Transportation Research Part C 67 (2016) 280–298

Contents lists available at ScienceDirect

Transportation Research Part C

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

Customizing driving cycles to support vehicle purchase and use decisions: Fuel economy estimation for alternative fuel vehicle users ☆

Jun Liu [a,*,1], Xin Wang [b,1], Asad Khattak [a]

[a] Department of Civil & Environmental Engineering, The University of Tennessee, United States
[b] Virginia Department of Transportation, Transportation & Mobility Planning Division, United States

ARTICLE INFO

Article history:
Received 5 April 2015
Received in revised form 5 January 2016
Accepted 23 February 2016
Available online 12 March 2016

Keywords:
Driving cycle
Alternative fuel vehicle
Fuel economy
Cluster analysis
Micro-trips

ABSTRACT

Wider deployment of alternative fuel vehicles (AFVs) can help with increasing energy security and transitioning to clean vehicles. Ideally, adopters of AFVs are able to maintain the same level of mobility as users of conventional vehicles while reducing energy use and emissions. Greater knowledge of AFV benefits can support consumers' vehicle purchase and use choices. The Environmental Protection Agency's fuel economy ratings are a key source of potential benefits of using AFVs. However, the ratings are based on pre-designed and fixed driving cycles applied in laboratory conditions, neglecting the attributes of drivers and vehicle types. While the EPA ratings using pre-designed and fixed driving cycles may be unbiased they are not necessarily precise, owning to large variations in real-life driving. Thus, to better predict fuel economy for individual consumers targeting specific types of vehicles, it is important to find driving cycles that can better represent consumers' real-world driving practices instead of using pre-designed standard driving cycles. This paper presents a methodology for customizing driving cycles to provide convincing fuel economy predictions that are based on drivers' characteristics and contemporary real-world driving, along with validation efforts. The methodology takes into account current micro-driving practices in terms of maintaining speed, acceleration, braking, idling, etc., on trips. Specifically, using a large-scale driving data collected by in-vehicle Global Positioning System as part of a travel survey, a micro-trips (building block) library for California drivers is created using 54 million seconds of vehicle trajectories on more than 60,000 trips, made by 3000 drivers. To generate customized driving cycles, a new tool, known as Case Based System for Driving Cycle Design, is developed. These customized cycles can predict fuel economy more precisely for conventional vehicles vis-à-vis AFVs. This is based on a consumer's similarity in terms of their own and geographical characteristics, with a sample of micro-trips from the case library. The AFV driving cycles, created from real-world driving data, show significant differences from conventional driving cycles currently in use. This further highlights the need to enhance current fuel economy estimations by using customized driving cycles, helping consumers make more informed vehicle purchase and use decisions.

☆ This article belongs to the Virtual Special Issue on Advances in alternative fuel vehicle transportation systems.

* Corresponding author.
E-mail addresses: jliu34@utk.edu (J. Liu), xin.wang@vdot.viginia.gov (X. Wang), akhattak@utk.edu (A. Khattak).

[1] These two authors contributed to this study equally.

http://dx.doi.org/10.1016/j.trc.2016.02.016
0968-090X/

1. Introduction

An alternative fuel vehicle (AFV) is a vehicle that runs on a fuel (e.g., battery electric) other than conventional petroleum fuels (gasoline or diesel) and also refers to any technology of powering an engine that does not involve solely petroleum (e.g., hybrid electric) (Wikipedia, 2014). Options for AFVs in market are vast but their penetration in fleets is still small, compared with conventional vehicles consuming gasoline or diesel. Enhanced energy security and cleaner travel are the major benefits that attract potential customers to transition from conventional vehicles to AFVs (Bunch et al., 1993; Nesbitt and Sperling, 1998; Struben and Sterman, 2008; Liu et al., 2015b). One of the most important aspects of vehicle purchase and use decisions concerns fuel economy.

Currently, the fuel economy is predicted by U.S. Environmental Protection Agency (EPA) using pre-designed standard driving cycles in a lab controlled condition. The precision of fuel economy estimation heavily relies on whether the driving cycle can represent real-life driving practices. EPA has designed various driving cycles, such as FTP (Federal Test Procedure, often called EPA75), HWFET (Highway Fuel Economy Driving Schedule), SFTP (Supplemental Federal Test Procedure), US06 (representing aggressive driving on highway), SC03 (representing hot ambient when AC is on) and C-FTP (representing city driving conditions in cold ambient temperature) (Davis et al., 2009; Berry, 2010), to account for various travel needs and driving contexts.

A study by Lin and Greene has shown that precision of fuel economy for individual drivers rather than bias may have limited the usefulness of EPA fuel economy ratings (Lin and Greene, 2011). To overcome the accuracy/precision issue, this study proposes the use of data from sensor/positioning technology and behavioral surveys. The question is: can a limited number of driving cycles represent trillions of vehicle trips in real-world, especially for real-world driving of AFVs? If driving practices in real-world are not similar across different vehicle groups (i.e., conventional vehicles vis-à-vis AFVs), then the answer would lean toward a “no”.

The use of standard driving cycles in a lab controlled condition to test all vehicles has its own drawback. One issue is that the standard test is based on deterministic driving cycles-it basically assumes all driving activities to be similar irrespective of drivers’ individual characteristics. But in real-world traffic condition, vehicles could be driven differently depending on individual’s driving styles. Another issue is that the current driving cycles do not consider the use of advanced driving aid technologies, e.g., cruise control. While in reality, a greater portion of drivers has applied these technologies to ease them from driving tasks. Moreover, there is substantial uncertainty about whether AFV users drive differently given AFVs having different engine performance, which can impact their fuel economy. How to design customized driving cycles in an appropriate manner, overcoming the lack of precision when using deterministic driving cycle, is thus of interest. The customized driving cycles for transition to AFVs should be able to: (1) represent real-world driving practices based on consumers’ individual characteristics; and (2) compare the fuel economy for consumers when they are driving AFVs versus conventional vehicles.

Previously, limited availability of data restrained the diversity and customization of driving cycles. Using “one-fit-all” pre-designed driving cycles was a good option. However, with increasing amounts of data generated by electronic sensors from various sources that include travelers, vehicles, infrastructure and the environment, referred to as “Big Data”, customizing driving cycles for individuals using gasoline vehicles or AFVs has become feasible. Using large-scale trajectory data merged with travel behavior information, this study aims to construct a practical methodology to customize driving cycles based on real-world driving data for various users and vehicles using different power systems. These customized driving cycles can be used to better estimate fuel economy for consumers based on their own driving style instead of using a “one-fit-all” pre-designed driving cycle. A more precise fuel economy estimate can potentially help consumers choose a more energy-efficient and cleaner vehicle. This study suggests a methodology that can help move government agencies (Environmental Protection Agency) and industry (vehicle manufacturers and energy related companies) toward driving cycles that are customized and based on local or regional conditions. The scope of this study is limited to developing an alternative way of generating individualized driving cycles using a large-scale real-world driving database. This should be a step toward estimating fuel economy, but how to estimate fuel economy using driving cycles is not within the scope of this paper.

2. Literature review

The US EPA uses 5 standard test cycles based on dynamometers in laboratory conditions to provide point estimates of vehicle fuel economy to consumers. This information is valuable in making vehicle purchase and use decisions. Other than capital costs, energy costs are heavily weighted when consumers make vehicle purchase decisions (Turrentine and Kurani, 2007; Greene, 2010; Lin and Greene, 2011). Driving cycles specified by DDS (Dynamometer Drive Schedule) are often used to estimate vehicle fuel economy which is highly associated with energy costs. Delucchi et al. compared the costs, including initial vehicle cost, operating and maintenance costs, and battery replacement costs, of Battery-powered Electric Vehicles (BEVs) with conventional vehicles (CVs) consuming gasoline (Delucchi and Lipman, 2001). They calculated vehicle energy use (a big component of operating costs) over a specified driving cycle – Federal Urban Drive Schedule (FUDS) which is used in conjunction with other driving cycles by the EPA. They reported that though BEVs have advantages in energy security and environment protection, the manufacturing cost for batteries must be lowered enough, in order for BEVs to be cost-competitive with gasoline CVs (Delucchi and Lipman, 2001). Lave et al. compared the fuel economy of hybrid vehicles (HEVs), the Toyota Prius, with CVs, Toyota Corolla, based on both urban and highway driving cycles (Lave and MacLean, 2002). They found significant smaller energy costs and emissions among HEVs. However, the HEVs' benefits from reduced energy costs and emissions are only a small fraction of the total cost including manufacturing costs. Prius would have a difficult time competing with Corolla given Corolla's already high fuel economy and lower emissions (Lave and MacLean, 2002).

Markel et al. examined the fuel consumption rates of CVs, HEVs and plug-in hybrid electric vehicles (PHEVs) over two standard driving cycles – UDDS and HWFET (Markel and Simpson, 2006). UDDS (Urban Dynamometer Driving Schedule) is designed for testing light-duty vehicles under city driving conditions and HWFET (Highway Fuel Economy Driving Schedule) represents the free-flow traffic condition on highways. They compared vehicle purchase and fuel consumption costs and benefits (i.e., reduced fuel consumption) of PHEVs relative to CVs. Though there are higher retail costs for PHEVs compared with CVs, PHEVs with a substantial amount of reduced lifetime energy costs offer still significant benefits to consumers. Markel et al. also mentioned that the fuel consumption rates on standard driving cycles may vary with actual in-use driving cycles (Markel and Simpson, 2006). Fontaras et al. reported that using standard driving cycles may under- or over-estimate fuel economy and emissions, if driving cycles do not properly represent real-world driving practices (Fontaras et al., 2008). They conducted fuel economy estimations for HEVs over pre-designed driving cycles, including a cold New European Driving Cycle—NEDC (the combined legislated driving cycle), a hot Urban Driving Cycle—UDC (urban sub-cycle of NEDC), the Artemis driving cycles (André, 2004), and real-world simulated driving cycles accounting for transient driving conditions. Compared with CVs, HEVs were found to have substantial fuel economy benefits in addition to reduced emissions under urban driving conditions (Fontaras and Samaras, 2007; Fontaras et al., 2008).

In addition to the driving cycles mentioned above, designed for estimating fuel economy (MPG-ratings) and vehicles emissions, EPA provides pre-designed driving cycles to test vehicles running under different driving conditions (EPA, 2013). FTP (Federal Test Procedure), often called EPA75 simulates the city driving conditions. Three additional SFTP (Supplemental Federal Test Procedure) are used to adjust the city and highway estimates to account for higher speeds, air conditioning use, and colder temperatures. They include US06 (representing aggressive driving on highway), SC03 (representing hot ambient when AC on) and C-FTP (representing city driving conditions in cold ambient temperatures). To account for more driving conditions, driving cycles to better represent local driving practices are also developed by EPA (EPA, 2013). The New York City Cycle (NYCC) features low speed stop-and-go traffic conditions. The California Air Resources Board LA92 Dynamometer Driving Schedule (often called the Unified driving schedule), was developed as a driving cycle having a higher top speed, a higher average speed, less idle time, fewer stops per mile, and a higher maximum rate of acceleration compared with FTP.

Researchers have realized that driving cycles should show different characteristics in different regions, given different contextual conditions coming from roadway geometry, land use and culture of driving. Studies were conducted to develop driving cycles to better represent local driving practices. Lin et al. created robust driving cycles for Los Angeles, called LA01 (Lin and Niemeier, 2002, 2003). They used a maximum likelihood estimation (MLE) partitioning algorithm Markov process to construct driving cycles for different levels-of-service on roadways. Tong et al. developed a driving cycle for Hong Kong, by extracting parts of the on-road speed data such that the summary statistics of the sample are close to the population (Tong et al., 1999). Following Tong et al., Hung et al. constructed Hong Kong driving cycles through a random selection process. They focused on getting reasonable cycle lengths and used more stringent criteria for selection of best driving cycles from candidate cycles. Driving cycles were selected by ensuring the assessment parameter was less than 5% different from the target mean values (Hung et al., 2007). Saleh et al. applied a similar methodology that Hung et al. used to select representative driving cycles from multiple driving cycles collected in Edinburgh, UK. The parameters used included speed, percentage time spent in cruise, accelerations, decelerations and idling, and their statistical validity over trip lengths (Saleh et al., 2009). Kamble et al. developed a driving cycle for Pune city in India (Kamble et al., 2009) and André et al. collected driving data from France, the UK, Germany and Greece, to estimate real-world European driving cycles (André, 2004; André et al., 2006).

Typically, assessment parameters are calculated to quantify driving characteristics of a cycle. Those parameters include average vehicle speed, average running speed, average acceleration and deceleration, and proportions of idling, acceleration, cruising and deceleration, average number of acceleration–deceleration changes, etc. Driving cycles are selected according to these parameters, using various methods such as a random selection process (Hung et al., 2007), Markov process (Lin and Niemeier, 2002, 2003), and micro-trips analysis (Kamble et al., 2009). However, real-world driving practices are represented imprecisely by a set of pre-designed driving cycles, because of the complexity of real-world driving, uncertain engine performance, vehicle age and characteristics, transient driver behaviors and variations in driving contexts (Ntziachristos and Samaras, 2000; Markel and Simpson, 2006). While substantial biases may not exist in the EPA driving cycles, their inaccuracy for individual drivers is a key issue. Fundamentally, pre-designed driving cycles may not represent the variations in real-world driving practices very well. Using large-scale trajectory data coupled with travel behavioral information, this study provides a practical methodology to customize driving cycles based on real-world driving data for various users and vehicles, i.e., conventional and alternative fuel vehicles that use different power systems.

3. Data description

This study pulls travel behavioral data from a large-scale travel survey database, generated by 2012–2013 California Household Travel Survey (CHTS) (Caltrans, 2013). The data were collected in 58 counties across the State of California.

The database contains information for driver, household, trip, and more importantly, second-by-second speed data. The speed trajectory data were processed and separated into micro-trips (defined as a continuous driving activity between two stops, one trip can contain one or multiple micro-trips). The data were collected by in-vehicle GPS (Global Positioning System) as well as OBD (On-Board Diagnostic) sensors during each trip.

The data cover various driving practices on different road types, made by vehicles of varied body types as well as different fuel types. Specifically, the database includes 54 million seconds of driving records, including 236,403 micro trips and 65,652 trips made by 2908 vehicles. These vehicles include 2253 conventional vehicles (CVs) consuming gasoline, 364 hybrid electric vehicles (HEVs), 109 battery electric vehicles (BEVs), 110 diesel vehicles and a small portion of vehicles consuming other alternative fuel types, such as natural gas and biofuels. These broad and diverse driving samples, with highly detailed operating information, constitute a rich large-scale database which allows for in-depth comparison and analysis through multiple lenses, e.g., vehicle fuel type, vehicle body type, micro-trip type, and many others.

4. Conceptual framework

The overall framework is illustrated in Fig. 1, where we conceptualize the system. The key proposed strategy deals with generating more precise information about fuel economy for individuals who are interested in purchasing conventional vehicles or AFVs. To design customized drive cycles, the overall approach uses case-based reasoning (Khattak and Kanafani, 1996; Khattak et al., 2006). This is done by clustering and joining micro-trips to create drive cycles, and vehicle simulations (e.g., the vehicle-specific power methodology used in the EPA MOVES model) to estimate the fuel economy for a driver. The idea is to use modern information technology to provide more precise information to consumers by matching the individual who desires a personal MPG estimate with similar individuals for whom driving patterns are already available. From a behavioral perspective, individuals who are considering purchase of vehicles will have a wider set of choices that include AFVs and more precise information about fuel economy. Depending on the context (income levels, roadway geometry, land use and culture of driving), if they adopt AFVs, people may also change their vehicle use decisions, e.g., use their vehicle more or less for various trip purposes. The outcomes of their decisions will be reelected in energy use, emissions and vehicle miles traveled. Policy-makers and policy-implementers (e.g., EPA) can provide strong behavioral incentives that can accelerate transition to AFVs.

5. Real-world driving practices between fuel types and with standard driving cycles

5.1. Creating equivalent groups

To explore how trips made by AFVs (i.e., BEVs and HEVs) are different from trips made by CVs consuming gasoline, this study compares real-world driving practices of BEVs, HEVs and CVs by selecting equivalent groups of users. Using equivalent groups given similar demographic characteristics, i.e., age, gender, and income, helps minimize the influences of driver demographics and highlights the effects of vehicle types on driving practices. Since there are only 109 BEVs in the database, the same number of vehicles were semi-randomly selected from 364 HEVs and 2253 CVs by one-to-one matching of demographics with BEV drivers. Eventually, each of the group has 106 vehicles, because some information was missing in three BEV observations. Table 1 shows the descriptive statistics of driver demographics in three selected vehicle groups and the entire sample. The age, gender, and household income in three selected groups have similar distributions indicating the samples are equivalent.

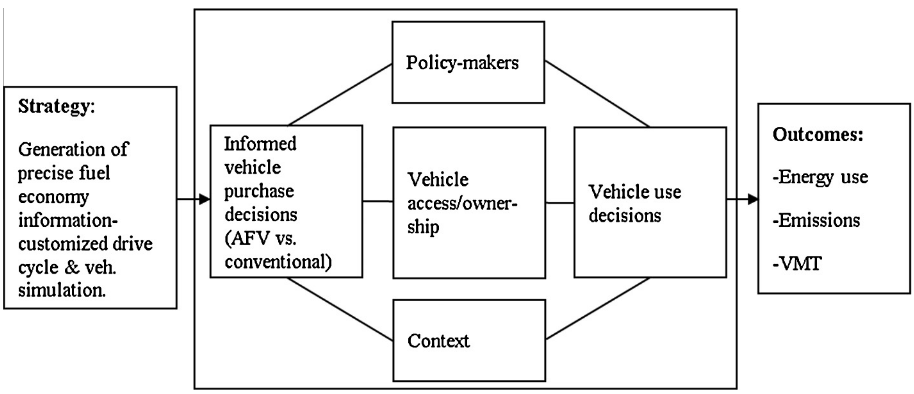

Fig. 1. Conceptual framework: strategies, behavior, and outcomes.

Table 1
Demographics of groups segmented by vehicle type.

Vehicle group	Demographics		*N*	Mean percent	Std. dev.	Min	Max
BEV (Battery Electric Vehicle)	Age (years)		106	49.415	10.403	16	71
	Gender [Male]		106	57.50	0.497	0	1
	Household Income	<74,999	106	3.80	0.191	0	1
		75,000–99,999	106	12.30	0.33	0	1
		100,000–149,000	106	26.40	0.443	0	1
		>150,000	106	57.50	0.497	0	1
HEV (Hybrid Electric Vehicle)	Age (years)		106	49.394	9.767	20	68
	Gender [Male]		106	57.50	0.497	0	1
	Household Income	<74,999	106	3.80	0.191	0	1
		75,000–99,999	106	12.30	0.33	0	1
		100,000–149,000	106	26.40	0.443	0	1
		>150,000	106	57.50	0.497	0	1
CV (Conventional Gasoline Vehicle)	Age (years)		106	49.415	10.403	16	71
	Gender [Male]		106	57.50	0.497	0	1
	Household Income	<74,999	106	3.80	0.191	0	1
		75,000–99,999	106	12.30	0.33	0	1
		100,000–149,000	106	26.40	0.443	0	1
		>150,000	106	57.50	0.497	0	1
All drivers	Age (years)		2908	48.804	13.49	16	88
	Gender [Male]		2908	48.00	0.5	0	1
	Household income	<74,999	2908	31.20	0.216	0	1
		75,000–99,999	2908	18.70	0.39	0	1
		100,000–149,000	2908	23.20	0.422	0	1
		>150,000	2908	26.90	0.443	0	1

5.2. How are conventional and alternative fuel vehicles used?

After controlling for driver demographics through study design, driving performance is compared across BEVs, HEVs and CVs. Fig. 2 presents the time spent on acceleration or deceleration by speed range in 0.5 mph increments, as well as the standardized time allocation percentages by speed bins. Time spent on accelerating or braking varies with speeds. Acceleration and deceleration are nearly equal in terms of time spent at all speed ranges. Major findings based on comparisons include:

- BEV trips have less time spent at high speeds (>60 mph) than equivalent groups.
- There are distinct spikes in BEV time use distribution, occurring at 55 mph, 60 mph and 65 mph. This implies that those are speeds at which cruise control is used. Based on the results, BEV drivers were more likely to use cruise control during driving than CV drivers.
- With speed increasing, more time is spent driving at constant speed. This outcome is more distinct for BEV and HEV groups compared with the CV group.

Note that the differences between vehicles' performance may be related to various factors, such as how a vehicle was used, e.g., for shopping or commuting, or other trip purposes. Furthermore, spatial and accessibility factors may also be in play. If a driver lives in a region next to an interstate, then there might be a greater chance for them to drive at higher freeway speeds compared with drivers who live far from interstates.

Given that driving cycles are essential to fuel economy estimation and emissions modeling, key parameters representing real-world driving cycles were selected as measurements to compare driving performance of each driving cycle (i.e., real-world vehicle trips). The parameter selections were made based on searching relevant studies that characterize driving cycles, as shown in Table 2. The distance over total duration is total average speed, therefore this study did not include distance in the analysis. The parameters selected include:

- Parameters describing the range and average magnitude of driving activities: maximum acceleration, maximum deceleration, average deceleration, average acceleration, root mean squared acceleration, maximum speed, total average speed, driving average speed, total cycle duration, and driving duration.
- Parameters representing time use during a trip: percent of time spent on idling, percent of time on acceleration, percent of time on deceleration, percent of time on cruise control.
- Events parameters: average number of acceleration/deceleration events per mile, and kinetic intensity.
- Volatility parameters to capture how drivers instantaneous driving decision changes during a trip: percentage of extreme acceleration or deceleration time (also called acceleration volatility score), maximum positive vehicular jerk (derivative of acceleration rate), average positive vehicular jerk, maximum negative vehicular jerk, average negative vehicular jerk and percentage of extreme vehicular jerk time (jerk volatility score). The calculation of volatility score using acceleration and

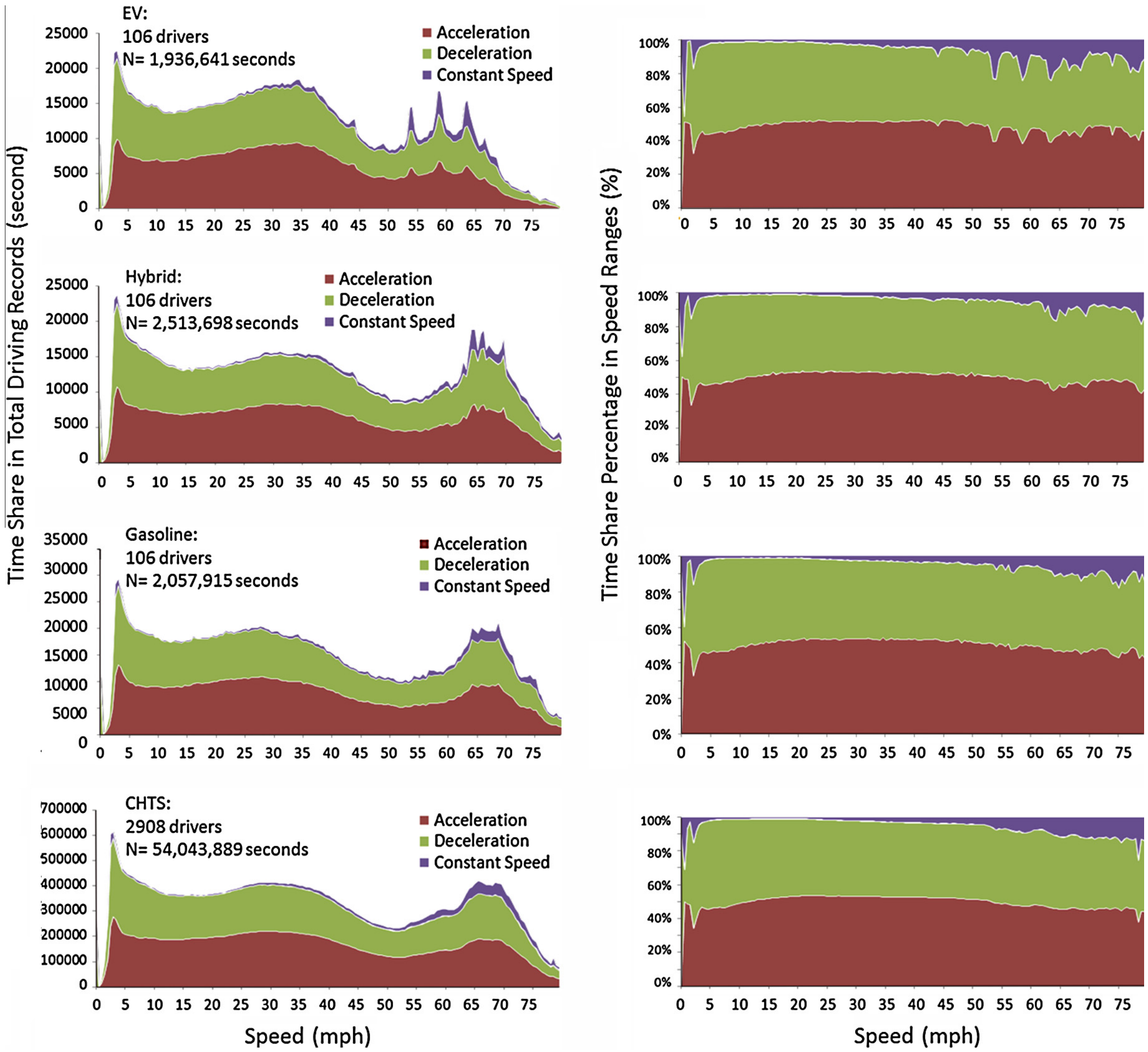

Fig. 2. Comparisons of acceleration–speed across time use.

jerk is based on previous studies (Liu et al., 2014, 2015a; Wang et al., 2015). Note that the driving volatility score is defined as the percentage of outlier acceleration events or vehicular jerk events during one trip. The threshold for identifying outlier events is established based on all 54 million seconds driving records collected in CHTS; it is two standard deviation above or below the mean acceleration or jerk for speed bins that range from 0 mph to 80 mph.

5.3. Are conventional and alternative fuel vehicles with driving cycles different?

Table 3 shows a comparison of 23 parameters used to quantify driving behavior and driving cycles at the trip level. Key differences between real-world driving trips of BEVs, HEVs and CVs in three equivalent groups include:

- Trips undertaken in BEVs are shorter (statistically significant at 95% level by a *t*-test), both in terms of total duration and driving duration, compared with HEV- and CV-trips made by similar drivers.
- BEVs have a statistically lower total average speed and driving average speed compared with HEVs and CVs.
- The average for maximum speeds of BEV trips is near 50 mph, which is lower than HEVs and CVs.
- BEV trips show higher average acceleration compared with HEVs, but lower than CVs. Maximum acceleration for BEVs is higher than that of HEVs and CVs; also BEVs are associated with higher variance in acceleration. However on average, BEV trips show lower magnitudes of average deceleration and (average) maximum deceleration compared with HEVs and CVs.
- Average vehicular jerk level is similar for BEVs, HEVs and CVs. But BEV drivers have higher maximum positive vehicular jerk.

Table 2
Parameter selection.

	EPA (1995)	Tong et al. (1999)	Q. Wang et al. (2008)	Kamble et al. (2009)	Brady and O'Mahony (2013)	Kent et al. (1978)	Lyons et al. (1986)	This study
Distance	✓				✓	✓	✓	
Total duration	✓			✓	✓	✓		✓
Driving duration	✓	✓		✓	✓	✓		✓
Total average speed	✓	✓	✓	✓	✓	✓	✓	✓
Driving average speed		✓	✓					✓
Maximum speed	✓				✓		✓	✓
Average acceleration	✓	✓	✓	✓	✓		✓	✓
Average deceleration	✓	✓	✓	✓	✓		✓	✓
Maximum acceleration	✓			✓	✓		✓	✓
Maximum deceleration	✓			✓	✓		✓	✓
Root mean square accel./decel.		✓	✓			✓	✓	✓
Average positive vehicular jerk	✓						✓	✓
Average negative vehicular jerk								✓
Maximum positive vehicular jerk								✓
Maximum negative vehicular jerk								✓
Root mean square jerk								✓
Acceleration/deceleration events		✓	✓					✓
Percent time on idling	✓	✓	✓	✓	✓	✓	✓	✓
Percent time on acceleration		✓	✓	✓	✓			✓
Percent time on deceleration		✓	✓	✓	✓			✓
Percent time on stable driving		✓	✓	✓	✓		✓	✓
Percent time on extreme accel./decel.								✓
Percent time on extreme veh. jerk								✓
Kinetic intensity/energy/power	✓	✓	✓				✓	✓

- The average acceleration/deceleration events per mile are similar for BEVs, HEVs and CVs.
- BEV drivers spend similar times on idling compared with HEVs and CVs. But spend more time on stable driving.
- The percent time spent on extreme acceleration or deceleration, or vehicular jerk is lower for trips taken in BEVs compared with HEVs and CVs. The finding is consistent with our previous study that compared extreme driving events (also called outlier events) between BEVs, HEVs and CVs (Liu et al., 2015b).
- BEV group shows similar kinetic intensity level compared with HEV and CV groups.

For the same driver, the driving performance may vary when they drive a HEV compared with a CV. For the same driver and same vehicle, driving performance may also vary under different driving conditions, e.g., free flow vs. congested traffic conditions. However, such information cannot be extracted from the cross-sectional data analyzed in this study. Instead, this study analyzes differences in driving performance between BEVs, HEVs and CVs and associates the performance with driver factors (e.g., driving behaviors), vehicle features (e.g., body type, engine, wheel drive), and the driving environments.

Since the comparison was made between equivalent groups (based on driver demographics), it is expected that the BEVs, HEVs, and CVs will have some similar driving characteristics, such as vehicular jerk. Even so, substantial differences can still be observed, in terms of, for instance, speeds, accelerations and time spent on acceleration. Studies have found that speed and acceleration are highly correlated with the energy consumption and emissions (Holm et al., 2007). Therefore, using the same or similar driving cycles to estimate the fuel economy for BEVs, HEVs and CVs may be inaccurate for individual cases; driving cycles that are customized for BEVs, HEVs and CVs specifically may lead to a more precise fuel economy estimation.

The driving patterns for real-world trips were also compared with four EPA specified driving cycles as well as California Driving Cycle (LA92) and New York City Cycle (NYCC). As a result, significant differences are found between BEVs, HEVs, CVs and existing driving cycles. Key findings include:

- The total durations and driving durations of BEV- and HEV-trips are generally longer than most EPA specified driving cycles, but shorter than FTP and LA92.

Table 3
Comparisons of 23 real-world driving performance parameters calculated at trip level.

Vehicle groups	BEV (N* = 2371)		HEV (N = 2652)		CV (N = 2397)		Regional–all veh. (N = 65,652)		Existing drive cycles					
Drive cycle parameters	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	US06	LA92	FTP	HWY	SC03	NYCC
Total duration (h)	0.261	0.231	0.299	0.298	0.274	0.309	0.262	0.303	0.167	0.399	0.521	0.213	0.167	0.166
Driving duration (h)	0.225	0.210	0.260	0.273	0.236	0.286	0.227	0.282	0.154	0.334	0.421	0.211	0.134	0.108
Total average speed (mph)	26.890	10.912	28.068	12.611	27.798	12.156	27.281	12.375	47.968	24.608	21.200	48.204	21.441	7.090
Driving average speed (mph)	27.224	10.885	28.382	12.589	28.141	12.126	27.617	12.349	51.850	29.399	26.202	48.584	26.624	10.917
Maximum speed (mph)	49.303	15.828	51.961	17.826	51.451	17.105	50.224	17.428	80.300	67.200	56.700	59.900	54.800	27.700
Average acceleration (ft/s^2)	2.128	0.676	2.074	0.649	2.218	0.713	1.460	0.468	2.198	2.207	1.676	0.637	1.648	2.036
Average deceleration (ft/s^2)	−2.193	0.643	−2.242	0.679	−2.382	0.764	2.141	0.687	−2.390	−2.473	−1.891	−0.725	−1.982	−1.985
Maximum acceleration (ft/s^2)	9.339	2.237	8.844	1.838	8.818	1.923	−2.314	0.741	12.320	10.120	4.840	4.693	7.480	8.800
Maximum deceleration (ft/s^2)	−9.938	2.355	−10.255	2.468	−10.372	2.470	8.674	1.928	−10.120	−12.907	−4.840	−4.840	−8.947	−8.653
Root mean square accel./decel. (ft/s^2)	1.465	0.430	1.463	0.439	1.557	0.483	−10.129	2.495	3.237	2.611	2.070	0.981	2.261	2.206
Average positive vehicular jerk (ft/s^3)	0.770	0.287	0.767	0.296	0.802	0.300	1.509	0.468	1.322	1.248	0.780	0.283	1.024	1.414
Average negative vehicular jerk (ft/s^3)	−0.596	0.203	−0.602	0.201	−0.626	0.203	0.786	0.305	−1.224	−1.189	−0.663	−0.266	−0.801	−1.276
Maximum positive vehicular jerk (ft/s^3)	6.479	2.084	6.354	2.049	6.400	2.194	−2.894	0.764	11.147	9.533	5.133	2.933	6.307	8.213
Maximum negative vehicular jerk (ft/s^3)	−2.942	0.810	−2.919	0.725	−2.935	0.757	6.240	2.214	−8.653	−12.320	−3.813	−2.347	−4.107	−6.160
Root mean square jerk (ft/s^3)	0.691	0.184	0.686	0.186	0.710	0.190	−0.613	0.210	1.819	1.519	0.926	0.367	1.178	1.502
Acceleration/deceleration events (no. per mile)	16.903	14.390	16.859	14.550	16.837	15.328	0.695	0.195	16.729	10.896	9.561	2.242	15.642	39.444
Percent time on idling	20.64	13.06	20.00	13.02	21.03	13.46	20.85	13.93	11.15	24.58	23.84	1.57	24.46	51.75
Percent time on acceleration	37.89	6.82	39.50	6.84	38.97	7.21	39.10	7.33	44.09	34.96	37.28	43.86	40.27	24.87
Percent time on deceleration	40.71	9.27	39.75	8.64	39.25	8.85	39.26	9.21	39.27	28.76	31.47	38.12	31.45	21.87
Percent time on stable driving**	5.60	7.85	4.76	6.16	4.41	6.16	4.57	6.34	5.49	7.38	3.52	16.45	2.16	0.00
Percent time on extreme accel./decel.	4.46	3.75	4.69	3.96	5.59	4.77	5.15	4.52						
Percent time on extreme vehicular jerk	4.79	4.11	4.80	3.91	5.32	4.30	5.00	4.18						
Kinetic intensity	3.295	8.526	3.350	5.499	3.298	5.364	3.681	22.884						

* N = Number of sampled trips in each group.
** Stable driving was defined by speed above 30 mph and acceleration less than 0.088 (ft/s^2).

- In terms of the total average and driving average speed, real-world trips are close to LA92 but still show noticeable differences (the total average speed of BEVs and HEVs is higher than LA92 cycle but their average driving speed is lower than LA92 cycle).
- The average for maximum speeds of BEV trips is substantially lower than four EPA standard driving cycles as well as LA92, except NYCC.
- In terms of the acceleration and deceleration, the average values for real-world trips are close to LA92 and US06. However, the magnitudes of the maximum values for BEVs, HEVs and CVs are summarily smaller than LA92.
- Real-world trips seem to have close magnitudes of vehicular jerk values with FTP cycle, noticeably smaller magnitudes than LA92, US06 and NYCC and SC03, and relatively larger magnitudes than HWY.
- BEV-, HEV- and CV-trips have more acceleration/deceleration events per mile than LA, FTP and HYW, and much less events than NYCC.
- In terms of time uses on driving modes (idling, accelerating, decelerating and stable driving), real-world trips spend less time on idling and stable driving but more time on accelerating and decelerating than LA92 and NYCC.

Overall, after controlling for certain driver demographics, the results show that BEV trips are shorter and calmer as indicated by less driving volatility and more stable driving, and HEV trips are longer and calmer. None of the existing driving cycles represent BEV and HEV driving characteristics well.

6. Building blocks of driving cycle design

To develop driving cycles, most previous studies have collected on-road data to develop real-world driving cycles. The methods for generating driving cycles generally rely on one trace that is the closest to the target cycle. Note that, the target cycle is a trace with mean values of characteristics (e.g., average and maximum speed, average and maximum acceleration or deceleration) of all observed traces. This method has been applied by EPA for generating the Urban Dynamometer Driving Schedule (UDDS) (EPA, 1995), Tong et al. (1999) applied the technique for generating a driving cycle for Hong Kong. Q. Wang et al. (2008) applied it for developing driving cycles in nine Chinese cities; and Saleh et al. (2009) applied the technique for designing a driving cycle for motorcycles in Edinburgh.

This study uses a similar methodology of matching the observed and target cycles. However, unlike the abovementioned studies, this study breaks down a trip further into micro-trips, defined as continuous driving activities between two stops (or idling). Kamble et al. (2009) and Fincher et al. (2010) introduced the idea of developing driving cycles with micro-trips. Using micro-trips increases the possibility of driving cycle matching the target cycle; it also allows greater flexibility in generating various driving cycles using limited driving data. In addition to taking advantage of the methods proposed by previous studies, this study introduces the idea of customizing driving cycles to specific users, based on their characteristics. Existing driving cycles are not sensitive to user (driver) characteristics or vehicle types. The study develops a method of searching appropriate cases (i.e., micro-trips) from a large-scale driving database, according to the user inputs (their socio-demographics, vehicle type and usage, etc.). Then customized driving cycles are generated by chaining and ranking all possible cycles for the vehicle-driver combination.

6.1. Micro-trips

In current travel surveys, a trip is defined as individuals moving from an origin to a destination (travel from one address to another address, with physical separation). Focusing on vehicular trips, driving can be interrupted several times during one trip, e.g., stops at intersections or stopped by traffic congestion. This makes it possible to further separate out one single trip into several micro-trips. Each micro-trip consists of continuous driving, forming the building blocks (EPA, 1995). Drivers often idle between two stops (Kamble et al., 2009). Given each micro-trip is a driving activity without interruption; it shows more homogeneous driving characteristics than an entire vehicular trip. Therefore micro-trips can suitably become cases representing base elements of a complete driving cycle. Only when several micro-cycles are chained together, a complete driving cycle can be created. Therefore it is critical to create a collection of cases and then to design the mechanism of how micro-trips can be chained together. As mentioned previously, CHTS database contains a large number of samples, including 236,403 micro-trips from 65,652 trips. This provides a sufficiently large data source to develop micro-trip case systems. It also allows us to learn how micro-trips are chained together for a complete trip.

6.2. Micro-trip clustering

After extracting the 23 driving parameters to quantify driving characteristics, qualitative analyses are needed for better structuring the micro-trips in the case system so they can be ready to select as elements for driving cycle design. To this end, rigorous clustering techniques were applied to group micro-trips based on various driving parameters extracted. The principle is to cluster similar micro-trips into one category meanwhile differentiating categories that are more different from each other.

To validate the clustering results, this study separated the data into two groups: experimental and randomly selected validation or hold-out group, before performing clustering. The micro-trips in the validation group were from 10% of 65,652 trips. The validation group has 24,024 micro-trips from the 6565 trips that were randomly selected. The remaining 212,379 micro-trips from the other 90% trips formed the experimental group. The micro-trips in the experimental group were first analyzed using the *K*-means clustering algorithm (Hartigan and Wong, 1979). The basic idea is: Given 212,379 observations (i.e., micro-trips) with a 23-dimentional real vector (i.e., 23 driving cycle parameters), *K*-means clustering aims to partition all observations into k ($\leqslant$212,379) clusters minimizing the within-cluster sum of squares. The objective function is:

$$J = \underset{k}{argmin} \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^{(j)} - c_j\|^2 \tag{1}$$

where $x_i^{(j)}$ = an observation (i.e., micro-trip) i in cluster j, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, k$. Note that x is a 23-dimentional real vector; n = the number of observations, equal to 212,379; k = the number of clusters, between 1 and 212,379; c_j = the center of cluster j; $\|x_i^{(j)} - c_j\|^2$ = the distance between an observation $x_i^{(j)}$ and the cluster center c_j.

Cubic Clustering Criterion (CCC) was used to compare the fit statistics of different clusters (SAS, 1983). Results shows that a 5-cluster structure has the best fit statistics (i.e., largest CCC = 0.62342). Fig. 3(a) shows the result of 5-cluster structure illustrated by micro-trip root mean square acceleration across micro-trip mean driving speed. Note that, since the micro-trip data are 23-dimentional (i.e., contains 23 parameters), the borders between clusters are not very clear when visualizing in two dimensions.

The clustering results were validated through the following steps:

(1) Using the 212,379 micro-trips as the response variable, and the 23 (driving performance) parameters as the explanatory variables a multinomial logistic (MNL) regression model was estimated (results are available from the authors).
(2) We then applied the MNL model estimates (i.e., coefficients of 23 parameters) to predict the possible clusters in which the 24,024 micro-trips in validation groups belong, as shown in Fig. 3(b) (also reproduced below).
(3) Next we obtained the number of clusters by clustering the 24,024 micro-trips based on the 23 parameters (which is similar to the procedure applied for the experimental group), as shown in Fig. 3(c).
(4) Finally, we calculated the percentage of correctly predicted observations in Fig. 3(b), compared to clusters in Fig. 3(c).

Fig. 3(b) and (c) shows that the clusters predicted based on experimental group are very similar to results of directly clustering micro-trips in the validation group. Specifically, 86% of observations in Fig. 3(b) and (c) match in terms of cluster

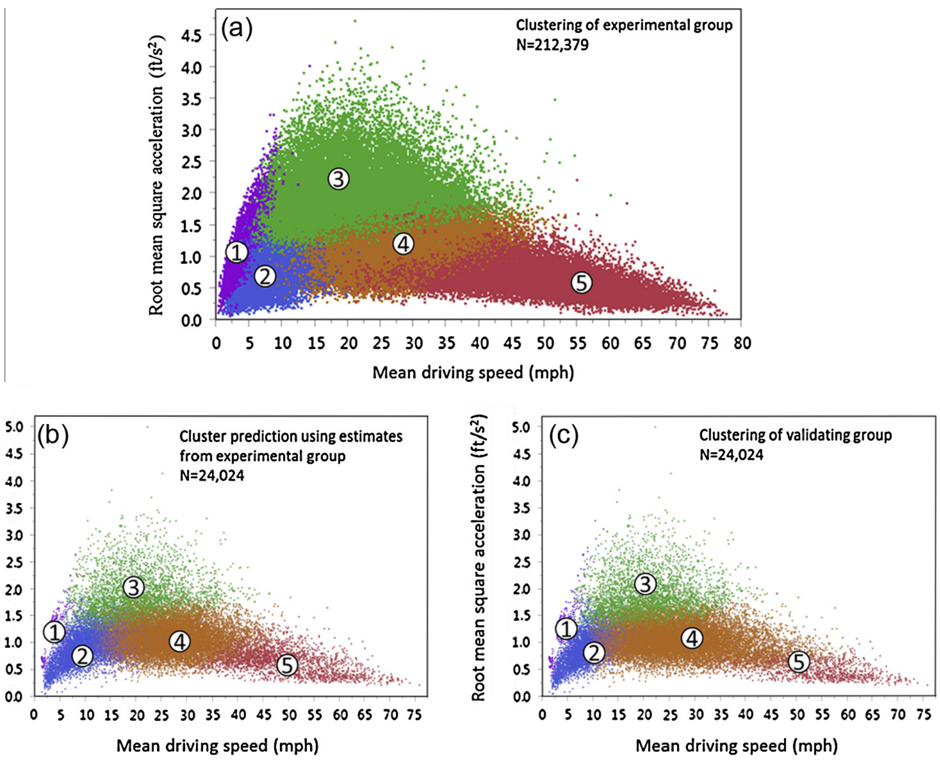

Fig. 3. Micro-trip clustering and validation (total N = 236,403).

number, indicating that there is approximately 86% chance that the clustering results of using micro-trips in the experimental group can predict the cluster of the validation micro-trip.

Sensitivity analysis was also conducted by choosing different clusters (e.g., 4-clusters, 6-clusters) and giving different sets of parameters. The empirical results showed that, in a 4-cluster structure, Cluster-1 and Cluster-5 remain the same, while Cluster-3 is partitioned into Cluster-2 and Cluser-4. In a 6-cluser structure, only Cluster-1 is divided into two clusters and other clusters remain the same. In general, the cluster structure is relatively stable in a 5-cluster structure.

The clustering with different sets of parameters was done by including all parameters but ''no speed-related factors'', ''no acceleration-related factors'', ''no vehicular jerk-related factors'', or ''no time-use factors''. The results showed that each set of parameters were associated with a different cluster structure, indicating all these factors have significant contributions in distinguishing between micro-trips. However, it is unclear how much information a parameter contributes to separating/distinguishing between two clusters. Therefore, this study applied Principal Component Analysis (PCA) to find out what parameters contribute more or provide equivalent information in characterizing micro-trips.

6.3. Characterizing micro-trips

To characterize micro-trips in clusters, PCA (Principal Component Analysis) was applied. PCA is capable of sorting out parameters that are more influential in forming a micro-trip cluster. PCA provide smaller independent linear combinations (principal components) of 23 variables. Fig. 4(a) shows the 5-cluster structure illustrated by first two principal components, explaining sizeable variance across observations at 36.9% and 21.8%, respectively. The borders between clusters are clearer than those in Fig. 3(a). The parameters that have a large weight on the first two components are used to characterize micro-trips in clusters. Fig. 4(b) presents the weight of each parameter on the first two components through their load matrix. Predictors having similar weights on the same principal are highly correlated, e.g., driving average speed, total average speed and maximum speed are highly correlated. Percent times on acceleration and on idling have the largest magnitudes (positive and negative) of weights on the first principal component. Root mean square acceleration and average deceleration have the largest weight on the second principal component. Therefore, these parameters can be used to characterize the four micro-trip clusters. Besides these four parameters, two intuitive parameters, trip duration and maximum speed, were also selected to represent the characteristics of five clusters. Table 4 presents the information of key parameters of micro-trips in five clusters. Note that, numbers in Table 4 are mean values of corresponding measures of all micro-trips in one cluster. Fig. 5 shows a sample trip containing five different micro-trips identified and labeled by the corresponding cluster number.

The following observations are in order regarding chaining micro-trips:

- Micro-trips in Cluster 1 have the lowest maximum speed and shortest duration, together with longer idling times (partly because of their low speed and short duration).
- Cluster 2 micro-trips also have a low speed but the speed is higher than Cluster 1. The duration is also longer than in Cluster 1. Clusters 1 and 2 micro-trips usually are the start or end leg of a trip and involve substantial local driving.
- Cluster 3 micro-trips have higher speeds than Clusters 1 and 2, but their speeds are still lower than 40 mph with medium levels of idling time. Cluster 3 has the highest average acceleration/deceleration among the five clusters.
- Cluster 4 micro-trips have higher speeds than the first three clusters. Clusters 3 and 4 micro-trips are mostly driven on collectors under different driving conditions.
- Cluster 5 micro-trips have the highest average speed, limited idling, largest deceleration and longest durations. The micro-trip represents driving on major arterials including freeways.

With clustered micro-trips and their general characteristics, we can encode each driving cycle: 1 denotes the cluster with the lowest average speed and 5 denotes the cluster with the highest average speed. Then a trip can be represented by a code sequence indicating the order of micro-trips in the chain. For instance, the trip shown in Fig. 5 has a code sequence 24351.

6.4. Case based system for driving cycle design

A Case Based System for Driving Cycle Design (CBDCD) is developed as a computer-aided machine learning tool. The system can take advantage of advanced modeling techniques to review, rank and synthesize micro-trip cases into a customized driving cycle by taking into account the qualitative (micro-trip cluster) and quantitative (performance parameter) information for each micro-trip. The designed driving cycle is selected by the degree of similarity between cases and the input. This methodology has the advantage of retaining the richness of historical big data of individual micro-trip cases, synthesizing new candidate driving cycles from existing cases, and eventually finding the best candidate driving cycle closest to the input from the user.

Fig. 6 shows the framework of the CBDCD. The CBDCD can be used to design two types of driving cycles:

(1) A customized driving cycle based on user information, given the user provides detailed information about self, such as demographics, commute trip information, vehicle fuel type, model, and year.
(2) A default driving cycle based on regional average if a consumer's information is not available in sufficient detail.

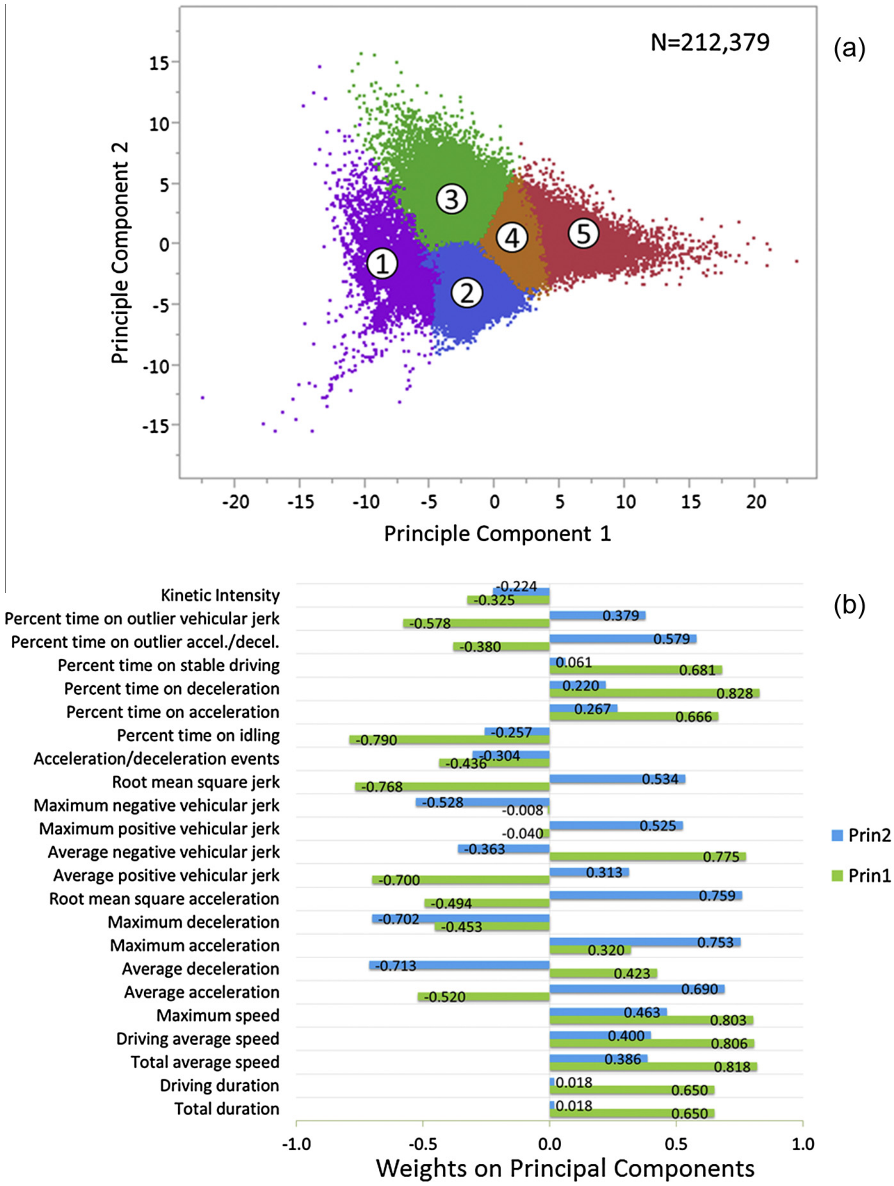

Fig. 4. Principal component analysis.

Table 4
Characterization of clusters.

Cluster #	Total duration (h)	Maximum speed (mph)	Percent time on idling	Percent time on acceleration	Root mean square acceleration (ft/s^2)	Average deceleration (ft/s^2)
1	0.002	4.047	93.11	2.11	1.035	−1.474
2	0.020	18.136	50.81	25.09	0.929	−1.410
3	0.018	33.791	42.76	29.63	1.896	−2.947
4	0.055	41.656	20.24	40.66	1.164	−1.819
5	0.314	69.866	6.24	42.46	0.654	−1.013

Key procedures of CBDCD are shown in Fig. 6 and detailed as follows:

(1) Establishing micro-trip library. This study used data from CHTS. The data were mainly documented into four files: GPS readings of every participating vehicle (speed, time series, vehicle number, etc.), trip descriptions (trip purpose, distance, time, etc.), vehicle information (vehicle model year, body type, fuel types, etc.) and driver socio-demographics

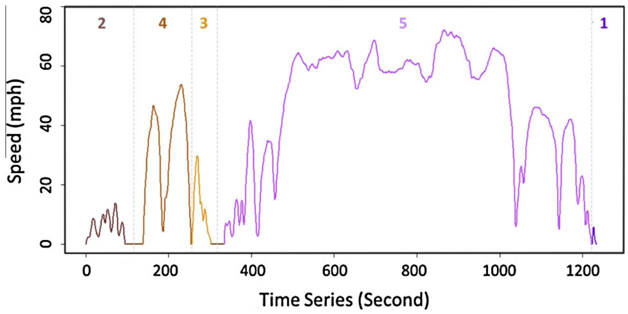

Fig. 5. Example of micro-trips clustered into a complete trip.

(driver gender, age, household income, etc.). All files were linked by vehicle identification numbers. Micro-trips are continuous driving between two stops (zero speeds), extracted from GPS readings and linked with vehicle and driver information. Micro-trips are clustered and coded as 1, 2, 3, 4, and 5, as demonstrated in Fig. 5. The vehicle and driver information are important user inputs that are required by CBDCD to customize driving cycles. The micro-trip library allows users to search for micro-trips that match the inputs, such as driver gender, age, vehicle body type and fuel type.

(2) Receiving user inputs. CBDCD allows users to specify three categories of information: driver (gender, age and household income), vehicle (fuel type, body type and vehicle age), and trip pattern (length or duration, and trip pattern). Note that, trip pattern is the code of micro-trips introduced above. Since this study found that the maximum speed is a critical parameter that features micro-trips, users may simply define the micro-trips based on speed limits of each road segments they plan to use. For example, the code 24532 represents that a trip where the driver may first take a residential road (speed limit – 20 mph), then an arterial road (speed limit – 45 mph), next an interstate segment (speed limit – 60 mph), followed by a distributor road (speed limit – 35 mph), and lastly a residential road. This is based on a user's knowledge of their trip pattern. If they do not know their trip pattern, then users may skip the trip pattern input. They can also skip other inputs as well, and default values will be used by the system.

(3) Searching for micro-trips. According to user inputs, CBDCD searches all possible micro-trips that match user inputs. Note that, if any information required above is not provided by the user, then the search will have a wider scope. For example, if no gender information is provided, then micro-trips of both male and female drivers in the database will be included in the search results, provided other information matched. However, the final output – the customized driving cycle, may be less personalized in such a case. Additionally, if no micro-trips in the library match user inputs, warnings are given. For example, if the driver age input is 25 years old and fuel type is electric vehicle, then the CHTS database may not return a match because it may not have 25 year olds owning/using an electric vehicle. At this point, users can continue to search micro-trips using information of driver age only, fuel type only, or terminate their search. With increasing amounts of data integrated in the library, this problem may go away and micro-trips can be found to match almost all feasible user input combinations.

(4) Chaining of micro-trips. If the trip-pattern is given by users, CBDCD chains micro-trips found from the library. If trip-pattern is given as 24532, then CBDCD randomly selects micro-trips coded (in the library) as 2, 4, 5, 3 and another 2, and then chains them together according to the order of the codes provided by users. Only chained micro-trips that have total durations within the range of trip duration given by users (±5 min) are kept as candidate driving cycles. The micro-trip chaining is repeated thousands of times (if the computational resources are available). As a result, thousands of candidate driving cycles can be generated.

(5) Ranking candidate driving cycles. Having candidate driving cycle generated, CBDCD characterizes them by the 23 driving parameters that use second-by-second GPS readings. The parameters are linked to the selected micro-trips. All candidate driving cycles are ranked by similarity scores. The similarity score is based on the sum of relative error between the 23 parameters of the candidate cycles and mean values of the 23 parameters of all candidate cycles. A cycle with mean values of the 23 parameters of all candidate cycles is called the target driving cycle. Given the values of target cycle parameters, the relative error of each parameter of the candidate cycle is calculated as follows:

$$\varepsilon_{\beta k} = \left| \frac{(M_{\beta k} - \overline{M}_k)}{\overline{M}_k} \right| \times 100\% \tag{2}$$

where $\varepsilon_{\beta k}$ = the relative error for the kth parameter (e.g., total average speed) of the candidate cycle β, β is the number of candidate cycles and $k = 1, 2, \ldots, N$. N is the total number of driving cycle parameters; $M_{\beta k}$ = the magnitude of the kth parameter of the candidate cycle β. $\overline{M}_k$ = the magnitude of the kth parameter of the target cycle.

Then, the similarity score of a candidate cycle can be calculated as follows:

$$S_\beta = 100\% - \frac{\sum_{k=1}^{N} \varepsilon_{\beta k}}{N} \tag{3}$$

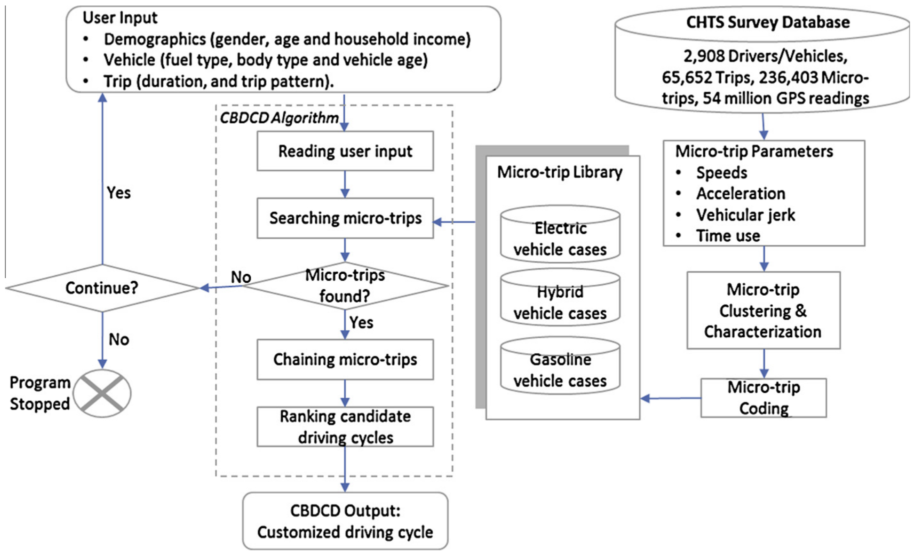

Fig. 6. Work flow of case based system for driving cycle design.

where S_β is the similarity score for candidate driving cycle β. The similarity score is calculated by using 100% minus the average relative errors coming from the parameters of candidate driving cycle β. The score ranges from zero to 100% (100% means that there are no errors and the candidate cycle matches with target cycle perfectly). The candidate cycle with the largest score is the best driving cycle that matches with the provided input information to the largest extent. Note that, the cycle parameters are weighted equally in the calculation. It will be reasonable to consider weighting the parameters differently in the calculation of similarity score. For instances hard accelerations or braking, captured by maximum acceleration or deceleration, may be weighted more than other parameters when calculating the sum of relative errors, considering their influences on fuel economy. The driving cycle that best matches with the parameters of the target cycle is selected as the final driving cycle.

(6) Displaying CBDCD outputs. Users can choose to display the final driving cycle and save second-by-second trajectory data for fuel economy and emissions estimation.

Fig. 6 shows the work flow of CBDCD. Specifically, CBDCD has the flexibility of generating different types of driving cycles according to the specification provided through user inputs. If no information entered, then CBDCD generates a driving cycle that is a default and it may be the average for a whole region. If only vehicle fuel type information provided (e.g., electric vehicle), then the driving cycle generated may serve as a regional cycle for vehicles of this fuel type in CHTS region. More information provided by the user generates more customized driving cycles.

7. Case study and validation

A case study of creating regional driving cycle for vehicle by different fuel types was conducted using the CBDCD system. Fig. 7(a) presents representative driving cycles for BEVs, HEVs and CVs produced by CBDCD, given micro-trip pattern sequence of 25432, and trips lasting 15–20 min. There are thousands of possible micro-trip pattern combinations. Driving cycles with the same pattern code represent those driving trips have the same number of stops, similar time spent on acceleration and deceleration, and are undertaken on similar roads. Eventually, three representative driving cycles were selected for BEVs, HEVs and CVs.

Fig. 7(b) presents further comparisons of these three driving cycles given specified micro-trip patterns. A more precise comparison of driving performance between different types of vehicles can be made, given the same travel needs (i.e., trip duration and number of stops) reflected by identical micro-trip patterns. Results show that for such a micro-trip pattern (25432), BEV cycle has a significantly higher percentage of time spent on stable driving than HEV and CV. CV cycle has a highest percentage of time spent on idling and BEV has the smallest percentage. HEV cycle has the most acceleration/deceleration events while BEV has the least acceleration and deceleration events. As for the average root square acceleration and maximum acceleration, CV cycle has the largest magnitudes while HEV cycle has the smallest magnitudes. Note that, the comparisons using specified micro-trip patterns, are not consistent with the general comparisons shown in Table 1, indicating the driving performance varies across various micro-trip patterns.

Possibly, an even more customized driving cycle can be generated given more potential user information. For instance, to compare the driving cycle of high-income drivers. Annual household income greater than $150,000, and age is between 40

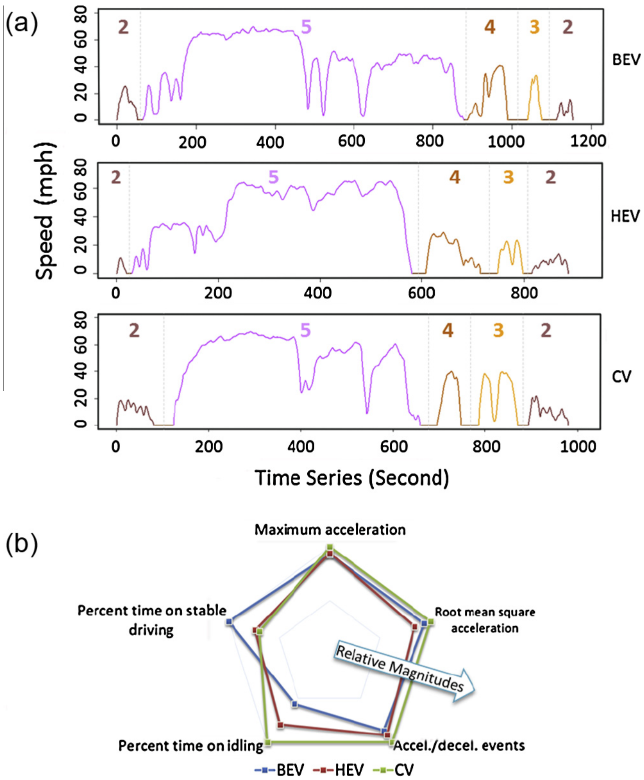

Fig. 7. Driving cycles given specified micro-trip patterns.

and 50 years were set as inputs; the system automatically searches the micro-trip database for sorting and ranking candidate micro-trip cases (by similarity scores). Fig. 8 shows driving cycles for target user groups making trips coded as 534 and 54.

A validation effort was undertaken to verify whether CBDCD can generate a driving cycle that matches a known driving cycle (called "trip" in the CHT Survey) in the validation group. Note that for the validation group, trips were randomly selected from 10% of trips in CHTS. One trip was randomly picked from the validation group for illustration. Fig. 9(a) presents the speed profile of this trip. The description of this trip can be obtained from CHTS database, as shown as follows:

- Vehicle/driver ID: 2692.
- Trip ID: 61214.
- Driver gender: Male.
- Driver age: 53 years old.
- Vehicle body type: Hatchback.
- Fuel type: Battery Electric.
- Vehicle age: 1 year.
- Household income: >$250,000.
- Trip duration: 28.6 min.
- Trip pattern: 354 (Note: This was roughly identified based on the number of stops and the maximum speed between two stops).

Then, above information was entered into the CBDCD as the specification for generating the customized driving cycle. Fig. 9(b) presents the driving cycle that was generated by CBDCD based on above information. Table 5 shows the relative errors between raw trip and CBDCD cycle. The mean error was 16.52%. If more information is available, such as driving style, the error might be lower. Additional validation work was done by repeating the method described above over 50 times. The mean error ranged from 15% to 30%. Note that parameters that are related to key factors (e.g., speed and acceleration) needed for fuel consumption estimation using Vehicle Specific Power methodology/equation (Andrei, 2001; Frey et al., 2007; Zhai et al., 2008) are associated with relatively smaller errors.

Although validation results show significant errors between real-world trips and the predicted cycle, these errors are much smaller than those between real-world trips with the pre-designed driving cycles (which range from 37% to 218%). Therefore, compared with the existing methods used (i.e., fixed driving cycles), the proposed method can more accurately

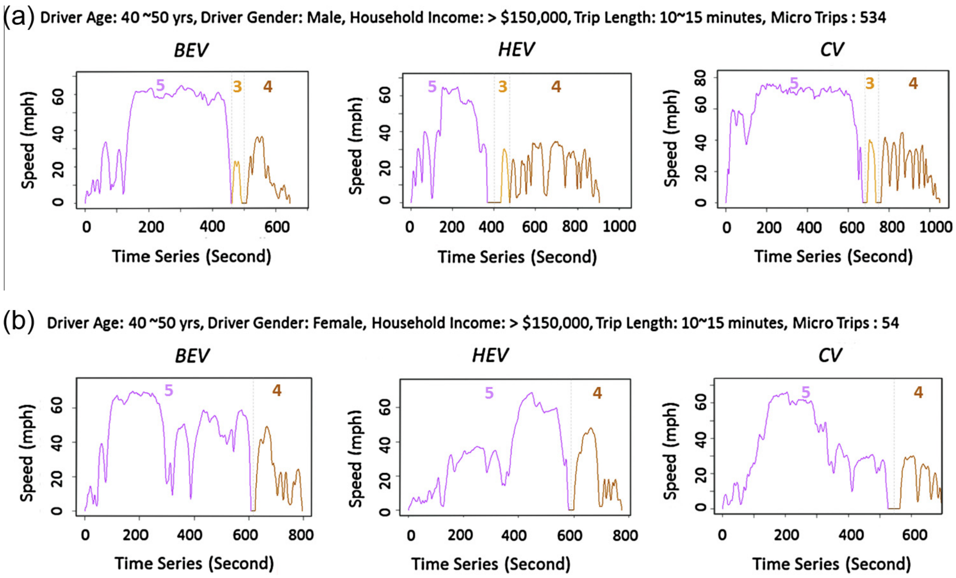

Fig. 8. Customized driving cycles for target user groups.

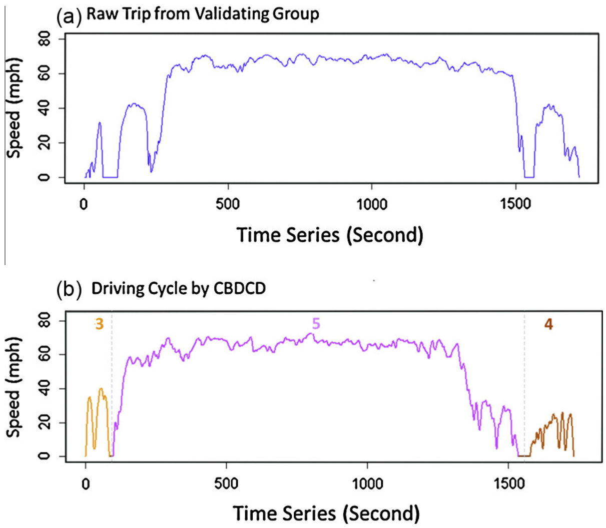

Fig. 9. Driving cycle validation.

predict driving cycles for individuals. Since the driving cycle plays a critical role in the vehicle fuel economy or consumption estimation, the driving cycle with smaller errors is likely to result in a more accurate fuel economy/consumption estimation for individuals.

Note that, this paper is not intended to change the current procedures of testing vehicle fuel economy/consumption by replacing the existing driving cycles designed by EPA. Currently, EPA uses the standard driving cycles to enforce legislation and require vehicle manufactures to comply with Corporate Average Fuel Economy (CAFE) and greenhouse gas (GHG) emissions regulations for light-duty vehicles (EPA, 2012). Therefore, using the standard driving cycles for fuel consumption/economy has some important advantages, e.g., the testing results are comparable across vehicle types and manufactures. The test results from the EPA standard driving cycles are also used to inform vehicle purchasers about expected fuel economy through window stickers. However, each sticker also contains a disclaimer stating that "Actual results will vary for many reasons, including driving conditions and how you drive and maintain your vehicle". The proposed customized driving cycles are meant to account for key factors associated with variations in driving patterns and hence more accurately predict actual fuel consumption for individuals.

Table 5
Relative errors between raw trip and CBDCD cycle.

Parameters	Raw trip	CBDCD cycle	Relative error (%)
Total duration (min)	28.633	31.317	9.37
Driving duration (min)	27.117	30.267	11.62
Total average speed (mph)	55.624	51.917	6.67
Driving average speed (mph)	56.000	53.718	4.08
Maximum speed (mph)	71.346	72.357	1.42
Average acceleration (ft/s^2)	0.581	0.645	10.99
Average deceleration (ft/s^2)	−0.593	−0.724	22.12
Maximum acceleration (ft/s^2)	7.059	7.100	0.57
Maximum deceleration (ft/s^2)	−8.669	−6.934	20.01
Root mean square accel./decel. (ft/s^2)	0.399	0.682	70.89
Average positive vehicular jerk (ft/s^3)	0.204	0.231	12.91
Average negative vehicular jerk (ft/s^3)	−0.203	−0.197	3.09
Maximum positive vehicular jerk (ft/s^3)	4.807	5.262	9.46
Maximum negative vehicular jerk (ft/s^3)	−2.407	−1.913	20.50
Root mean square jerk (ft/s^3)	0.292	0.212	27.45
Acceleration/deceleration events (no. per mile)	7.823	7.233	7.54
Percent time on idling	6.75	5.22	22.76
Percent time on acceleration	39.06	41.72	6.83
Percent time on deceleration	53.96	36.14	33.03
Percent time on stable driving	17.00	16.92	0.43
Percent time on extreme accel./decel.	1.89	2.48	30.79
Percent time on extreme vehicular jerk	2.93	2.59	11.57
Kinetic intensity	0.263	0.357	35.99
Average error			16.52

8. Fuel economy estimation

After customizing driving cycles for one type of vehicle given a target user group, fuel economy can be estimated specifically for this type of vehicles driven by a narrowed group of drivers. There are two options for using customized driving cycles to estimate fuel economy: (1) Applying Vehicle Specific Power (VSP) equation to calculate fuel consumption. VSP is defined as the instantaneous power per unit mass of the vehicle and is a function of vehicle speed, acceleration, road grade, aerodynamic drag, and tire rolling resistance (Andrei, 2001; Frey et al., 2007; Zhai et al., 2008). How to obtain fuel consumption using VSP equation can be found from many studies (Andrei, 2001; Frey et al., 2007; H. Wang et al., 2008; Zhai et al., 2008; Song et al., 2013); and (2) Using the customized driving cycles to predict MPG ratings based on dynamometer tests (Delucchi and Lipman, 2001; Lave and MacLean, 2002; André, 2004; Markel and Simpson, 2006; Fontaras et al., 2008) and vehicle simulation.

The current practice of using fixed (non-customized) driving cycles to provide estimates of fuel economy (e.g., 18 MPG for city and 22 MPG for highway, given a specific vehicle brand) may be unbiased, but imprecise (Lin and Greene, 2011). Using the customized driving cycles, customers considering AFVs can obtain more precise MPG or MPGe (equivalent MPG) because of the differences between driving cycles of BEV, HEV and CV fuel types. Thus customized driving cycles can provide more precise information to consumers when they are deciding which type of vehicles to purchase and how to use it.

9. Limitations

This study depends heavily on vehicle trajectory data collected by in-vehicle GPS. To some extent the accuracy and availability of location data constrain the analysis. Some other critical information from the survey remains unknown to the researchers due to privacy concerns, e.g., the geo-codes for the trips are not available. Missing geographically referenced information for trips prevents the researchers from extracting some contextual factors, e.g., whether the roadway used on a particular trip is a major arterial, a collector or a local road. Therefore, the micro-trip clustering completely depends on using driving performance data without considering the surrounding contextual factors. Availability of such information will further enrich the study. Further, driving style was not reported in the database; it can be a significant parameter to characterize driving cycles. Though data used for this study are large-scale, the study is still limited by its sample size and temporal–spatial constraints. For example, driving practices across seasons cannot be analyzed because drivers were observed only for a month or two. Additionally, economic impacts are not considered in the method proposed by this study. CBDCD has not been tested or verified/validated in real-life fuel economy estimation, which is the next step.

The price of gasoline can change substantially making AFVs less attractive in terms of fuel savings. Owing to the complexity of political, social, and economic factors involved, this issue is beyond the scope of this paper.

10. Conclusions and continuing research

Knowledge about how AFVs perform in real-world is important for assessing their real fuel economy and for the realization of their benefits in terms of fuel savings and emissions reduction. This study contributes by analyzing information about early adopters of AFVs. A critical component of fuel economy and emissions is the driving cycle. Given the shortcoming of using existing "one-fits-all" driving cycles for all types of vehicles, this paper creates a practical tool based on a comprehensive case-based reasoning system which can design customized driving cycles for individual users. The input information can be highly flexible, depending on different needs of users. The system takes advantage of emerging data mining and machine learning techniques to create driving cycles and relies heavily on a large-scale vehicle trajectory data collected in surveys.

Before proposing the methodology for customizing driving cycles, this study compares real-world driving performance of BEVs, HEVs and CVs in equivalent groups. Results show heterogeneous driving performance across these three fuel types. Further, the real-world driving performance is clearly different from the characteristics of existing standard driving cycles. Thus, customizing driving cycles based on large-scale real-world driving practices can improve the precision of fuel economy for individuals considering purchase of conventional or alternative fuel vehicles.

Given the large range of real-world driving performance by various drivers and vehicles, this study extracts information about micro-trips from the CHTS data, described by 23 driving performance parameters. The micro-trips are further grouped through machine learning techniques, such as principal component analysis and cluster analysis. Clustering of micro-trips helps separate a complete driving cycle into several building blocks, where individuals face various driving contexts, e.g., local roads, collectors, and major arterials. These micro-trip building blocks provide a collection of cases that form the basis of designing high-quality driving cycles. A Case Based System for Driving Cycle Design (CBDCD) is proposed by embedding the case collection with algorithms which have the capability to review, sort cases and eventually synthesize micro-trip cases into candidate driving cycles. A final driving cycle is selected based on similarity with driving characteristics and attributes of a specific user/consumer (user's age, gender, income, commute trip distance and duration, etc.). In this way, a cycle is customized to a user, and represents real-world driving performance. Fuel economy of a target vehicle can be estimated by applying VSP equation to the driving cycle or using dynamometer tests and simulations. Customizing driving cycles can improve precision and help transitioning from the standard EPA driving cycles by taking advantage of large-scale data and relatively new and novel computational/analysis methods.

The CBDCD can also provide default driving cycle design using regional average data without detailed consumer information. Auto manufactures can potentially use regional driving cycles to provide consumers with more precise estimation of fuel economy (customized to the region). This can still potentially help consumers understand the benefits of AFVs and help them make more informed vehicle purchase decisions (Greene, 2011; Greene et al., 2013).

In the future, a validation study is needed to evaluate the precision of fuel economy estimated using customized driving cycles through field tests. The micro-trip database should be expanded to cover other parts of the country. This study uses a database from the California household travel survey (Caltrans, 2013). Trajectory data from other regional surveys, e.g., Atlanta regional survey (Atlanta Regional Commission, 2011), can also be merged into the current micro-trip database; other data sources which are increasingly available publically or privately can also be merged, e.g., Naturalistic Driving Study (NDS) Data from the second Strategic Highway Research Program (Campbell, 2012). Alternative methods for estimating fuel economy (statistical or simulation based) should be explored. Further research should also do an in-depth exploration of the relationships between fuel economy, micro-trip patterns, vehicle fuel types, and user characteristics; and how the fuel economy estimated based on personal information influences vehicle purchase decisions.

Acknowledgements

The National Renewable Energy Laboratory provided the data used in this study. Statistical software JMP and R was used for machine learning (*K*-means Clustering and PCA). The research was supported through the TranLIVE University Transportation Center grant sponsored by the Research and Innovative Transportation Administration, of US DOT. Special thanks are extended to the following entities for their support at University of Tennessee: Transportation Engineering & Science Program and Initiative for Sustainable Mobility. The views expressed in this paper are those of the authors, who are responsible for the facts and accuracy of the information presented herein.

References

André, M., 2004. The ARTEMIS European driving cycles for measuring car pollutant emissions. Sci. Total Environ. 334, 73–84.
André, M., Joumard, R., Vidon, R., Tassel, P., Perret, P., 2006. Real-world European driving cycles, for measuring pollutant emissions from high-and low-powered cars. Atmos. Environ. 40, 5944–5953.
Andrei, P., 2001. Real World Heavy-duty Vehicle Emissions Modeling. West Virginia University.
Atlanta Regional Commission, 2011. Atlanta Regional Commission (ARC) Regional Travel Survey http://www.atlantaregional.com/transportation/travel-demand-model/household-travel-survey (accessed 08-01-2014).
Berry, I., 2010. The Effects of Driving Style and Vehicle Performance on the Real-World Fuel Consumption of U.S. Light-Duty Vehicles. Massachusetts Institute of Technology.
Brady, M.J., O'Mahony, M., 2013. The development of a driving cycle for the greater Dublin area using a large database of driving data with a stochastic and statistical methodology. In: Irish Transport Research Network Conference. Trinity College, Dublin, Ireland.
Bunch, D.S., Bradley, M., Golob, T.F., Kitamura, R., Occhiuzzo, G.P., 1993. Demand for clean-fuel vehicles in California: a discrete-choice stated preference pilot project. Transport. Res. Part A: Policy Pract. 27, 237–253.
Caltrans, 2013. California Household Travel Survey http://www.dot.ca.gov/hq/tsip/otfa/tab/chts_travelsurvey.html (accessed May 2nd 2014).
Campbell, K.L., 2012. The SHRP 2 Naturalistic Driving Study: Addressing Driver Performance and Behavior in Traffic Safety. TR News.
Davis, S., Diegel, S., Boundy, R., 2009. Transportation Energy Data Book: Edition 28, ORNl-6984. Oak Ridge National Laboratory, Oak Ridge, Tennessee www.cta.ornl.gov/dataS.

Delucchi, M.A., Lipman, T.E., 2001. An analysis of the retail and lifecycle cost of battery-powered electric vehicles. Transport. Res. Part D: Transp. Environ. 6, 371–404.

EPA, 1995. Final Technical Report on Aggressive Driving Behavior for the Revised Federal Test Procedure Notice of Proposed Rulemaking. U.S. Environmental Protection Agency, Washington, DC.

EPA, 2013. Dynamometer Drive Schedules http://www.epa.gov/nvfel/testing/dynamometer.htm (accessed March 3rd 2014).

Fincher, S., Palacios, C., Kishan, S., Preusse, D., Perez, H., 2010. Modifying Link-Level Emissions Modeling Procedures for Applications within the MOVES Framework. DTFH61-09-C-00028. Federal Highway Administration, U.S. Department of Transportation, Washington, DC.

Fontaras, G., Pistikopoulos, P., Samaras, Z., 2008. Experimental evaluation of hybrid vehicle fuel economy and pollutant emissions over real-world simulation driving cycles. Atmos. Environ. 42, 4023–4035.

Fontaras, G., Samaras, Z., 2007. A quantitative analysis of the European Automakers' voluntary commitment to reduce CO_2 emissions from new passenger cars based on independent experimental data. Energy Policy 35, 2239–2248.

Frey, H.C., Rouphail, N.M., Zhai, H., Farias, T.L., Gonçalves, G.A., 2007. Comparing real-world fuel consumption for diesel-and hydrogen-fueled transit buses and implication for emissions. Transport. Res. Part D: Transp. Environ. 12, 281–291.

Greene, D.L., 2010. How Consumers Value Fuel Economy: A Literature Review.

Greene, D.L., 2011. Uncertainty, loss aversion, and markets for energy efficiency. Energy Econ. 33, 608–616.

Greene, D.L., Evans, D.H., Hiestand, J., 2013. Survey evidence on the willingness of US consumers to pay for automotive fuel economy. Energy Policy 61, 1539–1550.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. Appl. Stat., 100–108

Holm, P., Tomich, D., Sloboden, J., Lowrance, C., 2007. Traffic analysis toolbox, vol. iv: Guidelines for Applying Corsim Microsimulation Modeling Software.

Hung, W., Tong, H., Lee, C., Ha, K., Pao, L., 2007. Development of a practical driving cycle construction methodology: a case study in Hong Kong. Transport. Res. Part D: Transp. Environ. 12, 115–128.

Kamble, S.H., Mathew, T.V., Sharma, G., 2009. Development of real-world driving cycle: case study of Pune, India. Transport. Res. Part D: Transp. Environ. 14, 132–140.

Kent, J., Allen, G., Rule, G., 1978. A driving cycle for Sydney. Transport. Res. 12, 147–152.

Khattak, A., Kanafani, A., 1996. Case-based reasoning: a planning tool for intelligent transportation systems. Transport. Res. Part C: Emerg. Technol. 4, 267–288.

Khattak, A.J., Dahlgren, J.W., McDonough, P., 2006. Tools for supporting implementation decisions of intelligent transportation system. Transport. Res. Rec.: J. Transport. Res. Board 1944, 41–50.

Lave, L.B., MacLean, H.L., 2002. An environmental-economic evaluation of hybrid electric vehicles: Toyota's Prius vs. its conventional internal combustion engine Corolla. Transport. Res. Part D: Transp. Environ. 7, 155–162.

Lin, J., Niemeier, D.A., 2002. An exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle. Atmos. Environ. 36, 5759–5770.

Lin, J., Niemeier, D.A., 2003. Estimating regional air quality vehicle emission inventories: constructing robust driving cycles. Transport. Sci. 37, 330–346.

Lin, Z., Greene, D., 2011. Predicting Individual Fuel Economy. 2011-01-0618, SAE Technical Paper.

Liu, J., Khattak, A., Wang, X., 2015a. Creating indices for how people drive in a region: a comparative study of driving performance. In: 94th Annual Meeting of the Transportation Research Board Washington, DC.

Liu, J., Khattak, A., Wang, X., 2015b. The role of alternative fuel vehicles: using behavioral and sensor data to model hierarchies in travel. Transport. Res. Part C: Emerg. Technol. 55, 379–392. http://dx.doi.org/10.1016/j.trc.2015.01.028.

Liu, J., Wang, X., Khattak, A., 2014. Generating real-time driving volatility information. In: 2014 World Congress on Intelligent Transport Systems, Detroit, MI.

Lyons, T., Kenworthy, J., Austin, P., Newman, P., 1986. The development of a driving cycle for fuel consumption and emissions evaluation. Transport. Res. Part A: Gen. 20, 447–462.

Markel, T., Simpson, A., 2006. Cost-benefit analysis of plug-in hybrid electric vehicle technology. In: 22nd International Electric Vehicle Symposium.

Nesbitt, K., Sperling, D., 1998. Myths regarding alternative fuel vehicle demand by light-duty vehicle fleets. Transport. Res. Part D: Transp. Environ. 3, 259–269.

Ntziachristos, L., Samaras, Z., 2000. Speed-dependent representative emission factors for catalyst passenger cars and influencing parameters. Atmos. Environ. 34, 4611–4619.

Saleh, W., Kumar, R., Kirby, H., Kumar, P., 2009. Real world driving cycle for motorcycles in Edinburgh. Transport. Res. Part D: Transp. Environ. 14, 326–333.

SAS, 1983. SAS Technical Report A-108, Cubic Clustering Criterion. SAS Institute Inc., Cary, NC.

Song, Y.-Y., Yao, E.-J., Zuo, T., Lang, Z.-F., 2013. Emissions and fuel consumption modeling for evaluating environmental effectiveness of ITS strategies. Discrete Dyn. Nat. Soc.

Struben, J., Sterman, J., 2008. Transition challenges for alternative fuel vehicle and transportation systems. Environ. Plann. B: Plann. Des. 35, 1070–1097.

Tong, H., Hung, W., Cheung, C., 1999. Development of a driving cycle for Hong Kong. Atmos. Environ. 33, 2323–2335.

Turrentine, T.S., Kurani, K.S., 2007. Car buyers and fuel economy? Energy Policy 35, 1213–1223.

U.S. Environmental Protection Agency, 2012. Joint Technical Support Document: Final Rulemaking for 2017–2025 Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards. EPA-420-R-12-901, U.S. Environmental Protection Agency (EPA) Ann Arbor, MI.

Wang, H., Fu, L., Zhou, Y., Li, H., 2008. Modelling of the fuel consumption for passenger cars regarding driving characteristics. Transport. Res. Part D: Transp. Environ. 13, 479–482.

Wang, Q., Huo, H., He, K., Yao, Z., Zhang, Q., 2008. Characterization of vehicle driving patterns and development of driving cycles in Chinese cities. Transport. Res. Part D: Transp. Environ. 13, 289–297.

Wang, X., Khattak, A., Liu, J., Masghati-Amoli, G., Son, S., 2015. What is the level of volatility in instantaneous driving decisions? Transport. Res. Part C: Emerg. Technol. 58, 413–427.

Wikipedia, 2014. Alternative Fuel Vehicle http://en.wikipedia.org/wiki/Alternative_fuel_vehicle (accessed February 10 2015).

Zhai, H., Frey, H.C., Rouphail, N.M., 2008. A vehicle-specific power approach to speed-and facility-specific emissions estimates for diesel transit buses. Environ. Sci. Technol. 42, 7985–7991.