Transportation Research Part D 137 (2024) 104507

Contents lists available at ScienceDirect

Transportation Research Part D

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

Using natural driving experiments and Markov chains to develop realistic driving cycles

J.D.K. Bishop [a], C.J. Axon [b,*]

[a] International Finance Corporation, Washington DC 20433, USA
[b] Institute of Energy Futures, Brunel University of London, Uxbridge, London UB8 3PH, UK

ARTICLE INFO

Keywords:
Driver characteristics
Driving behaviour
Drive cycle
Driving metrics
Fuel economy
Markov chain

ABSTRACT

The main purpose of driving cycles is to estimate accurately on-road fuel use and the associated emissions of greenhouse gases and other air pollutants by vehicles. Conventionally, driving cycles are developed using micro-trips, Markov chains, or hybrid approaches, with accuracy determined by comparing metrics of the candidate cycles with the observed data. Through a natural driving experiment, we suggest traffic and road topology have a dominant role in influencing individual driving styles, more so than driver age or gender, or vehicle characteristics. Using experimental data and a Markov chain approach, we make three contributions to driving cycle development. First, we identify a reduced set of 26 metrics which materially influence fuel economy. Second, we assess the trade-offs in accuracy between reproducing vehicle dynamics and fuel economy. Finally, we identify the impact of natural driving variability on the accuracy of candidate cycles.

1. Introduction

National emissions inventories require realistic estimates of actual fuel use by different types of vehicles. To estimate emissions due to the operations of vehicles, drive cycles are used to represent typical or standardised driver behaviours. The more accurately such cycles capture real-world driving behaviour, the better the estimates of environmental impacts. Standardised driving cycles are also important for the type-approval of new vehicles developed by manufacturers. Historically, regulated driving cycles overestimated fuel economy and underestimated the associated emissions under real-world driving (Franco et al., 2013). For example, the gap between real-world fuel economy and that achieved under the New European Driving Cycle (NEDC) grew from less than 10 % in 2001 to a maximum of 37 % in 2016, dropping to 30 % in 2019 (Dornoff et al., 2024). Studies on plug-in hybrid vehicles (PHEV) found real-world fuel consumption was up to five times higher than type-approval values on average (Plotz et al., 2022, 2021). Some discrepancy is to be expected given the differences between test procedures and real-world driving. In particular, the all-electric range was shorter, and the engine used more fuel, than under type-approval testing. The consequence is PHEVs tend to drive more using the engine than under the test procedure, resulting in higher fuel use and associated emissions. Isenstadt et al. (2022) found PHEV fuel consumption in the US is on average up to 67 % higher than the EPA fuel label.

Some of the shortcomings in the accuracy of regulated driving cycles were addressed in the development of the World Harmonised Light Duty Driving Test Cycle (WLTC) and associated Test Procedure (WLTP). The WLTC is based on 750,000 km of driving data collected in 2010–2011 using multiple vehicle types on different road types and conditions across North America, Europe and Asia

* Corresponding author.
E-mail address: Colin.Axon@brunel.ac.uk (C.J. Axon).

https://doi.org/10.1016/j.trd.2024.104507
Received 2 July 2024; Received in revised form 30 September 2024; Accepted 4 November 2024
Available online 9 November 2024
1361-9209/

(Tutuianu et al., 2014). The WLTC and WLTP improved on the original regulated driving cycles, however, as of 2022, a discrepancy of 14 % remained between real-world and type-approval fuel economy (Dornoff et al., 2024).

The WLTC training data was at least six years old when the WLTP was implemented in Europe in 2017. This difference between data collection and cycle introduction is relevant because we postulate a spatial–temporal distribution of driving habits, based on characteristics of both the vehicle and driver. For example, 40 local driving cycles (de Andrade et al., 2020) showed large differences in average speed and acceleration, supporting the argument of unique spatial characteristics of driving. A study of regional driving in the US, comprising over 10,000 vehicles and 1.3 million vehicle miles travelled concluded regional fuel economy varies due to differences in driving patterns (Borlaug et al., 2020). We can expect the age and performance characteristics of vehicles to afford different driving behaviours, causing the driving cycle to change through time. This change may be due to the higher torque output of modern engines allowing faster acceleration; or aids to support eco-driving, such as gear-shift indicators, auto start/stop, fuel economy indicators and electronic driving modes. However, the repeatability of the WLTP and other test procedures is based on narrow conditions, such as no road-grade or auxiliaries, including air-conditioning or headlights. Consequently, the restrictions of the WLTP may erode the ability to predict real-world fuel economy using the WLTC.

1.1. Driving cycle metrics

Much of the development of real-world cycles is focused on capturing driving behaviour by collecting on-road measurements from many vehicles to create a body of data (called a corpus) and reproducing it in a candidate cycle of a shorter duration. Studies have investigated the optimum driving cycle length to return the smallest overall error, compared to the training data. Research by Desineedi et al. (2020) compared the average error of bus driving cycles constructed using clustered micro-trips and Markov chains by time of day. In general, average error across 12 driving cycle metrics approached a minimum for driving cycles longer than 2000 s. A range of metrics have been used to assess the accuracy of the candidate cycle against the corpus, including: average velocity, acceleration and deceleration; proportion of time spent accelerating, decelerating, idling and cruising; and number of starts/stops per km (Barlow et al., 2009). Most candidate cycles capture the behaviour of the corpus to within 5 %, but the challenge is to create a candidate cycle that can reproduce accurately both the fuel economy and emissions of the vehicle. In particular, the speed profile is related to emissions but may not reproduce them completely. Instead, a candidate cycle incorporating changes in vehicle specific power (VSP) may capture most of the activity upon which (light-duty) vehicle emissions are dependent (Carslaw et al., 2013; Carslaw and Rhys-Tyler, 2013; Lee and Frey, 2012; Yu et al., 2010; Zhang et al., 2014). In an exhaustive study of the metrics, characteristic parameters, and performance values of driving cycles, Quirama et al. (2021) hypothesised that two or three metrics were sufficient to characterise the corpus well. Although a unique pair and triple returned the lowest error, up to 15 pairs and triples returned similar performance. Additionally, different metrics were found to return the smallest error across the two regions and different pollutants.

Most driving cycle research focuses on reproducing the speed-time characteristics of observed data, but only recently has fuel-use and emissions data been exploited to construct representative driving cycles. Quirama et al. (2021) introduced an energy-based micro-trip method to reproduce on-road energy use and emissions using 15 buses measured on similar routes to those used by Huertas et al. (2018), with specific fuel consumption as one of the metrics to find the best candidate driving cycle. Earlier work by Bishop et al. (2019) used the elastic net approach on 29 driving cycle metrics, nine vehicle attributes and two ambient conditions to determine those which influenced most the trip average fuel use by and associated emissions of nitrogen oxides (NOx) from Euro 6 petrol and diesel passenger cars. Sangeetha and Bose (2021) introduced a method (termed the Real-World Cycle Condenser) to consider engine parameters that have a role in emissions formation by identifying families of micro-trips with similar driving cycle metrics, selecting a representative trip within a family and then eliminating the remaining trips. 'Condensed' cycles are generated until one is found which meets a threshold of 2 % error, compared to the corpus. Then, a vehicle system simulator was used to assess the fuel use and emissions of the corpus and the condensed cycle and verified on a chassis dynamometer. In general, fuel use and CO_2 emissions were reproduced well, while the synthesised cycle failed to capture trip-level NOx and CO emissions as accurately.

1.2. Driving cycle development

The micro-trip method remains the most popular in the literature for developing driving cycles. Some authors have used a hybrid micro-trip and Markov chain approach to create candidate cycles. In one study, the full training data of 459 vehicles, sampled across 17,000 days and 3.3 million km was categorized into seven speed clusters of 10–20 km/h each (Ma et al., 2019). These clusters were used to create a seven-by-seven transition probability matrix with which the authors created candidate driving cycles for peak and off-peak driving. In another study, the driving of 40 electric taxis was monitored across six months, with the data split into micro-trips and assigned to a nine-by-nine transition probability matrix based on average velocity intervals of 10 km/h. Twelve driving cycle metrics were used to calculate the root mean squared (RMS) error and identify the best candidate driving cycle (Wang et al., 2019). These hybrid micro-trip Markov chain approaches may be necessary for managing larger amounts of data. However, it remains fundamentally a micro-trip method with the 'memory' concept of the Markov chain, rather than random selection, to determine the next micro-trip in the sequence.

The use of n-1 Markov chains is an alternative to the micro-trip method (Bishop et al., 2012; Lee et al., 2011; Lee and Filipi, 2011) and assumes that the probability of moving from one speed to another is not independent. From the data comprising the corpus, a transition probability matrix embeds the likelihood of moving between any pair of consecutive (speed) states. Subsequently, the Markov chain approach has been used by Qiu et al. (2022) to compare with the performance of a genetic algorithm using 1768 micro-trips classified by k-means into three clusters, with the data grouped into 5 km/h bins for the transition probability matrix. The Markov chain approach was found to capture the trend of real-world driving, but the authors' spectral clustering-based method was more accurate at reproducing the values of eight commonly-used driving cycle metrics to within 10 %. Peng et al. (2020) used k-means to divide 11,619 micro-trips into three clusters. Transition probability matrices were developed for each cluster and Markov chains used to generate candidate cycles of various lengths. Six driving cycle metrics and the speed-acceleration frequency distribution of the candidate driving cycles were compared to the corpus, with the 'target' cycle being the candidate with the sum of the smallest error. A min–max Ant Colony Optimisation with Markov chain has been used (Cui et al., 2022) where the journey through the transition probability matrix mirrors the foraging behaviour of an ant swarm. The authors used the Comprehensive Modal Emissions Model to estimate fuel consumption of the best candidate driving cycle, compared to what was observed in the real-world.

These studies illustrate hybrid approaches, where clustering or micro-trip methods are used to pre-process the data and feed into a Markov chain process. Such pre-processing may help to manage the quantity of data being used to generate the Markov transition probability matrix, but introduces subjectivity, such as the choice of the number of clusters or the identification of important components (through principal component analysis, PCA) on which to evaluate how well the candidate driving cycle reproduces the corpus. The different mathematical and computational approaches each have their advantages and disadvantages, but what connects many of the studies described is that they collect data using specific routes, cities, fleets, or (professional) drivers. Comparing synthesised drive cycles with the corpus may yield information about the efficacy of the analytical method chosen, but the usefulness of the drive cycle itself is dependent on how well the corpus captures the natural variability of real driving.

Markov processes have complemented other methods such as clustering or micro-trips, to yield more accurate driving cycles. However, we suggest the Markov chain approach is sufficiently robust to accommodate the corpus directly, removing the need for data pre-processing and allowing a closer link to be drawn between the data captured during the experiment and the best candidate driving cycle.

In summary, recent literature has introduced new approaches to driving cycle development across different vehicle types. Our focus is to fill two gaps in the literature: first, a widely varying number of driving cycle metrics continues to be used to determine how well a candidate cycle captures the dynamics of the observed data; and second, there is little evidence to show how the final candidate driving cycle reproduces the observed fuel use. We build on earlier real-world driving cycle development (Bishop et al., 2012) in three ways. First, we use data collected from different passenger vehicles and drivers as part of a natural experiment to reflect the breadth in the range of driving styles and vehicle types. We collect driving data, including fuel use, at high frequency and maintain this granularity throughout our analysis. Second, we use the elastic net approach to inform a materiality assessment of driving cycle metrics to identify those most influential on fuel economy. Finally, we identify those candidate driving cycles which reproduce accurately both the driving cycle metrics and observed fuel use in the corpus.

2. Method

The data used to develop the driving cycles was collected from 2011 to 2018 through a voluntary study of members of staff at Brunel University of London and the University of Cambridge. The collection, management and storage of data was approved by Ethics Committees at both universities. Each participant was provided with an OBDII dongle which was paired with the DashCmd or Torq apps, both of which were available to Apple and Android users. Table 1 indicates the vehicle characteristics and total distance contributed to the study; each driver making multiple trips. Vehicle make and model were recorded by the participants, while physical and engine characteristics were derived from Automobile Catalog.[1] The rolling resistance coefficient (C_{rr}) was assumed to be 0.0059 for all participating cars.

Over the course of the study, Driver 1 and Driver 4 changed vehicles a number of times. We treat the trips undertaken in each vehicle separately because the vehicle type might afford different behaviour. Fig. 1 outlines the routes taken by the participants. A driver's speed profile is influenced by the driver demographics, the capability of the vehicle and the external constraints imposed by the topology of the road network, and traffic conditions. Having participants who frequently made the same trip captured the inherent variability of traffic conditions. The corpus contains trips from the most common types of road: city-centre, suburban, country, and highway.

2.1. Building the corpus

Engine channel data was captured and recorded at the natural rate (often greater than 1 Hz). The raw data was filtered using integer timestamps in seconds to return a 1 Hz set of data for each trip, driver and vehicle. The data was pre-processed to account for three types of error: 1) loss of GPS signal; 2) a logged trip starting or ending at non-zero speed; and 3) no fuel use recorded.

The loss of GPS signal resulted in entries with zero latitude, longitude, and speed. Entries meeting these three conditions were identified. The velocity in these entries was replaced with an interpolated velocity based on the entries immediately preceding and following. This procedure led to 483 s of driving being interpolated (0.54 % of the total). On occasion, the logger started recording only once the vehicle was moving, leading to a non-zero start velocity. In this case, the data between the start of the file and the first zero velocity entry was truncated. A similar procedure was followed for the data between the last zero velocity entry and the end of the file. This procedure led to 6437 s of driving being excluded from the corpus (7.3 % of the total). Fuel use was not recorded for Driver 8,

[1] https://www.automobile-catalog.com/.

Table 1
Physical and trip characteristics of study vehicles.

Vehicle characteristic	Units	Driver 1	Driver 2	Driver 3	Driver 4	Driver 5	Driver 6	Driver 7	Driver 8	Driver 9	Driver 10	Driver 11
Model year		2011	2009	2001	2013	2012	2006	2017	2017	2018	2018	2018
Make and model		Ford Focus	Mercedes Benz C180	Porsche 911	Ford Focus	Ford Focus	Audi A3	Hyundai i30	Mercedes Benz C200	Hyundai Tuscon	Citroen Picasso	Vauxhall Corsa SRi
Vehicle mass (m)	kg	1282	1405	1375	1276	1263	1330	1322	2015	1497	1297	1585
Frontal area (A_f)	m^2	2.26	2.17	1.91	2.36	2.25	2.01	2.09	2.17	2.44	2.39	2.05
Drag coefficient (C_d)		0.32	0.25	0.30	0.27	0.31	0.33	0.3	0.27	0.3	0.3	0.27
Peak power	kW	80	115	221	74	70	103	74	135	97	96	55
Peak torque	Nm	240	230	350	170	230	320	134	300	161	230	130
Engine volume	cc	1560	1597	3387	999	1560	1968	1368	1991	1591	1199	1398
NEDC fuel use	l/100 km	4.4	6.7	12	4.8	4.2	5.5	5.4	6.8	7.2	5	5
	MJ/100 km	164	213	381	152	157	206	172	216	229	159	178
Fuel type		Diesel	Petrol	Petrol	Petrol	Diesel	Diesel	Petrol	Petrol	Petrol	Petrol	Petrol
Reference no.[†]		1232960	1550540	2866790	1776185	1592855	1190765	2525600	2504240	2761880	2513150	2769680
Total driving distance	km	54	172	7	42	188	19	50	N/a	156	341	113

[†] https://www.automobile-catalog.com.

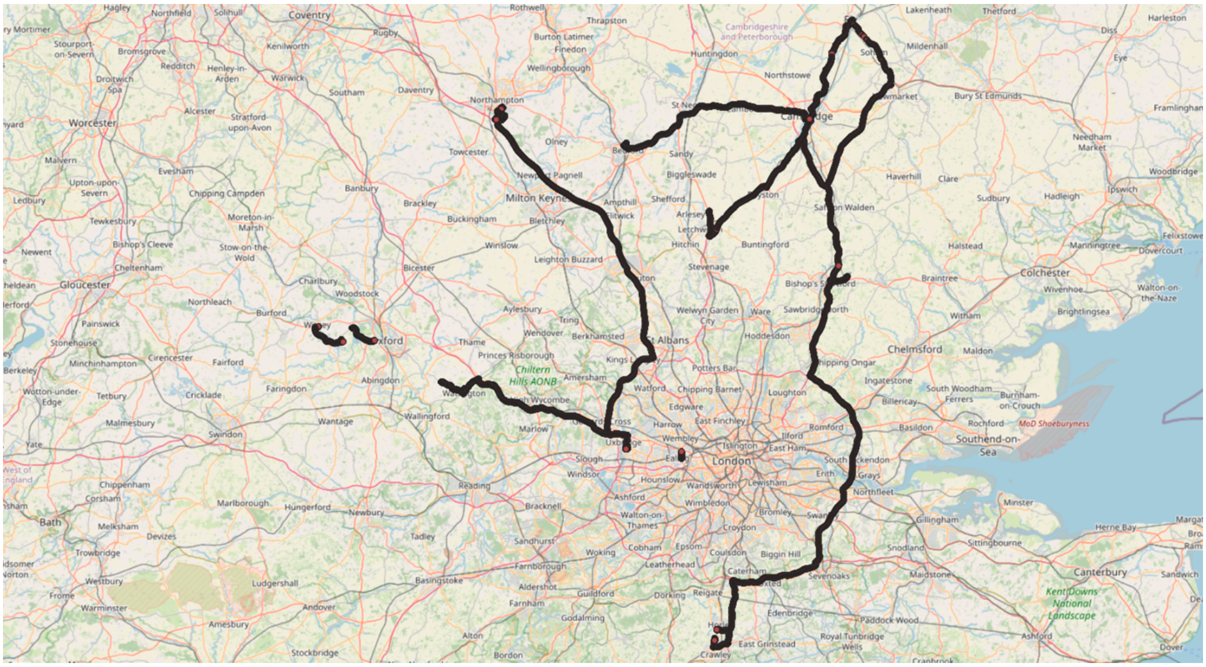

Fig. 1. GPS trace of the corpus. Each trace represent multiple trips.

equivalent to 2622 s of driving being excluded from the corpus (2.6 % of the total). In total, the corpus comprised 1142 km of driving over 55,969 s, with a total of 9059 s cleaned as above. GPS coordinates associated with the cleaned data were mapped to a digital elevation model.[2] The slope between two consecutive coordinates was taken as the quotient of the rise (altitude) over the run (horizontal distance travelled in one second based on the velocity). The slope was calculated and included in the estimate of VSP, but the impact on fuel use was not considered important because over 99 % of slope values were between −0.5 % and 0.5 %.

For each vehicle, the trips were appended into a single velocity–time profile, on which 29 metrics were calculated, given in Table 2. Specific power was estimated for each trip per vehicle based on the vehicle characteristics in Table 1. The corpus is illustrated in Fig. 2 using the profiles of speed and specific power. The four anomalous spikes seen in the specific power profile of the corpus (Fig. 2b). These do not influence the results because the Markov chain approach favours higher frequency transitions and is insensitive to low frequency data anomalies, a notable advantage of this method.

Vehicle recorded fuel use in litres per second which was converted to fuel use per km based on the total volume of fuel and the total distance travelled. The accuracy of fuel use derived from the OBD is dependent on how well the vehicle estimates fuel use and measures distance. Pavlovic et al. (2021) tested light and heavy vehicles and concluded fuel use estimated via the OBD was within 7 % of what was observed in the laboratory, with accuracy increasing with increased driving time. This value was converted to MJ/100 km using the lower specific energy of motor spirit and gas/diesel oil, at 31,791 MJ/m^3 and 37,385 MJ/m^3, respectively. The observed fuel economy of the corpus was 231 MJ/100 km.

The velocity–time data for each vehicle was appended to give a new single velocity–time profile representing the full study data. Vehicle-specific information is lost when the training data is converted to a frequency and transition probability matrix. A representative vehicle was developed based on the relative proportion of driving data, yielding a new mass, C_d and C_{rr} of 1402 kg, 0.29 and 0.006, respectively. This representative vehicle was used to estimate the specific power in each candidate driving cycle.

2.2. *Creating the frequency and transition probability matrices*

Markov chains introduce memory by ensuring the probability of achieving a state in time *t* is a function of the states at time *t*-1, *t*-2, …,*t*-n. Using one state per km/h, with the speed rounded to the nearest one km/h, an n-1 transition frequency matrix is created where the entry at (*i,j*) represents the number of times state *j* follows state *i*. The transition *frequency* matrix is transformed into the transition probability matrix using the method devised by Bishop et al. (2012). The code is written in Matlab (version R2023a) using the Statistics and Machine Learning Toolbox and runs on a i7-7700 processor (Windows 10 Pro) with 256 kB L1 cache and 32 GB of RAM.

The n-1 Markov chain considers the pair representing the velocity one second in the past and the current velocity. The square frequency matrix was populated with the number of times a specific n-1 pair was observed. Fig. 3 illustrates a heat map of the transition probability matrix which was derived from each row element divided by its row sum.

[2] Available at https://www.gpsvisualizer.com.

Table 2
List of driving cycle metrics, elastic net coefficients, their product and the material metrics denoted with * in the last column.

#	Driving cycle metric	Abbreviation	Units	Initial value	Elastic net coefficient	Metric Coefficient	
1	Running velocity	v run	m/s	20.6	–33.4	−688	*
2	Average velocity	avg v	m/s	18.3	5.24	95.9	*
3	Number of accelerations and decelerations per km	acc-dec/km	#	3.97	1.19	4.72	
4	Average vehicle specific power	avg VSP	kW/kg	7.75	2.85	22.1	*
5	Variance of VSP	VSP var	kW/kg	13.0	−20.5	−266	*
6	Root mean squared of acceleration	RMS acc	m/s^2	0.70	−12.9	−9.03	*
7	Number of stops per km	Stops/km	#	0.33	−25.5	−8.42	*
8	Maximum VSP	Max VSP	kW/kg	376	0.26	97.8	*
9	Average acceleration	Avg accn	m/s^2	0.45	112	50.4	*
10	Average deceleration	Avg decn	m/s^2	−0.45	48.8	–22.0	*
11	Proportion of time spent decelerating	% dec	%	22.5	−5.26	1.18	*
12	Proportion of time spent accelerating	% acc	%	23.6	67.8	16.0	*
13	Proportion of time spent cruising	% cruise	%	47.0	96.6	45.4	*
14	Proportion of time spent idling	% idle	%	6.94	–23.0	−1.60	*
15	Relative positive acceleration	RPA	m/s^2	0.15	31.7	4.76	
16	Relative negative acceleration	RNA	m/s^2	−0.14	−50.0	7.00	*
17	Relative positive speed squared	RPSS	m^2/s^2	25.5	45.9	1170	*
18	Relative positive speed cubed	RPSC	m^3/s^3	720	−0.45	−324	*
19	Kinematic intensity	KI		0.20	135	27.0	*
20	Aerodynamic velocity	V aero	m/s	26.8	19.1	512	*
21	Characteristic acceleration	Char acc	m/s^2	0.14	34.4	4.82	*
22	Maximum acceleration	Max acc	m/s^2	0.89	106	94.3	*
23	Maximum deceleration	Max dec	m/s^2	−0.89	119	−106	*
24	Maximum velocity	V max	m/s	33.5	−11.1	−372	*
25	Interquartile range of acceleration	IQR accn	m/s^2	0.89	152	135	*
26	Interquartile range of velocity	IQR v	m/s	20.1	−2.39	−48.04	*
27	95th percentile of the product of velocity and acceleration	va95	m^2/s^3	28.0	5.62	157	*
28	Median jolt	median jolt	m/s^3	0	135	0	
29	Interquartile range of jolt	IQR jolt	m/s^3	1.79	−166	−297	*

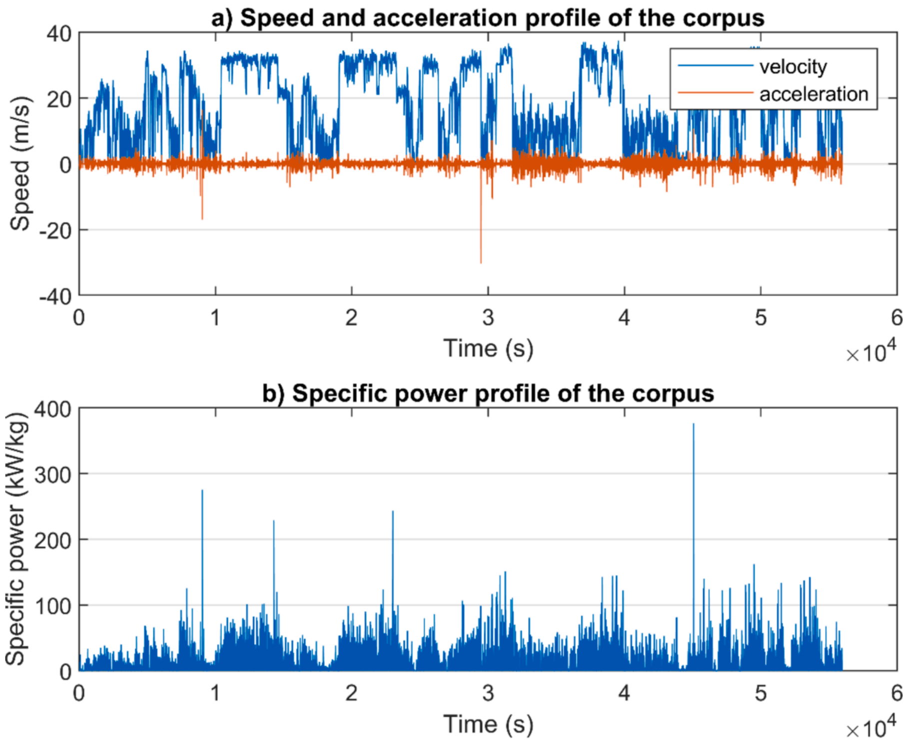

Fig. 2. Plot of corpus showing a) speed and acceleration, and b) specific power.

2.3. Identifying the key driving cycle metrics

The driving cycle metrics are based on velocity, acceleration or a combination of the two. Therefore, there is correlation across the metrics implying that a simplified subset might capture the variability of driving behaviour evident in 29 driving cycle metrics (Table 2) given in Bishop et al. (2019). By using PCA, other studies assume independence of variables, and therefore the metrics. In contrast, the elastic net method does not assume independence, identifying correlated variables and applying coefficients to balance the number of variables with accuracy. The elastic net was applied to the complete set of individual trips that comprise the corpus. The independent variable is each driving cycle metric, and the dependent variable is trip level fuel use (MJ/100 km).

The group of independent variables was split into a training set and a validation set, based on Matlab's cvpartition cross-validation, with a training set of 70 % and a test set of 30 % of the observations. The 'standardize' and 'intercept' flags were set to false: if 'standardize' was true, the elastic net would attempt to fit models to predictor data standardized to have mean and variance of zero and one, respectively; and a true 'intercept' flag would return a linear model which did not pass through the origin. The cvpartition and elastic net method returned a reduced set of lasso variables and associated linear elastic net coefficients which yielded an estimate of fuel economy. There was a different set of lasso variables, elastic net coefficients and estimated fuel economy for each cvpartition run. Therefore, the cvpartition approach was run 500 times which yielded 500 different lasso fuel economy values. The set of elastic net coefficients which delivered estimated fuel economy closest to the corpus was chosen.

A 'materiality' assessment was conducted to determine which metric and elastic net coefficient pairs most influenced the fuel economy estimate. Whether the input data was standardized or not prior to applying the elastic net, the result was the same. The product of the elastic net coefficient and corresponding metric was either negative or positive, depending on the sign of the individual components. The material metric and elastic net coefficient pairs were chosen by ordering negative and positive products, largest to smallest (Fig. 6). To estimate the evolution of fuel economy, the cumulative sum was taken of metrics, ordered in terms of the product of their value and elastic net coefficient, from the most negative to the least positive. The set of material metrics corresponded to those for which the cumulative sum first came within 5 % of the corpus. The remaining metrics were not considered material as their impact was less than 5 % of the observed fuel economy.

2.4. Create driving cycles and identifying the best candidates

A random walk of *m* steps, beginning at state 1 (or idle), is performed through the probability transition matrix. The resulting vector of length *m* is a candidate driving cycle and the absolute deviation between each material metric (Table 2) of the random walk and the corpus is calculated. In most studies, arbitrary 5 % or 10 % thresholds are used to determine the best random walk, or candidate driving cycle. This process was repeated in powers of 10, from zero (representing one walk) to five (100,000 walks) and lengths from 500 to 5000 steps (each step is one second) in blocks of 500 steps. Evaluating candidate driving cycles across a range of repetitions and lengths shows how accuracy changes with varying input parameters. For each repetition and driving cycle length, we identified the five candidate driving cycle(s) with the smallest:

- sum of absolute deviations across driving cycle material metrics, and
- absolute deviation from the corpus fuel economy.

Optimizing using these two dimensions tests if the candidate driving cycle with the smallest sum of absolute deviations across

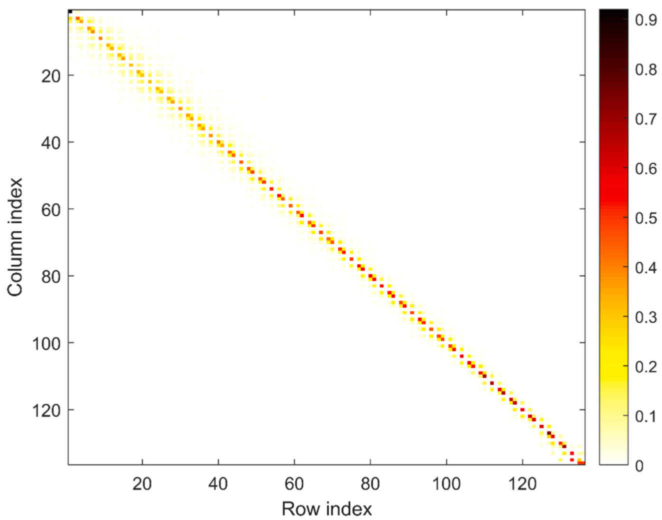

Fig. 3. Heat map of transition probability matrix representing the likelihood of the participants drivers changing speed (at one km/h granularity).

driving cycle metrics is also the one with the smallest deviation between estimated and observed fuel economy.

3. Results and discussion

There are three main results which we present and discuss in this section. First, we show how driver variability is expressed in the corpus. Second, we identify the most important driving cycle metrics to estimate fuel economy by applying the elastic net process. Third, we run the Markov chain process and identify the best candidate driving cycle. We discuss the trade-off between meeting driving cycle metrics and fuel economy and the difficulty in reproducing all metrics accurately.

3.1. The corpus and driver variability

The variation in the corpus was shown in Fig. 2. Fig. 4 illustrates the variation in the magnitude of driving cycle metric by driver. Across the metrics, the boxes are overlapping in most cases and there is little vertical spread in the median value (red horizontal line) by driver. This comparison shows how similar the driving characteristics were, despite differences across driver demographics and

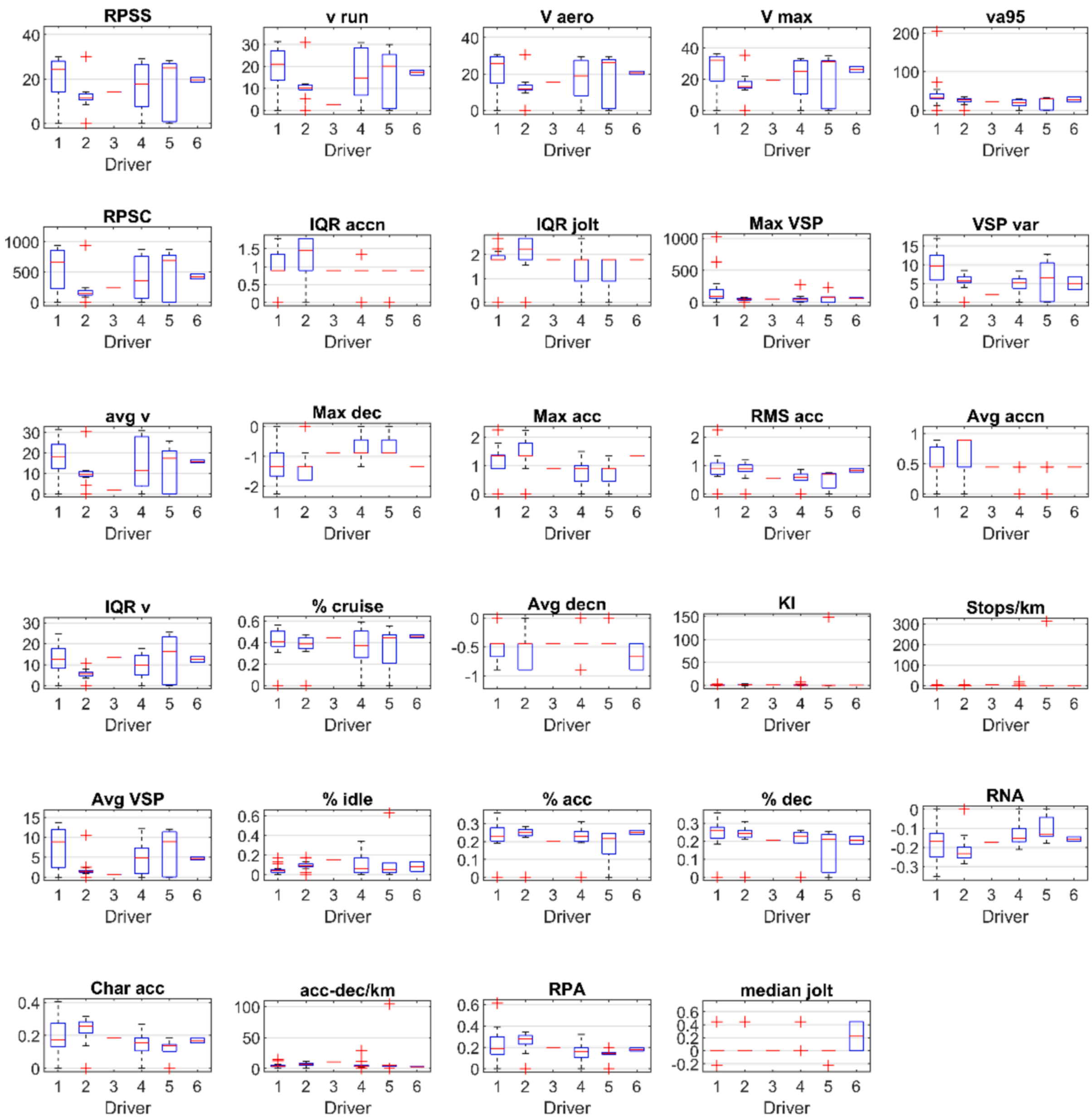

Fig. 4. Variation of magnitude of driving cycle metric, by driver.

vehicle attributes (Table 1). Therefore, we suggest each driver's data is largely representative of the other because the effect of the external traffic and road network characteristics are dominant.

This dominance of exogenous conditions implies that the driving cycles we have developed will remain representative until the traffic and road network characteristics change materially. For example, the influence of individual driving styles may become more prominent in areas where road capacity improvements reduce congestion. Similarly, the influence of driver behaviour and vehicle attributes may be muted in areas where road closures and new traffic patterns push more vehicles onto the same route, increasing congestion. We did not investigate the extent of the change in external conditions, such as slope, influencing materially driving characteristics. Vehicle specific power is the only driving cycle metric incorporating road slope.

3.2. Identifying key driving cycle metrics to estimate fuel economy

The elastic net process excluded characteristic acceleration (coefficient of zero) from the reduced set of indicators, resulting in all 29 driving cycle metrics being relevant for reproducing the corpus' fuel economy accurately (Fig. 5).

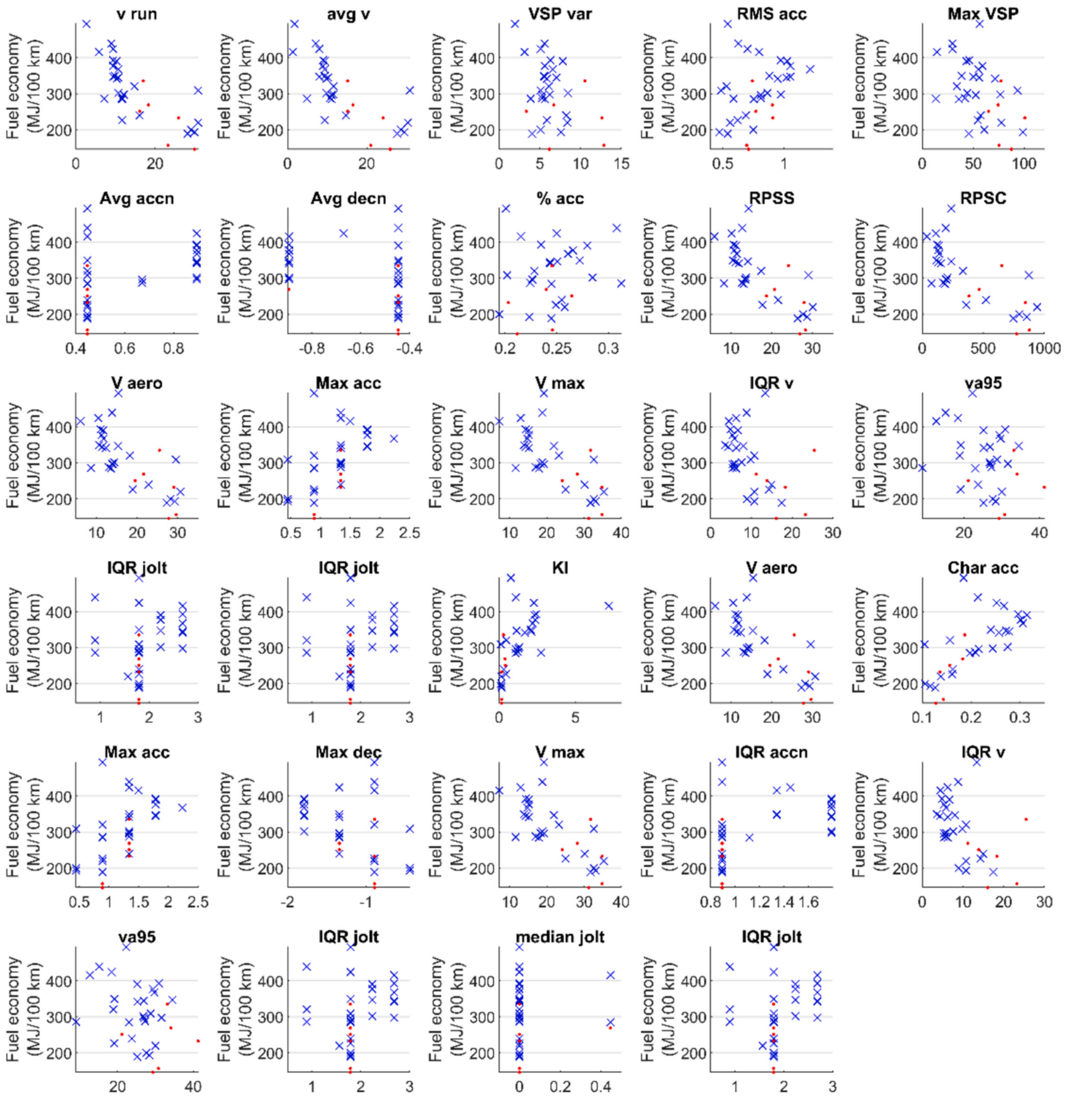

Fig. 5. The set of driving cycle metrics identified by the elastic net process, with trips using petrol vehicles indicated with blue crosses and trips using diesel vehicles indicated with red dots.

Table 2 illustrates the 29 driving cycle metrics and coefficients which yielded the smallest absolute deviation from the corpus fuel economy. The linear combination of these variables and elastic net coefficients resulted in a fuel economy of 229 MJ/100 km, which is 0.66 % lower than the observed fuel economy of the corpus.

The materiality analysis determined the impact of each metric and elastic net coefficient pair on the estimate of fuel economy. Ordering the most positive and negative pairs of driving cycle metrics and corresponding elastic net coefficients identifies the material metrics for determining fuel economy (Fig. 6) and listed in Table 2. Therefore, while 29 metrics were identified as contributing information to the estimated of fuel economy, only 26 of them were considered material. The fuel economy predicted by the material driving cycle metrics, indicated by the * in the last column of Table 2, was 220 MJ/100 km, which is 4.7 % lower than the corpus fuel economy.

This fuel economy estimate is farther away from the corpus value than the estimate using 29 metrics. However, the error is marginal and the reduction of complexity of the reduced set is an advantage in terms of computation, for practical implementation, and for ease of communication. These 26 metrics were used to identify the best candidate driving cycle.

3.3. Understanding the distribution of errors

To illustrate the varying degrees of accuracy of candidate driving cycles we use the case of 100,000 random walks through the transition probability matrix. Fig. 7 displays how well the candidate driving cycles met both the corpus driving dynamics, represented by driving cycle metrics, and fuel economy. Fig. 7a shows the number of driving cycle metrics met to within 5 % of the corpus, with the vertical line at 29 driving cycle metrics, indicating where all metrics are within 5 %. The distribution shows that 91 % of cycles met up to only five driving cycle metrics to within 5 %. Fig. 7b shows the distribution of fuel economy for the same candidate driving cycles, with the solid vertical line indicating the corpus fuel economy value, and the dotted vertical lines showing the corpus value ±5 %. Now, almost 3500 candidate cycles reproduce the observed fuel economy accurately. There are two important conclusions from this analysis. First, many more candidate driving cycles reproduced observed fuel economy accurately than the corpus driving cycle metrics. Secondly, there may not be a single, 'best' candidate, suggesting that drive cycle analysis should be looking for a family of cycles of equal or similar accuracy.

3.4. Identifying the best candidate driving cycles

Each best candidate cycle for the combination of repetitions and duration has a corresponding fuel economy. Relating the driving cycle metrics to observed fuel economy allows evaluation of the accuracy of a candidate driving cycle on this dimension and test if reproducing the driving cycle metrics closely translated into an accurate estimate of fuel economy.

Choosing the five best candidate cycles for one to 100,000 repetitions across driving cycles ranging from 500 s to 5000 s duration yielded a family of 300 cycles. There was one candidate cycle of length 4500 s with 19 metrics met to within 5 % of the corpus, illustrated in Fig. 8a. Our method returned candidate driving cycles with the three VSP metrics to within 5 % of the corpus, implying slope did not have a material impact on the accuracy of our approach. This observation aligns with (Desineedi et al., 2020) who found candidate driving cycle accuracy improved for durations longer than 2000 s. Fig. 8b indicates how the number of metrics met to within 5 % evolved with cycle length. Accuracy increased with driving cycle length, from fewer than five metrics for 500 s to over 10 metrics met to within 5 % at 5000 s. In general, accuracy did not increase after 2000 s. Accuracy improved quickly with the number of

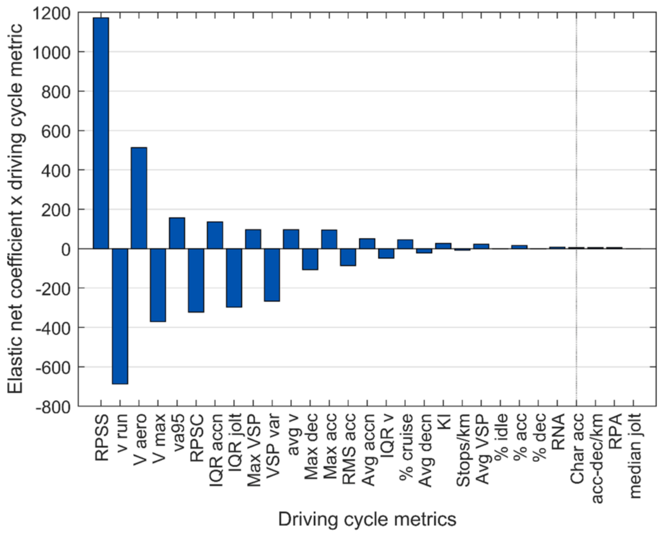

Fig. 6. Product of metric and elastic net coefficients, ordered by most positive/negative to least positive/negative. The vertical dotted line indicates the number of material driving cycle metrics.

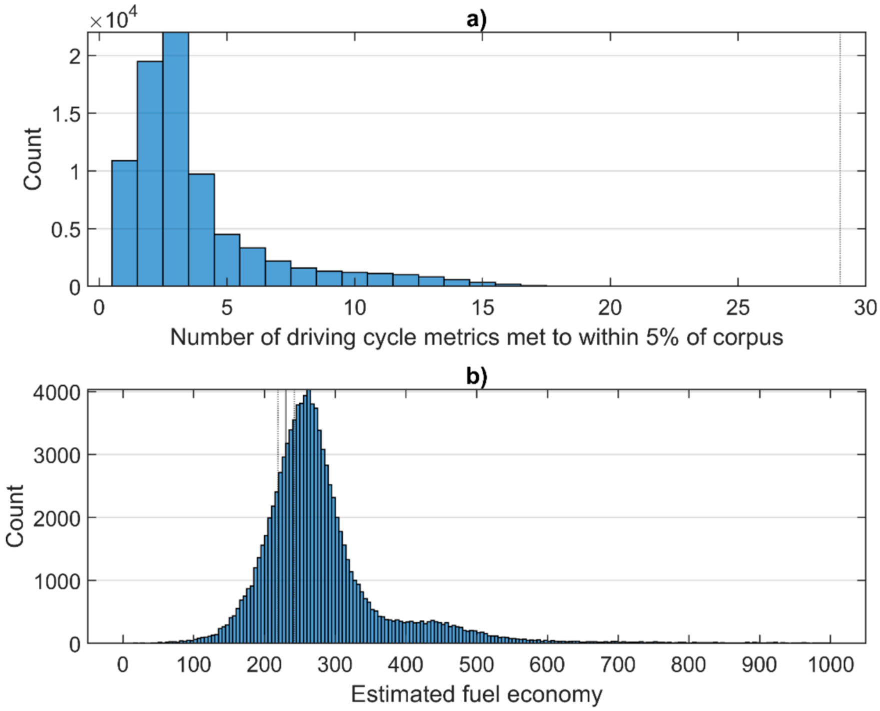

Fig. 7. Distribution of errors in driving cycle metrics and fuel economy across the combinations of sequence length (from 500 s to 5000 s) and repetitions (from 1 to 100,000). Subplots a) and b) indicate absolute values.

repetitions, from less than five metrics for a single run to over 10 metrics at 100,000 repetitions (Fig. 8c). The candidate cycle with 19 driving cycle metrics met to 5 % yielded estimated fuel economy of 293 MJ/100 km, equivalent to 27 % higher than the corpus. Fig. 8d and Fig. 8e indicate how the best candidate cycles based on driving cycle metrics yielded significant over- and under- estimates of fuel economy, relative to the corpus (black dashed line). The spread of fuel economy estimates tended to decrease with increasing cycle length and increasing repetition settling to a value which was higher than the corpus. Finally, Fig. 8f and Fig. 8g show that the variation in the sum absolute error between candidate driving cycles and the corpus decreased slightly with cycle length, and more substantially with the number of repetitions.

In contrast, 107 candidate cycles delivered fuel economy to within one decimal place of the observed value. Two of the 107 candidate cycles are illustrated in Fig. 9a, confirming our earlier conclusion that it is easier to reproduce the corpus fuel economy than its driving cycle metrics. Fig. 9b and Fig. 9c indicate that relative few metrics are met to within 5 % of the corpus despite increasing cycle duration and number of repetitions, respectively. Optimizing for fuel economy shows less spread (over- and under-estimations) for increasing cycle duration, while decreasing substantially with increasing number of repetitions (Fig. 9d and Fig. 9e, respectively). Despite this accuracy for fuel economy, these candidate driving cycles reproduce fewer than five metrics to within 5 % of the corpus, implying that most cycles have characteristics differing substantially from the corpus.

Fig. 10 illustrates the significant increase in computation time for 100,000 repetitions, for a marginally larger number driving cycle metrics to within 5 % and improvement in fuel economy estimates.

3.5. Assessing the trade-off between meeting driving cycle metrics and fuel economy

The trade-off in accuracy between driving cycle metrics and fuel economy is seen in Fig. 11. The solid vertical line indicates corpus fuel economy and the solid horizontal line indicates 26 driving cycle metrics met to within 5 %. The closer the candidate driving cycles are to the intersection of the two lines indicated, the more accurate they are in both dimensions.

Only one candidate cycles met 19 metrics to within 5 % of the corpus. In Fig. 11, we illustrate the trade-off between meeting driving cycle metrics and fuel economy more clearly by showing the 36 cycles which met at least 15 driving cycle metrics to within 5 % of the corpus. The blue crosses indicate the horizontal spread of fuel economy accuracy, from 201 MJ/100 km up to 386 MJ/100 km, for at least 15 metrics met to within 5 % of the corpus. The orange circles demonstrate that the best candidate cycles (based on fuel economy only) meet fuel economy accurately, but with only seven or fewer metrics matching the corpus to within 5 % in general.

The conclusion is choosing a candidate driving cycle based on number of metrics met to within 5 % can yield accurate fuel economy; however, choosing a candidate cycle based on fuel economy is unlikely to yield a large number of driving cycle metrics met to within 5 % of the corpus. There are families of best candidate cycles depending on which dimension is optimized. However, compared to the corpus, the two best candidate driving cycles overall meets 15–16 driving cycle metrics to within 5 % and fuel

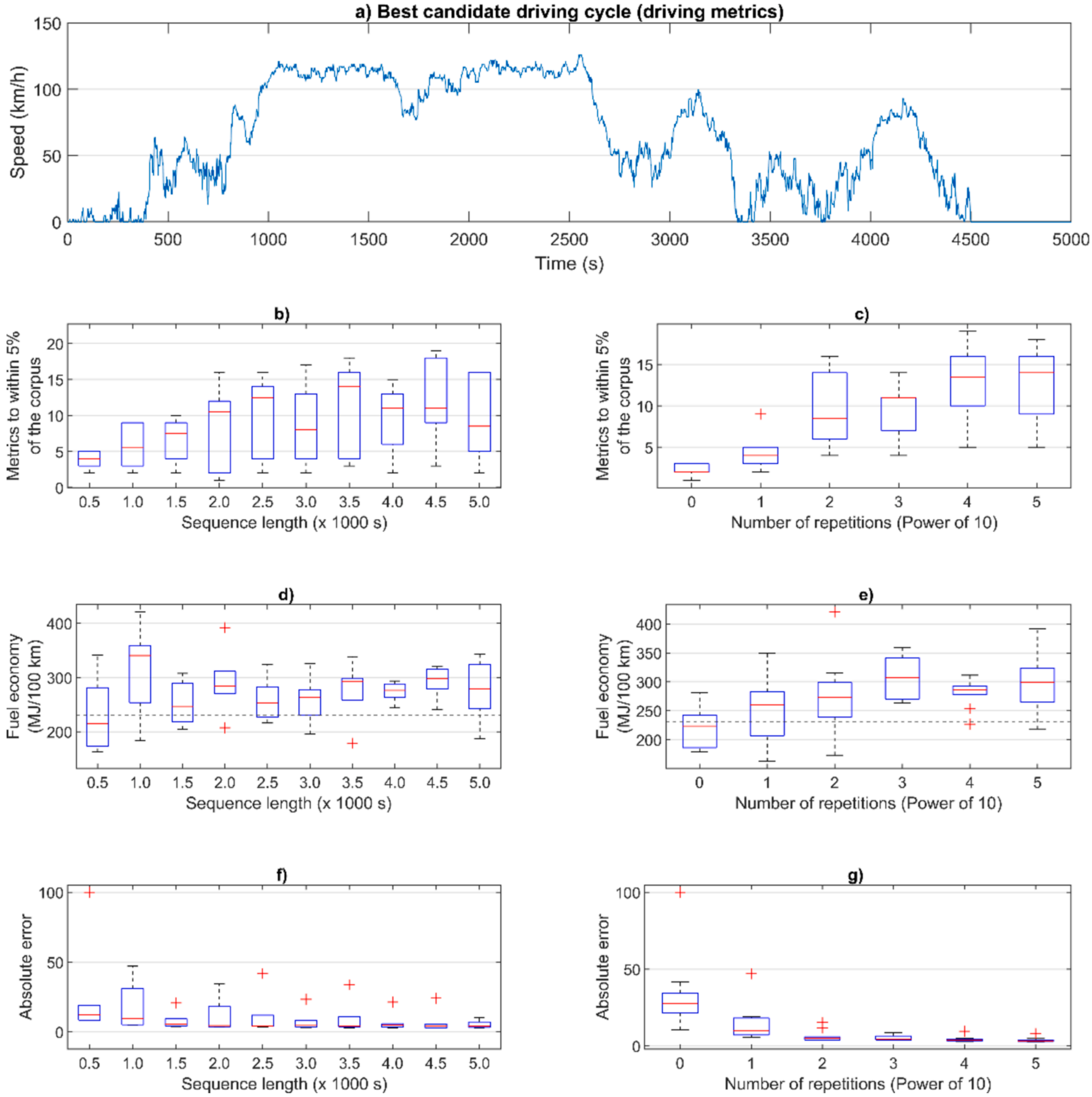

Fig. 8. Best candidate driving cycles based on driving metrics.

economy to within 5 % (two blue crosses within the vertical lines of Fig. 11. Our approach reduced the 29 initial metrics to a set of 26, material to predicting fuel economy. However, only 19 of the 26 material metrics (or 17 of the initial 29 driving cycle metrics as illustrated by the blue crosses in Fig. 11) are met to within 5 % in the family of best candidate driving cycles.

There were some metrics which were difficult to reproduce accurately. To illustrate, we show the 36 candidate driving cycles which met at least 15 metrics to within 5 %. Fig. 12a and Fig. 12b highlight how well these candidate cycles met each metric. The cycles meeting the most metrics to within 5 % of the corpus are to the left, with accuracy decreasing towards the right. This change in accuracy is illustrated by the appearance of more yellow and red squares, indicating errors of greater than 5 % and greater than 10 %, respectively. Errors greater than 10 % persist across all cycles for: the number of accelerations and decelerations per km, average acceleration, average deceleration, and the IQR of acceleration. Some of these metrics are present in the material set (Fig. 12b).

The number of accelerations and decelerations per km reflects the number of speed changes per km travelled. As the transition probability matrix views transitions at resolution of 1 km/h, a transition from 50 km/h to 49 km/h and back to 50 km/h in a single km will influence this metric, even if the driver was intending to drive at a constant speed of 50 km/h. This jitter (statistical noise) might arise from small dips and bumps (or potholes) in the road surface, or the driver shifting their foot slightly on the accelerator pedal. The corpus shows an average of 4.0 acceleration and deceleration events per kilometre. Therefore, one or two additional changes per

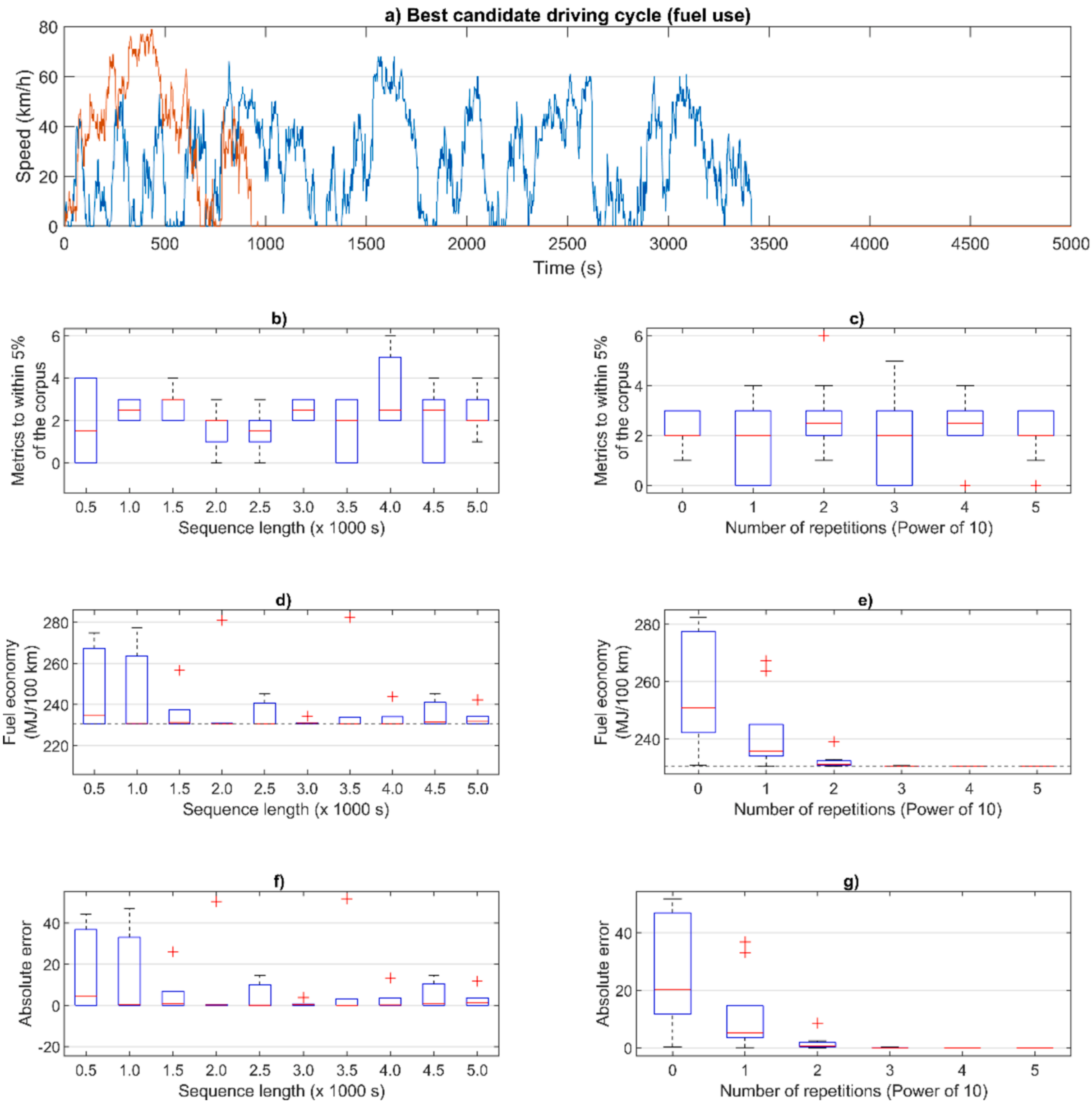

Fig. 9. Best candidate driving cycles based on fuel economy.

kilometre in a candidate driving cycle represents a significant error relative to the corpus, but not a material change in driving behaviour. For example, the number of acceleration and deceleration event per kilometre in two best candidate cycles in Fig. 8a is 6.4–6.6 per km.

The other metrics which are consistently unmet are, or are derived from, acceleration e.g. jolt. Average acceleration and deceleration across the corpus is 0.45 m/s^2 and −0.45 m/s^2, respectively. Therefore, reproducing these metrics, and those derived from them, may be difficult based on their relatively low magnitude.

We distinguish between the noise from the naturally occurring brief changes in speed when trying to maintain a target and the less frequent, but prolonged, acceleration or deceleration events implying driver intent. Fig. 13 shows the number of seconds of consecutive accelerations and/or decelerations in the corpus: more than 90 % of events are shorter than four seconds in duration, with 51 % of events lasting one second only. Therefore, we can conclude most of the speed transitions are noise, with the tail of the distribution more indicative of driver intent. The volume of noise arises, in part, because we use a 1 km/h resolution transition probability matrix. A less granular transition probability matrix would smooth the noise and may represent driver intent more explicitly but deliver a less useful candidate driving cycle.

The test vehicles received type approval under the NEDC, with the combined fuel use in litres/100 km which was converted to MJ/

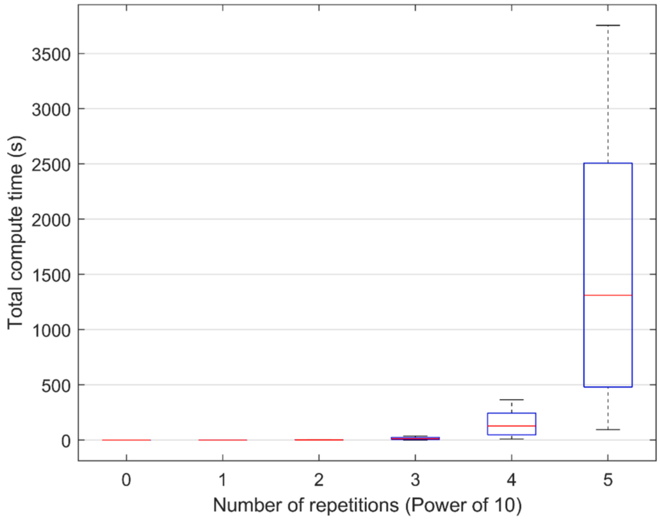

Fig. 10. Evolution of computer time with number of repetitions.

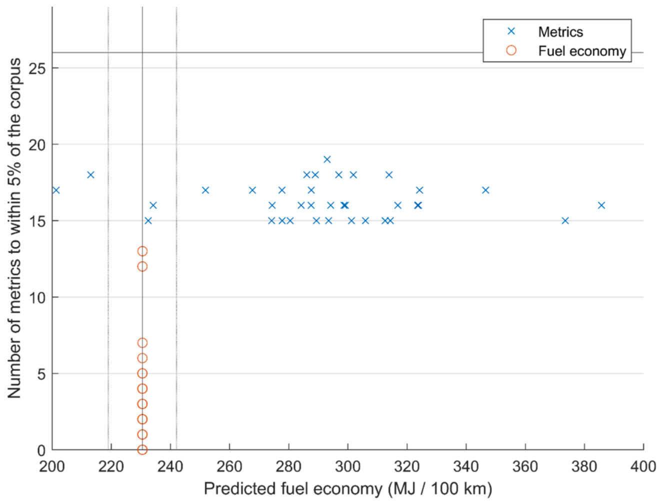

Fig. 11. Trade-off between candidate cycles meeting driving cycle metrics and the fuel economy of the corpus.

100 km (Table 1). The median fuel-use observed in our experiments was 26 % higher than expected under the NEDC combined cycle, ranging from 10 % higher for the 2012 Ford Focus (vehicle 5) to 137 % higher for the 2013 Ford Focus (vehicle 4) (Table 3). Although we do not have comprehensive driving data for every vehicle, this experiment confirms the observations by Dornoff et al. (2024) that type-approval tests, particularly the NEDC, do not reflect real-world fuel use accurately.

4. Conclusion

We introduced a real-world driving cycle generated using a Markov chain method which reproduced accurately the two dimensions of metrics and fuel economy of a natural driving experiment (corpus). We identified 1) a reduced set of metrics which influenced fuel economy materially, 2) trade-offs in accuracy between reproducing vehicle dynamics and fuel economy, and 3) and the impact of natural driving variability (noise) on the accuracy of candidate cycles.

A large set of driving cycle metrics have been used by other studies; we systematically rationalized our set from the initial 29 metrics to a set of 26 material to predicting fuel economy accurately. The elastic net assessment of the materiality of driving cycle metrics is an important contribution to the field of driving cycle development. To date, driving cycle metrics have been used to describe the corpus, with little discussion as to which, if any, were important.

Selecting candidate cycles based on accuracy of reproducing driving cycle metrics or predicted fuel economy yielded different results. In general, it was more difficult to reproduce accurately the driving cycle metrics of the corpus, illustrated by only one candidate cycle meeting 19 driving cycle metrics to within 5 % of the corpus. However, 107 candidate cycles delivered fuel economy

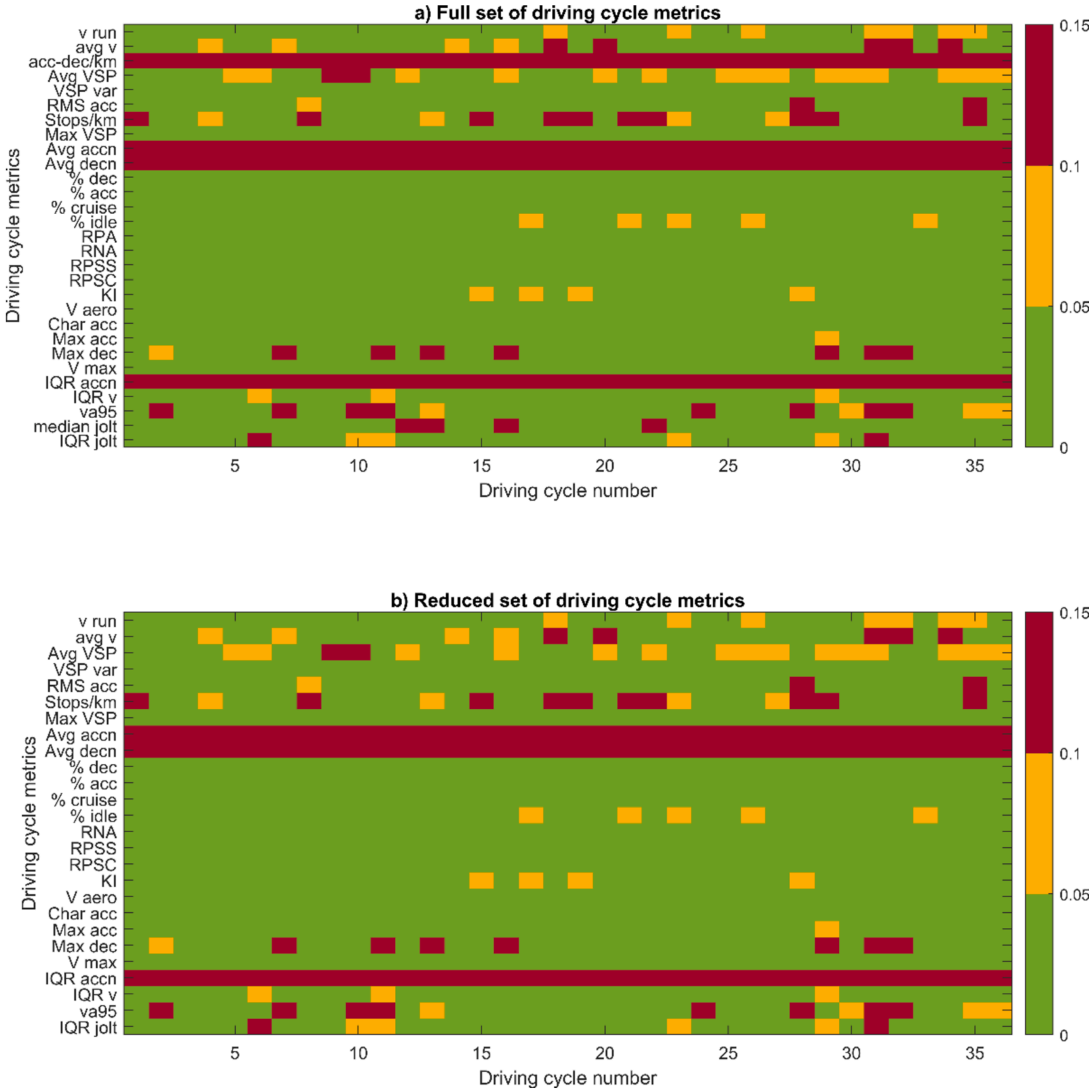

Fig. 12. For the 36 candidate cycles accuracy of meeting a) full set of driving cycle metrics and b) material set of driving cycle metrics. Green illustrates errors of up to 5 %, yellow indicates errors of 5–10 % and red indicates errors greater than 10 %.

within one decimal place of what was observed in the corpus, indicating a trade-off in accuracy, depending on the dimension being optimized.

We found it was difficult for the Markov chain to reproduce accurately the corpus acceleration and acceleration-derived metrics. These acceleration properties have small absolute values, meaning that significant error can arise from immaterial changes in driving behaviour. In general, we found noise dominated the acceleration and deceleration events, masking driver intent. However, as metrics are generally distance- or time-based, this noise is smoothed-out in the aggregation. Investigating the impact of noise is beyond the scope of our work but could be an interesting future study.

The driver-vehicle pairs represent a range of gender, age, extent of driving experience, and vehicle capability. Although the main limitations of this work are number of drivers and the overall sample size, in most cases, the observed driving style was similar across the participants. It transpires that the slope of the roads in our study was unimportant, but this could be addressed in a study with greater topographical diversity. The Markov approach is flexible to accommodate new data and ensures the most common behaviours, captured as observed transitions, are reflected in the candidate cycles. It remains an open question as to the minimum quantity of data to reproduce real-world driving cycles. We suggest that the real-world driving behaviour is determined more by the road topology and presence of other drivers, than the performance potential of the vehicles. This reversion to the mean is useful when considering the

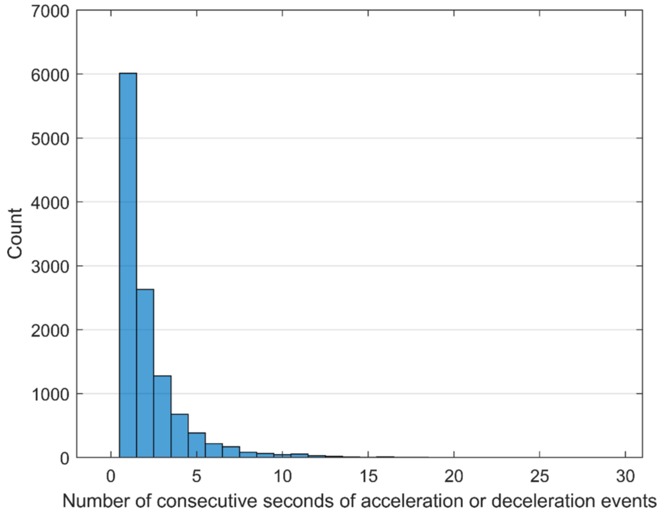

Fig. 13. Distribution of acceleration-deceleration events.

Table 3
Differences between the type approval and observed fuel use for the study vehicles.

Make and model	Model year	Reference no.†	NEDC fuel use (MJ/100 km)	Observed fuel use (MJ/100 km)	Difference (%)
Ford Focus	2011	1232960	164	232	41
Mercedes Benz C180	2009	1550540	213	266	25
Porsche 911	2001	2866790	381	494	30
Ford Focus	2013	1776185	152	360	137
Ford Focus	2012	1592855	157	173	10
Audi A3	2006	1190765	206	259	26
Hyundai i30	2017	2525600	172	240	40
Mercedes Benz C200	2017	2504240	216	n/a	n/a
Hyundai Tuscon	2018	2761880	229	309	35
Citroen Picasso	2018	2513150	159	195	23
Vauxhall Corsa SRi	2018	2769680	178	197	11

† https://www.automobile-catalog.com.

design of future studies.

CRediT authorship contribution statement

J.D.K. Bishop: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **C.J. Axon:** Writing – review & editing, Project administration, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Barlow, T.J., Latham, S., McCrae, I.S., Boulter, P.G., 2009. A reference book of driving cycles for use in the measurement of road vehicle emissions (No. PPR354). TRL, Wokingham, UK.

Bishop, J.D.K., Axon, C.J., McCulloch, M.D., 2012. A robust, data-driven methodology for real-world driving cycle development. Transp. Res. Part D: Transp. Environ. 17, 389–397. https://doi.org/10.1016/j.trd.2012.03.003.

Bishop, J.D.K., Molden, N., Boies, A.M., 2019. Using portable emissions measurement systems (PEMS) to derive more accurate estimates of fuel use and nitrogen oxides emissions from modern Euro 6 passenger cars under real-world driving conditions. Appl. Energy 242, 942–973. https://doi.org/10.1016/j.apenergy.2019.03.047.

Borlaug, B., Holden, J., Wood, E., Lee, B., Fink, J., Agnew, S., Lustbader, J., 2020. Estimating region-specific fuel economy in the United States from real-world driving cycles. Transp. Res. Part D: Transp. Environ. 86, Article no: 102448. Doi: 10.1016/j.trd.2020.102448.

Carslaw, D.C., Rhys-Tyler, G., 2013. New insights from comprehensive on-road measurements of NOx, NO2 and NH3 from vehicle emission remote sensing in London. UK. Atmos. Environ. 81, 339–347. https://doi.org/10.1016/j.atmosenv.2013.09.026.

Carslaw, D.C., Williams, M.L., Tate, J.E., Beevers, S.D., 2013. The importance of high vehicle power for passenger car emissions. Atmos. Environ. 68, 8–16. https://doi.org/10.1016/j.atmosenv.2012.11.033.

Cui, Y., Zou, F., Xu, H., Chen, Z., Gong, K., 2022. A novel optimization-based method to develop representative driving cycle in various driving conditions. Energy 247, 123455. https://doi.org/10.1016/j.energy.2022.123455.

Andrade, G.M.S. de, Araújo, F.W.C. de, Santos, M.P.M. de N., Magnani, F.S., 2020. Standardized Comparison of 40 Local Driving Cycles: Energy and Kinematics. Energies 13, Article no: 5434. Doi: 10.3390/en13205434.

Desineedi, R.M., Mahesh, S., Ramadurai, G., 2020. Developing driving cycles using k-means clustering and determining their optimal duration. Transp. Res. Procedia 48, 2083–2095. https://doi.org/10.1016/j.trpro.2020.08.268.

Dornoff, J., Valverde-Morales, V., Tietge, U., 2024. On the way to 'real-world' CO2 values? The European passenger car market after 5 years of WLTP (White Paper). International Council on Clean Transportation, Berlin, Germany.

Franco, V., Kousoulidou, M., Muntean, M., Ntziachristos, L., Hausberger, S., Dilara, P., 2013. Road vehicle emission factors development: A review. Atmos. Environ. 70, 84–97. https://doi.org/10.1016/j.atmosenv.2013.01.006.

Huertas, J.I., Giraldo, M., Quirama, L.F., Díaz, J., 2018. Driving Cycles Based on Fuel Consumption. Energies 11, Article no. 3064. Doi: 10.3390/EN11113064.

Isenstadt, A., Yang, Z., Searle, S., German, J., 2022. Real World Usage of Plug-in Hybrid Vehicles in the United States. The International Council on Clean Transportation.

Lee, T.-K., Adornato, B., Filipi, Z.S., 2011. Synthesis of Real-World Driving Cycles and Their Use for Estimating PHEV Energy Consumption and Charging Opportunities: Case Study for Midwest/U.S. IEEE Trans. Veh. Technol. 60, 4153–4163. Doi: 10.1109/TVT.2011.2168251.

Lee, T.-K., Filipi, Z.S., 2011. Synthesis of real-world driving cycles using stochastic process and statistical methodology. Int. J. Veh. Des. 57, 17–36. https://doi.org/10.1504/IJVD.2011.043590.

Lee, T., Frey, H.C., 2012. Evaluation of Representativeness of site-specific fuel-based vehicle emission factors for route average emissions. Environ. Sci. Technol. 46, 6867–6873. https://doi.org/10.1021/es204451z.

Ma, R., He, X., Zheng, Y., Zhou, B., Lu, S., Wu, Y., 2019. Real-world driving cycles and energy consumption informed by large-sized vehicle trajectory data. J. Clean. Prod. 223, 564–574. https://doi.org/10.1016/j.jclepro.2019.03.002.

Pavlovic, J., Fontaras, G., Broekaert, S., Ciuffo, B., Ktistakis, M.A., Grigoratos, T., 2021. How accurately can we measure vehicle fuel consumption in real world operation? Transp. Res. Part D: Transp. Environ. 90, 102666. https://doi.org/10.1016/j.trd.2020.102666.

Peng, Y., Zhuang, Y., Yang, Y., 2020. A driving cycle construction methodology combining k-means clustering and Markov model for urban mixed roads. Proc. Inst. Mech. Eng. Part J. Automob. Eng. 234, 714–724. https://doi.org/10.1177/0954407019848873.

Plotz, P., Moll, C., Mock, P., 2021. Form lab-to-road: real-world fuel consumption and CO2 emissions of plug-in hybrid electric vehicles. Environ. Res. Lett. 16. https://doi.org/10.1088/1748-9326/abef8c.

Plotz, P., Link, S., Ringelschwender, H., Keller, M., Moll, C., Bieker, G., Dornoff, J., Mock, P., 2022. Real-world usage of plug-in hybrid vehicles in Europe: a 2022 update on fuel consumption, electric driving, and CO2 emissions. International Council on Clean Transportation, Berlin, Germany.

Qiu, H., Cui, S., Wang, S., Wang, Y., Feng, M., 2022. A clustering-based optimization method for the driving cycle construction: a case study in Fuzhou and Putian China. IEEE Trans. Intell. Transp. Syst. 10, 18681–18694. https://doi.org/10.1109/TITS.2022.3160275.

Quirama, L.F., Giraldo, M., Huertas, J.I., Tibaquirá, J.E., Cordero-Moreno, D., 2021. Main characteristic parameters to describe driving patterns and construct driving cycles. Transp. Res. Part D: Transp. Environ. 97, Article no: 102959. Doi: 10.1016/j.trd.2021.102959.

Sangeetha, R.T., Bose, Ibrahim, M., 2021. Condensation of Real-World Drive Cycle into Synthetic Drive Cycle - An Innovative Method to Predict Real Driving Emissions. Presented at the SAE WCX Digital Summit, pp. 2021-01–0602. Doi: 10.4271/2021-01-0602.

Tutuianu, M., Marotta, A., Heinz, S., Ericsson, E., Haniu, T., Ichikawa, N., Ishii, H., 2014. Development of a World-wide Worldwide harmonized Light duty driving Test Cycle (WLTC) (No. UN/ECE/WP.29/GRPE/WLTP-IG). United Nations, Geneva, Switzerland.

Wang, Z., Zhang, J., Liu, P., Qu, C., Li, X., 2019. Driving Cycle Construction for Electric Vehicles Based on Markov Chain and Monte Carlo Method: A Case Study in Beijing, in: Energy Procedia, Innovative Solutions for Energy Transitions. Presented at the 10th Int. Conf. on Applied Energy (ICAE2018), 22-25 Aug, Hong Kong, China, pp. 2494–2499. Doi: 10.1016/j.egypro.2019.01.389.

Yu, L., Wang, Z., Shi, Q., 2010. PEMS-based approach to developing and evaluating driving cycles for air quality assessment (Research Report No. SWUTC/10/169300-1). Texas Southern University, Houston, Texas, USA.

Zhang, S., Wu, Y., Liu, H., Huang, R., Un, P., Zhou, Y., Fu, L., Hao, J., 2014. Real-world fuel consumption and CO2 (carbon dioxide) emissions by driving conditions for light-duty passenger vehicles in China. Energy 69, 247–257. https://doi.org/10.1016/j.energy.2014.02.103.