Transportation Research Part D 65 (2018) 540–550

Contents lists available at ScienceDirect

Transportation Research Part D

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

# How much vehicle activity data is needed to develop robust vehicle specific power distributions for emission estimates? A case study in Beijing

Zhiqiang Zhai[a], Guohua Song[a,*], Lei Yu[b]

[a] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China
[b] Texas Southern University, 3100 Cleburne Avenue, Houston, TX 77004, USA

## 1. Introduction

Traffic emissions have become one of the main sources of air pollution in China's urban areas in recent years. In order to inform emission reduction policymakers, Beijing has been developing a system for real-time traffic emission estimates at the link level since 2012, employing the methodology framework of the Motor Vehicle Emission Simulator (MOVES) model. This is not an easy task as there are over 5.6 million vehicles registered in Beijing (data through the end of 2017), and their emissions vary widely depending on source types (vehicle types, ages, emission standards, etc.) and vehicle activities. The MOVES model utilizes the concept of the vehicle specific power (VSP) (Jiménez-Palacios, 1999) binning method to develop emission rates that reflect the emission characteristics of each source type and to develop VSP distributions (i.e. operating mode distributions) that represent the average vehicle activities on roads (U.S. EPA, 2016). The VSP distributions are defined as the time fractions spent in various power demands. On-road emissions of each source type are calculated by multiplying emission rates by VSP distributions.

However, large errors in emission estimates can result when using the default VSP distributions in MOVES, especially in non-U.S. cities (Alam and Hatzopoulou, 2016). Hence, it is recommended that instrumented vehicles be employed to collect local vehicle activity data to develop specific VSP distributions for each city, as described in the U.S. Environmental Protection Agency (EPA) guidelines (U.S. EPA, 2016). Over the past couple of years, Beijing collected hundreds of millions of GPS data points for light duty vehicles (LDVs), buses, and trucks and developed a comprehensive database of VSP distributions in order to model the on-road vehicle activities for emission estimates. China now plans to utilize this emission estimate system in other cities; thus, for each city, a large amount of vehicle activity data needs to be collected in order to develop the database of VSP distributions.

In this context, modelers face an important question: How much vehicle activity data is needed to develop robust VSP distributions for emission estimates? Collecting vehicle activity data is expensive and time-consuming and this can present major budget challenges. If the VSP distributions are developed with insufficient data and, thus, not robust enough. This study aims to address this question through analysis of the VSP distributions developed by using vehicle activity data of various sample sizes.

## 2. Literature review

Vehicle activities are one of the most critical inputs for calculating traffic emissions. Typically, three methods are used to represent the average vehicle activities on roads: driving cycles, second-by-second speed profiles, and VSP distributions. Existing emission models, such as MOBILE (U.S. EPA, 2000), EMFAC (U.S. CARB, 2006), COPERT (Ntziachristos et al., 2009), and HBEFA (Haan and Keller, 2004), and multiple studies (Bishop et al., 2012; Duarte et al., 2016; Kamble et al., 2009; Tong et al., 1999; Wang et al., 2008; Yu et al., 2008) in recent decades have employed driving cycles related to traffic situations (e.g. facility types, average

* Corresponding author.
*E-mail address:* ghsong@bjtu.edu.cn (G. Song).

https://doi.org/10.1016/j.trd.2018.09.004

speeds, and levels of service) to represent vehicle activities for emission estimates. This method is often questioned because there are problems with using driving cycles to represent real-world vehicle activities (Pelkmans and Debal, 2006). The emission models of VT-Micro (Rakha et al., 2004) and CMEM (Barth et al., 2000) use second-by-second speed profiles to estimate instantaneous emissions. This method is suitable only under simulated conditions because, in the real world, it is not possible to obtain the speed profiles of all vehicles on a link. The MOVES (U.S. EPA, 2016) and IVE (U.S. EPA, 2008) models categorize vehicle activities into various bins of power demands and employ the VSP distributions to represent the average vehicle activities.

However, in MOVES, the default VSP distributions, which are developed with 49 driving schedules, contribute to uncertainties in the resulting emission estimates when running MOVES directly. Some modelers (Alam and Hatzopoulou, 2016; Chamberlin et al., 2013; Talbot et al., 2014) calibrated the VSP distributions obtained from real-world data with the defaults in MOVES and indicated that they were significantly different and that the emissions derived using the local VSP distributions are significantly improved. Thus, it is recommended to replace the MOVES defaults with locally developed VSP distributions.

Three data sources have been frequently used to develop the VSP distributions for traffic emission estimates: (1) outputs of traffic simulation models, such as VISSIM, PARAMICS, and INTEGRATION (Abou-Senna et al., 2013; Abou-Senna and Radwan, 2013; Wang et al., 2017, 2015; Xie et al., 2012; Xu et al., 2016); (2) driving simulators (Li et al., 2015a; Sun et al. 2016; Zhao et al. 2015); and (3) real-world data (Bar-Gera, 2007; Li et al., 2015b; Wei et al., 2012; Yao et al., 2013; Song and Yu, 2009). Studies have often questioned whether the driving behaviors in traffic simulations differ from real-world situations (Ko et al., 2018; Song et al., 2015). The advancement of GPS technology has made big data for vehicle activities in the real world available and affordable in recent years. Thus, many researchers have employed GPS devices to collect vehicle activity data in the real world and have developed local VSP distributions to embed into the MOVES database for emission estimates (Alam and Hatzopoulou, 2016; Sentoff et al., 2015). Frey et al. (2006) compared the real-world vehicle speed profiles for specific road types, revealing that there are often significant similarities when such profiles are grouped with respect to average speed. Thus, VSP has been categorized into corresponding road types and average speeds to develop facility- and speed- specific VSP distributions in follow-up studies (Frey et al., 2017; Frey and Zhang, 2007; Tanvir et al., 2018). Song et al. (2012) developed speed- specific VSP distributions for urban restricted access roads and indicated that the VSP distributions are similar to the normal distribution. Zhai et al. (2017) validated the temporal and spatial consistency of facility- and speed-specific VSP distribution. In many studies, the VSP binning method, which is defined by an equal VSP interval of 1 kW/ton, as shown in Eq. (1), was often employed for the purpose of reducing both the aggregation errors and the computational complexity.

$$\forall\ VSP \in [n, n+1),\quad VSP\ \ Bin = \#n,\quad n\ \ \text{is}\ \ \text{an}\ \ \text{integer} \tag{1}$$

However, the majority of researchers have analyzed the characteristics of VSP distributions on the assumption that the data samples used in their studies were “enough”, and some were not concerned about sample size. Basically, the VSP distribution is developed by counting the number of VSP values in each bin. The VSP distributions developed by using 300 s of vehicle activity data, for example, are significantly different than those developed by using 12,000 s of data, as illustrated in Fig. 1. Large errors can result from the use of small sample sizes for VSP distributions when conducting emission estimates.

This study collected over 30 million second-by-second GPS data points for LDVs and developed the facility- and speed-specific VSP distributions by using various sample sizes. The $CO_2$, CO, $NO_X$, and HC emissions derived from the VSP distributions of each sample size were calculated and compared to determine how much data is needed to develop robust VSP distributions for given estimated errors.

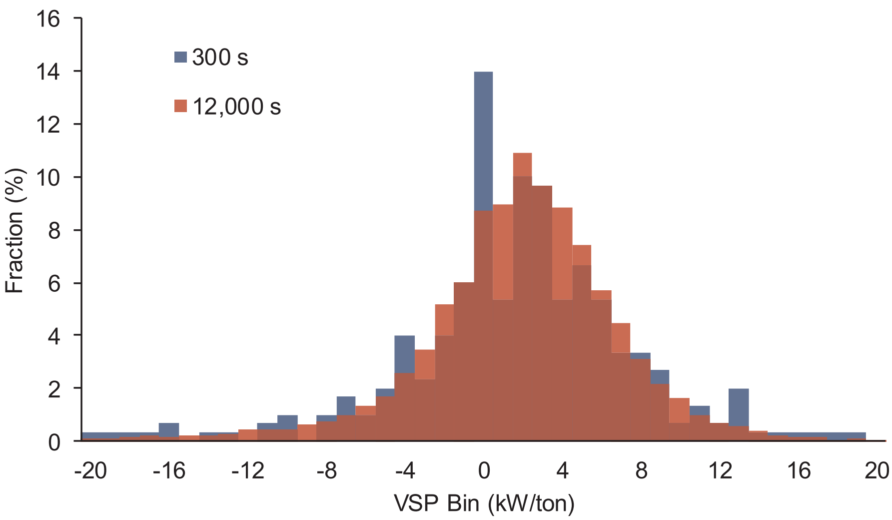


**Fig. 1.** VSP distributions developed by using 300 s and 12,000 s of data.