Atmospheric Environment 142 (2016) 351–359

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

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
NO_2

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.

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), lavuy@tx.technion.ac.il (Yuval), dbroday@tx.technion.ac.il (D.M. Broday).

http://dx.doi.org/10.1016/j.atmosenv.2016.08.015
1352-2310/

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; Shekarrizfard 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 (Antoniou et al., 2011). For example, Castro et al. (2012) fitted 5000 taxies 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 $\times$ 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 $\times$ 250 m^2, 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