse nature of these sensors, but we hope to significantly improve these estimates as we add more sensors to our system. In reality, each coordinate of the spatial grid has a concentration level that depends on the dispersion at the current time step and decaying levels of background concentrations from the previous runs of the dispersion model.

In general, partnering agencies will likely want to analyze data from our system using their own data analytics, and produce visualizations that best communicate the results they are interested in to their respective audiences. Thus, our dispersion visualization system is limited to producing relatively simple offline visualizations. We have used the R language plotting commands to produce spatially interpolated heat-maps of the dispersion model results. An example of this is provided in the next section. While these visualizations could be provided as a web map service in the future, our current focus is on modeling and systems requirements for providing accurate data feeds.

7 EXPERIMENTAL RESULTS AND APPLICATIONS

7.1 Experimental setup

In this section, we present some preliminary results from the prototype system we have built. The emissions estimation system is fully operational and computes real-time emissions estimates for the entire Bay Area. Figure 1 shows the real-time emissions estimates as presented by the Mobile Millennium visualizer. The colored highway links are an indicator function of the quantity of emissions. The web interface allows users to select and observe information on emissions data for CO_2, CO, and particulate matter on each road segment and the information from regional air pollution monitors provided by the CARB feed.

7.2 Numerical results

As an example of data produced by our system, we generated one hour heat maps for the entire Bay Area, Figure 2, and the San Jose area, Figure 3, on August 9, 2010 during the afternoon peak hour of 5:00–6:00 PM. The receptor grid (not shown for clarity) was spatially interpolated with inverse distance weighting. Background regional concentrations are not added into these maps. In both figures, Q is th normalized CO concentration

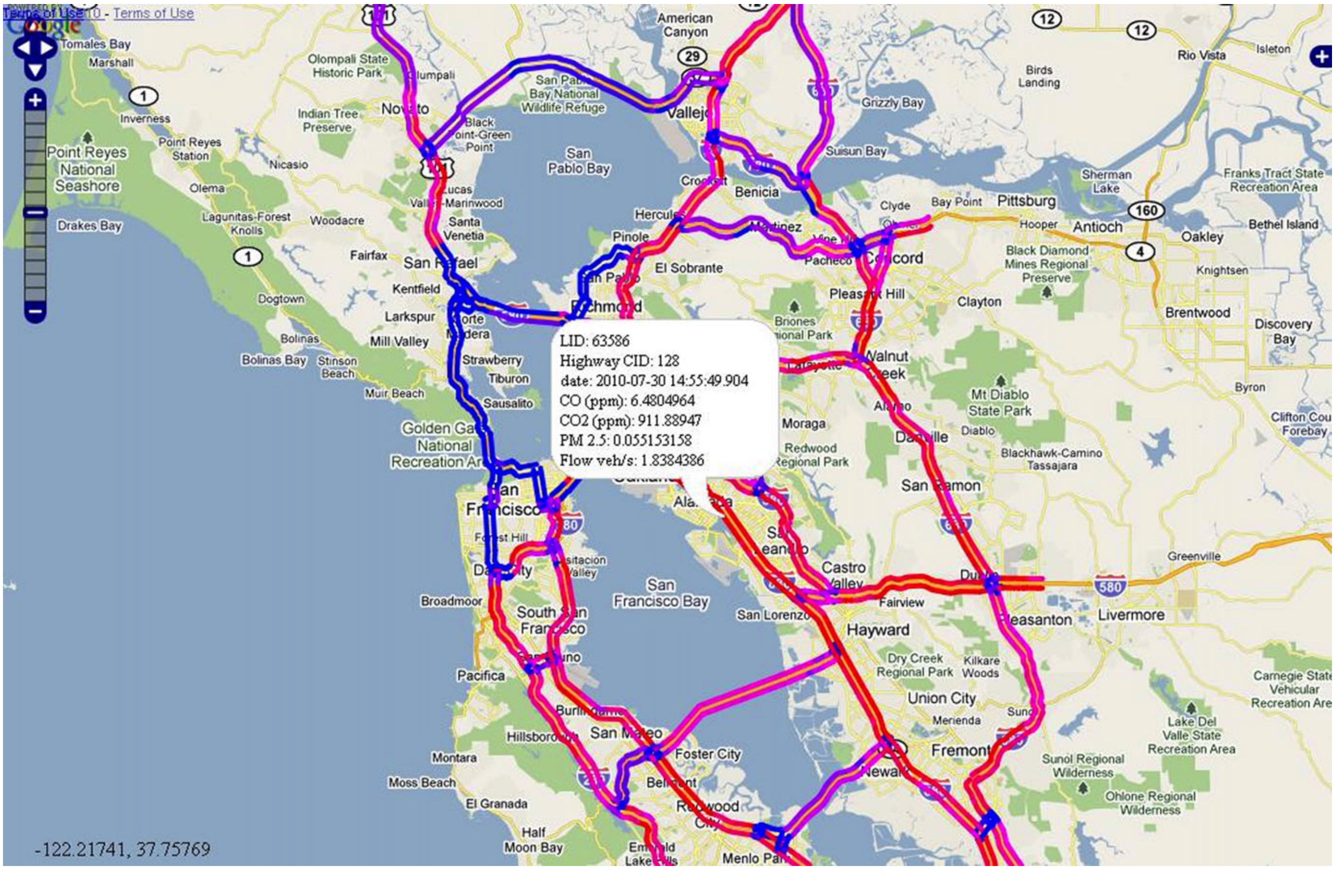

Fig. 1. Mobile Millennium emissions visualizer. The interface allows users to select a geographic region and a pollutant type of interest. The roads are colored based on the pollution emissions level on each segment. Detailed statistics about each segment can be viewed by clicking on the road segment.

(across the range of values present in the given map), which helps to see the full range of variation in concentrations across the regions. Also, presenting normalized concentrations to the [0,1] scale has the added benefit of not misleading the viewer that these are equivalent to pollutant concentrations that would be measured in communities; they are not equivalent because these maps do not include the regional background concentration and represent estimates of roadway-attributable relative concentrations. This information is of particular importance for transportation and land use planning. The maps illustrate the higher concentrations near and downwind of the major roadways (roadways shown in black), which is generally consistent with studies of near-roadway pollution concentrations that indicate elevated concentrations primarily in the downwind direction over a 300–500 ft distance (California Air Resources Board, 2005). Also note that both maps illustrate relatively high concentrations in specific locations, which supports the idea that our system may have use in understanding environmental justice issues and mitigating pollution impacts in areas that may show relatively high concentrations over time. The hot spots clearly reflect nuanced relationships between emissions, wind patterns, and downwind concentrations produced by our system that are not captured by current one-size-fits-all health protective guidance being promoted by CARB.

7.3 Applications

While this project is still in its early stages and the systems infrastructure is being improved over time, we are already in the process of exploring possible user studies with local, regional, and state agencies to better understand the needs and design choices for a unified visualization tool beyond the heat maps shown in the previous section. Maps can be powerful tools for identifying and communicating the extent of environmental justice problems, but selecting the right attributes and scale are an important part of the analysis. For example, as illustrated in the previous section, different map scales may

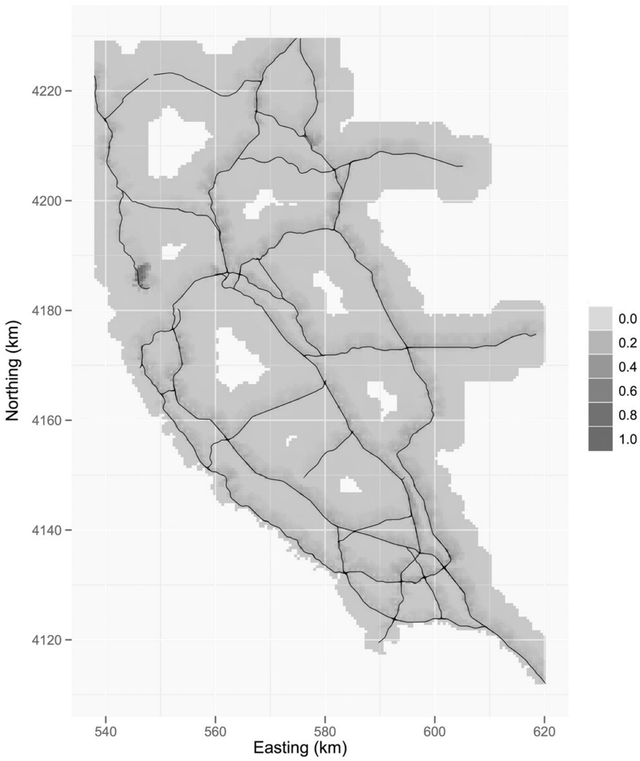

Fig. 2. Dispersion heat map of carbon monoxide (CO) levels for the San Francisco Bay Area with the intensity indicating the relative CO level. Note the high concentration levels at the south end of the Golden Gate Bridge due to heavy congestion at the toll booths.

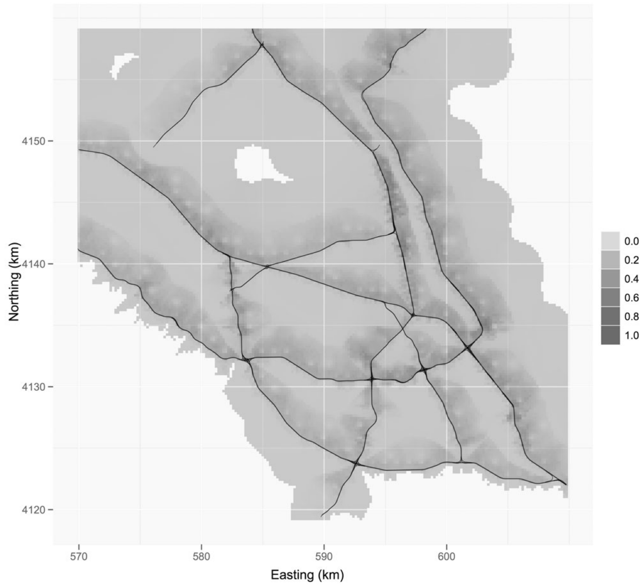

Fig. 3. Dispersion heat maps of carbon monoxide (CO) levels for Greater San Jose with the intensity indicating the relative CO level. The heat map indicates which communities are more prone to higher pollutant levels based on wind directions.

be helpful in identifying hotspots that naturally exist at different spatial scales. Moreover, to the degree that our tools can document ongoing changes in traffic pollution-related environmental justice issues, they have the potential to be greatly utilized since both California state and federal policies on transportation infrastructure require the consideration of environmental justice issues.

Another potential application of the system, which we have already presented to the California Department of Transportation, is the possibility of a hotspot detection tool for transportation planning. The hotspot tool would more clearly highlight the roadway links that contribute to the highest exposures in near-roadway populations. Such a tool might help identify co-benefits to transportation management projects that would not only improve traffic flow in these problem areas, but in doing so, could also reduce air pollutant emissions and their associated negative health impacts.

Finally, within California, with the passage of Senate Bill SB375 to create more sustainable communities there will be a need for tools that can track changes in regional transportation that occur throughout time, as many Bay Area communities adapt new forms of land use (e.g., mixed use, higher densities, and/or transportation-oriented developments) as well as promote alternative modes of transport that do not involve private car use. There is a need to track the progress of these changes and to identify areas of improvement and of need for improvement.

8 CONCLUSIONS AND FUTURE WORK

We have presented the theory, models, and tools needed to implement a real-time roadway emissions estimation and dispersion system, and described a prototype implementation for the San Francisco Bay Area. We have made several simplifying assumptions in the interest of first demonstrating the feasibility and applications of such a system. The current system which estimates the contribution of highway traffic to air pollution will enable transportation engineers and public health officials to better analyze the health impact of highway traffic. Moreover, the system can begin to serve as a platform for further integration of other sources of air pollution information (e.g., stationary sources and/or off-road mobile sources such as air, rail, and off-shore vessels) and more sophisticated models of air pollution dispersion (e.g., those that better account for regional atmospheric chemistry). Additionally, future deployment of mobile air quality sensors carried by smartphones and fleet vehicles will enable the validation of these types of models and allow for data assimilation algorithms that can improve the accuracy of model-based estimates using measurement data. We are also currently working on building more accurate flow models and on increasing our computing power to run higher resolution dispersion models. As traffic data improves to allow richer models of arterial and local traffic, these may also be included in our system. Although making these changes and additions will require careful consideration of the limitations and applicability of different models and their data requirements, we have designed our system to be modular to better accommodate these future changes.

There are numerous potential applications of our system, which still need to be evaluated in practice. As stated in the beginning of this article, our system provides opportunities for a much more detailed understanding of local-scale air pollution which could be used by air quality agencies and health officials to better inform communities of health risks. While providing health risk information is important, perhaps even more importantly this system provides a tool for those in the fields of transportation engineering and land use planning to better identify and manage the health impacts of their decisions. We encourage these agencies to work with us to evaluate our system. At the same time, we are in the process of extending our system to incorporate nation-wide data, which will create a tremendous opportunity to understand the relationships between traffic patterns and air pollution outside the San Francisco Bay Area.

ACKNOWLEDGMENTS

The authors would like to thank Daniel Edwards, Pierre-Emmanuel Mazaré, Jack Reilly, Joe Butler, and the systems team at the California Center for Innovative Transportation for their contributions.

REFERENCES

Aziz, H. M. A. & Ukkusuri, S. V. (2012), Integration of environmental objectives in a system optimal dynamic traffic assignment model, *Computer-Aided Civil and Infrastructure Engineering*, **27**, 494–511.

Benson, P. (1979), CALINE-3. A versatile dispersion model for predicting air pollution levels near highways and arterial streets, FHWA/CA/TL-79/23, California Department of Transportation, Sacramento, CA.

Benson, P. (1984a), CALINE3 line source dispersion model, FHWA-CA-TL-84-15, California Department of Transportation, Sacramento, CA.

Benson, P. (1984b), CALINE4—a dispersion model for predicting air pollutant concentrations near roadways, FHWA-CA-TL-84-15, California Department of Transportation, Sacramento, CA.

Benson, P. (1986), Evaluation of the CALINE4 line source dispersion model for complex terrain application, *Transportation Research Record*, **1058**, 7–13.

Benson, P. (1992), A review of the development and application of the CALINE3 and CALINE4 models, *Atmospheric Environment*, **26B**(3), 379–90.

Cadle, S. H., Chock, D. P., Monson, P. R. & Heuss, J. M. (1977), General Motors Sulfate Dispersion Experiment: experimental procedures and results, *Journal of Air Pollution Control Association*, **27**, 33–8.

California Air Resources Board (2005), *Air Quality and Land Use Handbook: A Community Health Perspective*. Available at: http://www.arb.ca.gov/ch/landuse.htm.

California Air Resources Board (2006), EMFAC 2007 User's Guide. Available at: http://www.arb.ca.gov/msei/onroad/downloads/docs/user_guide_emfac2007.pdf.

Claggett, M. & Miller, T. (1979), *Carbon Monoxide Monitoring and Line Source Model Evaluation Study for an Urban Freeway and Urban Intersection*, Final Report, Illinois Environmental Protection Agency, Springfield, IL.

Daganzo, C. (1995), The cell transmission model, part II: network traffic, *Transportation Research Part B*, **29**(2), 79–93

ENV4-CT96-0201 (1998), System for planning and research in towns and cities for urban sustainability (SPARTACUS), EC Environment and Climate Research Programme (1994–1998): Research Theme 4, EC DGXII, Brussels.

Evensen, G. (2007), *Data Assimilation: The Ensemble Kalman Filter*, Springer Verlag, Berlin, Heidelberg.

Godunov, S. (1959), A difference method for the numerical calculation of discontinuous solutions of hydrodynamical equations, *Mathematics Sbornik*, **47**(3), 271–306.

Gomes, G., Munoz, L., Yi, J., Toy, C., Cinnamon, S., Horowitz, R. & Alvarez, L. (2001). Meso-microscale traffic simulation of an AHS control architecture, in *Proceedings of the 2001 American Control Conference*, Arlington, VA, pp. 1806–11.

Gualtieri, G. & Tartaglia, A. (1998), Predicting urban traffic air pollution: a GIS framework, *Transportation Research Part D*, **3**, 329–36.

Health Effects Institute (2010), *Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects*, vol. 17, Health Effects Institute Special Report, Health Effects Institute. Available at: http://pubs.healtheffects.org/view.php?id=334.

Herrera, J. C., Work, D., Ban, X., Herring, R., Jacobson, Q. & Bayen, A. (2010), Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment, *Transportation Research C*, **18**, 568–83.

Hoh, B., Gruteser, M., Herring, R., Ban, J., Work, D., Herrera, J., Annavaram, M., Jacobson, Q. & Bayen, A. (2008), Virtual trip lines for distributed privacy-preserving traffic monitoring, Annual International Conference on Mobile Systems (MOBISYS), Brekenridge, CO.

Holmes, N. & Morawska, L. (2006), A review of dispersion modeling and its application to the dispersion of particles: an overview of different dispersion models available, *Atmospheric Environment*, **40**(30), 5902–28.

Karner, A. A. & Niemeier, D. E. D. (2010), Near-roadway air quality: synthesizing the findings from real-world data, *Environmental Science and Technology*, **44**(14), 5334–44.

Karppinen, A., Kukkonen, J., Elolahde, T., Konttinen, M., Koskentalo, T. & Rantakrans, E. (2000), A modelling system for predicting urban air pollution: model description and applications in the Helsinki metropolitan area, *Atmospheric Environment*, **34**, 3723–33.

Leveque, R. (1992), *Numerical Methods for Conservation Laws*, Birkhauser Verlag, Basel, Switzerland.

Lightwill, M. & Whitham, G. (1955), On kinematic waves. II. A theory of traffic flow on long crowded roads, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **229**(1178), 317–45.

Namdeo, A., Mitchell, G. & Dixon, R. (2002), TEMMS: an integrated package for modelling and mapping urban traffic emissions and air quality, *Environmental Modeling and Software*, **17**, 179–90.

Pierce, T, Isakov, V., Haneke, B. & Paumier, J. (2008), Emission and Air Quality Modeling Tools for Near-Roadway Applications, EPA/600/R-09/001, U.S. Environmental Protection Agency, Office of Research and Development, Washington DC. Available at: http://www.epa.gov/airscience/pdf/Pierceetal_EPA2008.pdf(2008).

Rodes, C. & Holland, D. (1979), NO_2/O_3 sampler siting study, 68-02-2292, Environmental Protection Agency, Research Triangle Park, NC.

Sharma, N. & Rao, K. C. K. C. (2004), Vehicular pollution prediction modelling: a review of highway dispersion models, *Transport Reviews*, **24**(4), 409–35.

Vardoulakis, S., Fisher, B., Pericleous, K. & Gonzalez-Flesca, N. (2003), Vehicular pollution prediction modelling: a review of highway dispersion models, *Atmospheric Environment*, **37**(2), 155–82.

Work, D., Blandin, S., Tossavainen, O.-P., Piccoli, B. & Bayen, A. (2010), A traffic model for velocity data assimilation, *Applied Mathematics Research eXpress (AMRX)*, **1**, 1–35.

Yura, E., Kear, T. & Niemeier, D. (1984), Using CALINE dispersion to assess vehicular PM2.5 emissions, *Atmospheric Environment*, **41**, 8747–57.

Zegeye, S., Schutter, S. D., Hellendoorn, J. & Breunesse, E., (2011), Nonlinear MPC for the improvement of dispersion of freeway traffic emissions, in *Proceedings of the 18th IFAC World Congress,* Milan, Italy, 10703–8.

Zhu, Y., Hinds, W. C., Kim, S., Shen, S. & Sioutas, C. (2002), Study of ultrafine particles near a major highway with heavy-duty diesel traffic, *Atmospheric Environment*, **36**(27), 4323–35.