Classification of road traffic and roadside pollution concentrations for assessment of personal exposure

Haibo Chen*, Anil Namdeo, Margaret Bell

Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK

Received 25 January 2007; received in revised form 19 March 2007; accepted 2 April 2007
Available online 29 May 2007

Abstract

Nowadays urban pollution exposure from road transport has become a great concern in major cities throughout the world. A modelling framework has been developed to simulate Personal Exposure Frequency Distributions (PEFDs) as a function of urban background and roadside pollutant concentrations, under different traffic conditions. In this paper, we present a technique for classifying roads, according to their traffic conditions, using the traffic characteristics and fleet compositions. The pollutant concentrations data for 2001 from 10 Roadside Pollution Monitoring (RPM) units in the city of Leicester were analysed to understand the spatial and temporal variability of the pollutant concentrations patterns. It was found that variability of pollutants during the day can be associated with specific road traffic conditions. Statistical analysis of two urban and two rural Automated Urban and Rural Network (AURN) background sites for particulates suggests that PM$_{2.5}$ and PM$_{10}$ are closely related at urban sites but not at rural sites. The ratio of the two pollutants observed at Marylebone was found to be 0.748, which was applied to Leicester PM$_{10}$ data to obtain PM$_{2.5}$ profiles. These results are being used as an element in the PEFDs model to estimate the impact of urban traffic on exposure.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Traffic classifications; Roadside pollution concentrations; PEFD; k-Means algorithm; RPM and AURN

1. Introduction

Urban air pollution is one of several major atmospheric pollution problems currently confronting the world’s population. The problem is growing because of rapidly increasing urban populations, unchecked urban and industrial expansion, and the phenomenal surge in the number and use of motor vehicles. Examining air quality information is important in understanding possible human exposure and potential impacts in health and welfare. The UK National Atmospheric Emissions Inventory (NAEI) report shows that road traffic is the largest emission source of many health-related air pollutants (AEAT/NETC, 2004) such as carbon monoxide (CO, 59%), nitrogen oxides (NOx, 45%), benzene (32%), 1,3-butadiene (75%) and primary PM$_{10}$ (25%), some of which contribute to the formation of ozone and secondary particles. Road traffic becomes an increasingly important sector as the particle size decreases. In 2002, it accounts 38% and 52% of PM$_{1.0}$ and PM$_{0.1}$ emissions, respectively. With road traffic emissions accounted for such a high proportion of the pollutants, and with the forecast of constantly growing traffic volume resulting in congestion and in turn exacerbating vehicle emissions (DoT, 1997), further refinement of the associations between air pollution, and in particular that related to traffic, and health is needed. A theoretical modelling exercise on a study of the relationship between industrial and traffic sources contributing to air quality objective exceedences in the UK was carried out, which showed that traffic emissions make a greater contribution to ground level concentrations of NO$_2$ than industrial sources per unit emission, and that street canyon conditions...
gives rise to higher NO\textsubscript{2} concentrations than an open road (Leksmono et al., 2006). Farias and ApSimon studied the temporal and spatial contributions to NO\textsubscript{x} concentrations from aircraft and traffic around Heathrow airport in West London (Farias and ApSimon, 2006). Their analysis showed that in the areas where people are exposed to air pollution, the impact of traffic emissions is larger than that of aircraft emissions.

Indeed, air quality problems experienced today in the UK are caused by pollution from road transport which is being addressed by the UK National Air Quality Strategy (DETR, 2000). Information about pollution may be obtained from two sources, namely monitoring sites (field data) and simulation (modelled data). A number of air quality simulation models have been developed, including two widely used models in the UK, namely ADMS-Urban (Carruthers and McHugh, 1996) and Airviro (SMHI, 2002), taking into account the whole range of relevant emission sources such as traffic, industrial, commercial, domestic and other less well-defined sources. When calibrated appropriately, these mesoscopic (or regional) models can be served as a useful tool for indirect estimation of air pollution levels and hence human exposure. Since a significant part of exposure of the population in many cities is caused by emissions from traffic in urban streets, the development of street pollution models has been a focus in exposure assessment. Examples of such models are STREET (Johnson et al., 1973), CPBM (Yamartino and Wiegard, 1986), OSPM (Hertel and Berkowicz, 1989; Berkowicz et al., 2006), CAR (Den Boeft et al., 1996), PANACHE (Tripathi, 1996), SPRAY (Tinarelli et al., 1994; Calori et al., 2006), Neural Network Models (Agirre-Basurko et al., 2006; Zito et al., 2007), and statistical distribution models (Gokhale and Khare, 2007). As the dispersion of street pollution is highly influenced by many factors such as the topology of the streets (street orientation, street width, height of buildings, etc.) and local wind turbulences, the calibration of these models to achieve accurate estimation is time-consuming and difficult. Geographic information systems (GIS) have been increasingly becoming a useful tool for the automatic map interpretation and representation of the street configuration in the assessment of air pollution exposure (Jensen et al., 2001; Beyea and Hatch, 1999).

The AURN sites serve as important sources of information about field pollution concentrations and hence indirect estimation of exposure at these fixed locations. Although there are over 1500 sites across the UK which monitor air quality, such coverage is not sufficient for the monitoring and modelling of street pollution. In Leicester, the local authority has installed 10 air quality units near the busy roads to measure the roadside pollution. Data from these units were used in this study.

Another serious limitation of the current air quality assessments of air pollution control policies is that they rely on estimates of outdoor concentrations rather than personal exposures (DoH, 1998). In our recent study on Reducing Urban Pollution Exposure from Road Transport (RUPERT), a modelling framework has been developed to simulate PEFDs of four pollutants (i.e. NO\textsubscript{2}, CO, PM\textsubscript{10} and PM\textsubscript{2.5}), based on time-activity patterns of different population groups (e.g. children, elderly, office and home works, etc.) in a range microenvironments (e.g. bedroom, kitchen, lounge, office, classroom, shop, bar/restaurant, transport, outdoors, large buildings, etc.) across a city, as a function of urban background and roadside pollutant concentrations, under different traffic conditions (Bell et al., 2004). In this way, the potential health benefits of traffic measures, designed to reduce the concentrations at the roadside and urban background locations, can be estimated by linking the modelling of roadside pollutant concentrations with the probabilistic modelling of population exposures. In order to achieve this, it is necessary to assign a pollution profile to each road type across a network.

The modelling framework in RUPERT combines new and existing models relating traffic and air pollution data, with particular emphasis of the impact of congestion, and a probabilistic model of personal exposure. The relationships between predicted PEFDs across a city and outdoor concentrations will provide a basis from which to estimate the potential health benefits of measures to reduce concentrations at roadside and urban background locations. In this paper, we present a technique for classifying roads based on the traffic characteristics and fleet compositions. The pollutant concentrations data for 2001 from 10 roadside pollution monitoring units in Leicester were analysed to understand the spatial and temporal variability of the pollutant concentrations patterns. It was found that variability of pollutants during the day can be associated with specific road traffic types. These results are being used as an element in the PEFDs model to estimate the impact of urban traffic on exposure.

2. Methodology

2.1. Study site

There are several air quality monitoring schemes (or systems) running in parallel in Leicester (Latitude: 52.38N, Longitude: 01.08W, population: 280 thousand inhabitants). Firstly, the AURN site maintained and run by DEFRA (Department for Environment Food and Rural Affairs) measures levels of O\textsubscript{3}, CO, SO\textsubscript{2}, PM\textsubscript{10} and NO\textsubscript{x}. The Leicester AURN site is classified as an Urban Centre. Urban Centre sites are non-kerbside sites located in an area representative of typical population exposure in town or city centre areas e.g. pedestrian and shopping areas. Leicester AURN site is located in a pedestrian piazza (i.e. an open space for pedestrians’ use) between eight and eleven-storey council offices. It is situated approximately 30 m from a three lane one-way road, which is subject to congestion at peak times. Sampling heights are typically within 2–3 m. Secondly, Leicester City Council (LCC) monitors air quality at roadside mainly using 10 RPMs and seven air quality monitoring stations (i.e. not at roadside). At other specific locations, regular monitoring is carried out for short-term periods of typically one month using a mobile van. Additional data from two AURN urban sites in London (Marylebone Road and Bloomsbury) and two rural sites in Rochester and Harwell have been analysed for...
better understanding of temporal and spatial variability of gaseous pollutants and size distribution of particulates.

LCC have installed 10 RPMs within the city of Leicester, as shown in Fig. 1. RPM sites are mainly located on major radial and ring roads representing busy and congested roads. RPM data is available at one-minute interval. For the RUPERT project, data from the ten RPMs were collected at one minute interval for the year 2001. CO and NO2 data were collected and then segregated and averaged to produce yearly and seasonal (winter and summer) weekday and weekend profiles. Geometric means and standard deviations have been used in developing a statistical description of these profiles as the data generally follow a geometric distribution.

Fig. 2 shows yearly profiles of CO at ten RPMs. These profiles represent average pollution over the year and therefore, without the day-to-day variation, help to illustrate the underlying relationships between roadside pollution and traffic conditions. CO profiles clearly show a diurnal variation consistent with traffic flow profiles indicating a strong relationship with the amount of traffic activity (LCC, 2001). The profiles are seen to fall into different categories, those displaying a dominant morning (W0625) or evening (W0158) peak or both, with profile W0552 being more pronounced than W1032. The profiles reflect not only the volume (W0158 has higher levels of CO throughout the day compared to W2626) but also the nature of traffic on the road. For example, a one-way street (W0158), a two-way radial with a dominant flow into (W0625) or out of city (W0948) or with busy commuter traffic at peak times in both directions (W0914).

Similar statistical analysis was carried out for data from seven air quality monitoring stations run by LCC. Particulates below 2.5 µm diameter are not monitored in Leicester at either AURN or LCC sites, though PM10 is monitored at AURN site. It was hypothesised that PM2.5 could be calculated from PM10. To test this, data were needed from sites where both PM10 and PM2.5 are simultaneously monitored. Only two urban and two rural sites in UK do this. The two urban sites are Marylebone Road and Bloomsbury in London and the two rural sites are at Rochester (Kent) and Harwell (Oxfordshire). Statistical analysis of the data from these four sites suggests that PM2.5 and PM10 are closely related at urban sites but not at rural sites as shown in Fig. 3. The ratio of the two pollutants observed at Marylebone was found to be 0.748, which was applied to Leicester PM10 data to obtain PM2.5 profiles.

2.2. The k-means algorithm

In order to classify the road types across Leicester, data including the morning average speed, the morning and afternoon peak vehicle flows and the daily average flow from the TRIPS (TRansport Improvement Planning System, Citilabs Europe, 2001) model were used and analysed using the k-means algorithm. This involves the parametric estimation of the probability density function and requires a priori statistical information such as the number of generators and the underlying functions within the data set. For most real-world applications, however, this type of information is unknown. Non-parametric techniques, such as Parzen windows (Parzen, 1962), solve this problem by obtaining an estimate of the density at every data point in the training set. However, these techniques are computationally expensive for large data sets and have the disadvantage of modelling any noise in the training set. Semiparametric estimation achieves this more parsimonious representation by applying kernel estimation methods, as used in the non-parametric techniques, to a smaller number of kernel functions. The number is less than the number of data points but still large compared to the probable number of generators within the data. In many practical applications, this approach has been shown to be an effective method for density estimation (Traven, 1991). Semiparametric estimation may be regarded as a clustering or partitioning of the data set in terms of cluster means and covariance matrices. The number of kernel functions used by the K-means algorithm (Mac-Queen, 1967) must be decided in advance. This algorithm aims at minimizing an objective function, in this case a squared error function.

\[
\text{The objective function } = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_{ij}^{*} - c_{j} \right\|^2
\]

Where \( \left\| x_{ij}^{*} - c_{j} \right\|^2 \) is a chosen distance measure between a data point \( x_{ij}^{*} \) and the cluster centre \( c_{j} \), and an indicator of the distance of the \( n \) data points from their respective cluster centres.

The calculation of the centres and membership of the clusters associated with the kernel functions can be implemented as follows.

1. Define the number \( K \) (say \( K = 4 \)) of kernel functions/ clusters.
2. Assign the data points at random to one of the \( K \) sets.
3. Compute the mean of the data points in each set.
4. Re-assign each data point to a new set according to which is the nearest mean.
(5) Re-compute the means of the sets.

(6) Repeat steps 4 and 5 until there is no further change in the grouping of the data points.

After the centres and memberships of the clusters are found, the k-means algorithm can then be used to classify road links into different groups. Each road link has a set of characteristic parameters such as travel speed, vehicle fleet compositions and vehicle fuel types. These parameters are used by the k-means algorithm to assign a road link to one of the groups which is the most similar to the road link. Other factors, such as the distribution between gasoline and diesel personal cars, and the fraction of cold-starts, may also have some impact. However, such data were not available in this study.

3. The data and results

3.1. Classification of road traffic

Although many UK cities use the demand responsive control system SCOOT (Split Cycle Offset Optimisation Technique, Hunt et al., 1982) to control and manage the road traffic in the network, the system only covers part of the road network, especially for busy urban roads excluding bus lanes. To overcome this problem, in this study, we use traffic data from survey and the TRIPS traffic model which covers both major and minor roads in the city, and the motorways passing by the city.

TRIPS is a transportation planning package which enables strategic as well as detailed analyses of multi-modal transportation networks. TRIPS provides a framework for implementing a wide range of travel demand forecasting models. In the TRIPS model used in this work, there were 3263 road links in the Leicester network. A number of output parameters from the model were used by the k-means algorithm to classify the roads in the network into several groups based on their traffic characteristics so that roadside pollution concentrations can be analysed to identify the hotspots of certain pollutants. A sensitivity test was carried out to evaluate the effects of the parameters to the classification of roads. This was done by including and excluding a parameter at a time and observing the changes made to the road classification. It turns out that there are strong correlations amongst these parameters (i.e. the classification wouldn’t be affected by excluding the parameters being tested), which implies that some of the parameters could be dropped from the modelling. The first ten most representative parameters chosen to use from the sensitivity test are the morning average speed, hourly flows in hours 9, 12, 18 and 24, and vehicle compositions including PET_WO_C (petrol cars without a catalyst), PET_W_C (petrol cars with a catalyst), LGV_D (diesel light goods vehicles), BUSES and HGV (heavy goods vehicles). These parameters are then normalised to 1 to remove their differences in magnitude.

Using the ten parameters, the K-means algorithm classified those links into six groups as shown in Table 1. Further analysis of the nature of the k-means classification reveals that road links in each group possess a certain level of similarity. For example, cluster 1 consists of motorway links, cluster
are bus lanes and cluster 3 are those roads with low flows such as residential roads. Clusters 2, 5 and 6 are urban roads with high, medium and low HGV flows, respectively. Of the 3263 road links in total and the numbers of links in the six clusters are 14, 755, 27, 14, 1140 and 1313, respectively. Their corresponding percentages are 0.4, 23.1, 0.8, 0.4, 34.9 and 40.2.

3.2. Analysis of roadside pollution concentrations

The RPMs were then grouped and classified using the k-means clusters to associate an average CO and NO2 profile with a particular road classification. If a link (for example, Newarke Street), fell into k-means Cluster 2 then the RPM located adjacent to it (in this case W0158), was classified as Category 2. RPMs at Newarke Street, Melton Road, Uppingham Road, Welford Road and Soar Valley Way were grouped into Category 2 whereas RPMs at Narborough Road, Hinckley Road, A50 New Parks, Norman/Wilton and A6 Ashtree Road were grouped into Category 5. Average profiles for Categories 2 and 5 were derived and are shown in Fig. 4. AURN profile was classified as Category 6. Table 2 shows the list of RPMs classification by road type.

4. Conclusions

This work has demonstrated that the roads across a city can be classified into categories that can be associated with essentially different daily pollutant profiles. The ten most sensitive variables governing the classification were the morning average speed, hourly flows in hours 9, 12, 18 and 24, and vehicle compositions including PET_WO_C, PET_W_C, LGV_D, BUSES and HGV. The urban roads with high HGV flows showed high levels of pollution with marked peaks in the pollution profile associated with traffic peaks. The urban roads with medium HGV showed medium levels of pollution.

Table 1

<table>
<thead>
<tr>
<th>Parameters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEED</td>
<td>0.9082</td>
<td>0.1905</td>
<td>0.2698</td>
<td>0.2041</td>
<td>0.5188</td>
<td>0.1764</td>
</tr>
<tr>
<td>Flow in Hour09</td>
<td>0.7736</td>
<td>0.3811</td>
<td>0.0560</td>
<td>0.3835</td>
<td>0.3657</td>
<td>0.3586</td>
</tr>
<tr>
<td>Flow in Hour12</td>
<td>0.9916</td>
<td>0.5357</td>
<td>0.0399</td>
<td>0.4980</td>
<td>0.5000</td>
<td>0.4935</td>
</tr>
<tr>
<td>Flow in Hour18</td>
<td>0.9416</td>
<td>0.4032</td>
<td>0.0456</td>
<td>0.5502</td>
<td>0.5319</td>
<td>0.5270</td>
</tr>
<tr>
<td>Flow in Hour24</td>
<td>0.9640</td>
<td>0.2125</td>
<td>0.0277</td>
<td>0.3129</td>
<td>0.3018</td>
<td>0.3035</td>
</tr>
<tr>
<td>PET_WO_C</td>
<td>0.7143</td>
<td>0.7203</td>
<td>0.5285</td>
<td>0.0000</td>
<td>0.7350</td>
<td>0.7413</td>
</tr>
<tr>
<td>PET_W_C</td>
<td>0.3316</td>
<td>0.3570</td>
<td>0.9506</td>
<td>0.0000</td>
<td>0.3420</td>
<td>0.3478</td>
</tr>
<tr>
<td>LGV_D</td>
<td>0.7143</td>
<td>0.8023</td>
<td>0.6376</td>
<td>0.0000</td>
<td>0.6069</td>
<td>0.6034</td>
</tr>
<tr>
<td>BUSES</td>
<td>0.0007</td>
<td>0.0122</td>
<td>0.0152</td>
<td>1.0000</td>
<td>0.0673</td>
<td>0.0639</td>
</tr>
<tr>
<td>HGV</td>
<td>0.7905</td>
<td>0.4534</td>
<td>0.1383</td>
<td>0.0000</td>
<td>0.2079</td>
<td>0.1789</td>
</tr>
</tbody>
</table>

Fig. 4. Summer weekdays profile of CO by road category (3 for roads with low flows, 2 and 5 for urban roads with high and medium HGV flows, respectively).
Table 2
Roadside pollution monitors and their classification by road type

<table>
<thead>
<tr>
<th>Station name</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newarkie Street</td>
<td>2</td>
</tr>
<tr>
<td>Melton Rd</td>
<td>2</td>
</tr>
<tr>
<td>Uppingham Rd</td>
<td>2</td>
</tr>
<tr>
<td>Welford Rd</td>
<td>2</td>
</tr>
<tr>
<td>Soar Valley Way</td>
<td>2</td>
</tr>
<tr>
<td>Narborough Rd</td>
<td>5</td>
</tr>
<tr>
<td>Hinckley Rd</td>
<td>5</td>
</tr>
<tr>
<td>A50 New Parks</td>
<td>5</td>
</tr>
<tr>
<td>Norman/Wilton</td>
<td>5</td>
</tr>
<tr>
<td>A6 Ashtree Rd</td>
<td>5</td>
</tr>
<tr>
<td>AURN</td>
<td>6</td>
</tr>
</tbody>
</table>

Category 2 = urban road with high HGV. Category 5 = urban road with medium HGV. Category 6 = sub-urban road with medium HGV.

with less pronounced peaks whilst the classifications with predominant cars were associated with the urban background. Statistical analysis of two urban and two rural AURN background site for particulates suggests that PM$_{2.5}$ and PM$_{10}$ are closely related at urban sites but not at rural sites. The ratio of the two pollutants observed at Marylebone was found to be 0.748, which was applied to Leicester PM$_{10}$ data to obtain PM$_{2.5}$ profiles. Finally, this work forms part of the PEFDs tool which can be used to simulate personal exposure and hence to assess the health impact of traffic.

Acknowledgements

We thank the Engineering and Physical Sciences Research Council (EPSRC) for financing the infrastructure of the Instrumented City Facility, DoH for sponsoring the RUPERT project and LCC for supporting the continuous data capture of traffic, air quality and meteorological conditions without their support this work would not have been possible. We also thank LCC for providing the TRIPS data. The views expressed in this paper are those of the authors and do not represent the view of the abovementioned organisations or any of the non-academic partners of this project.

References


