



Aggregated GPS tracking of vehicles and its use as a proxy of traffic-related air pollution emissions



Shimon Chen^{*}, Shlomo Bekhor, Yuval, David M. Broday

Department of Civil and Environmental Engineering, Technion, Israel Institute of Technology, Haifa 3200003, Israel

HIGHLIGHTS

- A methodology for examining ICT-derived traffic volumes is presented.
- Whereas some problems were found, in general ICT data were reliable.
- Traffic was reported in cells devoid of roads, due to scattering from adjacent cells.
- ICT data have better spatiotemporal availability than more traditional data sources.
- The ICT-based traffic volumes were successfully used as a proxy for NO₂ emissions.

ARTICLE INFO

Article history:

Received 18 April 2016

Received in revised form

3 August 2016

Accepted 5 August 2016

Available online 6 August 2016

Keywords:

Air quality modelling

Aggregated tracking of vehicles

Information and communication technology

Optimised dispersion model

NO₂

ABSTRACT

Most air quality models use traffic-related variables as an input. Previous studies estimated nearby vehicular activity through sporadic traffic counts or via traffic assignment models. Both methods have previously produced poor or no data for nights, weekends and holidays. Emerging technologies allow the estimation of traffic through passive monitoring of location-aware devices. Examples of such devices are GPS transceivers installed in vehicles. In this work, we studied traffic volumes that were derived from such data. Additionally, we used these data for estimating ambient nitrogen dioxide concentrations, using a non-linear optimisation model that includes basic dispersion properties. The GPS-derived data show great potential for use as a proxy for pollutant emissions from motor-vehicles.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Motor-vehicles are the largest contributors to urban air pollution in the developed world. This results from the ever-increasing abundance of motorized vehicles in urban areas (Fenger, 2009); the proximity of vehicular emissions to the population (Karppinen et al., 2000); and the difficulty in controlling emissions from internal combustion engines (Wang et al., 2004). Diesel-powered

vehicles generally emit more nitrogen oxides and particulate matter than petrol-powered vehicles, while carbon monoxide and volatile organic compounds emissions are lower (Song, 2000; McAllister et al., 2011). Heavy-duty vehicles are usually both diesel-powered and burn fuel at a higher rate than private cars (Gaffney and Marley, 2009; Schipper, 2008). Thus, in order to reliably assess exposure to air pollution it is important to account for both traffic activity patterns and fleet composition. Most common air quality models, such as Chemistry-Transport Models (CTM) and Land Use Regression (LUR), make use of traffic-related variables as input (Hennig et al., 2016).

Some studies considered only geographical attributes of the road network (e.g., category and size) to proxy emissions (Hoek et al., 2008). This approach can lead to large model errors, as observed by Johansson et al. (2015). A more accurate proxy could be obtained when accounting for the traffic volume (Beelen et al., 2013; Janssen et al., 2008). Indeed, some studies used detailed

Abbreviations: AQM, Air Quality Monitoring; ATV, Aggregated Tracking of Vehicles; CBS, Central Bus Station; CTM, Chemistry Transport Model; ICT, Information and Communication Technology; LUR, Land Use Regression; ODM, Optimised Dispersion Model; PSFM, Percentage Span From Mean; PSFMwA, Percentage Span From Mean with Adjacent cells; TA, Traffic Assignment.

^{*} Corresponding author.

E-mail addresses: shimi@technion.ac.il (S. Chen), sbekhor@technion.ac.il (S. Bekhor), lavyuy@tx.technion.ac.il (Yuval), dbroday@tx.technion.ac.il (D.M. Broday).

traffic counts to estimate emissions (e.g. Barth and Boriboonsomsin, 2009; Pratt et al., 2014). However, such data are rarely available with sufficient accuracy and detail across large areas, and are also biased towards roads with higher traffic volumes. For example, Beelen et al. (2007a,b) tried to obtain traffic intensities for the entire Netherlands, in order to use them for exposure assessment. Combining inputs from all the municipalities turned out to be an enormous effort and resulted in a lot of missing data. Similarly, at any given time 30% of the inductive loop detectors in California are malfunctioning (Herrera et al., 2010).

Another commonly used method to estimate traffic intensities is the use of Traffic Assignment (TA) models (e.g. Dhondt et al., 2012; Yuval et al., 2013; Shekarzifard et al., 2015). TA models require population travel demand information, which is traditionally derived from census data, questionnaires and road sensors (Calabrese et al., 2011; Bekhor and Shem-Tov, 2015). These data sources are often limited in spatio-temporal scope (Calabrese et al., 2011). For example, relatively high TA model errors have been found on low-volume roads in Florida (Lan et al., 2005), and relatively large night-time errors were reported in Berlin (Rieser et al., 2007).

A possible alternative is to use data streams from emerging Information and Communication Technologies (ICT), such as GPS and cellular tracking (Antoniu et al., 2011). For example, Castro et al. (2012) fitted 5000 taxis with GPS devices and successfully estimated road capacities. Bekhor et al. (2013) used data from GPS-equipped vehicles tracked by Decell Technologies to estimate free-flow driving speeds. Although not yet ubiquitous, the proportion of vehicles with integrated GPS systems is increasing (Castro et al., 2012). GPS devices have a reported accuracy of about 10 m (Djuknic and Richton, 2001). Mobile phone triangulation data are less accurate but can also be used to estimate traffic parameters and are more pervasive in the population (Bar-Gera, 2007). Bekhor and Shem-Tov (2015) showed high correlations between traffic patterns inferred from cellular data and more traditional surveys. However, some important issues were highlighted, including under-estimation of traffic during morning rush-hour and inaccuracies in regions with a sparse distribution of antennas. A prototypical combined approach was presented by Herrera et al. (2010), using GPS-enabled mobile phones to gather traffic data. The temporal coverage of ICT data is not limited, since these devices normally operate continuously. In addition, ICT can provide data directly on a grid, which unlike data on road segments do not require a conversion process (such as employed by Yuval et al., 2013; Pratt et al., 2014) for use as emission proxies. However, gridded data may be less accurate as it does not represent the exact location of the emissions. In addition, there may be sampling biases associated with an uneven distribution of personal ICT devices among the population.

Recent studies (Liu et al., 2013b; Pratt et al., 2014; Gariazzo et al., 2016; Dewulf et al., 2016) used ICT tracking of individuals' trajectories to assess their personal exposure over a modelled pollution map. Etyemezian et al. (2003) used on-board GPS receivers to correlate driving conditions and road dust emissions. Other studies (Liu et al., 2013a; Borrego et al., 2016) used Aggregated Tracking of GPS-equipped Vehicles (ATV) to estimate driving behaviour and incorporate it into an emission factors calculation. However, no spatial air quality modelling scheme has been reported to date using this revolutionary data source as a proxy for vehicle volumes or traffic flows.

In this work we study an ATV dataset that covers most of the populated area of Israel in a 250 m × 250 m grid and use it to proxy traffic emissions, extending our previously published optimised dispersion work (Yuval et al., 2013).

2. Materials and methods

2.1. ATV traffic data

During 2012, Decell Technologies collected GPS data from a fleet of vehicles fitted with GPS tracking devices. Their sample included more than 100,000 vehicles of various types, including a high percentage of the heavy trucks in Israel. There were approximately 2 million vehicles in Israel at 2012 (Bekhor et al., 2013). Decell extrapolated the vehicle volumes from their sample to the full population using traffic counts obtained from the Israeli Central Bureau of Statistics. These counts were performed using pneumatic road tubes at selected road segments of the main highways in Israel during one week at 2012 (Central Bureau of Statistics, 2015).

We obtained the 2012 yearly mean traffic data from Decell. Vehicle volumes were given in 125,733 grid cells of 250 × 250 m², covering about 37% of the land area of Israel and 86% of its population. We refer the reader to the web version of this article for an interactive view of the study area. Each grid cell contained a distinct mean volume (average number of vehicles passing through the cell per hour) for buses, trucks and private vehicles, with trucks defined as vehicles over 5 tonnes (excluding buses) and minibuses counted as private vehicles. For each of the three vehicle types, mean volumes were provided for 11 daily time windows: 00:00–03:00, 03:00–06:00, 06:00–07:00, 07:00–08:00, 08:00–09:00, 09:00–12:00, 12:00–15:00, 15:00–18:00, 18:00–20:00, 20:00–22:00 and 22:00–24:00. These time windows provide more granular separation in hours with higher temporal variability (i.e. rush-hours) and coarser separation for times with reduced traffic activity. Moreover, separate data were obtained for weekdays (Sunday–Thursday in Israel), Fridays and Saturdays (the weekend in Israel). Additional data granulation has been made for routine and vacation (Jewish holidays and summer) periods, for a total of 198 data per grid cell. This granulation was chosen in order to optimise the trade-off between the temporal resolution of the dataset and the sample size of the signals used for each datum.

2.2. ATV data validation

Errors in the input data to an air quality model have the potential to propagate and cause inaccuracies in exposure estimation. As ICT data have never before been used for this purpose, we find it particularly important to put them through scrutiny and document the results.

2.2.1. Daily patterns

Grid cells located on routes between residential and business areas are expected to show peak traffic volume during the morning and afternoon rush-hours. Several such grid cells were selected and examined. In addition, a more quantitative analysis was performed using magnetic loop detector traffic counts at 5-min intervals for Highway 20 (H20) from the 15 to the 23 of April 2011. This period included a routine weekend, a routine weekday, three holiday weekdays and a holiday weekend. The linear correlations between the traffic counts and the ATV data were calculated, with the high-resolution traffic counts averaged over the ATV data time windows. Only results for the routine weekday (April 17) and the routine Saturday (April 16) are presented.

2.2.2. Spatial patterns

It is expected to find higher traffic volumes in grid cells that intersect with main roads, and lower volumes as the road priority decreases. However, the difference between traffic volumes in different municipalities could, in principal, be greater. We chose to test the spatial fit of the ATV data to the Israeli road network in

three different municipalities. These included two large cities (Tel-Aviv and Haifa) and one suburban region (Lev Hasharon). Road categories were extracted from OpenStreetMap and assigned a class according to their official definition (OpenStreetMap Contributors, 2016) and manual sampling of roads from each category. Table 1 details the interpreted type and assigned class for each OpenStreetMap road category. Each grid cell was assigned the highest class of road segment intersecting it. The Spearman correlation (Spearman, 1904) was calculated between the ordinal classes and the continuous traffic volumes in each municipality.

Daily traffic counts for 423 intersections in Tel Aviv were obtained from the municipality. Each count was taken on a single workday, once in each location, from 07:00 to 19:00 during 2008–2012. Of these counts, 395 had separate entries for trucks, buses and private vehicles. The spatial Pearson correlations of the traffic volumes of different vehicle types were compared between the counts and the ATV dataset. For this comparison, the 395 closest unique ATV grid cell centroids were chosen. ATV traffic estimates were summed over the whole day from 07:00 until 19:00.

2.2.3. Traffic flow conservation

The number of vehicles travelling in sequential segments of a road should remain constant between intersections. We located isolated road segments of five highways (H2, H4, H6, H34 and H70). Each of these segments is: (a) roughly parallel to one of the ATV data grid axes; and (b) has no neighbouring non-agricultural roads within 750 m of it (the width of three grid cells). These roads are marked on the [Supplementary interactive map](#) available with the online version of this article. Five grid cells along each of the five road segments were selected. The percentage difference between the traffic volume in each grid cell to the segment mean was calculated:

$$\delta_{ij} = \frac{V_{ij} - \bar{V}_j}{\bar{V}_j} \cdot 100\% \quad (1)$$

where V_{ij} is the total vehicle volume in the i th grid cell along the j th road segment and \bar{V}_j is the arithmetic mean of the total vehicle volume in the 5 grid cells along the j th road segment. For each road segment, the Percentage Span From Mean (PSFM) was defined as:

$$\text{PSFM}_j = \max(\delta_{1j}, \dots, \delta_{5j}) - \min(\delta_{1j}, \dots, \delta_{5j}) \quad (2)$$

In order to account for possible scatter of ATV data around the highway, the PSFM was calculated also while accounting for traffic volume in the two directly adjacent cells to each chosen grid cell, normal to the road direction. In this way, each of the five datapoints along the road segment represents the sum of three adjacent cells, and the calculated PSFM value for the segment in this case is designated the Percentage Span From Mean with Adjacent cells (PSFMwA).

2.2.4. Locations with known traffic patterns

Unique traffic patterns are expected at certain landmarks. In industrial areas we expected a large proportion of trucks that transport goods. Around central bus stations, we expected an increased number of buses. In highly Jewish religious areas, little to no traffic is expected during the Shabbat (Jewish weekly day of rest). At the same time, we expected high traffic volumes around popular recreational areas of the secular community. The ATV dataset was tested against these assumptions.

To assess the reliability of truck volume estimates, the ratio of truck to private vehicle volumes was compared between different land-use types. Three polygon categories were extracted from

OpenStreetMap: residential, industrial and quarry areas. Industrial and quarry polygons were considered industrial and compared as one group against the residential polygons. A total of 323 industrial and 1549 residential polygons were located. Of these, 227 industrial and 1147 residential polygons had traffic in them according to the ATV data. In each polygon, the ratio between the mean private volume and the mean trucks volume was calculated. The time period chosen for this comparison was 9–12, as we expect trucks to be at industrial locations during this time (and less so at night). The statistical significance of the difference between the polygons with different land-use categories was tested using a one-tailed Mann-Whitney test.

Bus traffic is expected to be maximal in the immediate vicinity of central bus stations (CBSs) and to decrease with increasing distance from them. Sixteen CBSs were selected and a 1.5 km buffer area was constructed around them (less than half the distance between the two most proximate CBSs). After filtering out grid cells with no intersecting roads, the Spearman correlation between the distance from the CBS centroid and the bus-to-total traffic volume ratio was calculated. The Spearman coefficient was selected because there was no reason to assume the relationship to be linear or to have any other known characteristics, except monotony.

The Meah Shearim neighbourhood in Jerusalem is characterised by an ultra-orthodox population, with virtually no traffic activity during the Shabbat, from Friday evening until Saturday evening (with the exact time varying throughout the year). In contrast, the Tel-Aviv port area is a popular recreational area for the secular community during Friday night. The temporal patterns of the ATV data in both locations were compared and contrasted against the aforementioned assumptions. The two locations are marked on an interactive map, available as [Supplementary information](#) with the on-line version of this article.

2.3. Use of ATV data as a proxy for traffic emissions

The optimised dispersion model (ODM) described by Yuval et al. (2013) was used to estimate ambient NO_2 concentrations based on the ATV data. This model uniquely combines proxies of traffic emissions with ambient concentrations (measured at AQM stations) and meteorological data, and produces high-resolution concentration maps. The concentration in each grid cell (C_i) is modelled as a superposition of contributions from all the M grid cells in the study area as follows:

$$C_i = p_1 + p_2 \sum_{j=1}^M \frac{T_j f(\theta_{ij})^{p_3}}{(D_{ij} + p_4)^{p_5}} \quad (3)$$

where T_j is a proxy for vehicular nitrogen oxides emissions in the j th cell (in this work it is the vehicle-type weighted ATV-based volume); D_{ij} is the euclidean distance between the i th (receptor) and j th (source) cells; θ_{ij} is the angle between the regional wind vector and the connecting vector from the j th to the i th cell; $f(\theta_{ij})$ is $\cos(\theta_{ij})$ for $\theta_{ij} \leq 90^\circ$ and zero otherwise; and p_{1-5} are unknown parameters which are optimised for a least-squares fit between the measured and modelled concentration at each timepoint (i.e. for every half hour). ODM should not be confused with more traditional atmospheric dispersion models (a recent description of these was made by Leelösy et al., 2014). Unlike ODM, these models attempt to follow exactly atmospheric physics, using some idealised assumptions. Instead, the ODM scheme incorporates some principals of pollutant transport and dispersion, in a simple and closed form, such that the model parameters can be easily optimised to fit the measured concentrations. For further details, see Yuval et al. (2013).

Table 1

List of OpenStreetMap road categories and their corresponding interpreted types and classes.

OpenStreetMap road category	Interpreted type	Assigned road class
Motorway, motorway link, trunk	Freeway	5
Trunk link, primary, primary link, secondary	Arterial	4
Secondary link, tertiary, unclassified	Collector and distributor	3
Tertiary link, residential	Local	2
Track, service, living street	Services and access	1

Yuval et al. (2013) modelled NO₂ concentrations over a 25 km wide coastal strip in central Israel. To compare our results with those of Yuval et al. (2013), we modelled the same pollutant over the same study area. Namely, the only difference from Yuval et al. (2013) is the use of ATV traffic volume estimates (provided on a spatial grid) rather than using traffic assignment model output (that were provided for distinct road segments and were transformed into gridded values). Twenty-two general monitoring stations were active in the study area during 2012 (near-road stations were not included for reasons discussed in Yuval et al., 2013). The monitoring data were obtained from the Technion Center of Excellence in Exposure Science and Environmental Health's air pollution monitoring data archive. The data pass quality assurance and quality control processing before being released for use.

To use the traffic volume estimates as proxies of emissions, the private and truck ATV traffic estimates were combined, taking the emission from trucks to be 3.6 times higher than from private vehicles (Pratt et al., 2014). The ODM was used to estimate NO₂ concentrations for all the half-hourly timepoints in 2012. Model performance was assessed by a leave-one-out complete cross-validation. Each half-hour was assigned the mean traffic volume in the corresponding ATV time window (Section 3.1) (i.e. at 13:30 the assigned traffic was of 12:00–15:00). For time-points that are exactly on the border between neighbouring time windows, the traffic volume from the earlier window was used (i.e. the traffic at 15:00 was assumed to be that of 12:00–15:00), since the reported monitoring data are averaged over the preceding half-hour. Due to computation limitations, the original ODM study of Yuval et al. (2013) was applied for in a 500 m × 500 m grid. To match the resolution of the current work and enable fair comparison of the two studies, the resolution of the ATV dataset had to be degraded. Each 250 × 250 m² cell was approximated to have a uniform traffic density, from which a sum was taken to the 500 × 500 m² cells.

Five statistics were used for assessing the quality of the results: the mean bias (MB), the mean error (ME), the normalised mean error (NME), the ratio of modelled values within a factor of two of the observations (FAC2), and the coefficient of determination (R^2).

3. Results and discussion

3.1. ATV data validation

3.1.1. Daily patterns

Fig. 1 depicts a common daily pattern at the busy Ziv junction in Haifa. A peak in private traffic is seen between 08:00–09:00 and a smaller peak is visible between 15:00–18:00. Private traffic volumes are much higher than truck volumes, which in turn are much higher than bus volumes. Truck traffic peaks at midday (12:00–15:00), while bus traffic peaks in the early morning (06:00–08:00). This is the expected pattern in an urban area.

In order to quantitatively evaluate the ATV daily patterns, they were compared to traffic counts in H20 from April 2011. Decell's ATV data is provided on a spatial grid. Often, each grid cell includes more than one road, so a perfect correlation between vehicle counts on a specific road segment and the gridded ATV value

cannot be expected. In addition, Decell's data are yearly means and counts from a specific week in April of a different year can reasonably deviate from it somewhat. Still, the ATV data are very similar to the traffic count data (Fig. 2). Different patterns can be observed for Saturdays and for weekdays in both the traffic counts and the ATV data.

3.1.2. Spatial patterns

Spearman correlations between ATV traffic volume estimates and the highest road class (see Table 1) in each grid cell were 0.79, 0.75 and 0.78 in Haifa, Tel Aviv-Yafo and Lev Hasharon, respectively. In all cases this correlation is statistically significant (p -value < 10^{-10}). For Tel Aviv-Yafo, the trend is also displayed in a box plot (Fig. 3).

Spatial linear correlations between the traffic volume of different vehicle types in Tel-Aviv are shown in Table 2. The correlations are consistently higher in the ATV database. In particular, all the vehicle types have high correlations with each other (–0.9) according to the ATV data, while the traffic counts show the highest correlation between trucks and private vehicles, with relatively low correlation between buses and private vehicles. We believe that the generally higher correlations in the ATV dataset can be explained by its: (a) much more coarse spatial separation (250 × 250 m² grid cells as opposed to specific intersections); and (b) yearly averaging of volumes (the counts are more sensitive to daily variations). Still, the similarity between the correlations in the ATV data is striking and may indicate bias in the estimation of bus volumes.

3.1.3. Traffic volume conservation

Variation of approximately 70% from the mean traffic volume was observed along isolated road segments of H2, when using only the grid cell in which the road segment actually passes (Fig. 4). However, when accounting for the two adjacent cells (in the transverse direction, perpendicular to the road), the deviation from the mean traffic volume decreased to only 1–2%. Similar patterns of decreased inconsistency of the traffic volume when including adjacent cells were observed in most other road segments (Table 3). It should be noted that the road with the highest deviations among those tested was H70, which is the only east-west (rather than north-south) road examined. Due to the morphology of Israel's road network, it was difficult to find additional suitable east-west roads for this test. H70 is also the highway with the lowest volumes among the five. The adjacent cells have no roads intersecting them and yet traffic is reported within them by the ATV database. Overall, 2.35% of the traffic in the ATV database is reported in cells that are devoid of road segments.

3.1.4. Locations with known traffic patterns

Statistics of the truck-to-private ratio in residential and industrial areas are shown in Table 4. The median ratio in industrial areas is 1.4 times higher than in residential areas. The difference between the medians is statistically significant (p -value < 10^{-10}).

Regarding bus traffic near CBSs, we report here only the results for 7–8 AM. Similar results were obtained in the other time windows. The bus traffic volume was negatively correlated with the

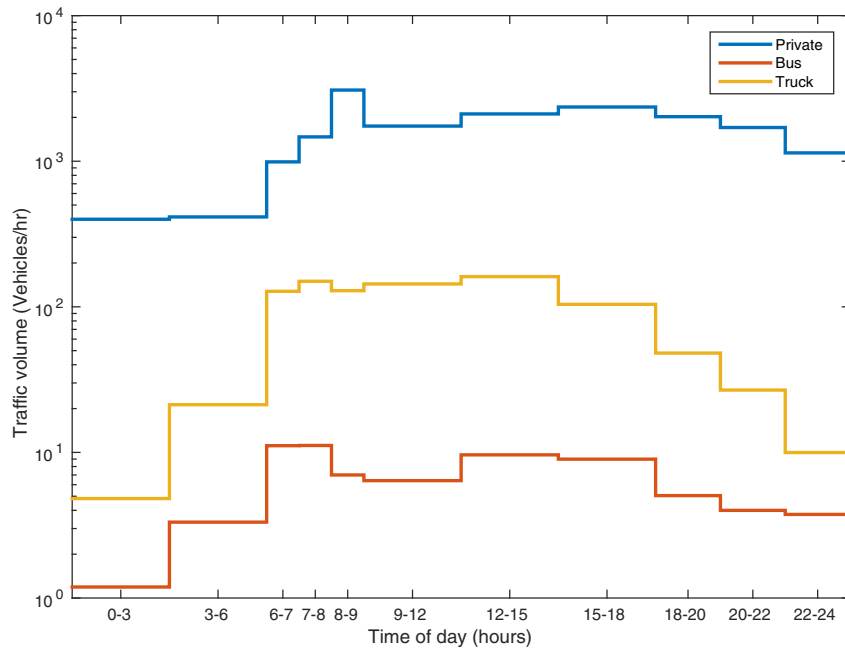


Fig. 1. Daily ATV traffic pattern at Ziv junction in Haifa.

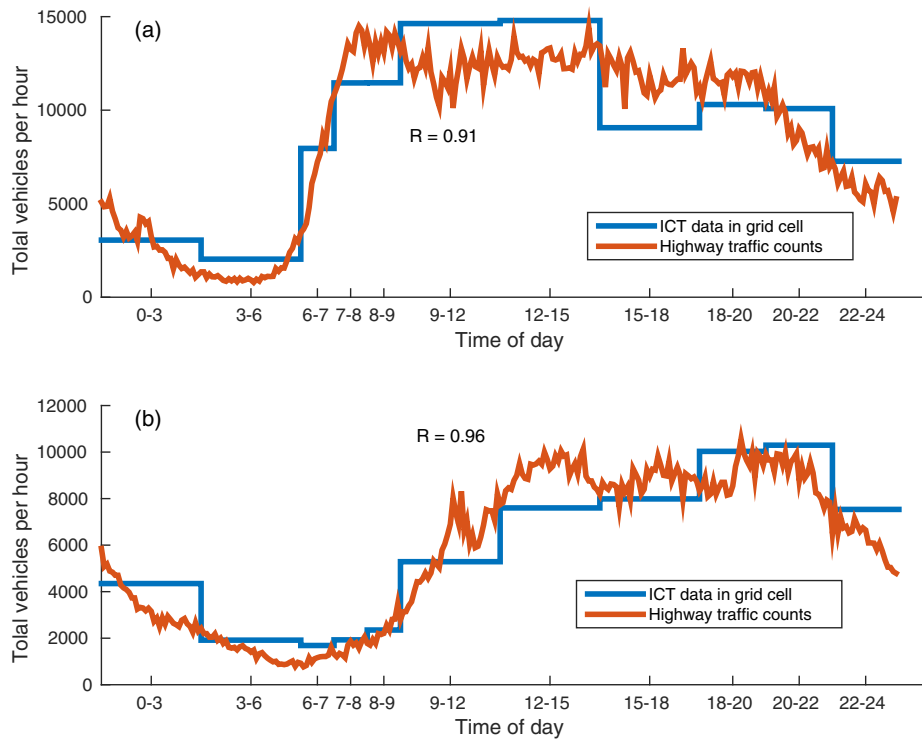


Fig. 2. Comparison of traffic counts and ATV data for H20 on (a) Sunday and (b) Saturday.

distance from each of the 16 CBSs examined. This correlation was statistically significant (p -value < 0.05) for exactly half (8) of the CBSs. The correlations were generally weak, with most of the values between 0 and -0.3 . The very weak correlations, together with the higher than expected correlation between buses and other types of vehicles (Table 2) lead us to suggest that the quality of bus data is questionable, thus we did not use the ATV bus volumes as input to the ODM.

In general, the different daily patterns of the traffic volume correspond well with the expected patterns. Specifically, in the ultra-orthodox neighbourhood in Jerusalem a sharp decrease in traffic volume is evident on Friday night, up until Saturday evening (Fig. 5). In contrast, the Tel-Aviv port recreational area is characterised by peak traffic in the late hours between Thursday and Friday and between Friday and Saturday.

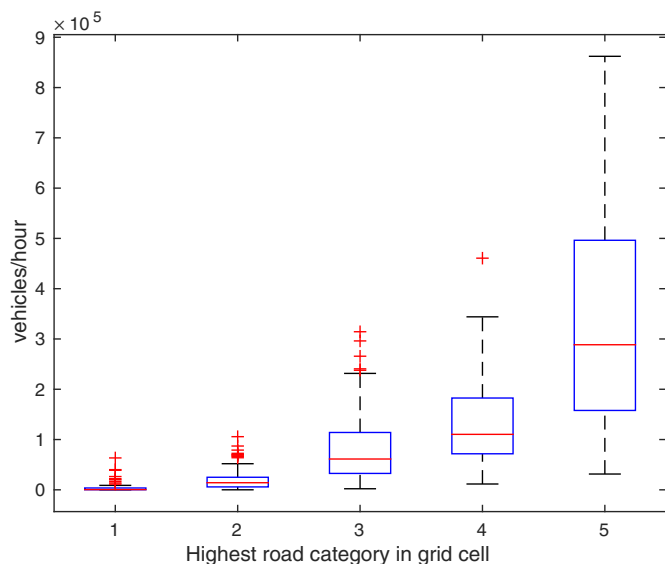


Fig. 3. Box plot of ATV traffic estimates and highest road class in grid cells. The red lines mark the medians, the blue horizontal lines mark the 25 and 75 percentiles and the black whiskers represent the maximum and minimum non-outlier values of each set. Outliers are defined as having a difference of more than 1.5 inter-quartile ranges from the blue box and are marked by red +. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Spatial Pearson correlation coefficients between the different vehicle types in the Tel-Aviv area in the ATV traffic volume data and among the traffic counts. Number of observations is 395 in both cases.

Traffic data source	Private-truck	Private-bus	Truck-bus
Counts	0.75	0.36	0.46
ATV	0.93	0.88	0.92

Table 3

Traffic conservation in selected road segments from 5 different highways in terms of the conservation measure Percentage Span From Mean (PSFM, see Section 3.1.3) and its counterpart that accounts for traffic volume also in the adjacent cells (PSFMwA, (%)). Two daily time periods are reported (8–9 and 22–24).

Highway	PSFM (8–9)	PSFM (22–24)	PSFMwA (8–9)	PSFMwA (22–24)
H2	68	72	2.4	1.0
H4	29	42	2.2	3.3
H6	53	61	5.1	14
H34	6.1	4.5	2.5	5.0
H70	230	230	26	30

Table 4

Properties of the distributions of truck-to-private ratios in residential and industrial grid cells. Columns are: number of observations (n), mean, median, standard deviation (STD) and skewness.

Landuse type	n	Mean	Median	STD	Skewness
Residential	1147	0.21	0.15	0.52	25.7
Industrial	227	0.48	0.21	1.32	6.99

with the exception of the mean absolute error. The model is slightly biased towards concentration over-estimation. Fig. 6 shows the change in coefficient of determination for an average day. The best fit is achieved during the morning and afternoon rush-hour traffic while poorer performance is observed at night. The poor performance during periods of lower traffic (e.g. nights and vacations) could indicate that either the traffic data is not as good during these times or that the role of traffic in ambient NO_2 during these times is lower. Notable NO_2 sources and processes not accounted for by the ODM are industrial stacks and re-circulation.

Average NO_2 concentration maps for various times of day are shown in Fig. 7. At the late night, relatively high concentrations are visible near the coastline whereas in the morning the high con-

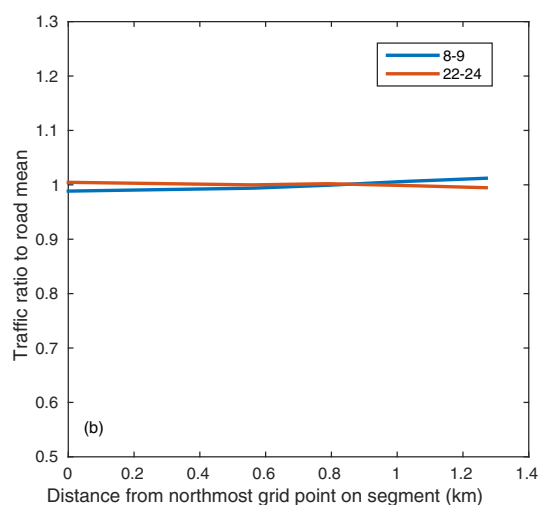
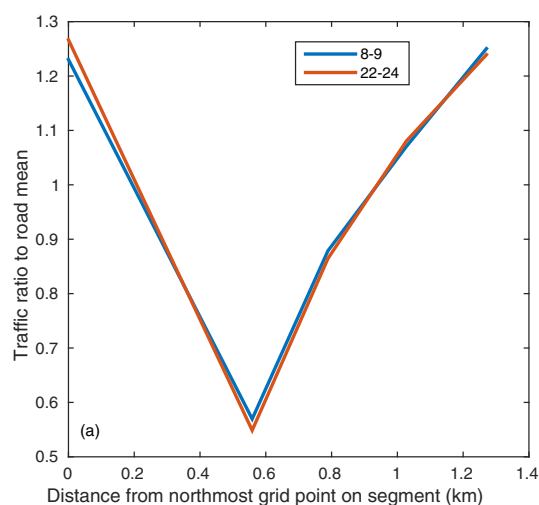


Fig. 4. (a) Ratio of traffic volume in a grid cell to the road segment mean traffic volume as a function of distance from the northernmost grid cell in the selected segment of H2 ($n = 5$ grid cells) for 08:00–09:00 (red) and 22:00–24:00 (blue). (b) is like (a) but with ATV traffic volume in the 2 adjacent cells in the east-west (perpendicular to the road) direction added to each cell. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Use of ATV data as a proxy for traffic emissions

Table 5 shows cross-validated performance measures of the ODM in different periods of the week and of the year. The model generally performs better when NO_2 concentrations are higher,

concentrations are more inland and at midday the high concentrations are limited to locations very proximate to the main roads. This corresponds well with our understanding of meteorology in the Israeli coastal plane and its dominant breeze cycle. At night, easterly winds prevail and carry pollutants towards the Mediterranean

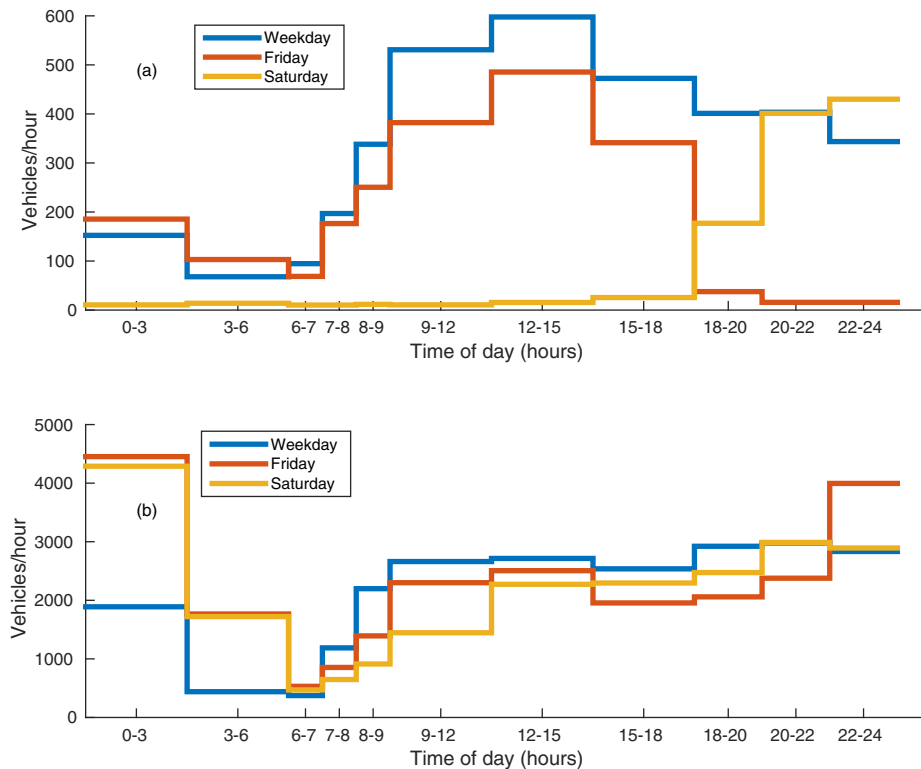


Fig. 5. ATV daily routine private traffic patterns at (a) the orthodox Jewish neighbourhood of Meah Shearim, Jerusalem and (b) a secular nightlife area in Tel-Aviv.

Table 5

ODM cross-validated performance measures for 6 temporal categories in 2012. The table provides the number of observation and modelled value pairs (n), the mean observed concentration value (Mean), the mean bias (MB), mean error (ME), normalised mean error (NME), the ratio of modelled values within factor of two from their corresponding observations (FAC2) and the coefficient of determination (R^2).

Type of period	Type of day	n	Mean (ppb)	MB (ppb)	ME (ppb)	NME (%)	FAC2	R^2
Routine (regular days)	Weekday	176,476	14.9	0.07	5.3	36	0.79	0.59
	Friday	34,956	10.3	0.07	4.1	40	0.76	0.55
	Saturday	34,596	8.3	0.08	3.6	44	0.73	0.52
Vacation (Jewish holidays and summer)	Weekday	60,710	9.9	0.09	4.6	46	0.71	0.36
	Friday	12,723	8.1	0.06	3.8	48	0.68	0.42
	Saturday	12,493	7.0	0.18	4.4	63	0.60	0.06

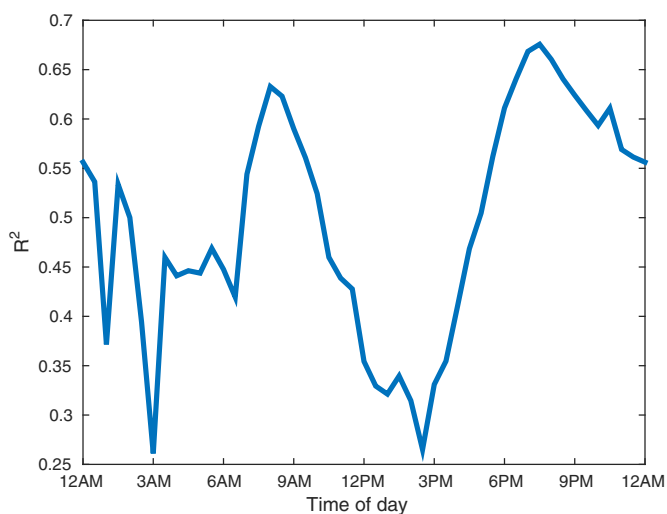


Fig. 6. Daily pattern of cross-validated coefficients of determination between ODM-derived and measured ambient nitrogen dioxide concentrations. Data from all half-hourly time-points and all monitoring stations is included.

while during the day on-land winds are observed. At mid-day, photochemical reaction rates are the highest, causing the lowest NO_2 concentrations away from main roads. For all time-points there are extreme concentrations around coordinates (195,675), which includes a segment of H6 that is reported by the ATV database to have extremely high traffic.

4. Conclusions

The daily patterns captured by the ATV dataset agree with our understanding of traffic patterns in Israel. The data are sensitive to different traffic characteristics during workdays and weekends. The very high correlation between the ATV data and the counts in highway 20 is particularly important, as they were taken in a single week in a different year than the ATV data were collected. This shows that yearly mean ATV traffic data are useful for high-resolution air quality modelling. The spatial patterns also fit well with our expectations, showing higher traffic volumes in grid-cells with primary roads intersecting them. Spatial patterns of truck traffic also fit well to our expectations, with substantially higher ratios of truck-to-private vehicle traffic in industrial areas. The use

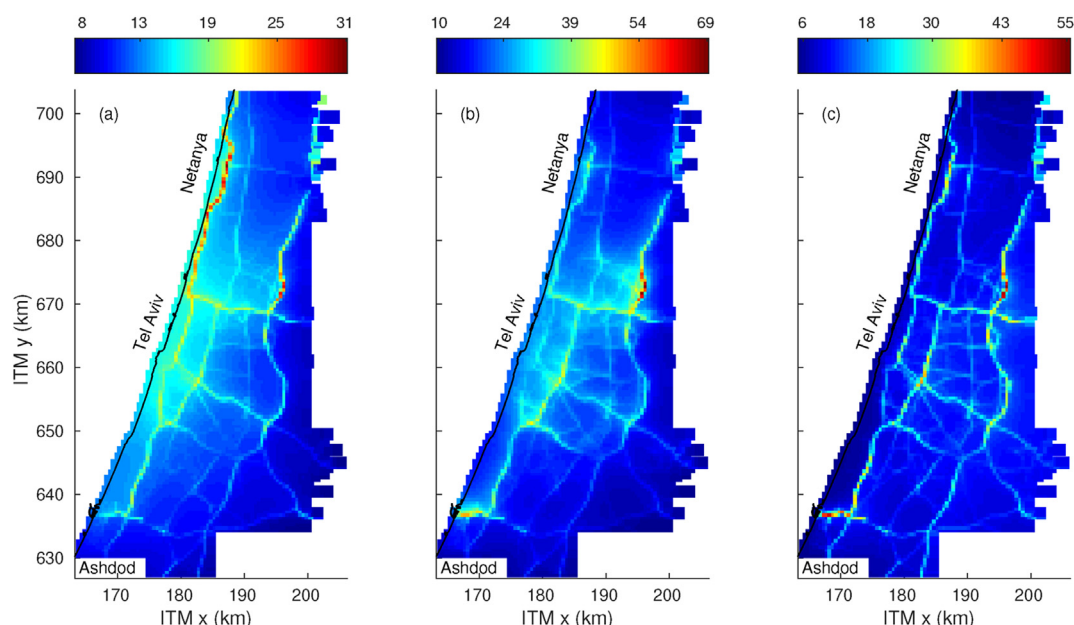


Fig. 7. Maps of NO₂ concentrations (ppb) for routine workdays ($n = 194$) obtained using the ODM for three time points: (a) 03:00, (b) 07:30, (c) 13:00. Color-coding is different for each of the maps.

of OpenStreetMap layers to classify data points by land-use categories could be easily reproduced in other locations. In contrast to trucks, the spatial patterns of bus volumes did not correspond with our expectations and seems to have too high correlations with private vehicle volumes. Hence, we chose not to use the bus data as input to the ODM.

Testing traffic volume conservation along road segments is complex due to the presence of several roads in each grid cell. In our analysis we minimised this effect by focusing on highway segments with only minor agricultural routes in the same grid cell. We found scattering of vehicles volumes to adjacent cells. The scattering is not uniform in space, as some locations suffer from greater scattering around the highway. This inaccuracy of the ATV data likely propagates errors when using them as a proxy of traffic emissions in an air quality model. In particular, a substantial exposure over-estimates may be attributed to individuals who live short distances (few grid cells) away from a highway.

Using the ODM with either ATV data (the current work) or with TA model output (Yuval et al., 2013) resulted in similar performance. Yet, using ATV data has several advantages over using TA model output, among them running the model for additional time windows of the day, thus enabling a higher temporal resolution. However, it should be noted that at times of low traffic (late night, Saturdays and vacations) the model performed not as well as in times of high traffic. We believe that this result reflects other contributions to NO₂ emissions, which in these periods become more dominant than in hours and places with dense traffic. We are currently improving the ODM to include point sources, which is expected to mitigate this issue. Additional advantages of ATV traffic signals over TA traffic estimates are their availability for wide areas and varying road types (rather than just the main roads), and the lack of need to convert line-segment (road) traffic into area sources of emission.

In summary, we have shown that traffic volume estimates obtained from ATV have value as an emission proxy for models that calculate the spatio-temporal distribution of traffic-related air pollutant concentrations. Though we showed this using one model (ODM), we believe that our results are relevant for other air quality

models that can account for vehicular activity, including LUR and CTM.

Acknowledgements

The research was supported by the Technion Center of Excellence in Exposure Science and Environmental Health (TCEEH).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.atmosenv.2016.08.015>.

References

- Antonioni, C., Balakrishna, R., Koutsopoulos, H.N., Oct. 2011. A Synthesis of emerging data collection technologies and their impact on traffic management applications. *Eur. Transp. Res. Rev.* 3 (3), 139–148.
- Bar-Gera, H., Dec. 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. *Transp. Res. Part C Emerg. Technol.* 15 (6), 380–391.
- Barth, M., Boriboonsomsin, K., 2009. Traffic congestion and greenhouse gases. *AC-CESS Mag.* 1 (35).
- Beelen, R., Hoek, G., Fischer, P., van den Brandt, P.A., Brunekreef, B., Mar. 2007a. Estimated long-term outdoor air pollution concentrations in a cohort study. *Atmos. Environ.* 41 (7), 1343–1358.
- Beelen, R., Hoek, G., van den Brandt, P.A., Goldbohm, R.A., Fischer, P., Schouten, L.J., Jerrett, M., Hughes, E., Armstrong, B., Brunekreef, B., Nov. 2007b. long-term effects of traffic-related air pollution on mortality in a Dutch Cohort (NLCS-AIR study). *Environ. Health Perspect.* 116 (2), 196–202.
- Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dédélé, A., Gražulevičienė, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgen, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., Jun. 2013. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. *Atmos. Environ.* 72, 10–23.
- Bekhor, S., Lotan, T., Gitelman, V., Morik, S., Jul. 2013. Free-flow travel speed analysis and monitoring at the national level using global positioning system measurements. *J. Transp. Eng.* 139 (12), 1235–1243.
- Bekhor, S., Shem-Tov, I.B., Apr. 2015. Investigation of travel patterns using passive

- cellular phone data. *J. Locat. Based Serv.* 9 (2), 93–112.
- Borrego, C., Amorim, J.H., Tchepel, O., Dias, D., Rafael, S., Sá, E., Pimentel, C., Fontes, T., Fernandes, P., Pereira, S.R., Bandeira, J.M., Coelho, M.C., Apr. 2016. Urban scale air quality modelling using detailed traffic emissions estimates. *Atmos. Environ.* 131, 341–351.
- Calabrese, F., Lorenzo, G.D., Liu, L., Ratti, C., Apr. 2011. Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Comput.* 10 (4), 36–44.
- Castro, P.S., Zhang, D., Li, S., Jun. 2012. Urban traffic modelling and prediction using large scale taxi GPS traces. In: Kay, J., Lukowicz, P., Tokuda, H., Olivier, P., Krüger, A. (Eds.), *Pervasive Computing*, No. 7319 in *Lecture Notes in Computer Science*. Springer, Berlin Heidelberg, pp. 57–72.
- Central Bureau of Statistics, Mar. 2015. Traffic Counts on Non-urban Roads 2009–2014. Tech. Rep. 1602. State of Israel, Jerusalem.
- Dewulf, B., Neutens, T., Lefebvre, W., Seynaeve, G., Vanpoucke, C., Beckx, C., Van de Weghe, N., 2016. Dynamic assessment of exposure to air pollution using mobile phone data. *Int. J. Health Geogr.* 15, 14.
- Dhondt, S., Beckx, C., Degraeuwe, B., Lefebvre, W., Kochan, B., Bellemans, T., Int Panis, L., Macharis, C., Putman, K., Sep. 2012. Health impact assessment of air pollution using a dynamic exposure profile: implications for exposure and health impact estimates. *Environ. Impact Assess. Rev.* 36, 42–51.
- Djuknic, G.M., Richton, R.E., 2001. Geolocation and assisted GPS. *Computer* 34 (2), 123–125.
- Etyemezian, V., Kuhns, H., Gillies, J., Chow, J., Hendrickson, K., McGown, M., Pitchford, M., Oct. 2003. Vehicle-based road dust emission measurement (III): effect of speed, traffic volume, location, and season on PM10 road dust emissions in the Treasure Valley, ID. *Atmos. Environ.* 37 (32), 4583–4593.
- Fenger, J., Jan. 2009. Air pollution in the last 50 years – from local to global. *Atmos. Environ.* 43 (1), 13–22.
- Gaffney, J.S., Marley, N.A., Jan. 2009. The impacts of combustion emissions on air quality and climate – from coal to biofuels and beyond. *Atmos. Environ.* 43 (1), 23–36.
- Gariazzo, C., Pelliccioni, A., Bolignano, A., Apr. 2016. A dynamic urban air pollution population exposure assessment study using model and population density data derived by mobile phone traffic. *Atmos. Environ.* 131, 289–300.
- Hennig, F., Sugiri, D., Tzivian, L., Fuks, K., Moebus, S., Jöckel, K.-H., Vienneau, D., Kuhlbusch, T.A.J., de Hoogh, K., Memmesheimer, M., Jakobs, H., Quass, U., Hoffmann, B., Mar. 2016. Comparison of land-use regression modeling with dispersion and Chemistry transport modeling to assign air pollution concentrations within the Ruhr area. *Atmosphere* 7 (3), 48.
- Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M., Aug. 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment. *Transp. Res. Part C Emerg. Technol.* 18 (4), 568–583.
- Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., Oct. 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* 42 (33), 7561–7578.
- Janssen, S., Dumont, G., Fierens, F., Mensink, C., Jun. 2008. Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmos. Environ.* 42 (20), 4884–4903.
- Johansson, L., Epitropou, V., Karatzas, K., Karppinen, A., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., Kompatsiaris, I., Feb. 2015. Fusion of meteorological and air quality data extracted from the web for personalized environmental information services. *Environ. Model. Softw.* 64, 143–155.
- Karppinen, A., Kukkonen, J., Elolähde, T., Kontinen, M., Koskentalo, T., Rantakrans, E., 2000. A modelling system for predicting urban air pollution: model description and applications in the Helsinki metropolitan area. *Atmos. Environ.* 34 (22), 3723–3733.
- Lan, C.-J., Li, J., Gu, X., Li, M.-T., Jung, R., Gan, A., 2005. Accuracy Standards and Data Coverage Requirements for Model Validation in FSUTMS. Tech. Rep. BD-432. Florida Department of Transportation.
- Leelőssy Jr., A., Molnár, F., Izsák, F., Havasi, Á., Lagzi, I., Mészáros, R., Aug. 2014. Dispersion modeling of air pollutants in the atmosphere: a review. *Cent. Eur. J. Geosci.* 6 (3), 257–278.
- Liu, H., Chen, X., Wang, Y., Han, S., Aug. 2013a. Vehicle emission and near-road air quality modeling for Shanghai, China. *Transp. Res. Rec. J. Transp. Res. Board* 2340, 38–48.
- Liu, H.-Y., Skjetne, E., Kobernus, M., 2013b. Mobile phone tracking: in support of modelling traffic-related air pollution contribution to individual exposure and its implications for public health impact assessment. *Environ. Health* 12, 93.
- McAllister, S., Chen, J.-Y., Fernandez-Pello, A.C., 2011. Diesel engines. In: *Fundamentals of Combustion Processes*. Mechanical Engineering Series. Springer, New York, pp. 227–241.
- OpenStreetMap Contributors, 2016. OpenStreetMap retrieved 2016-01-19.
- Pratt, G.C., Parson, K., Shinoda, N., Lindgren, P., Dunlap, S., Yawn, B., Wollan, P., Johnson, J., May 2014. Quantifying traffic exposure. *J. Expo. Sci. Environ. Epidemiol.* 24 (3), 290–296.
- Rieser, M., Nagel, K., Beuck, U., Balmer, M., Rümenapp, J., Dec. 2007. Agent-oriented coupling of activity-based demand generation with multiagent traffic simulation. *Transp. Res. Rec. J. Transp. Res. Board* 2021, 10–17.
- Schipper, L., 2008. Automobile fuel economy and CO2 emissions in industrialized countries: troubling trends through 2005–2006. In: *Transportation Research Board 87th Annual Meeting*.
- Shekarzifard, M., Valois, M.-F., Goldberg, M.S., Crouse, D., Ross, N., Parent, M.-E., Yasmin, S., Hatzopoulou, M., Jul. 2015. Investigating the role of transportation models in epidemiologic studies of traffic related air pollution and health effects. *Environ. Res.* 140, 282–291.
- Song, C., Sep. 2000. Diesel emissions and reduction. In: *Chemistry of Diesel Fuels*. CRC Press, pp. 36–41.
- Spearman, C., 1904. The proof and measurement of association between two things. *Am. J. Psychol.* 15 (1), 72–101.
- Wang, L.K., Pereira, N.C., Hung, Y.-T., Li, K.H., 2004. Vehicle air pollution and its control. In: *Air Pollution Control Engineering*, vol. 1. Springer, pp. 446–463.
- Yuval, Bekhor, S., Broday, D.M., Nov. 2013. Data-driven nonlinear optimisation of a simple air pollution dispersion model generating high resolution spatiotemporal exposure. *Atmos. Environ.* 79, 261–270.