PERGAMON

Atmospheric Environment 36 (2002) 5759–5770

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

An exploratory analysis comparing a stochastic driving cycle to California's regulatory cycle

Jie Lin[a], Debbie A. Niemeier[b,*]

[a] *Center for the Environment, Harvard University, Cruft Lab 216, 19 Oxford Street, Cambridge, MA 02138, USA*
[b] *Department of Civil and Environmental Engineering, University of California, One Shields Avenue, Davis, CA 95616, USA*

Received 23 May 2002; received in revised form 11 August 2002; accepted 20 August 2002

Abstract

As the fundamental building block of the emissions estimation process, a driving cycle needs to be representative of real-world driving behavior. The driving cycle construction method becomes crucial for generating a representative driving cycle. In this paper, we revisit the Unified Cycle's (i.e., the LA92 driving cycle) construction method. The California Air Resources Board's Unified Cycle used a "microtrips" approach, a speed–acceleration frequency distribution plot, and a quasi-random selection mechanism to build the driving cycle. There is concern that the Unified Cycle does not reflect the true driving patterns due to the identified flaws in the construction methodology. Treating a driving trace as a stochastic process, we construct a new driving cycle (LA01) with the same driving data originally used to build the Unified Cycle. We then compare the two driving cycles with the sample data set with respect to the durations and intensities of the modal events. The new driving cycle is found to better replicate the modal events observed in the sample data. A comparison of average road power values between the sample data, LA01, and the Unified Cycle also confirms the effect of fine-scale driving on emissions. These differences result from the different construction approaches and can be expected to affect emissions inventory estimation.
© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Driving cycle; Microtrips; Modal events; Transition probability; Emissions inventory

1. Introduction

Generally speaking, vehicle emissions are estimated by multiplying an emission factor with an associated travel activity as part of the mobile emissions estimation process (e.g., MOBILE by US Environmental Protection Agency or US EPA, and EMFAC by California Air Resources Board or CARB). A model year emission factor is a base emission rate adjusted by various correction factors related to off-cycle conditions such as temperature, speed or soak time, and the type of fuel used.

*Corresponding author. Tel.: +1-530-752-8918; fax: +1-530-752-7872.

E-mail addresses: jielin@deas.harvard.edu (J. Lin), dniemeier@ucdavis.edu (D.A. Niemeier).

For EPA's MOBILE series models, the basic emission rates are calculated using the tailpipe emissions of a vehicle following the speed trace of the federal test procedure (FTP) driving cycle on a laboratory dynamometer In MOBILE6, adjustments for driving other than that encompassed in the FTP are derived using facility and level of service, which is a proxy for congestion, correction factors. For California's mobile emissions model EMFAC2000, emission rates are derived by testing a vehicle on a driving cycle known as the Unified Cycle (or the LA92 driving cycle) for the base emission rate and the Unified Correction Cycles for off-cycle speed corrections (California Air Resources Board, 2000).

The FTP was developed in the early 1970s using a speed-time trace intended to represent vehicle operations in the Los Angeles urban area for the purpose of

1352-2310/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S1352-2310(02)00695-7

measuring light-duty vehicle (LDV) exhaust emissions (Austin et al., 1993). This cycle was later adopted by US EPA and became the standard driving cycle for the purposes of (initially) new vehicle certification programs and (subsequently) emissions inventory development.

The Unified Cycle was built on chase car data known as the LA92 data, which were collected on the road network in the Greater Metropolitan Los Angeles area between April and May of 1992 using a refined chase car protocol (Austin et al., 1993, for a summary of the issues associated with chase car studies see Morey et al., 2000). The Unified Cycle was designed to address concerns related to the FTP's lack of high speed and high acceleration activities typically associated with contemporary driving patterns. This cycle has been used to derive basic emission rates in the California emission model, EMFAC2000.

Many studies have commented on the problems inherent in the existing driving cycles. Among the problems noted are the underestimation of acceleration and cruise activities between 64 and 80 $km\,h^{-1}$ and above 96 $km\,h^{-1}$ (US Environmental Protection Agency, 1995a), underestimation of the time spent in cold transient mode and thus the emissions (Venigalla et al., 1995), and overestimation of the time at stop and at cruise between 40 and 56 $km\,h^{-1}$ in the FTP (St. Denis et al., 1994). These problems may arise jointly or separately from two possible sources: the lack of representative driving data with respect to the driving behavior encapsulated within a cycle and/or a problematic cycle construction methodology. Most studies have focused on the representativeness of the driving contained within a given cycle, while a few, mostly European, studies have examined how construction methods can affect the ultimate composition of the driving cycle (e.g., Andre, 1996; Kuhler and Karstens, 1978; Milkins and Watson, 1983).

Recognizing the fundamental role of a driving cycle in the emissions inventory estimation practice, in this paper we examine the driving cycle construction methodology used to create the Unified Cycle. As part of this discussion we identify a number of limitations to the Unified Cycle construction method. We then construct a new driving cycle (denoted LA01) using a new stochastic approach developed in Lin and Niemeier (2002). We complete the paper by comparing LA01, the Unified Cycle, and the LA92 data used to construct both cycles.

2. History of the Unified Cycle

The Unified Cycle was developed using driving data collected in Los Angeles, California in spring 1992, known as the LA92 driving data. The driving data consisted of 102 runs and 100,709 seconds of route-based second-by-second speed traces recorded while following randomly selected vehicles (i.e., target vehicles) operating on a mix of routes designed to represent all travel occurring in Los Angeles (Austin et al., 1993). Data on the facility (roadway) type and congestion level (designated by six levels of service from A to F) were simultaneously recorded. "Composite" driving data, that is blended chase and target vehicle data, were both used to generate the speed–acceleration frequency distribution plot (SAFD) and the driving cycle. Forty-seven percent of the data came from the chase car and 53% from the target vehicle.

To construct the cycle, the sample data were first divided into "microtrips" bounded by two consecutive stops (Austin et al., 1993). Although chase car data only accounted for about half of the sample data, they were used to generate the set of microtrips and the target vehicle data were ignored. A total of 833 microtrips resulted from the sampling of 102 different routes.

The microtrip selection mechanism was crucial to the cycle building procedure. The Unified Cycle utilized a "quasi-random" approach for microtrip selection (Austin et al., 1993). First, "seed" microtrips were selected completely at random to complete a seeding period as part of the first 120-s start period. Subsequent microtrips were then randomly selected such that they improved the match to the sample's SAFD. Each time a microtrip was selected, it was removed from the microtrip set. The remaining microtrips were scanned and subsequent microtrips were again selected such that they improved the match of the cycle's SAFD to the sample's SAFD. This procedure repeated until a pre-determined driving cycle time was achieved. A driving cycle constructed this way is not entirely a random combination of microtrips (i.e., quasi-random) because the microtrips were, instead of at random, selected to incrementally improve the cycle SAFD, and therefore this reduces the total number of all possible combinations of microtrips if otherwise selected at random (Austin et al., 1993). At the end of the process, some 18,000 cycles were created by repeatedly combining a subset of the microtrips such that each cycle was approximately 20 min long. The final driving cycle (i.e., the Unified Cycle) best matched the SAFD of the entire data set within 22% of the sum of differences (Austin et al., 1993).

Fig. 1 shows the speed profile of the final Unified Cycle and Table 1 compares it with the sample dataset with respect to global statistics such as average and maximum speed, maximum acceleration rate, and percent idle time. The Unified Cycle's average speed (39.6 $km\,h^{-1}$), maximum speed (108.1 $km\,h^{-1}$) and acceleration (+3.1 $m\,s^{-2}$) are consistently lower than those observed in the sample (45.7 $km\,h^{-1}$, 129.2 $km\,h^{-1}$ and 3.6 $m\,s^{-2}$). The Unified Cycle has about 16% of idle time, higher than that represented in the sample data (11.8%).

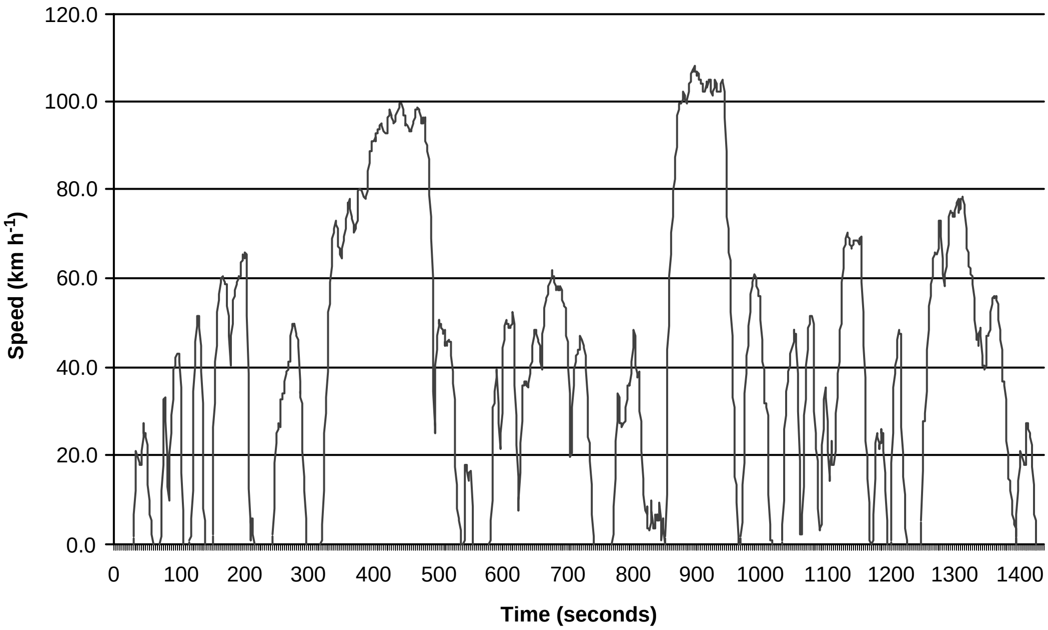

Fig. 1. Speed profile of the Unified Cycle.

Table 1
Basic statistics of the Unified Cycle

	Unified Cycle	Sample data
Avg. speed ($km\,h^{-1}$)	39.6	45.7
Max. speed ($km\,h^{-1}$)	108.1	129.2
Max. accel ($m\,s^{-2}$)	3.1	3.6
Idle (%)	16.4	11.8
Stops per mile	1.5	1.3

Fig. 2 displays the percent differences in SAFD between the Unified Cycle and the sample data set. A positive value indicates a higher frequency observed in the Unified Cycle and a negative value denotes a higher frequency observed in the sample. The plot shows that the Unified Cycle has 2.5% more occurrences at speeds and accelerations close to zero than observed in the sample data, which is consistent with the higher percent of idle time in the Unified Cycle shown in Table 1. The maximum difference in the remaining areas is no >0.8 percentage points, indicating a fairly good match.

A major concern related to the microtrip selection procedure is that the microtrip approach does not replicate a modal activity's actual frequency, duration and intensity. For example, suppose there are two microtrips: one follows an average driving pattern and the other contains some irregular driving. We further assume that both microtrips equally incrementally improve the cycle's SAFD during the course of cycle construction, meaning that adding either of the two microtrips to the driving cycle equally reduces the sum of differences between the cycle's and the sample's SAFDs, regardless of where the reduction in difference occurs. Theoretically speaking, there is an equal probability that either microtrip can be chosen using the Unified Cycle construction method. Intuitively, however, irregular driving rarely occurs in reality, and therefore is less likely to be presented in a driving cycle. The Unified Cycle construction method is unable to differentiate the two microtrips based only on the SAFD, because it does not differentiate between modal events (e.g., cruise, idle, acceleration and deceleration) within a microtrip nor does it consider the temporal interdependence of modal activities. The ability to identify and replicate particular modes of engine operation through a combination of speed and acceleration events is more useful and practical for improving estimation of vehicle emissions (An et al., 1997; Barth et al., 1996, 1997). Although there are different approaches that could be used to overcome these problems, the most straightforward is to develop the driving cycle based on the stochastic process it actually comprises.

3. Stochastic approach to driving cycle construction

In Lin and Niemeier (2002) we illustrated a new stochastic method developed for constructing a driving cycle. The stochastic method divides a speed trace into modal events (e.g., cruise, idle, acceleration, or

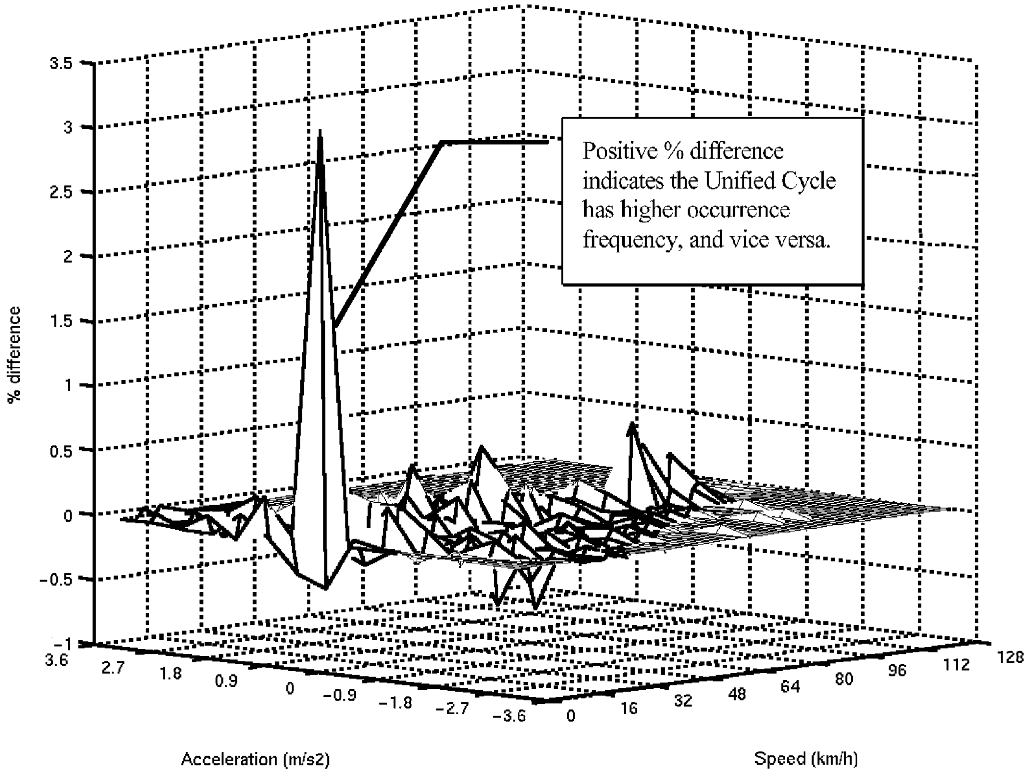

Fig. 2. Difference in SAFD between the sample data and the Unified Cycle.

deceleration) and describes the sequencing of those events using Markov process theory. That is, the occurrence of modal event k is dependent upon the previous modal event, $k-1$.

First, a route-based speed trace is partitioned into snippets of varying durations and intensities by using a maximum likelihood estimation (MLE) approach of mixture decomposition (Symons, 1981). The partitioning of a speed trace can be viewed as estimating an unknown cluster membership, x_i, for each of the instantaneous observations, y_i's (acceleration rates) within a speed trace of length n, where $i = 1, \dots, n$,

$$f(y_i|\vec{\theta}) = \sum_{g=1}^{G} \pi_g f(y_i|\vec{\theta}_g, x_i = g, \forall y_i \in C_g). \tag{1}$$

Here $f(y_i)$ defines the density function of observation y_i. G is the total number of clusters and for each cluster $g(g = 1, \dots, G)$, π_g is the probability of y_i being in cluster g. By definition $\sum_g \pi_g = 1$ and $P\{x_i = g\} = \pi_g$. C_g is the collection of observations, y_i's, in cluster g. The unknown parameter vector, $\vec{\theta}_g$, comprises of the mean and variance of the observations, y_i's.

Let $\vec{y}$ denote the vector of n observations y_i's, and $\vec{\theta}$ be the corresponding vector of parameters. Under the assumptions of multivariate normality and equal variance (σ_g) of the observations, y_i's, by cluster, the likelihood function derived from Eq. (1) can be written as

$$L(\vec{\theta}|\vec{y}) = \prod_{g=1}^{G} (\pi_g^{n_g} \sigma_g^{-n_g}) \exp\left\{ -\frac{1}{2} \sum_{g=1}^{G} \sum_{C_g} \frac{(y_i - \mu_g)^2}{\sigma_g^2} \right\}, \tag{2}$$

where n_g is the number of observations in C_g. The product of $\pi_g^{n_g}$ over G clusters is the likelihood that n_g observations are assigned to cluster $g(g = 1, \dots, G)$. Eq. (2) gives the likelihood that the ith observation $y_i, (i = 1, \dots, n)$, is assigned to cluster $g(g = 1, \dots, G)$.

After all the traces are segmented, the snippets are then clustered into modal event bins by applying the MLE algorithm once again. At this step the cluster variables include snippets' average, minimum, and maximum speeds and acceleration rates. A modal event bin, labeled by speed and acceleration, can contain hundreds, even thousands of snippets of similar speed and acceleration characteristics but differing event durations. All the modal event bins combined define a set of states and form the state space of a Markov process.

The modal event selection process in the new methodology employs the same SAFD matching mechanism as used in the current driving cycle construction methods. However, unlike the Unified Cycle method, which used the quasi-random microtrip selection technique, the sequence of modal events in the new method is determined by a transition probability defined in Eq. (3). That is, given the current modal operating event, the modal event with the highest transition probability is more likely to be chosen next.

$$P_{rs} = P\{Z_\tau = s | Z_{\tau-1} = r\}, \tag{3}$$

where $\{Z_\tau\}, \tau = 1, 2, \ldots, T$, is a stochastic process with a state space S, and $r, s \in S$. The physical interpretation of the transition probability p_{rs} is the probability of the next modal event being chosen from modal event bin s given that the current event is drawn from modal event bin r. The maximum likelihood estimate (MLE) of the transition probability, p_{rs}, is (Lee et al., 1970)

$$p_{rs} = \frac{N_{rs}}{\sum_s N_{rs}}, \tag{4}$$

where N_{rs} is the total number of modal events in bin s whose previous modal event is from bin r.

Once the transition matrix containing all of the transition probabilities from one modal event bin (row) to another (column) is computed, the next event bin number (i.e., state) is determined according to the distribution of the row transition probabilities (conditioned on the same state). This is realized by a random number generator that generates the next event bin number based on the probability distribution determined by the respective row transition probabilities. After the next event bin number is determined, a modal event of any length (in seconds) that improves the match to the SAFD of the sample data and matches the end speed of the previous modal event with its starting speed (typically within $\pm 0.2\,\mathrm{km\,h^{-1}}$) is chosen from that modal event bin. This allowed us to constrain any cycle differences to those due to the cycle construction method itself. The selection process is repeated until the desired cycle length is achieved.

Any number of driving cycles can be created during the construction procedure. Evaluation of the candidate driving cycles often involves examining a set of the cycle's global statistics, such as average speed, maximum and minimum acceleration, the second moment speed and acceleration as root mean square (RMS) values, and % idle. Watson et al. (1982) also used the positive kinetic energy (PKE) per unit distance as a measure of acceleration work for the Melbourne driving cycle. As an alternative to PKE, Gammariello and Long (1993) calculated the total driving power requirements involving the kinetic, road load, and grade power for the Unified Cycle and the FTP. Engine power is another measure formulated as a function of vehicle speed, acceleration, vehicle mass, and road grade angle (An et al., 1997).

In the new method, we continue to use average speed, maximum speed, maximum acceleration, and idle time (%), but we add road power as a surrogate for vehicle emissions. Road power captures the combined effects of rolling resistance, or friction, aerodynamic resistance, or air drag, and acceleration resistance (Stoeckenius et al., 2000). We use the following equations to calculate road power (provided by the Sierra Research, Inc.):

$$\begin{aligned}\text{Road power} = {} & (86.3 V_t + 0.0459 V_t^3 \\ & + 317(V_t - V_{t-1}) V_t)/1000, \quad \text{if} \quad V_t > V_{t-1}\end{aligned}$$

which is simplified to

$$\text{Road power} = (86.3 V_t + 0.0459 V_t^3)/1000, \quad \text{if} \quad V_t < \text{or} = V_{t-1} \tag{5}$$

where V_t is the instantaneous speed at time t. The cycle with the smallest performance value (PV) is chosen as the final representative driving cycle,

$$\begin{aligned}\mathrm{PV} = {} & w_1|\Delta\bar{v}| + w_2|\Delta\bar{a}| + |w_3\Delta v_{\max}| + w_4|\Delta v_{\min}| \\ & + w_5|\Delta a_{\max}| + w_6|\Delta a_{\min}| + w_7|\Delta\%_{\mathrm{idle}}| \\ & + w_8|\Delta\bar{P}_d| + w_9|\Delta v_{95}| + w_{10}|\Delta a_{95}| + w_{11}|\Delta P_{95}|,\end{aligned} \tag{6}$$

where $\bar{v}$ is the average speed, $\bar{a}$ the average acceleration/deceleration rate (mph/s), $v_{\max}$ the maximum speed (mph), $v_{\min}$ the minimum speed (mph), v_{95} the speed at the 95th percentile, $a_{\max}$ the maximum acceleration rate, $a_{\min}$ the minimum acceleration rate (including negative values), a_{95} the acceleration/deceleration rate at the 95th percentile, $\%_{\mathrm{idle}}$ the % idle time, $\bar{P}_d$ the average road power, P_{95} the road power at the 95th percentile, and $w_{1,2,\ldots,11}$ the weights.

The delta (Δ) differences in the equation are between the driving cycle and the sample chase car dataset. By assigning different values to the weighting coefficients ($w_1, w_2, \ldots, w_{11}$), the parameters can be weighted differently based on the importance of the parameter. In this particular study, all the weights were set to unity.

The benefits of this stochastic method are that partitioning of the modal events is an endogenous outcome of the speed profile without arbitrarily breaking a speed trace according to zero speeds or stops, and therefore maintains the integrity of a vehicle operation mode. The Markov modal events reveal the stochastic nature of the driving data. The ability to replicate modal events and driving variability is a significant contribution to the emissions estimation practice. Finally, the stochastic property of the new approach marks a fundamental difference between the existing driving cycle methods and the newly proposed method.

To demonstrate how the new method might impact cycle construction, we apply the new method to the LA92 driving data used to create the Unified Cycle using the two cycles, and compare the modal characteristics between the Unified Cycle, the newly created driving cycle (named LA01), and the sample data. LA01 and the comparison results are presented next.

4. Results

Six modal event bins were established during the creation of LA01 (Table 2). The number of bins requires a balancing between the number of bins and the data that are available to populate them. The majority of modal events (over 85%) are in the first three bins with relatively low average speeds ($<50\,km\,h^{-1}$). The average acceleration rates are nearly zero. A modal event bin corresponds to particular driving characteristics. For instance, Bin 6 events exhibit a relatively long average duration (181 s), a high average speed ($95.3\,km\,h^{-1}$), and virtually zero average acceleration rate with a small standard deviation ($0.13\,m\,s^{-2}$), suggesting that Bin 6 comprises mainly long high-speed steady-state events.

Table 3 is the transition matrix containing the transition probabilities of the event bins. For instance, given that the current state is Bin 1 ($8.7\,km\,h^{-1}$), the next most likely state is Bin 2 with an average speed of $30.3\,km\,h^{-1}$ and a transition probability of 57.9%. If the current state is Bin 2 ($30.3\,km\,h^{-1}$), the next likely state is to draw a modal event from either Bin 1 (48.9%) or Bin 3 (38.1%). The transition probabilities suggest that after a low-speed modal event is selected the next state from which an event may be selected is either acceleration to a higher-speed state or deceleration to an even lower-speed state, and that a high-speed state is generally likely to be followed by a lower-speed state.

At the end of the cycle construction process, the seven top candidate driving cycles were selected for evaluation. The cycles' performance statistics are listed in Table 4. The cycle with the lowest PV value was chosen to be the final LA01. A comparison between the sample and the cycles, and among the candidate cycles demonstrates an (expected) consistency in replicating the sample driving characteristics among the top candidate driving cycles. The PV value is a reasonable, although not perfect, performance indicator in identifying the representative driving cycles. We would also like to point out that the PV value is subject to those parameters cycle developers value the most in a driving cycle.

LA01 has similar speed and acceleration statistics with the Unified Cycle (Table 5), but much lower percent of idle time (8.9%) than the Unified Cycle (16.4%). Its percent of idle time is close to that observed in the sample (11.8%). LA01 has 96 modal events at an average duration 14 s compared to 16 microtrips at an average duration of 75 s in the Unified Cycle. The speed profile of LA01 is shown in Fig. 3. Visual comparison to the Unified Cycle shown in Fig. 1 suggests that LA01 contains a greater number of microtransient activities, which of course provides greater information about driving variability, meaning more frequent changes among modal events and generally shorter durations of the events. Quantitatively, there are 0.28 accelerations and decelerations per second in LA01 and 0.18 per second for the Unified Cycle. We believe that such a difference is caused by the different cycle construction methodologies—a modally based stochastic approach versus a non-modally based microtrip approach.

We plotted the percent differences in SAFD between LA01 and the Unified Cycle (Fig. 4). Positive difference values indicate higher frequencies observed in the Unified Cycle and negative difference values indicate higher frequencies observed in LA01. The largest difference in frequency between the two cycles is

Table 3
Transition matrix

Bin no. From	To						
	1	2	3	4	5	6	Total
1	29.5	57.9	12.7	0.0	0.0	0.0	100.0
2	48.9	6.4	38.1	6.5	0.0	0.0	100.0
3	16.4	45.6	16.1	15.8	5.2	0.9	100.0
4	0.5	37.4	47.1	0.0	5.1	9.9	100.0
5	0.0	4.9	56.9	16.7	2.8	18.8	100.0
6	0.0	0.0	17.6	33.3	40.2	8.8	100.0

Table 2
Summary of modal event bins in the sample data set

Bin No.	No. of events	Speed ($km\,h^{-1}$)		Acceleration ($m\,s^{-2}$)		Duration (s)	
		Avg.	Std. dev.	Avg.	Std. dev.	Avg.	Std. dev.
1	1515 (32.4%)	8.7	6.9	0.00	0.90	17	26
2	1375 (29.4%)	30.3	5.0	−0.13	1.18	12	16
3	1191 (25.4%)	47.8	6.0	−0.06	0.69	20	23
4	382 (8.2%)	63.9	3.9	−0.05	0.42	36	44
5	132 (2.8%)	74.8	4.5	−0.08	0.61	36	32
6	87 (1.9%)	95.3	6.9	−0.02	0.13	181	206

Table 4
Assessement of top candidate cycles' performance

	Duration (s)	Avg. speed ($km\,h^{-1}$)	Avg. accel. ($m\,s^{-2}$)	Max. speed ($km\,h^{-1}$)	Min. accel. ($m\,s^{-2}$)	Max. accel. ($m\,s^{-2}$)	% idle	Avg. road power (kW)	PV
Sample	100709	45.7	0.00	129.2	−6.79	6.91	11.8	5.17	
LA01	1381	40.9	0.00	114.9	−3.44	2.89	8.9	8.76	48.5
Cycle 1	1104	30.4	−0.01	72.9	−6.31	3.61	7.4	7.00	128.0
Cycle 2	1113	29.8	0.00	105.7	−5.15	2.90	24.4	6.98	90.4
Cycle 3	1099	28.7	−0.01	72.9	−5.01	4.31	6.6	6.19	126.7
Cycle 4	1172	40.9	−0.01	79.2	−3.62	3.97	5.2	7.64	111.3
Cycle 5	1112	32.6	−0.01	76.6	−5.01	3.61	8.4	7.49	117.6
Cycle 6	1122	32.8	−0.01	79.2	−5.01	3.61	6.8	7.52	116.9

Table 5
Basic statistics of the driving cycles

	Unified Cycle	LA01 cycle
Avg. speed ($km\,h^{-1}$)	39.6	40.9
Std. dev. speed ($km\,h^{-1}$)	31.7	31.5
Max. speed ($km\,h^{-1}$)	108.1	114.9
Std. dev. acceleration ($m\,s^{-2}$)	0.8	0.5
Max. deceleration ($m\,s^{-2}$)	−3.9	−3.4
Max. acceleration ($m\,s^{-2}$)	3.1	2.9
Idle (in % of time)	16.4	8.9
Cycle duration	1436	1381
Number of trip segments	16[a]	96[b]

[a] Number of microtrips.
[b] Number of modal events.

observed in the zone near zero speed and acceleration, where a positive value of about +2.5% indicates that much higher idle or quasi-idle activities are observed in the Unified Cycle. LA01 produces slightly more occurrences at high speeds ($>96\,km\,h^{-1}$) than the Unified Cycle. These seemingly minor differences may contribute to significant increases in emissions, an issue we will address later in the discussion section.

Given the importance of the vehicle operation modes to the emissions estimation practice, we conducted a modal event comparison among the sample data, LA01, and the Unified Cycle. A modal event is derived from MLE partitioning. For comparison purposes, we re-classified the modal events into three acceleration categories: average acceleration rates between $\pm 0.45\,m\,s^{-2}$, $> +0.45\,m\,s^{-2}$, and $< -0.45\,m\,s^{-2}$. Within each acceleration category, the modal events are further broken into eight speed classes each with an increment of $16\,km\,h^{-1}$. The reason for re-classification is to identify the differences in driving (i.e., frequency, duration, and intensity of a modal event) at finer scales between the sample data set, LA01, and the Unified Cycle. Furthermore, historical dynamometer data suggests that CO, VOC, and NO_x have the highest emission rates at the high acceleration mode and the lowest emission rates at the low deceleration modes (Cernuschi et al., 1995). In addition, previous research has shown that a modal event's frequency, duration, and intensity can also reflect emissions variability between drivers (Holmén and Niemeier, 1998).

Table 6 shows the sample data modal event distributions by speed–acceleration category. Over 90% of the modal events occurred at an average speed $<64\,km\,h^{-1}$. In particular, about one-fourth of the total modal events were observed at speeds of $<16\,km\,h^{-1}$; another one-fourth of the modal events had average speeds between 16 and $32\,km\,h^{-1}$. That means that half of the modal events were reflective of local street driving and/or driving in congestion, either of which is associated with high vehicle emissions. Approximately 65% of the modal events have the mildest acceleration rates between $\pm 0.45\,m\,s^{-2}$, meaning that the other 35% of the modal events were associated with relatively high acceleration and deceleration rates, potentially causing higher vehicle emissions.

To statistically quantify the significance of the differences in MLE modal event distributions between each of the two driving cycles and the sample data, we apply the two-sample Kolmogorov–Smirnov test between the sample data and LA01 cycle, and between the sample data and Unified Cycle. The null hypothesis of the tests is that two independent samples come from identical distributions (population) (Steel et al., 1997). An attractive feature of the K–S test is that the distribution of the K–S test statistic itself does not depend on the underlying cumulative distribution function being tested. Another advantage is that it is an exact test as opposed to the chi-square goodness-of-fit test that depends on an adequate sample size for the approximations to be valid (Chakravarti et al., 1967). The significance levels are 0.931 for the sample data—LA01 comparison and 0.261 for the sample data—Unified Cycle comparison (Table 7). Although the

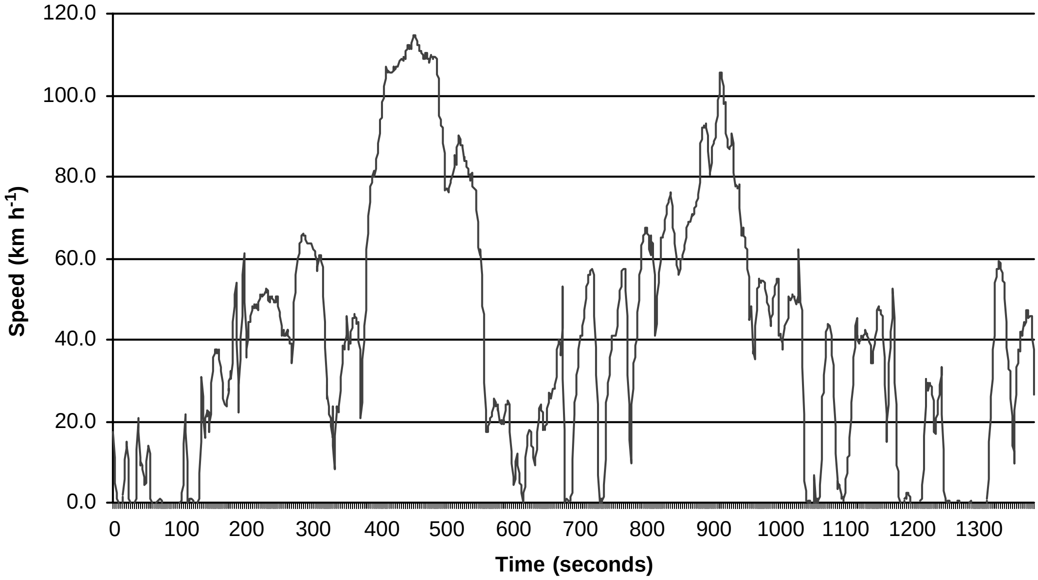

Fig. 3. Speed profile of LA01.

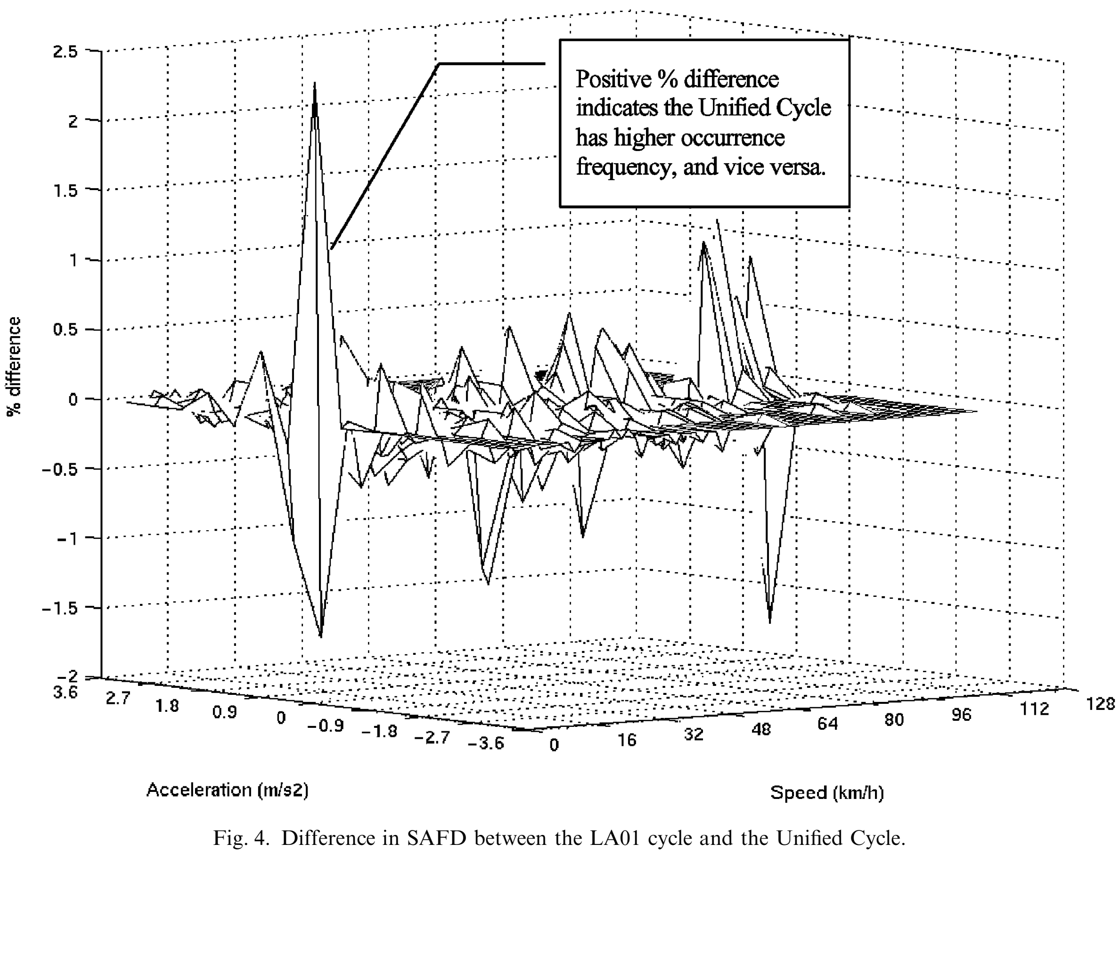

Fig. 4. Difference in SAFD between the LA01 cycle and the Unified Cycle.

statistics indicate that neither driving cycle is statistically different from the sample data with respect to their modal event distributions, a much higher similarity level is observed between the LA01 cycle and the sample data. This suggests that LA01 is a better fit to the sample data than the Unified Cycle.

Table 8 tabulates the average durations of all modal events at various acceleration categories for the sample data, LA01, and the Unified Cycle. For events with acceleration rates within $\pm 045\,\mathrm{m\,s^{-2}}$, LA01 replicates the average event durations better than the Unified Cycle, when the average event speed is $<80\,\mathrm{km\,h^{-1}}$. However, the duration variances become very large at speeds $>80\,\mathrm{km\,h^{-1}}$, suggesting that another factor, perhaps congestion, begins to play a role in defining the duration of modal events, because congestion is likely to result in greater variations in travel speed and the duration with which speeds can be maintained. For events at average acceleration rates $>0.45\,\mathrm{m\,s^{-2}}$ or $<-0.45\,\mathrm{m\,s^{-2}}$, the observed average duration is generally short even in the sample dataset. Again, both cycles replicate the average durations of modal events relatively well at lower speeds and poorly at higher speeds.

Table 6
Frequencies of modal events[a] in the sample data

Speed ($\mathrm{km\,h^{-1}}$)	$<-0.45\,\mathrm{m\,s^{-2}}$	$[-0.45, +0.45]\,\mathrm{m\,s^{-2}}$	$>+0.45\,\mathrm{m\,s^{-2}}$
<16	1.5	18.9	5.2
[16,32)	7.2	10.8	7.2
[32,48)	5.0	13.9	4.9
[48,64)	2.0	14.0	1.3
[64,80)	0.4	5.0	0.4
[80,96)	0.0	1.4	0.0
[96,112)	0.0	0.7	0.0
⩾112	0.0	0.0	0.0

[a] Total number of modal events in the sample data is 4560.

Table 7
Two-sample K–S test of similarity in modal event distribution

Test pair	No. of events	K–S value	Similarity level (2 tailed)
Sample	4560	0.541	0.931
LA01	96		
Sample	4560	1.009	0.261
UC	125		

5. Discussion

We have described the Unified Cycle construction methodology. We have also described a stochastic approach to driving cycle construction and compared the new driving cycle (LA01) with the Unified Cycle. Both LA01 and the Unified Cycle closely replicate the global (or overall) characteristics associated with speed and acceleration observed in the sample driving data. On the other hand, the comparisons between the two cycles showed major differences with respect to modal events' frequencies, durations, and intensities. Generally speaking, LA01 consists of the modal events with frequencies and durations most closely resembling those of the sample data than does the Unified Cycle, after controlling for the intensities of the events. Given the design of the study, we can reasonably conclude that those differences result from the different driving cycle construction methodologies.

The stochastic approach addresses a number of major problems noted in the Unified Cycle construction procedures. For example, the Unified Cycle is able to preserve global characteristics, such as average speed (Austin et al., 1993), but local characteristics, such as small acceleration and decelerations, tend to be lost as demonstrated in our study. Those small timescale activities are important because they are likely to result

Table 8
Average duration of modal events (in seconds)

Speed ($\mathrm{km\,h^{-1}}$)	$\pm 0.45\,\mathrm{m\,s^{-2}}$			$>+0.45\,\mathrm{m\,s^{-2}}$			$<-0.45\,\mathrm{m\,s^{-2}}$		
	Sample	LA01	UC	Sample	LA01	UC	Sample	LA01	UC
<16	26±31	22	17	5±3	3	4	5±3	2	2
[16,32)	17±20	20	10	5±3	3	6	5±4	4	3
[32,48)	23±22	22	12	6±3	4	5	6±4	3	4
[48,64)	30±28	23	24	9±5	3	3	7±5	4	4
[64,80)	45±41	50	39	10±4	3		11±5		4
[80,96)	99±124	38	149	3±1			2±0		3
[96,112)	291±257	87	77						
⩾112	128±46								

Table 9
Average road power of modal events (in kW)

Speed ($km\,h^{-1}$)	$\pm 0.45\,m\,s^{-2}$			$>+0.45\,m\,s^{-2}$			$<-0.45\,m\,s^{-2}$		
	Sample	LA01	UC	Sample	LA01	UC	Sample	LA01	UC
<16	1.6±1.5	0.9	0.2	3.5±2.6	7.5	6.6	1.1±1.2	0.9	0.7
[16,32)	7.4±3.2	4.5	8.5	14.4±6.0	13.5	16.3	4.0±3.9	15.7	11.7
[32,48)	9.4±7.0	8.6	15.5	20.5±8.0	24.6	34.0	6.8±8.7	38.6	49.2
[48,64)	11.8±8.2	10.5	21.4	24.3±8.7	48.9	65.3	7.7±5.1	55.3	66.6
[64,80)	14.8±6.5	12.7	12.0	35.0±9.0	65.9		9.3±2.0		145.2
[80,96)	23.2±11.6	23.7	17.5	15.4±4.8			18.1±0.0		347.0
[96,112)	24.6±6.8	25.3	24.4						
>112	29.7±0.4								

in increased emissions. The problem with a microtrip is that it does not differentiate significantly between modal operating conditions within a microtrip and may result in a driving cycle with disproportional representation of a set of modal events. The new method defines segments based entirely on the physical speed profile and re-creates a sequence of modal events, taking into account not only the frequencies of occurrence but also the occurrence order of the modal events (through the Markov chain property). LA01 is found to replicate the vehicle operation modes in the sample data set generally better than the Unified Cycle with respect to frequency, average duration and intensity. The intent of our study was not to comment on the suitability of the underlying data, but rather to comment on the ability of the Unified Cycle to adequately replicate the driving behavior observed in the sample data.

We have also discussed the differences in SAFDs and characteristics of modal events. The impact of such differences on emissions can be non-trivial. To demonstrate the effects on emissions, we calculated the road power values for each modal event in the sample data, LA01, and the Unified Cycle. Again, the modal events were categorized into the eight speed subclasses and three acceleration subclasses (Table 9). Within the acceleration subclass of $\pm 0.45\,m\,s^{-2}$, LA01 represents the average road power values of the sample's modal events better than the Unified Cycle, where the average road power values seem to be either overestimated by 65% and 81.4% (i.e., for speed subclasses [32,48) and [48,64) $km\,h^{-1}$, respectively) or underestimated by 87.5% and 24.6% (i.e., for speed subclasses [0,16) and [80,96) $km\,h^{-1}$). For the acceleration subclasses $>0.45\,m\,s^{-2}$ or $<-0.45\,m\,s^{-2}$, again, LA01 generally produces better average road power values than the Unified Cycle when the event's average speed is low. In the higher speed subclasses, the average road power values of either cycle are much higher than those observed in the sample, which may be partly related to the statistically insufficient sample sizes of those particular modal events. Nonetheless, the Unified Cycle produces even higher average road power values in those subclasses than LA01.

Closer resemblance of average road powers indicates better emission estimates for LA01 than for the Unified Cycle. Such differences correspond to the fact that more fine-scale modal events are observed in the new driving cycles. It is believed that enrichment during microtransient operation is one of many causes of CO, NO_x and HC emission increases. The US EPA found a statistically significant correlation between delta throttle position (DTP), which measures microtransient operation and the emission differences between the stoichiometric tests (US Environmental Protection Agency, 1995b). For instance, the same study revealed that aggressive and microtransient driving combined contributed to a 57.8% increase in NO_x (measured in g/mile) (US Environmental Protection Agency, 1995b). Quantitative and direct evidence of such differences requires a dynamometer test of the new driving cycles.

We also have concerns about the use of SAFD. The SAFD is designed to capture the frequencies of second-by-second (or instantaneous) speed–acceleration pairs. However, it does not contain information about what the speed and acceleration are at the next second. In other words, matching a cycle's SAFD with the sample SAFD only guarantees the occurrence frequencies of individual speed and acceleration, and does not necessarily reflect the sequence of the modal events in the cycle. Therefore, the use of transition probability ensures that a driving cycle replicates a variety of modal events proportional to those observed in the sample dataset. However, there are concerns related to the reliability and robustness of the transition probability estimation. Referring back to Eq. (4), the MLE estimate of a transition probability is simply the fractional occurrence of observed state j from state i. The estimates will be statistically more reliable if the numbers of the observed cases are sufficiently large. Hence the robustness of the new approach also relies on the size of the sample. Future investigation is needed into the threshold sample size at which reliable transition probabilities can be achieved.

We used a single PV to define how well a driving cycle behaves by a series of descriptive statistics in the new cycle construction method. We recognize that such a simple combination of measures may not necessarily define the most representative driving cycle. It is likely that other important measures are not included in the performance measurement, or that there is a better way to measure a driving cycle's performance. Further research needs to be conducted on how to select the "best cycle" from a set of cycles.

6. Conclusion

Driving cycles are important features in the vehicle emission estimation process. Therefore, a "good" driving cycle is crucial for the entire practice. In this paper, we have discussed the construction procedure used to develop the California Air Resources Board's Unified Cycle. We have also described a new stochastic method of driving cycle construction and demonstrated how new driving cycles compare to California's current regulatory cycle. The MLE partitioning algorithm associates a segment with a specific modal operating condition, and at the same time preserves finely resolved driving variability within the segment. Transition probabilities describe the contingency of one modal event on another. These improvements represent important advances in the development of stochastic driving cycle construction methods.

There are limitations to the new method. For instance, the performance value (PV) defined in the new cycle construction methodology may not necessarily give us the best representative driving cycle. The segment-to-segment speed match criterion in the new method is also somewhat rudimentary. For example, we expect that less speed differential tolerance should be allowed at higher speeds, where even small changes in acceleration patterns can produce enrichment. There is also the reliability concern about transition probabilities related to sample sizes. Therefore, further research should be undertaken to determine the optimal set of performance characteristics that should be used to select the "best" cycle from a set of cycles for it is critical from a vehicle testing resource perspective to limit the overall number of cycles. Research should be conducted on sampling strategies and the amount of data that is required and statistically sufficient to generate reliable transition probabilities of modal events. Finally, additional dynamometer testing should be undertaken to assess how much unit emission rates change under the new cycles. This would also provide useful information on the overall number of cycles that should be created.

Acknowledgements

This research was partially funded under the UC Davis-Caltrans Air Quality Project (CAMP). We would like to thank Jeff Long of the California Air Resources Board, Tom Carlson of Sierra Research, Inc. and Mike Brady of the California Department of Transportation for their help and comments. We would also like to thank the anonymous reviewers, whose comments were especially helpful. The views expressed in the paper are those of the authors alone.

References

An, F., Barth, M., Norbeck, J., Ross, M., 1997. Development of comprehensive modal emissions model: operating under hot-stabilized conditions. Transportation Research Record 1587. National Research Council, Transportation Research Board, Washington, DC, pp. 52–62.

Andre, M., 1996. Driving cycle development: characterization of the methods. SAE Technical Paper Series, 961112, pp. 1–13.

Austin, T.C., DiGenova, F.J., Carlson, T.R., Joy, R.W., Gianolini, K.A., Lee, J.M., 1993. Characterization of driving patterns and emissions from light-duty vehicles in California. Final Report, contract no. A932-185, California Air Resources Board.

Barth, M., An, F., Norbeck, J., Ross, M., 1996. Modal emissions modeling: a physical approach. Transportation Research Record 1520. National Research Council, Transportation Research Board, Washington, DC, pp. 81–88.

Barth, M., Younglove, T., Wenzel, T., et al., 1997. Analysis of modal emissions from Diverse in-use vehicle fleet. Transportation Research Record 1587. National Research Council, Transportation Research Board, Washington, DC, pp. 73–84.

California Air Resources Board, 2000. EMFAC2000 Technical Support Documentation, July 2000. http://www.arb.ca.gov/msei/msei.htm.

Cernuschi, S., Giugliano, M., Cemin, A., Giovannini, I., 1995. Modal analysis of vehicle emission factors. Science of the Total Environment 169, 175–183.

Chakravarti, I.M., Laha, R.G., Roy, J., 1967. In: Handbook of Methods of Applied Statistics, Vol. I. Wiley, New York, pp. 392–394.

Gammariello, R.T., Long, J.R., 1993 An emissions comparison between the Unified Cycle and the federal test procedure. Presented at the Specialty Conference The Emission Inventory: Perception and Reality, Pasadena, CA.

Holmén, B.A., Niemeier, D.A., 1998. Characterizing the effects of driver variability on real-world vehicle emissions. Transportation Research Part D 3, 117–128.

Kuhler, M., Karstens, D., 1978. Improved driving cycle for testing automotive exhaust emissions. SAE Technical Paper Series 780650.

Lee, T.C., Judge, G.G., Zellner, A., 1970. Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North-Holland Publishing Company, Amsterdam, London, pp. 23–25.

Lin, J., Niemeier, D.A., 2002. Estimating regional air quality vehicle emission inventories: constructing robust driving cycles. Transportation Science, forthcoming.

Milkins, E., Watson, H., 1983. Comparison of urban driving patterns. SAE Technical Paper Series 830939.

Morey, J.E., Limanond, T., Niemeier, D.A., 2000. Validity of chase car data used in developing emissions cycles. Journal of Transportation and Statistics 3, 15–28.

St. Denis, M.J., et al., 1994. Effects of in-use driving conditions and vehicle/engine operating parameters on "off-cycle" events: comparison with federal test procedure conditions. Journal of Air and Waste Management Association 44, 31–38.

Steel, R.G.D., Torrie, J.H., Dickey, D.A., 1997. Principles and Procedures of Statistics: A Biometrical Approach. McGraw-Hill Series in Probability and Statistics, McGraw-Hill Companies, Inc., United States of America, ISBN 0-07-061028-2.

Stoeckenius, T., Pollack, A., Carlson, T., 2000. Speed Correction Factor Improvement Study Estimating Sample Size Requirements for the Chase Car Study. Prepared for Sierra Research for submittal to Mike Brady, Caltrans Environmental Program, Sierra/Caltran Task Order No. 1, December 21, 2000.

Symons, M.J., 1981. Clustering criteria and multivariate normal mixtures. Biometrics 37, 35–43.

US Environmental Protection Agency, 1995a. Proposed regulations for revisions to the federal test procedure for emissions from motor vehicles. Report No. 2060-AE27, January, Office of Air and Radiation, Office of Mobile Sources.

US Environmental Protection Agency, 1995b. Support document to the proposed regulations for revisions to the federal test procedure: detailed discussion and analysis. Office of Air and Radiation, Office of Mobile Sources Certification Division, January 31, 1995.

Venigalla, M., Miller, T., Chatterjee, A., 1995. Alternative operating mode fractions to federal test procedure mode mix for mobile source emissions modeling. Transportation Research Record 1472. National Research Council, Transportation Research Board, Washington, DC, pp. 35–44.

Watson, H.C., Milkins, E.E., Braunsteins, J., 1982. The development of the Melbourne Peak Drive Cycle. SAE/ARRB Second Conference on Traffic Energy and Emissions, Melbourne paper 82148.