UC Riverside

2017 Publications

Title

A comparative study on data segregation for mesoscopic energy modeling

Permalink

https://escholarship.org/uc/item/43p069h9

Journal

Transportation Research Part D: Transport and Environment, 50

ISSN

13619209

Authors

Li, Weixia
Wu, Guoyuan
Zhang, Yi
et al.

Publication Date

2017

DOI

10.1016/j.trd.2016.10.029

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Transportation Research Part D 50 (2017) 70–82

Contents lists available at ScienceDirect

Transportation Research Part D

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

A comparative study on data segregation for mesoscopic energy modeling

Weixia Li [a], Guoyuan Wu [b,*], Yi Zhang [a], Matthew J. Barth [b]

[a] Department of Automation, Tsinghua University, Beijing 100084, China
[b] Bourns College of Engineering – Center for Environmental Research & Technology (CE-CERT), University of California, Riverside, 92507 CA, USA

A R T I C L E I N F O

Article history:
Available online 9 November 2016

Keywords:
Mesoscopic energy consumption model
Data segregation
Operating mode

A B S T R A C T

On-road vehicles have been considered as one of the major contributors to energy consumption and air pollutant emissions. In order to quantify the corresponding environmental impacts, great efforts have been dedicated to the microscopic and macroscopic modeling for vehicle energy consumption and emissions. However, the mesoscopic modeling research that is focused on estimating trip-based energy consumption and is critical to some ITS applications (e.g., environmentally-friendly navigation), is relatively deficient. This study aims to investigate the effects of different data segregation methods on the mesoscopic modeling for vehicle energy consumption. A variety of novel methods, including the so-called conditional operating mode based method, have been proposed and evaluated using field data. Based on real-world data, statistical analyses have demonstrated the superior performance of enhanced models (i.e., conditional operating mode/VSP based models) in estimating vehicle energy consumption on a trip basis, compared to the other four models (velocity binning, time snipping, distance snipping and VSP based models) tested in this study.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Transportation activities have been contributing to a significant amount of energy consumption and emissions in daily lives. According to China Vehicle Emission Control Annual Report 2015, vehicle population has increased up to 246 million in China in 2014, resulting in over 45 million tons of pollution annually, including more than 6 million tons of nitrogen oxides (NOx), around 0.6 million tons of particulate matter (PM), 4 million tons of hydrocarbon (HC) and 34 million tons of carbon monoxide (CO) emissions (MEPPRC, 2015). In the United States, the transportation sector accounted for 26.5% of greenhouse gas (GHG) emissions in 2014, where CO_2 emissions increased by 16% from 1990 to 2014 (USEPA, 2016a). Such ever-increasing pressure on the environment has raised worldwide awareness, and therefore numerous studies have been dedicated to reducing vehicle energy consumption and emissions (e.g., via the development of intelligent transportation systems or ITS). Among these studies, one of the critical issues for evaluating the effectiveness of different ITS applications or strategies is how to accurately model the energy consumption and emissions.

Over the years, a number of energy consumption and emission models have been proposed across various levels of temporal resolution for different purposes. Some models used in evaluating network-wide energy consumption and emissions are classified as macroscopic models, represented by MOBILE (EPA, 2002) and EMFAC (CARB, 2016). These models present

* Corresponding author.
E-mail addresses: wx-li11@mails.tsinghua.edu.cn (W. Li), gywu@cert.ucr.edu (G. Wu), zhyi@mail.tsinghua.edu.cn (Y. Zhang), barth@cert.ucr.edu (M.J. Barth).

http://dx.doi.org/10.1016/j.trd.2016.10.029
1361-9209/© 2016 Elsevier Ltd. All rights reserved.

outstanding performance in estimating region-wide emission inventories according to aggregated network parameters, but may not be suitable for evaluating operational effects of many Intelligent Transportation Systems (ITS) at the vehicle level. On the other hand, microscopic models which relate energy consumption and emissions with vehicle activities (e.g., second-by-second speed) prove to be much more accurate. Examples of microscopic models include the Comprehensive Modal Emission Model (Barth et al., 2000), VT-Micro fuel consumption and emission model (Ahn et al., 2002), and Passenger car and Heavy duty Emission Model (Hausberger et al., 2002). Besides the macroscopic and microscopic models, U.S. EPA's latest emission model, the Motor Vehicle Emission Simulator or MOVES (USEPA, 2016b), is a comprehensive multi-scale estimating model which can take vehicle operating modes (at the microscale level) into consideration, and aggregate them spatially and temporally to provide macroscale energy and emission estimates.

However, none of the models stated above can be directly applied to the occasions where estimates of link-based or trip-based energy consumption and emissions for a specified vehicle rely on some mesoscopic parameters, such as average link velocity. A notable example is the so-called environmentally-friendly navigation whose objective is to find the route with least energy consumption or pollutant emissions based on real-time traffic information as well as road characteristics, such as gradient (Barth et al., 2007). Another potential application is personalized vehicle refueling assistance which can be used to warn drivers to refuel in time based on trip-based fuel consumption estimation/prediction (even with driver behavior modeled implicitly). From the perspective of traffic operators and policy makers, such mesoscopic parameter based models are very useful to estimate/predict region-wide fuel consumption of motor vehicles based on aggregated traffic conditions such as link (regarded as small trips) average speed. For these applications, neither macroscopic nor microscopic models are suitable for energy consumption estimation/prediction. This is especially true for microscopic models which depend highly on second-by-second trajectory data. In practice, this type of data is very costly to obtain for all vehicles, even along a corridor, not to mention for a city or at a larger regional level. However, the aggregated traffic information such as link average speed may be readily available for a large scale network from real-time traffic information providers (e.g., INRIX, Google). Therefore, researchers have developed mesoscopic models which estimate link-level energy and emission factors by aggregating vehicles' operating states over trajectory snippets (Kang et al., 2011; Boriboonsomsin et al., 2012; Yao and Song, 2013). Nevertheless, performance of such models may vary significantly with the way to segregate vehicle trajectories into snippets, e.g., by distance, by duration, or by operating state. To the authors' best knowledge, there is no study to systematically evaluate the impacts of different segregation methods on the model effectiveness. To address this issue, this paper proposes a framework for comparative evaluation of the effects of various trajectory segregation strategies including the one by Operating Mode (by conditioning on speed) which is first proposed in this study. In addition, validation is conducted with the real-world data collected in Beijing, China. Probe vehicles were used to collect second-by-second trajectory data to construct the proposed mesoscopic model in the following manner: (1) probe vehicles' second-by-second fuel consumption was estimated by using a microscopic model (e.g., CMEM in this study); and (2) both fuel consumption and activity data of probe vehicles were aggregated to relate the trip-based (or segment-based) fuel consumption with average travel speed along the segment which is assumed to be the surrogate of real-world traffic states (e.g., average speed that can be available from INRIX, Google).

The rest of this paper is organized as follows: Section 2 elaborates the details in methodology, including research framework, data collection and processing efforts, model parameter selection, and segregation methods. The description of mesoscopic models based on different data segregation methods is presented in Section 3, followed by statistical analyses on the models' performance. The last section concludes this study along with some discussion.

2. Methodology

As aforementioned, mesoscopic energy and emission models are critical for some major ecological ITS (e.g., eco-routing navigation). A state-of-the-practice procedure to establish such models is illustrated in Fig. 1: (a) The activity data (i.e., second-by-second speed profiles) of GPS-instrumented vehicles are collected from the real-world experimentation; (b) Based on the latitude, longitude and time-stamp information, a spatial and temporal matching is conducted to associate with roadway characteristics (e.g., road type, gradient) and traffic conditions (if available); (c) The speed profiles are broken down into snippets based on certain segregation rule (e.g., by some spatial or temporal span, by speed level); (d) For each snippet, the overall vehicle-specific energy consumption and pollutant emissions (normalized by travel distance), and some aggregate driving statistics (the most widely used is average speed) are then calculated. Depending on the data availability, the energy consumption and pollutant emissions may be estimated from the measurement by on-board sensors or outputs by the calibrated microscopic energy and emission models. Also, the average speed may be evaluated by the speed profiles of testing vehicles or information of other traffic; (e) The mapping of average speed onto normalized energy/emissions (i.e., the mesoscopic model) is developed across a variety of driving scenarios, such as road type and road grade. At the heart of mesoscopic model development, the segregation method may significantly affect the model performance. Therefore, this study is focused on the design of different segregation methods and the comparison of their impacts on model performance.

2.1. Research approach

Before the elaboration of various segregation methods, the research framework for this study is depicted in Fig. 2. The whole framework is composed of three stages, i.e. data collection, method development and impact evaluation. In the data

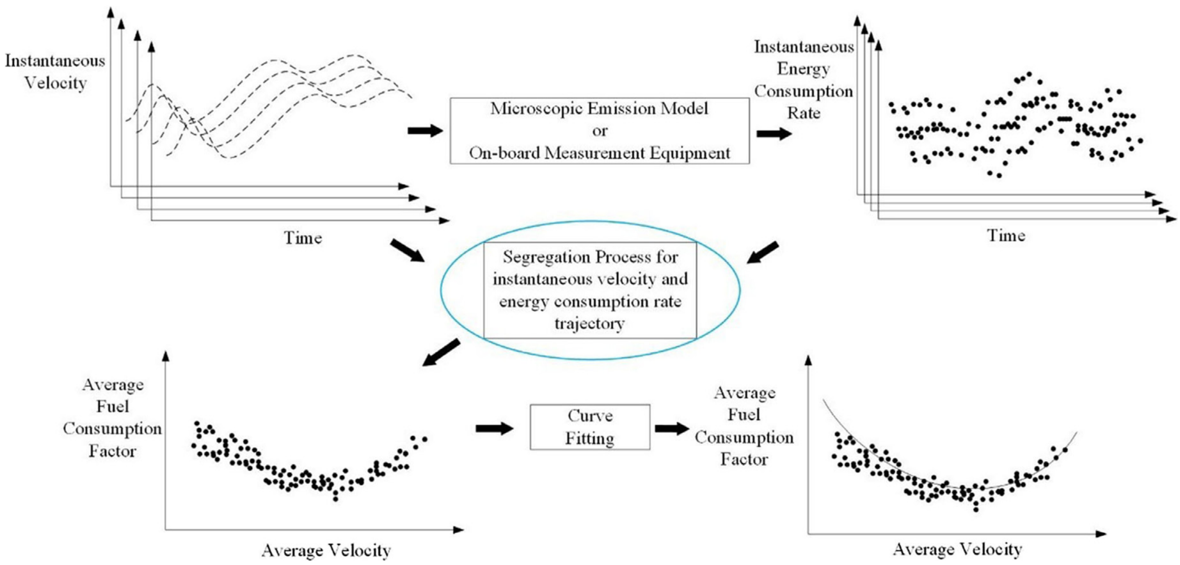

Fig. 1. Typical mesoscopic energy modeling approach.

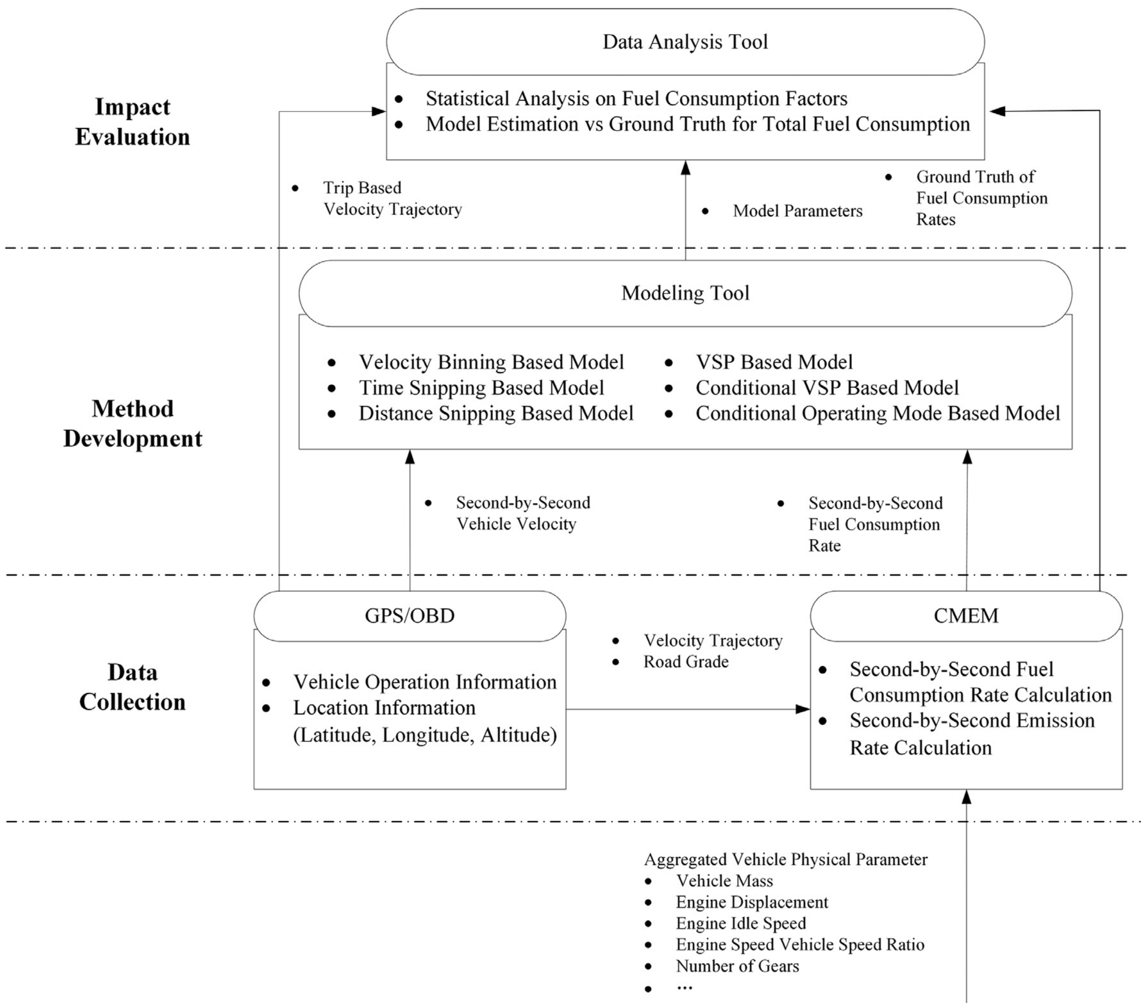

Fig. 2. Research framework for evaluating the impacts of snippet segregation methods.

collection part, real world activity data were collected using probe vehicles instrumented with GPS data loggers and on-board diagnostics (OBD) systems. The probe vehicles were driven on road by following the traffic, which ensured the collected data well represent the general traffic states. The microscopic emission model, CMEM, which has been calibrated was adopted to estimate the corresponding fuel consumption. In the method development part, six snippet segregation methods were developed via MATLAB to establish mesoscopic fuel consumption models based on second-by-second vehicle velocity and fuel consumption rate. When evaluating the effectiveness of different methods, the Cross Validation technique was applied where the dataset was partitioned into training portion and testing portion in a rotational manner and the statistical analysis was performed on the samples from all partitions (Picard and Cook, 1984). More specifically, dataset from each test run was used as the validation set while the remaining datasets were used as the training set based on leave-one-out cross-validation (LOOCV), and this was repeated on all ways in the partitioning of the original dataset.

2.2. Road test and data collection

In this study, road tests were carried out along the 4th Ring Road in Beijing (highlighted in bold in Fig. 3), one of the busiest expressways in the downtown area which highly represents the typical traffic patterns during the associated time periods. The length of road segment is around 65 km, and the designed speed is 80 km/h.

In order to take full consideration of the temporal variation of traffic conditions, 22 test runs during both off-peak hours (e.g., 21:00–23:00) and peak hours (e.g., 15:00–19:00) were conducted between October 2013 and December 2013. And Fig. 4 presents two typical traffic velocity patterns collected for off-peak hour (subfigure on the top) and peak hour (subfigure on the bottom). For all the road tests, the second-by-second velocity and location information (including latitude, longitude and altitude) were collected. It is noted that the test route is quite flat in this study, therefore the road grade effect was neglected when modeling the fuel consumption rate. However, when applying the methodology to areas with significant variations in road terrain, road grade can be modeled as another independent variable, as detailed in the authors' previous work (Barth et al., 2007).

2.3. Fuel consumption rate estimation

As stated above, the comprehensive modal emission model (CMEM) is adopted to estimate vehicle fuel consumption rate. CMEM is a microscopic model designed to predict second-by-second tailpipe emissions and fuel consumption for a wide variety of vehicle/technology categories. This model predicts emissions and fuel consumption according to specific physical process, and the associated analytical functions that relate the vehicle's driving states with energy/emissions have been developed based on a large amount of real-world testing data. CMEM was developed and validated using an extensive data set of second-by-second vehicle trajectory, fuel consumption, and emission data from over 340 vehicles. It classifies vehicles into 24 technology categories, and aggregated fuel consumption parameters were quantified for each category (Barth et al., 2000). This vehicle/technology aggregation is necessary since predicting fuel consumption for a vehicle class rather than an individual vehicle is usually more useful in practical applications. Therefore, energy consumption estimated using CMEM is capable of representing the generalized characteristics for aggregated vehicle/technology family. It is noted that the CMEM model may not cover all the newest vehicle categories and technologies (e.g. improved closed loop engine control). However, this may not be a critical issue in this study since an established vehicle category/technology is selected for the comparisons of fuel consumption.

The second-by-second vehicle fuel consumption estimation in CMEM is based on a physical, power-demand modal modeling approach, in which the entire fuel consumption process is decomposed into different steps, such as tractive power estimation, powertrain parameter calibration, and engine power calculation, whose analytical representations have been developed to capture the associated characteristics. The modeling process used in CMEM is deterministic rather than descriptive. It is based on casual parameters or variables (such as engine size and aerodynamic drag coefficient) rather than based on simply observing the effects and assigning them to statistical bins. Therefore, CMEM provides more understandings for the variations in fuel consumptions among vehicles, types of driving and other conditions. For detailed information about CMEM, please refer to Barth et al. (1996).

2.4. Modeling parameter selection

Previous studies indicate various exogenous factors, such as traffic conditions, roadway characteristics, weather, and even driver behavior may affect the vehicle fuel consumption rate, but velocity and acceleration are believed to be two decisive ones (Barth et al., 1996; Ahn et al., 2002). Preliminary analysis has been conducted on the dataset to evaluate the influence of velocity and acceleration on fuel consumption rate. As depicted in Fig. 5, the fuel consumption rate goes up when velocity and acceleration increase. In the deceleration process, little fuel is consumed as expected.

In order to include velocity and acceleration patterns in the modeling process to reflect comprehensive effects of the two parameters, several strategies are proposed and evaluated in this study, inspired by the snippet analysis in previous studies (Boriboonsomsin et al., 2012; Yao and Song, 2013). In the aforementioned research, the time based snipping was conducted for data aggregation, and vehicle specific power (VSP), a parameter which has been proved to be highly related to vehicle energy consumption (Jimenez-Palacios, 1998) and widely used in characterizing energy consumption features, was consid-

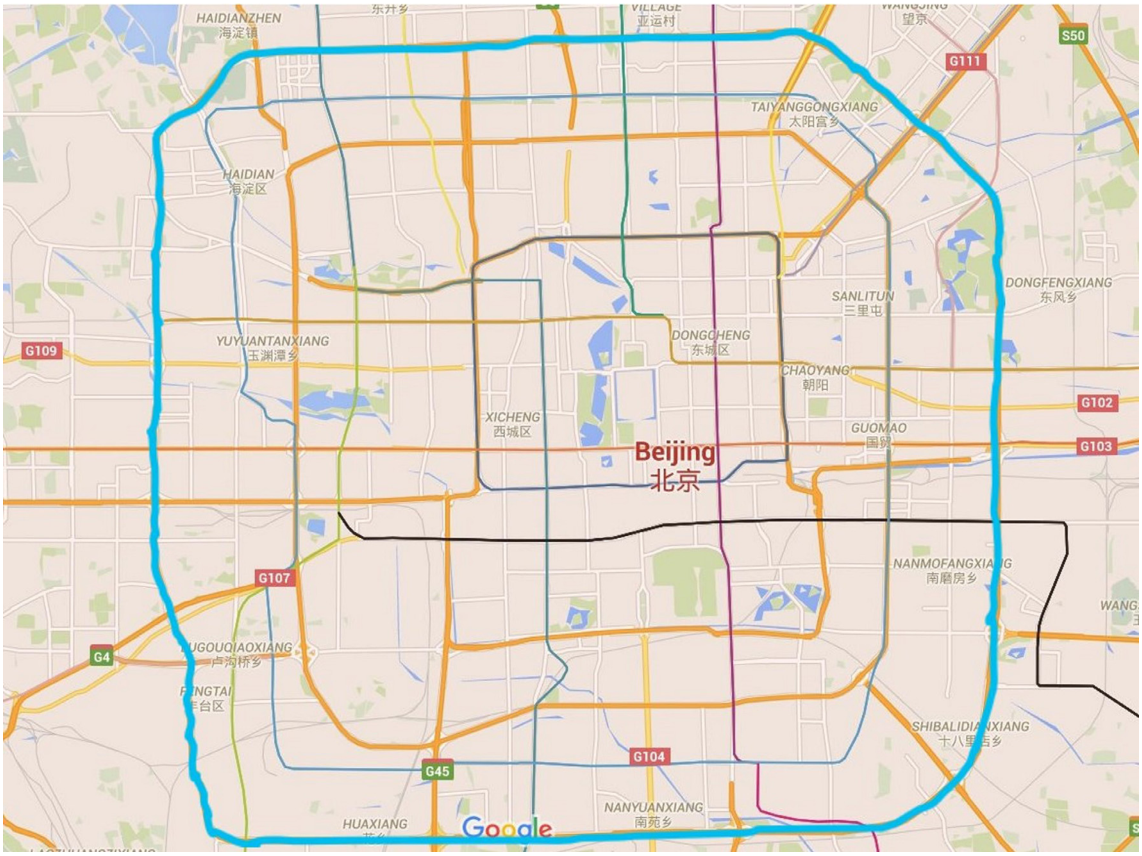

Fig. 3. Road test scenario – the 4th Ring Road in Beijing.

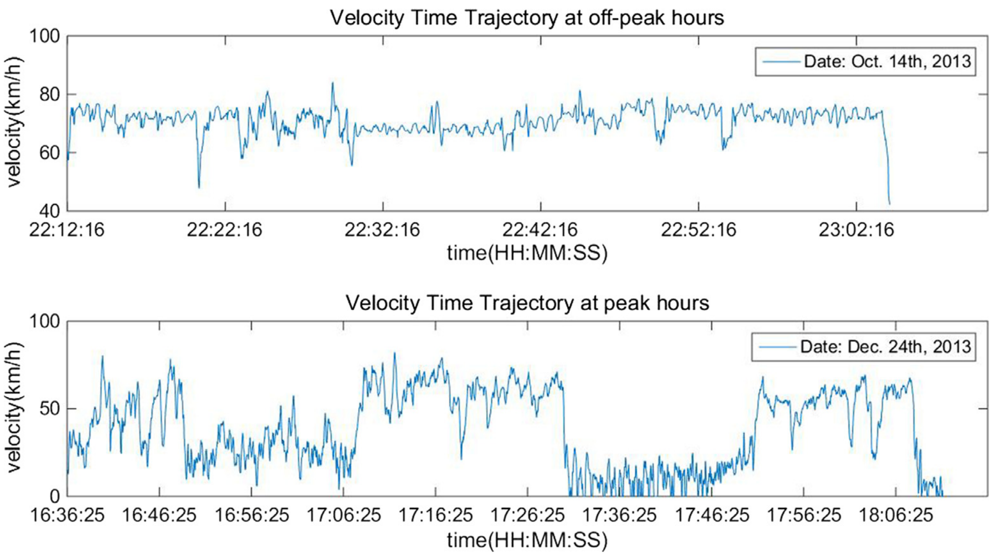

Fig. 4. Typical traffic velocity patterns of Beijing Ring Road.

ered in aggregation process in the latter one. For comparative purpose, these two methods are both applied to the dataset collected in this study. In addition, different segregation methods including basic velocity binning and new explanatory variables including the popular parameter "operating mode" used in USEPA's latest emission model (Koupal et al., 2003), are explored in the study. Major parameters related to both velocity and acceleration in this study include: (1) snippet average

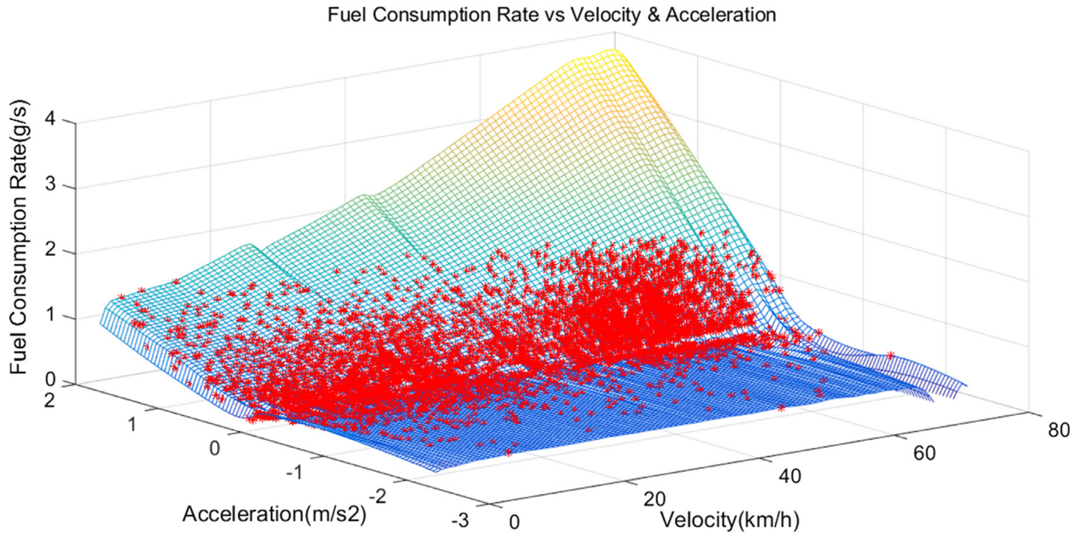

Fig. 5. Distribution of fuel consumption rate verses velocity and acceleration.

velocity, (2) VSP, and (3) operating mode (OpMode). The first one is the aggregated velocity measurements of smaller snippets that contain continuous vehicle operation information for a specified time or distance in each trip. The other two are directly defined as functions of velocity and acceleration. For example, VSP is defined as the instantaneous power per unit mass of the vehicle and it can be calculated by the instantaneous velocity and acceleration. Since power is always directly related to fuel consumption, VSP is believed to be effective in revealing vehicle fuel consumption characteristics. For light duty gasoline vehicles, the simplified formula used in calculating VSP is presented in Eq. (1) (Jimenez-Palacios, 1998).

$$\mathrm{VSP} = v * (1.1 * a + 0.132) + 0.000302 * v^3 \tag{1}$$

Operating mode is another parameter highly related to vehicle instantaneous power (USEPA, 2016b). It is defined on the basis of VSP, speed and acceleration. The concept of operating mode is used in the state of the art emission simulator MOVES, which is proved to be effective in aggregating vehicle fuel consumption and emissions. However, for mesoscopic mathematic modeling, operating mode based method is adopted for the first time in this study. The definition of operating mode is listed in Fig. 6.

2.5. Snippet segregation methods

Based on the homogeneity of key metrics, six methods were developed and compared in this study: (1) velocity binning; (2) time-based snipping; (3) distance-based snipping; (4) vehicle specific power based snipping; (5) conditional VSP based snipping; and (6) conditional operating mode based snipping. More detailed description of each segregation method is presented in the following section, and the notations to be used are listed hereinafter.

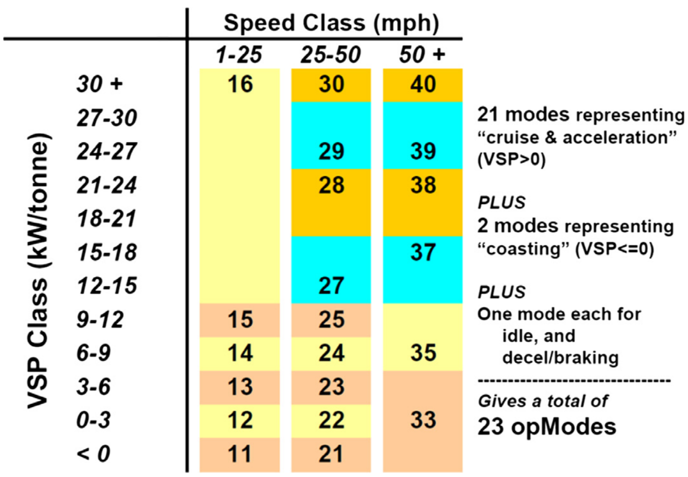

Fig. 6. Overview of VSP based Operating Mode (Warila et al., 2011).

v_j is the jth instantaneous velocity in the velocity array, km/h.
a_j is the jth instantaneous acceleration in the acceleration array, m/s^2.
VSP_j is the vehicle specific power corresponding to v_j and a_j, KW/tonne.
$OpMode_j$ is the jth operating mode corresponding to VSP_j, v_j and a_j.
FF_j is the jth instantaneous fuel consumption factor (i.e., distance-based) corresponding to instantaneous velocity v_j, kg/km.
FR_j is the jth instantaneous fuel consumption rate corresponding to instantaneous velocity v_j, g/s.
FR_{VSP_j} is the jth reconstructed fuel consumption rate corresponding to VSP_j, g/s.
FR_{OpMode_j} is the jth reconstructed fuel consumption rate corresponding to $OpMode_j$, g/s.
$\overline{FR_{VSPi}}$ is the average fuel consumption rate for the VSP bin i, g/s.
$\overline{FR_{CVSPi,k}}$ is the average fuel consumption rate for the conditional VSP bin (i,k), g/s.
$\overline{FR_{OpMode_i}}$ is the average fuel consumption rate for the operating mode i, g/s.
$\overline{FR_{COpModei,k}}$ is the average fuel consumption rate for the conditional operating mode (i,k), g/s.
$\bar{v}_i$ is the ith average velocity, km/h.
$\overline{FF}_i$ is the ith average fuel consumption factor corresponding to average velocity $\bar{v}_i$, kg/km.

It is noted that all aggregated fuel rate data are mean value based in this study, but median values can also be alternatives.

2.5.1. Velocity binning (VB) based segregation

In this method, activity data are separated according to the proximity of velocity value, i.e., binned by instantaneous velocity. The bin interval was selected as 1 km/h in this study, and the segregation process can be described by Eqs. (2)–(4). For the ith bin, the average velocity $\bar{v}_i$ is simply the arithmetic mean of all samples in the bin, while the average fuel consumption factor $\overline{FF}_i$ is the arithmetic mean of all fuel consumption factors calculated by Eq. (4). As expected, this segregation method requires the least effort for data processing, and considers the second-by-second data samples to be independent of each other.

$$FF_j = \frac{FR_j}{v_j} * 3.6 \quad j = 1, 2, \ldots \tag{2}$$

$$\bar{\mathrm{v}}_\mathrm{i} = \frac{\sum \mathrm{v_j}}{\mathrm{n_i}} \quad \forall\ \mathrm{v_j} \in [\mathrm{i}-1, \mathrm{i}) \quad \mathrm{i} = 1, 2, \ldots, \lceil \max(\mathrm{v_j}) \rceil \tag{3}$$

$$\overline{FF}_i = \frac{\sum FF_j}{n_i} \quad \forall\ v_j \in [i-1, i) \quad i = 1, 2, \ldots, \lceil max(v_j) \rceil \tag{4}$$

where n_i is the number of second-by-second data samples located within the velocity interval $(i-1, i]$. $\lceil\cdot\rceil$ is a ceiling operator that rounds a number to the next larger integer.

2.5.2. Time-snipping (TS) based segregation

To address the dearth of correlation among data samples, a simple segregation method can be based on the time proximity, i.e., the vehicle activity data were divided into smaller snippets with fixed time interval. Sensitivity analysis on snipping time intervals including 10 s, 30 s, 60 s, 90 s and 120 s were conducted, and the correlation coefficients between the aggregated fuel consumption and speed were compared in the study. It turned out that using the interval of 60 s may provide the best model performance (the quantification of model performance will be elaborated in the following). Therefore, the activity data were segregated into 60 s time interval. Eqs. (5) and (6) present the formulas to calculate average velocity and fuel consumption factor after segregation. It is apparent that such method has taken into account specific driving patterns and implicitly addresses the impacts of acceleration on fuel consumption.

$$\overline{v_i} = \frac{\sum_{j=60*(i-1)+1}^{60*i} v_j}{60} \quad i = 1, 2, \ldots, \left\lfloor \frac{max(j)}{60} \right\rfloor \tag{5}$$

$$\overline{\mathrm{FF}}_\mathrm{i} = \frac{\sum_{\mathrm{j}=60*(\mathrm{i}-1)+1}^{60*\mathrm{i}} \mathrm{FR_j}}{\sum_{\mathrm{j}=60*(\mathrm{i}-1)+1}^{60*\mathrm{i}} \mathrm{v_j}/3.6} \quad \mathrm{i} = 1, 2, \ldots, \left\lfloor \frac{\max(\mathrm{j})}{60} \right\rfloor \tag{6}$$

2.5.3. Distance-snipping (DS) based segregation

Instead of segregating the vehicle activity data according to the time proximity, another heuristic method is distance-snipping based segregation. There are two typical ways of distance-snipping: (1) based on link; and (2) based on fixed distance. Due to the lack of link information, the fixed distance segregation method was selected in this study. Sensitivity analyses for distance interval (including 0.1 km, 0.5 km, 1 km, 2 km, 5 km and 10 km) indicate that 1 km was a reasonable choice in this study. The average velocity and fuel consumption factor for each snippet were calculated according to Eqs. (7) and (8). Similar to TS based segregation, DS based method also considers driving cycle features to some extent.

$$\overline{v_i} = \frac{\sum_{j=k_{i-1}+1}^{k_i} v_j/3.6}{k_i - k_{i-1}} * 3.6 \quad i = 1, 2, \ldots, \left\lfloor \frac{\sum v_j/3.6}{1000} \right\rfloor \tag{7}$$

$$\overline{FF_i} = \frac{\sum_{j=k_{i-1}+1}^{k_i} FR_j}{\sum_{j=k_{i-1}+1}^{k_i} v_j/3.6} \quad i = 1, 2, \ldots, \left\lfloor \frac{\sum v_j/3.6}{1000} \right\rfloor \tag{8}$$

where $k_0 = 0$ and k_i is subject to the following constraints for $i \geqslant 1$.

$$\begin{cases} \sum_{j=k_{i-1}+1}^{k_i} v_j/3.6 \leqslant 1000 \\ \sum_{j=k_{i-1}+1}^{k_i+1} v_j/3.6 > 1000 \end{cases} \tag{9}$$

2.5.4. Vehicle specific power (VSP) based segregation

As a function of speed, acceleration, and other vehicle-specific factors, the vehicle specific power (VSP) may be used as a metric for binning the activity data. For each bin, fuel consumption rate was aggregated based on the distribution of fuel consumption rate within the VSP bin with Eq. (10). And the aggregated fuel consumption rates were used to characterize corresponding VSP bins. On the basis of VSP distribution of each trip, fuel consumption rates were reconstructed with aggregated values through Eq. (11). The average velocity and fuel consumption factor were calculated by Eqs. (5) and (12), respectively. Fig. 7 describes the process of VSP based segregation method.

$$\overline{FR_{VSPi}} = \begin{cases} \sum \frac{FR_j}{m_i} & \forall\ VSP_j \in [i-1, i) \quad i = 1, 2, \ldots \\ \sum \frac{FR_j}{m_i} & \forall\ VSP_j < 0 \quad\quad\quad i = 0 \end{cases} \tag{10}$$

$$FR_{VSP_j} = \begin{cases} \overline{FR_{VSP0}} & if\ VSP_j \leqslant 0 \\ \overline{FR_{VSPi}} & if\ VSP_j \in [i-1, i) \end{cases} \tag{11}$$

$$\overline{FF_i} = \frac{\sum_{j=60*(i-1)+1}^{60*i} FR_{VSP_j}}{\sum_{j=60*(i-1)+1}^{60*i} v_j/3.6} \quad i = 1, 2, \ldots, \left\lfloor \frac{max(j)}{60} \right\rfloor \tag{12}$$

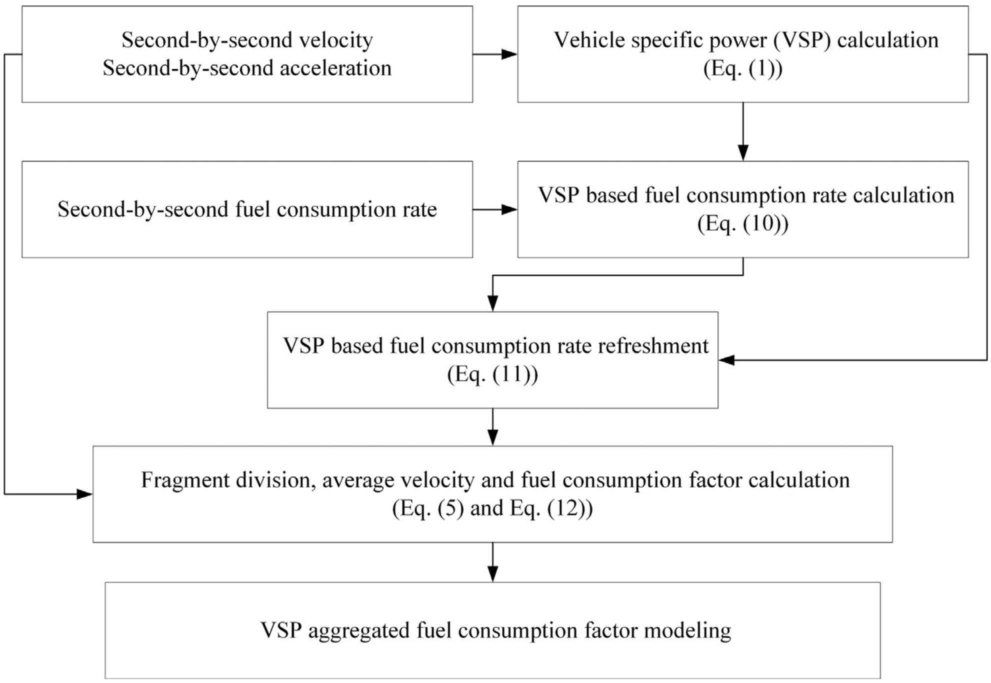

Fig. 7. Calculation process for VSP based fuel consumption factor.

where m_i stands for the number of VSP index that located within the specified VSP interval $[i-1, i)$ or $(-\infty, 0)$.

2.5.5. Enhanced vehicle fuel consumption modeling based on conditional VSP (CVSP)

The VSP based segregation method may be improved by adding velocity information in the modeling process. The index of conditional VSP was defined for the first time according to VSP interval and velocity interval through Eq. (13). The conditional expectation of fuel consumption rate for each combination of VSP interval and velocity interval was then calculated with Eq. (14). After that, second-by-second fuel consumption rate was further reconstructed according to the calculated conditional VSP in each trip. The same time snipping and fuel consumption factor calculation process for VSP based segregation was applied to obtain aggregated fuel consumption factor and velocity based on the distribution of conditional VSP. The block diagram for this method was rather similar to Fig. 7, but Eqs. (1) and (10)–(12) were replaced with Eqs. (13)–(16) for calculating the conditional VSP based fuel consumption rate and factor.

$$\mathrm{CVSP_j} = \begin{cases} (\mathrm{i,k}) & \text{if } \mathrm{VSP_j} \in [\mathrm{i}-1, \mathrm{i}) \ \text{ and } \ \mathrm{v_j} \in [\mathrm{k}-1, \mathrm{k}) \quad \mathrm{i} = 1, 2, \ldots \\ (0, \mathrm{k}) & \text{if } \mathrm{VSP_j} < 0 \text{ and } \mathrm{v_j} \in [\mathrm{k}-1, \mathrm{k}) \end{cases} \tag{13}$$

$$\overline{\mathrm{FR}_{\mathrm{CVSPi,k}}} = \frac{\sum \mathrm{FR_j}}{\mathrm{l_i}} \quad \forall\ \mathrm{CVSP_j} = (\mathrm{i,k}) \tag{14}$$

$$\mathrm{FR}_{\mathrm{CVSP_j}} = \overline{\mathrm{FR}_{\mathrm{CVSPi,k}}} \quad \text{for } \mathrm{CVSP_j} = (\mathrm{i,k}) \tag{15}$$

$$\overline{\mathrm{FF_i}} = \frac{\sum_{\mathrm{j}=60*(\mathrm{i}-1)+1}^{60*\mathrm{i}} \mathrm{FR}_{\mathrm{CVSP_j}}}{\sum_{\mathrm{j}=60*(\mathrm{i}-1)+1}^{60*\mathrm{i}} \mathrm{v_j}/3.6} \quad \mathrm{i} = 1, 2, \ldots, \left\lfloor \frac{\max(\mathrm{j})}{60} \right\rfloor \tag{16}$$

where l_i stands for the number of conditional VSP index that equals to (i, k).

2.5.6. Conditional operating mode (COpMode) based segregation

As aforementioned, operating mode is a strong indicator of vehicle instantaneous power, which is further relevant to fuel consumption (USEPA, 2016b). Therefore, a method taking operating mode as the explanatory variable is proposed for the first time in this study. The segregation process resembles conditional VSP based one where the conditional operating mode was defined in Eq. (17) in combination with Fig. 6. Based on this newly defined parameter, the conditional operating mode based fuel consumption rate and factor were calculated with Eqs. (18)–(20).

$$\mathrm{COpMode_j} = (\mathrm{i,k}) \text{ for } \mathrm{OpMode_j} = \mathrm{i} \text{ and } \mathrm{v_j} \in [\mathrm{k}-1, \mathrm{k}) \tag{17}$$

$$\overline{\mathrm{FR}_{\mathrm{COpModei,k}}} = \frac{\sum \mathrm{FR_j}}{\mathrm{l_i}} \quad \forall\ \mathrm{COpMode_j} = (\mathrm{i,k}) \tag{18}$$

$$\mathrm{FR}_{\mathrm{COpMode_j}} = \overline{\mathrm{FR}_{\mathrm{COpModei,k}}} \text{ for } \mathrm{COpMode_j} = (\mathrm{i,k}) \tag{19}$$

$$\overline{\mathrm{FF_i}} = \frac{\sum_{\mathrm{j}=60*(\mathrm{i}-1)+1}^{60*\mathrm{i}} \mathrm{FR}_{\mathrm{COpMode_j}}}{\sum_{\mathrm{j}=60*(\mathrm{i}-1)+1}^{60*\mathrm{i}} \mathrm{v_j}/3.6} \quad \mathrm{i} = 1, 2, \ldots, \left\lfloor \frac{\max(\mathrm{j})}{60} \right\rfloor \tag{20}$$

In summary, Fig. 8 describes the proposed snippet segregation methods from a perspective of evolution in information richness/relevance to the fuel consumption. Each dashed rectangle represents one information richness level. The left most rectangle contains the least information which is no other than instantaneous velocity. The second left rectangle involves

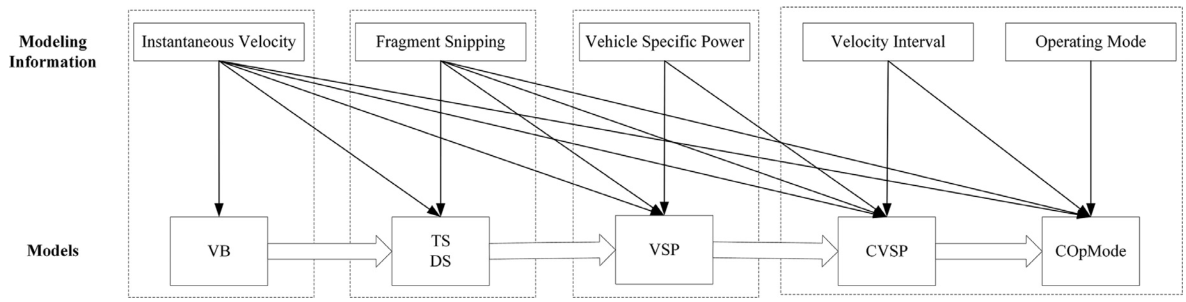

Fig. 8. Modeling evolution.

driving pattern information through snipping. The third rectangle adds VSP information to the previous methods. The right involves the most information with further addition of velocity interval. It is expected that the model performance can be improved by involving more information in the snippet segregation process.

2.6. Data fitting

For comparison purpose, the same model form is selected for fitting the aggregated data obtained from different snippet segregation methods. According to the authors' previous work, logarithmic function is proved to be effective in describing the relationship between energy consumption factor and velocity (Boriboonsomsin et al., 2012). Stepwise regression analysis with MATLAB (R2014b) shows that the 4th order logarithmic function is applicable, and the regression models are presented in Eq. (21).

$$\mathrm{logFF} = a * v^4 + b * v^3 + c * v^2 + d * v + e \tag{21}$$

3. Results and analyses

3.1. Fuel consumption models

Fuel consumption models based on different segregation methods are presented in Fig. 9. The blue points represent aggregated fuel consumption factors under different average velocities while the red lines are fitted by the associated regression models. Coefficients corresponding to the models in Fig. 9 are presented in Table 1, respectively.

Regression analyses demonstrate that the distance-snipping (DS) based model has rather low adjusted R^2 which means the model is not appropriate in explaining the relationship between vehicle fuel consumption and velocity. It can be

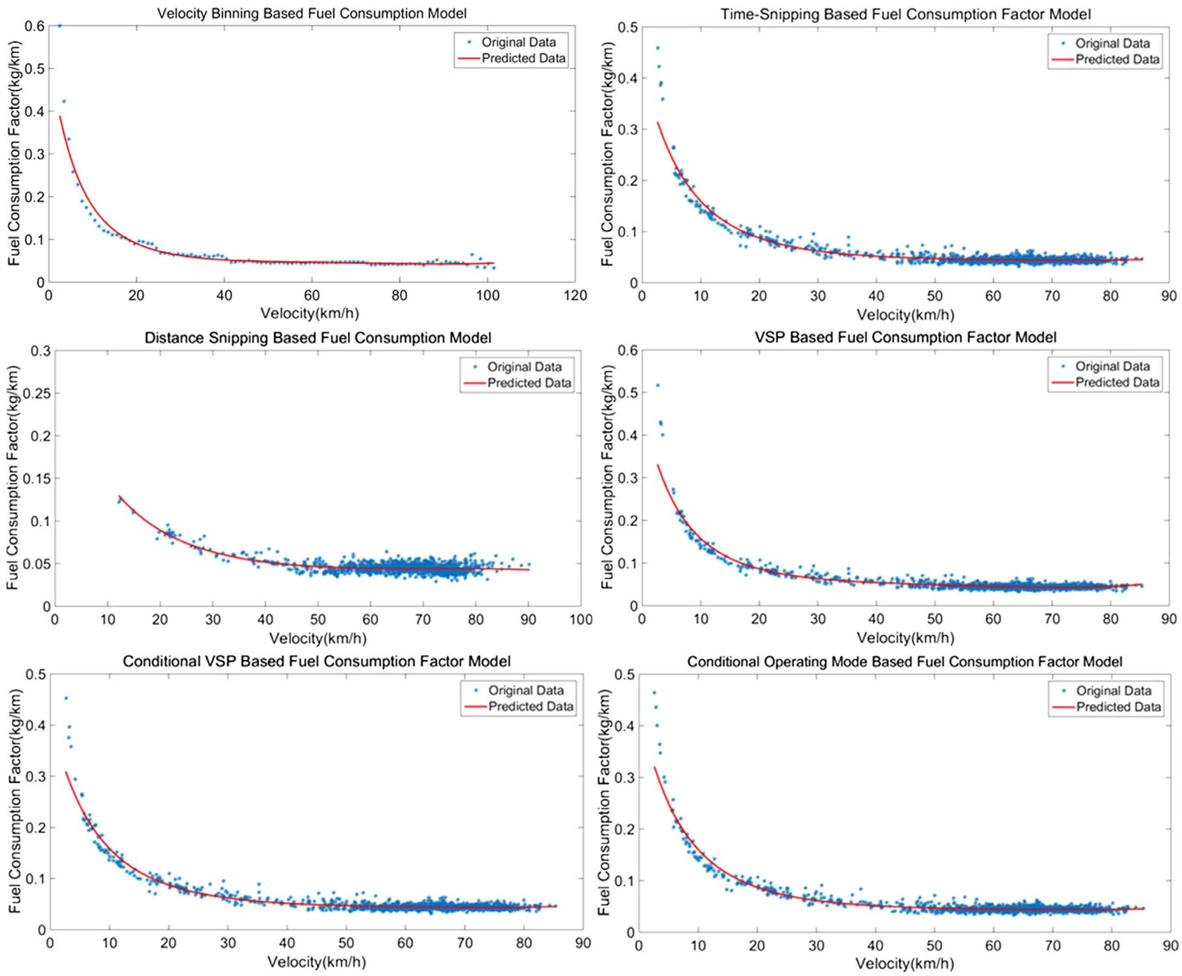

Fig. 9. Fuel consumption models.

Table 1
Coefficients for fuel consumption models.

Models	a	b	c	d	e	$\bar{R}^2$	RMSE
VB	$4.66 * 10^{-8}$	$-1.28 * 10^{-5}$	0.0013	−0.0603	−0.2688	0.9017	0.0243
TS	$4.53 * 10^{-8}$	$1.15 * 10^{-5}$	0.0011	−0.0528	−0.3724	0.9336	0.0097
DS	0	$-2.27 * 10^{-6}$	$4.84 * 10^{-4}$	−0.0344	−0.5365	0.6986	0.0049
VSP	$9.29 * 10^{-8}$	$-1.96 * 10^{-5}$	0.0016	−0.0612	−0.3306	0.9191	0.0104
CVSP	$4.19 * 10^{-8}$	$-1.08 * 10^{-5}$	0.0011	−0.0516	−0.3828	0.9386	0.0091
COpMode	$4.63 * 10^{-8}$	$-1.18 * 10^{-5}$	0.0012	−0.0539	−0.3621	0.9369	0.0095

observed that the aggregated data points based on distance snipping is less scattered when compared to others (most of data samples concentrate in the velocity range from 50 km/h to 80 km/h). The time-snipping (TS) based model which does not take any other parameter into consideration in modeling process also presents favorable performance. In light of these observations, data points are aggregated in time rather than in distance for the VSP-, CVSP- and COpMode-based models in this study. As presented in Table 1, all other models except DS based one have demonstrated satisfactory results in the sense of adjusted R^2 and root mean square error (RSME). For the VB based model, however, the instantaneous fuel consumption data is simply binned every 1 km/h, resulting in a very sparse data samples as noted in Fig. 9. This might be the main reason that high adjusted R^2 is obtained. Therefore, models based on TS, VSP, CVSP and COpMode exhibit satisfactory performance in terms of adjusted R^2 (between 0.92 and 0.94) and RMSE (ranging from 0.0091 to 0.0104).

In the following, statistical analyses were conducted from two aspects: (1) evaluation of fuel consumption factors estimated from different models; and (2) comparison of the trip-based fuel consumption from different models. It is noted that the ground truth was assumed to be the outputs from CMEM.

3.2. Statistical analyses on fuel consumption factors

In order to carry out statistical analyses on fuel consumption factors, all the tested trips were decomposed into large amount of 60 s segments, and fuel consumption factor (fuel consumption normalized by traveling distance) for each segment was estimated with the six different models. The Kolmogorov-Smirnov or K-S test (Wilcox, 2005) was applied to each model pair to analyze the difference in estimated fuel consumption factors from different models, and the results are presented in

Table 2
K-S test results for different fuel consumption models.

K-S test result (p value)	VB	TS	DS	VSP	CVSP	COpMode
VB	–	1 ($1.52 * e^{-54}$)	1 ($8.02 * e^{-60}$)	1 ($1.05 * e^{-25}$)	1 ($1.63 * e^{-47}$)	1 ($2.18 * e^{-46}$)
TS	1 ($1.52 * e^{-54}$)	–	1 ($9.94 * e^{-14}$)	1 ($5.48 * e^{-9}$)	0 (0.52)	0 (0.06)
DS	1 ($8.02 * e^{-60}$)	1 ($9.94 * e^{-14}$)	–	1 ($4.96 * e^{-14}$)	1 ($9.94 * e^{-14}$)	1 ($6.34 * e^{-18}$)
VSP	1 ($1.05 * e^{-25}$)	1 ($5.48 * e^{-9}$)	1 ($4.96 * e^{-14}$)	–	1 ($5.91 * e^{-7}$)	1 ($5.91 * e^{-7}$)
CVSP	1 ($1.63 * e^{-47}$)	0 (0.52)	1 ($9.94 * e^{-14}$)	1 ($5.91 * e^{-7}$)	–	0 (0.09)
COpMode	1 ($2.18 * e^{-46}$)	0 (0.06)	1 ($6.34 * e^{-18}$)	1 ($5.91 * e^{-7}$)	0 (0.09)	–

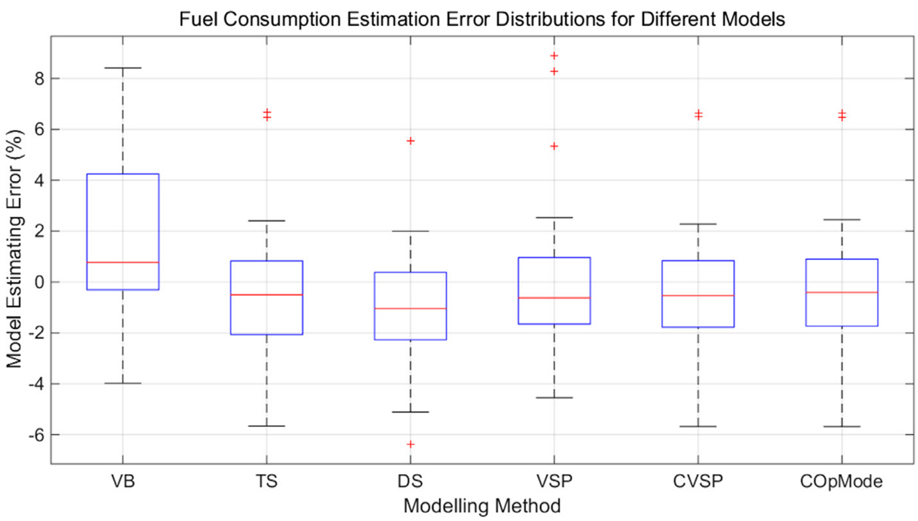

Fig. 10. Estimation error for different models.

Table 3
Measurement of effectiveness for model estimation error.

Statistics	VB	TS	DS	VSP	CVSP	COpMode
Mean	0.019	−0.00228	−0.00775	0.00272	−0.0022	−0.00158
Variance	0.00113	0.000854	0.000802	0.00120	0.000839	0.000843
Median	0.00771	−0.00498	−0.0104	−0.00621	−0.00535	−0.0041

Table 2. According to Table 2, statistically significant differences between almost all model pairs except for three pairs, i.e., TS versus CVSP, TS versus COpMode, and CVSP versus COpMode, are observed. What's more, the K-S test results may imply that models with different information richness levels are quite different from each other. For example, there is statistical evidence to show the VB model (information richness level 1) outputs are different from outputs from other models (information richness level $\geqslant 2$), and the same situation goes between the outputs from VSP model (information richness level 3) and TS/DS model (information richness level 2) or CVSP/COpMode model (information richness level 4). Therefore, adding more information in the modeling process will make difference in fuel consumption factor estimation in the statistical sense.

3.3. Comparison of trip-based fuel consumption estimates

As stated above, the cross validation technique is applied to the comparison of fuel consumption estimation on a trip basis. The boxplots of estimation errors for different models are illustrated in Fig. 10 and the basic statistics of estimation errors are presented in Table 3. It can be observed that velocity binning (VB) based model has largest estimation error statistics (in terms of mean, variation and median) for trip based fuel consumption. The possible reason is that the modeling process only considered instantaneous velocity and the acceleration/deceleration pattern was not properly modeled. The conditional operating mode (COpMode) based model outperforms the others with the smallest estimation error statistics, while time snipping (TS) based model exhibits satisfactory performance even it utilizes relatively less information when compared to the VSP based model or the CVSP based one.

4. Conclusions

This research is focused on an important issue, i.e., data segregation methods, in mesoscopic energy modeling. We developed a generalized framework for evaluating the effectiveness of different snippet segregation methods in estimating both fuel consumption factor and trip based fuel consumption, using the real-world data collected in Beijing. For data segregation method comparison, an evolutionary process has been carried out to increase the information richness for data segregation. In addition, novel modeling methods which take into consideration the operating mode and metrics conditional on velocity show promising statistical results in terms of trip-based fuel consumption estimation/prediction.

In this study, we used the outputs from CMEM as the ground truth due to the limitation in resource availability, but direct measurements will be definitely preferable to quantify the statistics of model estimation errors in the future. As aforementioned, the findings from this research will be valuable to guide the development of many environmental-focused ITS applications, such as eco-routing navigation, in which the trip based fuel consumption can be estimated/predicted based on the candidate routes and available aggregated traffic information (such as link speed) via the mesoscopic model, and then the optimal route will be recommended to the driver to obtain best fuel efficiency. Also, the model can provide advanced driving assistance such as refueling warning for drivers, and region-wide monitoring of motor vehicle energy consumptions for traffic operators.

The mesoscopic energy consumption model established in the study was validated using field data collected from freeways. As a future step, the proposed methodology will be further validated on a larger geographical scale, e.g., not only for freeways but also for arterial roads, to examine its applicability. Unlike freeways, arterial roads with signalized intersections may cause discontinuities to traffic where some aggregation methods (e.g., distance snipping based) proposed in this study may be sensitive to the model parameters. In this case, link-based segregation methods might provide more reliable results, where each link represents a block between two consecutive signals and the segregation methods will apply to the links rather than fixed time interval or distance. This will require the availability of road network topology and a significant amount of effort of map-matching, which can be one of the future research topics.

Acknowledgement

This work was supported by National Basic Research Program of China (973 Project) 2012CB725405. The authors are also grateful for the fruitful discussion with Dr. George Scora.

References

Ahn, K., Rakha, H., Trani, A., Van Aerde, M., 2002. Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. J. Transport. Eng. 128 (2), 182–190.

Barth, M., An, F., Norbeck, J., Ross, M., 1996. Modal emissions modeling: a physical approach. Transport. Res. Rec.: J. Transport. Res. Board 1520, 81–88.
Barth, M., An, F., Younglove, T., et al, 2000. Development of a Comprehensive Modeal Emissions Model: NCHRP Project 25-11 Final Report.
Barth, M., Boriboonsomsin, K., Vu, A., 2007. Environmentally-friendly navigation. In: 2007 IEEE Intelligent Transportation Systems Conference. IEEE, pp. 684–689.
Boriboonsomsin, K., Barth, M.J., Zhu, W., Vu, A., 2012. Eco-routing navigation system based on multisource historical and real-time traffic information. IEEE Trans. Intell. Transport. Syst. 13 (4), 1694–1704.
CARB, 2016. EMFAC2014 Web Database <http://www.arb.ca.gov/emfac/2014/>.
Hausberger, S., Rodler, J., Sturm, P., 2002. Emission factors for HDV and validation by tunnel measurements. VKM-THD Mitteilungen, vol. 81, pp. 93–100.
Jimenez-Palacios, J.L., 1998. Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing Doctoral dissertation. Massachusetts Institute of Technology.
Kang, J., Ma, T., Ma, F., Huang, J., 2011. Link-based emission model for eco routing. In: 2011 11th International Conference on ITS Telecommunications (ITST). IEEE, pp. 207–212.
Koupal, J., Cumberworth, M., Michaels, H., Beardsley, M., Brzezinski, D.J., 2003. Design and implementation of MOVES: EPA's new generation mobile source emission model. Ann Arbor 1001, 48105.
Ministration of Environmental Protection of the People's Republic of China, "China Vehicle Emission Annual Report", 2015. <http://wfs.mep.gov.cn/dq/jdc/zh/201601/P020160115523794855203.pdf>.
Picard, R.R., Cook, R.D., 1984. Cross-validation of regression models. J. Am. Stat. Assoc. 79 (387), 575–583.
U.S. Environmental Protection Agency, 2016a. Inventory of U.S. greenhouse gas emissions and sinks: 1990–2014 (draft), EPA 430-R-16-002, February.
U.S. Environmental Protection Agency, 2016b. MOVES2014a, <https://www3.epa.gov/otaq/models/moves/>.
U.S. Environmental Protection Agency, 2002. User's Guide to MOBILE6.1 and MOBILE6.2: Mobile Source Emission Factor Model.
Warila, J., Nam, E., Landman, L., Kahan, A., 2011. Light-Duty Exhaust Emission Rates in MOVES2010. MOVES Workshop, Ann Arbor.
Wilcox, R., 2005. Kolmogorov–smirnov test. In: Encyclopedia of Biostatistics.
Yao, E., Song, Y., 2013. Study on eco-route planning algorithm and environmental impact assessment. J. Intell. Transport. Syst. 17 (1), 42–53.