Transportation Research Part D 62 (2018) 699–714

Contents lists available at ScienceDirect

Transportation Research Part D

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

Analyzing spatiotemporal traffic line source emissions based on massive didi online car-hailing service data

Daniel(Jian) Sun[a,b,*], Kaisheng Zhang[c], Suwan Shen[d]

[a] State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Shanghai 200240, China
[b] China Institute of Urban Governance, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Shanghai 200240, China
[c] Shanghai Urban Construction Design & Research Institute (Group) Co., Ltd, No. 3447 Dongfang Road, Pu-Dong New District, Shanghai 200215, China
[d] Department of Urban and Regional Planning, University of Hawaii, Manoa, No. 2424 Maile Way, Honolulu, HI 96822, United States

A R T I C L E I N F O

Keywords:
On-line car-hailing service
NO_x emission
Fuzzy C-means clustering
Spatial analysis
Built environment

A B S T R A C T

Nowadays, the massive car-hailing data has become a popular source for analyzing traffic operation and road congestion status, which unfortunately has seldom been extended to capture detailed on-road traffic emissions. This study aims to investigate the relationship between road traffic emissions and the related built environment factors, as well as land uses. The Computer Program to Calculate Emissions from Road Transport (COPERT) model from European Environment Agency (EEA) was introduced to estimate the 24-h NOx emission pattern of road segments with the parameters extracted from Didi massive trajectory data. Then, the temporal Fuzzy C-Means (FCM) Clustering was used to classify road segments based on the 24-h emission rates, while *Geographical Detector* and *MORAN's I* were introduced to verify the impact of built environment on line source emissions and the similarity of emissions generated from the nearby road segments. As a result, the spatial autoregressive moving average (SARMA) regression model was incorporated to assess the impact of selected built environment factors on the road segment emission rate based on the probabilistic results from FCM. It was found that short road length, being close to city center, high density of bus stations, more ramps nearby and high proportion of residential or commercial land would substantially increase the emission rate. Finally, the 24-h atmospheric NO_2 concentrations were obtained from the environmental monitor stations, to calculate the time variational trend by comparing with the line source traffic emissions, which to some extent explains the contribution of on-road traffic to the overall atmospheric pollution. Result of this study could guide urban planning, so as to avoid transportation related built environment attributes which may contribute to serious atmospheric environment pollutions.

1. Introduction

With the increasing development of motorization and urbanization, transportation has gradually become one of the main sources of air pollution in many cities throughout the world. For example, according to statistics, on-road transportation accounts for about 47% of CO, 42% of NO_x, and 18.4% of PM emissions, respectively in Europe (Kousoulidou and Sarmaras, 2008); and 27.5% of total greenhouse gas in U.S., in which 41.9% results from passenger cars (Mausami and Weitz, 2017). It also contributes to about 19% of

* Corresponding author at: State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Shanghai 200240, China.
E-mail address: danielsun@sjtu.edu.cn (D. Sun).

https://doi.org/10.1016/j.trd.2018.04.024

nitrogen oxide emissions in Canada, rising up to 85% in Montreal (IJC, 2012). In China, it is estimated that transportation related activities account for approximately 46% of NO_x, 78% of CO, and 83% of HC in metropolises (Wang et al., 2008). Consequently, it's important to identify the hotspot and the mechanism of on-road emissions to provide guidance for built environment and land use during the planning stage.

Traditional emission measurements are based on laboratory-similar bench test (Silva et al., 2015), remote sensing (Ning and Chan, 2007), and tunnel test (Zhang et al., 2015). However, the emitting characteristics of motor engines extracted from these studies were only applicable to experimental environment. They are expensive and impossible to be used for dynamic monitoring or multi-vehicle test. In recent years, the portable emission measurement system (PEMS) has been widely used in mobile monitoring, which is easy to assemble. Defries et al. (2014) found that PEMS was more compatible to chassis testing than other mobile methods in detecting common pollutants. Gallus et al. (2016) also disclosed that the correlation tests between particle number PEMS and chassis dynamometer tests are in good agreement. However, a set of PEMS could just be used for one-vehicle at one time, taking months or even years to complete the regional data collection.

In such case, simulation models and detailed estimation formulas are developed to simplify the emission estimation. For example, in U.S., Environmental Protection Agency (EPA)'s MOtor Vehicle Emission Simulator (MOVES) is a state-of-the-science emission modeling system that estimates emissions for mobile sources at the national, county, and project level for criteria air pollutants, greenhouse gases, and air toxics. The emission calculation is based on predefined speed bins and vehicle specific power (VSP) distribution, which is suitable from microscopic to macroscopic. Abou-Senna et al. (2013) integrated MOVES and VISSIM microscopic simulation to predict emissions from vehicles on limited-access highways. Liu et al. (2013) input vehicle operational parameters extracted from GPS trajectories into MOVES for estimating vehicle emission factors in Shanghai. International Vehicle Emission (IVE) model is also a VSP-based algorithm, which generated 60 VSP bins to calculate emissions. The model was evaluated by utilizing an available dataset of remote sensing measurements on a large number of vehicles in Hangzhou, China, and found that it performed well on estimating HC and CO emissions (Guo et al., 2007). Nagpure et al. (2011) applied the fundamental equations of IVE model to estimate the emission rates from the light-duty commercial vehicles, and revealed the increases of CO and VOCs and a decrease of NO_x emission rate along with altitude. Computer Program to Calculate Emissions from Road Transport (COPERT) is another widely used emission model developed by European Environment Agency (EEA) using speed as the main independent variable (Broderick and O'Donoghue, 2007; Lozhkina and Vladimir, 2015). Since COPERT complies with the European emission standard which is similar to the one adopted in China, the package is capable for emission estimation in Chinese cities (Lang et al., 2016). Bellasio et al. (2007) implemented COPERT to calculate the emission inventory of Sardinia, Italy, and proved its reliability by comparison with field experiments. However, all above referred models require sufficient data when applied to a citywide level, which is unfortunately difficult to be obtained in practical.

On the country, vehicle-based GPS data as a source of driving trajectory record linking to the estimation and influencing mechanism of vehicular emissions, have gained increased attentions (Liu et al., 2009), which covers city wide road network continuously and consequently is believed more accurate. However, studies based on data from the new emerging car-hailing mode are scarce, while most similar studies were focused on taxi vehicle trajectories. Wang et al. (2008) utilized GPS to collect the driving pattern data including 140 taxis, and combined them with IVE to estimate the total emissions of CO, VOC, $NO_{x,}$ and PM from vehicles in Shanghai. However, this study was just of coarsely resolved data collection with limited vehicle number and land scope. Nyhan et al. (2016) obtained GPS trajectory data from over 15,000 taxis in Singapore, which enabled to quantify the instantaneous driving cycle parameters in high spatio-temporal resolution with the input of real-time emission model in calculating CO_2, NO_x, VOCs, and PM. Luo et al. (2016) extracted speed information from taxi GPS data of Shanghai to quantify spatial and temporal hotspot emissions using COPERT model, disclosing a distribution of dual-core cyclic structure. However, their study failed to distinguish vehicles of different type and emission standard, and the final the output contains solely plotting-based figures.

The existing literatures about on-road emissions mainly were associated with three gaps as follows. First, the majority of studies requires huge amount of survey data as the input for the simulation stage. Since some research are in a rather large scale, the temporal difference was neglected, by just analyzing day, month or year as a whole. Second, most studies just estimate emissions from the original GPS sources without further reckoning on various vehicle types and emission standards. Finally, the results are mainly qualitative plotting or simple statistics without explanation with additional information such as land use and built environment. In fact, from the perspective of sustainable human settlements, it's necessary to explore the correlation between the road emissions and the air pollution concentrations, which should be the ultimate objective in analyzing on-road emissions.

The purpose of this study is to investigate the 24-h line source emission pattern of the road network so as to analyze the problematic segments with continuous or unconventional high emissions. Formulas from COPERT were selected to estimate NO_x emission of different vehicle types and emission standards. The geographical methods including *geographical detector*, *MORAN's Index* and spatial regression were applied to analyze the spatial-temporal characteristics, as well as impacts from the built environment factors in the vicinity. In comparison, the previous study of the same authors (Zhang et al., 2017) has focused on discussing the spatiotemporal distribution and the cause of road congestion based on taxi GPS trajectories, revealing how land use and built environment affect road congestions. The main objective is to propose a modeling frame to combine Fuzzy C-Means (FCM) clustering and spatial regression, while this paper mainly focuses on road emissions, and utilizes COPERT model to calculate emission factor of road traffic, which is relatively innovative for the transportation big data analytics. One of the key innovation is to approximate the traffic flow composition for all roads by incorporating the surveyed results from 40 selected segments (both primary and secondary arterials), which extends the applicability of the sampled Didi on-line car-hailing data. Then, the 24-h trend pattern of spatial correlation between road emissions and air pollution concentration was further analyzed to emphasize the spatial and temporal relationship of NOx traffic emission to the overall atmospheric air quality. More specifically, the two sub-objectives of this study are