Analysis of Three Independent Real World Driving Data Studies: A Data Driven and Expert Analysis Approach to Determining Parameters Affecting Fuel Economy

Stewart A Birrell[1]*, James Taylor[1], Andrew McGordon[1], Joonwoo Son[2], Paul Jennings[1]

[1]WMG, University of Warwick, Coventry. CV4 7AL, United Kingdom

[2]HumanLAB, DGIST (Daegu Gyeongbuk Institute of Science & Technology), Daegu, South Korea

*Corresponding Author: S.Birrell@Warwick.ac.uk; +44 (0) 24 7657 3752

Abstract

It is well established that individual variations in driving style have a significant impact on vehicle energy efficiency. The literature shows certain parameters have been linked to good fuel economy, specifically acceleration, throttle use, number of stop/starts and gear change behaviours. The primary aim of this study was to examine what driving parameters are specifically related to good fuel economy using a non-homogeneous extended data set of vehicles and drivers over real-world driving scenarios. The analysis presented in this paper shows how three differing independent studies looking at the same factor (i.e. the influence of driver behaviour on fuel efficiency) can be evaluated, and, despite their notable differences in country, environment, route, vehicle and drivers, can be compared on broadly similar terms. The data was analysed in two ways; firstly, using expert analysis and the second a purely data driven approach. The various models and experts agreed that a combination of at least one factor from the each of the categories of vehicle speed, engine speed, acceleration and throttle position were required to predict the impact on fuel economy. The identification of standard deviation of speed as the primary contributing factor to fuel economy, as identified by both the expert and data driven analysis, is also an important finding. Finally, this study has illustrated how various seemingly independent studies can be brought together, analysed as a whole and meaningful conclusions extracted from the combined data set.

Highlights

- Analysis of three independent real-world driving studies, from two different countries, using three different vehicles and driving scenarios with 112 participants
- Combined approach to use expert analysis and purely data driven approach to identify driving parameters which relate to good fuel economy
- Standard deviation of speed was the primary contributing factor to fuel economy
- Illustrated how various three independent studies can be brought together, analysed as a whole and meaningful conclusions extracted from the combined data set

Keywords

Statistical analysis; Data-driven analysis; Real-world driving; Fuel economy; Driver behaviour

1. Introduction

It is well established that individual variations in driving style have a significant impact on vehicle energy efficiency, whether this is the fuel economy of an internal combustion engine vehicle (ICEV) or the allowable range of a hybrid or full battery electric vehicle (xEV). Regardless of the powertrain type it will always be of value to use the energy of the vehicle more efficiently. This may even be more crucial for an electric vehicle due to the limited energy of the battery (EC, 2009). One way energy efficiency can be improved is by adopting an eco-driving style; it has been suggested that fuel savings in an ICEV from adopting eco-driving techniques are between 5-15% (see Young et al, 2011 for a discussion). Reviewing the advice presented by numerous organisations (including the AA, RAC, Institute of Advanced Motorists, Energy Saving Trust, US Department of Energy to name but a few, as well as the references cited within this paper) has identified several key factors which contribute to an eco-driving style, and include:

- Plan ahead, anticipate traffic flow and keep a suitable following distance to help maintain a constant speed, while avoiding sharp braking and stops
- Use smooth but positive acceleration to reach high gears and desired cruising speeds sooner.
- Use engine braking (without changing down through the gears) for smooth deceleration, minimising the use of the foot brake where appropriate
- Change gear up as soon as possible (between 2000 and 2500 rpm) and consider the use of block gear changes (i.e. 3rd-5th) where appropriate
- Use uniform throttle positions with no more than half throttle used
- Obey speed limits.

In contrast to eco-driving, fuel economy has been shown to decrease by between 20 and 40% when driving in real-world rural or urban environment when a self-selected ‘aggressive’ driving style is adopted (i.e., heavy accelerations and braking, greater throttle use, higher engine speeds, later gear changes and faster driving speeds). However, aggressive driving on higher speed motorways resulted in a non-significant increase in fuel consumption of between 7 and 12%. In addition, for all conditions in spite of the aggressive driving style adopted journey times were unaffected (De Vlieger, 1997; De Vlieger et al., 2000). Further research conducted by El-Shawarby et al., (2005) showed that exploiting the vehicle’s maximum acceleration capabilities can use up to 60% more fuel than mild or normal acceleration. More specifically, Waters and Laker (1980) demonstrated that the optimal acceleration rate in a highly controlled test track scenario was 0.07 g (0.69 m/s^2), with fuel consumption increasing by 20% as acceleration increased up to 0.18 g (1.765 m/s^2) – which would equate to a 0-60 mph time of approximately 15 seconds. Speed oscillations during cruising of 5 km/h also increased fuel use, by 30% at 40 km/h and by 20% at 120 km/h (Waters and Laker, 1980).

A limited number of studies presented in the literature have examined the relationship between driving parameters (speed, acceleration etc.) which were significantly correlated with improved fuel economy in eco-driving verses baseline driving experimental groups, i.e. the aspects of eco-driving which related to improved fuel economy. Johansson et al (2003) found certain characteristics of driving behaviour that were significantly correlated with good fuel economy, such as avoiding unnecessary stops, low deceleration levels, minimising the use of 1st and 2nd gears, increased use of top gear, and block changing gears where possible. Gonder et al (2011) report that higher than average acceleration values among city trips corresponded with higher fuel consumption, and for highway journeys a higher average speed was the strongest indicator for poor fuel consumption. They also suggest that reducing acceleration rates results in some fuel savings, but reducing acceleration and decelerations altogether (i.e. stop/start cycles) saves a larger amount of fuel (Gonder et al, 2011).

Evans (1979) shown that minimizing stops led to fuel efficiency improvements twice that of simply reducing speed (13.9% verses 6.4%), which was achieved at no cost to journey time. This suggests that driving style improvements should focus on reducing the number of stops and not just the rate of acceleration out of a stop. This can be facilitated by planning ahead to anticipate traffic flow and could also be referred to as reducing the standard deviation (SD) of speed of a specific journey or sector.

Other eco-driving recommendations concern limiting throttle (or accelerator pedal) use to 50% of travel, while still being 'positive' to minimise engine inefficiencies (Johansson et al, 1999; van de Burgwal and Gense 2002), and changing gear before the engine speed reaches the point at which engine torque is at its highest (normally around 3,000 rpm for a gasoline fuelled engine), thereby avoiding driving at excessively high engine speeds (Johansson et al, 1999). Johansson and colleagues (1999) concluded that those who adhered to the above two recommendations, relating to throttle and engine speed, to a 'greater extent' exhibited significantly improved fuel consumption and lower emissions on average than those who followed the rules to a 'lesser extent'.

Specifically relevant to this current paper was the work by Ericsson (2001) who in a two-week instrumented vehicle study with 45 participants derived 16 independent driving factors from over 19000 driving patterns in real-traffic urban driving. When considering their effects on fuel consumption and emissions it was found that nine driving parameters were rated as having an important effect (Table 1).

Table1: Factors (or parameters) which Ericsson (2001) found to have the most significant impact on fuel economy, and their categories within.

Category	Factor	Effect on Fuel Economy and Emissions	Estimated Order Effect
Acceleration	Acceleration with strong power demand	↓	1
	Extreme acceleration	↓	2
	Acceleration with moderate power demand	↓	6
Engine Speed	Late gear changing from 2nd and 3rd	↓	4
	Engine speed > 3500 rpm	↓	6
	Moderate engine speeds in 2nd and 3rd	↑	5
Vehicle Speed	Speed oscillation	↓	6
	Stop factor	↓	3
	Speed 50-70 km/h	↑	6

The literature above has shown certain parameters have been linked to good fuel economy, specifically acceleration, throttle use, number of stop/starts and gear change behaviours. During the literature search for this paper the authors could not find any studies which defined driving characteristics which were correlated to good fuel economy using data collected from numerous initially independent studies. Johnasson et al (1999) did compare the effects of eco-driving tuition on different students driving three different routes in Sweden; however, these routes were specifically developed to be similar in nature and used the same vehicle. Limited research has been conducted when investigated cross-cultural comparisons regarding driver's hazard perceptions (Cheng Lim et al, 2013) and driving skill (Ozkan et al, 2006) to name a few, but not driving behaviours relating to fuel economy.

The primary aim of this study was to examine what driving parameters are specifically related to good fuel economy using a non-homogeneous extended data set of vehicles and drivers over real-world driving scenarios. The data will be analysed in two ways; firstly, using expert analysis and the second a purely data driven approach. The fact that the data set comprises three entirely different studies with three initially different aims, using different vehicles and participants, in three geographically different locations (two in the UK and one in Daegu, South Korea), presents a unique methodology. Results from which could be considered to be more applicable to cross-cultural comparisons; however the challenges of post-hoc analysis on non-uniform methodologies will also be discussed.

2. Methodology

2.1. The Three Studies

As mentioned above data from three separate studies were used for the analysis, table 2 presents a summary of the scenario driven, vehicle used and participants involved. The WMG study was conducted in Warwickshire in the UK, with the principle aim of collecting data for the verification of a driver model generated for the SAVE project. The MIRA study was conducted in conjunction with the TeleFOT EU project and investigated the effects of using in-vehicle information systems (IVIS), with the baseline (or no-feedback) condition providing the data. The DIGST study was conducted in Daegu in South Korea and again evaluated the use of IVIS, with data from the baseline condition being used. Whilst the three different institutions collected data for different purposes there were certain similarities, specifically a project goal to evaluate the effect of driving behaviour on fuel economy. Other similarities were a mixed route, real-world driving scenario; an instrumented vehicle collecting many driving performance and behaviour; and a tightly controlled and rigorous methodology being adopted.

Table 2: Overview of study parameters

		WMG	**MIRA**	**DGIST**
Scenario	Length (miles)	27.3	28.3	13.0
	Avg Time (mins)	65.8	50.1	31.7
	> 60 mph	17.9%	40.6%	45.7%
	>30mph<60mph	47.0%	50.3%	54.3%
	<30mph	35.1%	9.2%	0.0%
Vehicle	Make	Land Rover	Ford	Hyundai
	Model	Freelander	Focus	Genesis
	Fuel Type	Diesel	Diesel	Petrol
	Engine Size (l)	2.2	1.6	3.3
	Transmission	Manual	Manual	Auto

	Stop Start	Y	N	N
Participants	Total	20	40	52
	Male / Female	14 / 6	30 / 10	26/26
	Average Age	35.8	41.9	43.9

2.2. Data Reduction and Normalisation

A total of 75 different driving and vehicle parameters were collected over the three studies. A limited number of which were specific to each individual study, for example headway data with the MIRA study and standard deviation of steering wheel angle with DGIST. Many parameters were present in two of the three studies; however, a total of 14 parameters (19% of the total) were present in all three. Fuel economy was selected as the predictor variable, and time was considered too unique to each test to include. The 12 selected parameters are shown in table 3:

Table 3: Twelve parameters common to all three studies.

Category	**Driving Parameter**
Acceleration	Maximum acceleration (m/s²) Maximum deceleration (m/s²) Maximum lateral acceleration (left) (m/s²) Maximum lateral deceleration (right) (m/s²)
Engine Speed	Average engine speed (RPM) Maximum engine speed (RPM)
Vehicle Speed	Average speed (mph) Maximum speed (mph) Standard deviation of speed
Throttle	Average throttle position (%) Maximum throttle position (%) Standard deviation of throttle position

As the data were collected from several different studies, they need to be normalised so that comparisons between the data can be made. A naïve approach to normalisation was chosen, to assume no underlying knowledge about the distribution of the test parameters. This normalisation would scale all the values so that they have a range of 0-1; this is done by dividing each value by an accepted maximum for the parameter. For parameters where vehicle specifications were available, this was used for the normalisation maximum, and other parameters were normalised using the maximum value observed in that study group independently of the other studies (values expressed as a percentage were not normalised). This allows the studies to be compared in relative terms, and applies no weighting to the parameters under study.

2.3. Expert Verses Statistical Analysis

Two independent methods of analysis were used. The first is termed expert assessment where two of the authors with ‘expert’ knowledge in driver behaviour reviewed the collected data using limited statistical techniques (i.e. only raw data and raw data plots), and ‘eye-balled’ the data to identify trends and possible relationships. The second is a purely data driven approach, using advanced statistical methods involving no knowledge of what is being processed. Both methods were conducted independently, and only when the analyses were completed were the results compared.

3. Results

3.1. Analysis with Expert Knowledge

The 12 parameters that were common to the three studies (table 3) were analysed by the authors with the benefit of expert knowledge. It should be noted here that the data is noisy as befitting a combination of three real world driving studies that were planned and executed independently. The 12 parameters were plotted in raw and normalised forms against fuel economy.

It was clear from the expert analysis that there were certain parameters that had a stronger effect on fuel economy. These relationships were complicated by the fact that there may appear to be observable correlations within the individual study data sets, but when these are plotted as a combined data set there is no relationship. Average vehicle speed is a good example of this as shown in Figure 1 below. When the raw data are plotted together on the same graph (figure 1a) there is a clear tendency for a higher average speed to yield a greater fuel economy, as indicated positive trend when considering the three discrete studies as one data set on the graph. In this instance, since each study has different vehicles and driving routes this may simply illustrate the relative fuel economy of the vehicles in question, a fact that is revealed when the normalised data is considered as shown in Figure 1b. In this case, the relationship of fuel economy to average speed is all but lost.

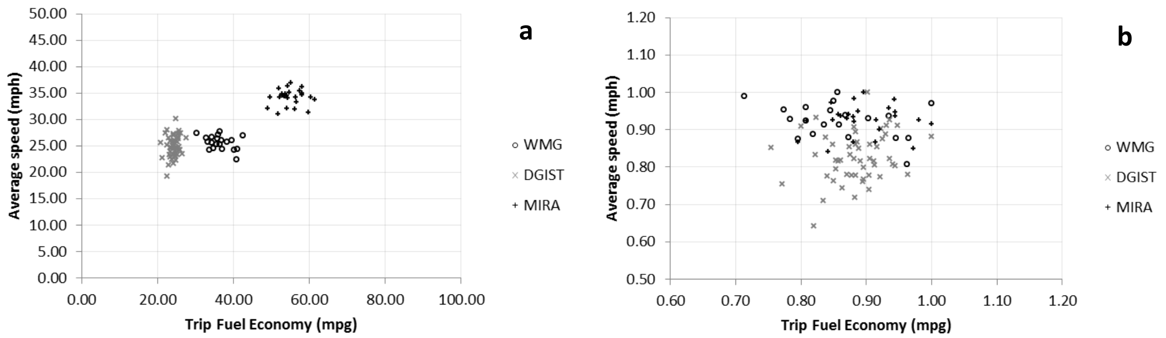

Figure 1: Example of raw (a) and normalised (b) data for average speed

The 12 common parameters identified within the three studies (table 3) can be categorised into four distinct parameter groups related to driver behaviour, once the dependent parameter, fuel economy, and the study-specific time to complete route have been removed:

- Acceleration
- Engine Speed
- Vehicle Speed
- Throttle

The first three categories of acceleration, engine speed and vehicle speed were also identified by Ericsson (2001) and are vehicle outputs. A throttle category was also identified within the parameters which are directly related to driver inputs.

The authors reviewed the raw and normalised data and identified meaningful relationships for each of the parameters against fuel economy, and these relationships are shown in table 4. With the benefit of expert knowledge, some of the parameters that appear to show a relationship to fuel economy could be removed. For example the graphical plots relating to maximum lateral acceleration (right) revealed an observable trend for a possible correlation to fuel economy. However, this could simply indicate the direction of the route (whether clockwise or anti-clockwise), or it could also be dominated by a particular cornering event on the route. For this reason maximum lateral acceleration was not considered one of the driving parameters selected from the expert analysis but maximum acceleration was.

Category	**Driving Parameter**	**Raw Data**	**Normalised**	**Order Effect**
Acceleration	Max acceleration	✓	?	✓
	Max deceleration	?	✗	
	Max lateral accel (L)	?	✗	
	Max lateral accel (R)	?	?	
Engine Speed	Ave engine speed	?	✓	
	Max engine speed	✓	✓	✓
Vehicle Speed	Ave speed	?	?	
	Max speed	✓	✓	✓
	SD of speed	✓	✓	✓
Throttle	Ave throttle position	✓	✓	✓
	Max throttle position	✗	✗	
	SD of throttle position	?	?	

Table 4: Results of the expert analysis. ✓ = Trend observable within the data which is supported by expert knowledge to have a likely effect on fuel economy; ? = Trend observable within the data but expert knowledge suggest may have limited impact on fuel economy; ✗ = No trend observable within the data.

For three of the categories, a single parameter that depends on fuel economy can be straightforwardly deduced from the data relationships. For the speed category, both maximum speed and standard deviation of speed show a relationship to fuel economy as shown in Figure 2.

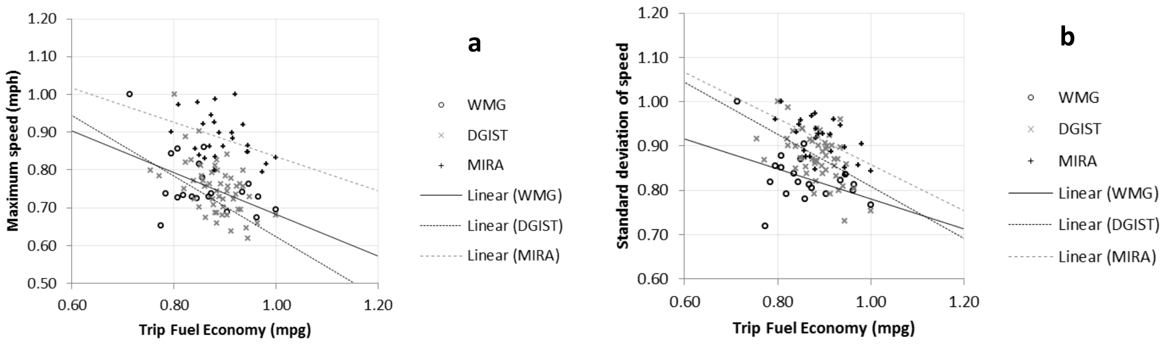

Figure 2: Fuel economy dependence on maximum vehicle speed (a) and standard deviation of speed (b).

With expert knowledge the authors were able to select standard deviation of speed as having a potentially richer description of the effect of driver behaviour on fuel economy. This was both based on the literature – which identified reducing the number of stop/starts as having a positive effect on fuel economy – but also other research conducted by the authors which identified standard deviation of speed as playing an important role in simulated fuel efficiency (McGordon et al, 2014). In addition, the maximum vehicle speed is a number representing a single occurrence during a journey, whereas the standard deviation of speed parameter captures behaviour from the whole journey. Of the parameters selected as shown in Table 4, two of them represent average values of the whole journey. Two parameters are maximum values; maximum acceleration is selected as average acceleration was not available, and may not be a useful parameter. For the engine speed parameter set, the average RPM did not show any relationship with fuel economy and therefore the maximum value was selected, which did show a relationship. Table 5 lists the parameters chosen using the expert knowledge method and Figure 3 shows the resultant dependencies of fuel economy on these parameters. Given the scatter of the points the lines should be treated as a guide to the eye.

Table 5: Parameters selected using the expert knowledge method

Maximum acceleration
Maximum engine speed
Standard deviation of speed
Average throttle

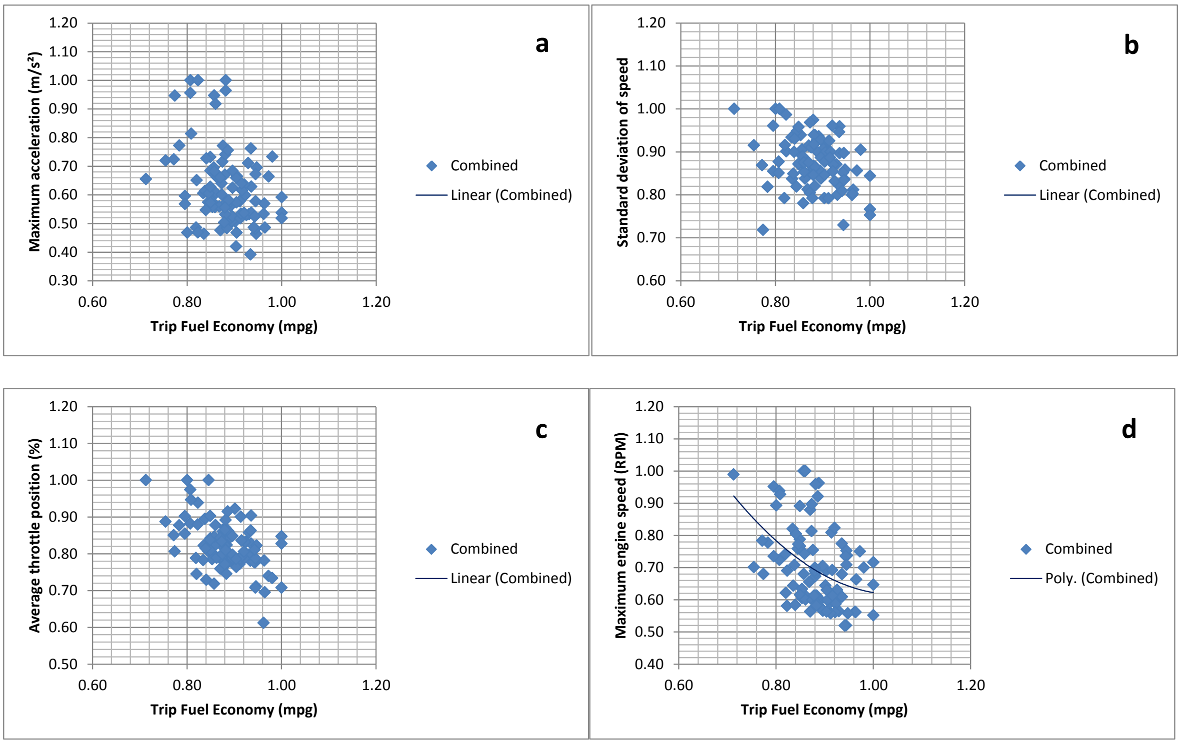

Figure 3: Fuel economy dependence on the four parameters selected through expert knowledge (a) maximum acceleration; b) Standard deviation of speed; c) Average throttle position; d) Maximum engine speed).

The relationships in figure 3 are broadly linear apart from the relationship between maximum engine speed and fuel economy which favours a quadratic fit. The graphs indicate that fuel economy can be increased by:

- Reducing maximum acceleration
- Reducing standard deviation of speed
- Reducing average throttle pedal position
- Reducing maximum engine speed

This analysis above has been performed using high level analysis methods, with the addition of expert knowledge. It has shown that relationships exist between certain parameters and fuel economy. The focus of the next part of this paper is on the ability of the data driven methods to provide an alternative, more robust, method of categorising the relationships.

4. Data Driven Analysis

4.1. Multiple Linear Regression

In order to illustrate the data-driven approach of analysis, prior knowledge of the data was limited to the details of the studies (i.e. vehicles used, number of participants etc.), but no insight into how parameters relate to each other.

When approaching a large data set from multiple sources, the first step is to construct a first pass predictor model to see if any immediate trends become apparent (Azzalini and Scarpa, 2012). With a large number of parameters for a predictor variable (fuel economy), the ultimate aim is to reduce the number of significant parameters to remove noise. A multiple linear regression (MLR) allows to data to be inspected, a prediction model to be constructed and the contribution from each parameter judged (Montgomery, 2013).

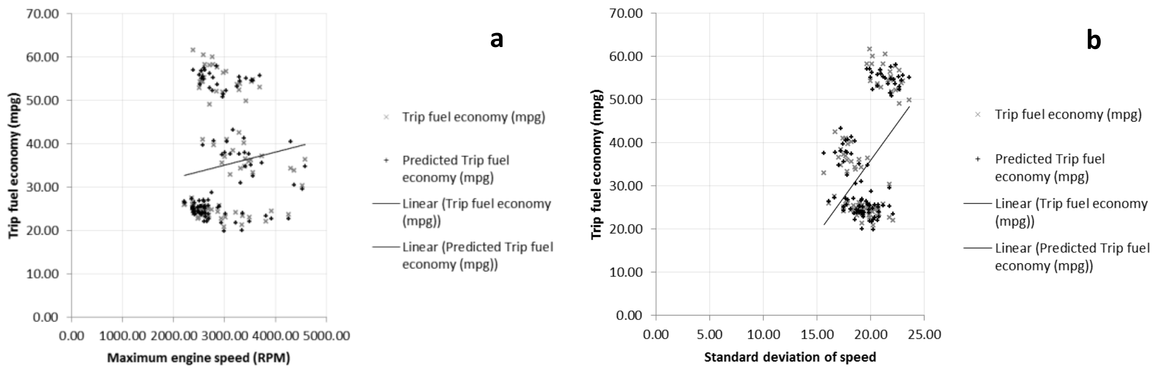

Figure 4: Raw data for maximum engine speed (a) and standard deviation of speed (b) again un-normalised trip fuel economy.

From applying MLR to the whole parameter set, the data from the three studies quickly become apparent, and it is clear that normalisation was required to make the data more directly comparable, as indicated in Figure 4. The R^2 value for the MLR predictor model for fuel economy in this case is 0.968, suggesting a high goodness-of-fit (where a value of 1 would indicate a perfect fit between the raw and predicted data). However, the inspection of the data reveals that this is a clear case of the model over-fitting the data; the predicted values generated by the model closely match the original data instead of aligning themselves to any kind of model. Also, the data are visually arranged in 3 groupings instead of lying on some kind of trend line or curve, showing that the model generated is acting as a classifier, not a predictor.

Normalisation was carried out as described in Section 2.2 and the MLR performed on the normalised data set; example results are shown in Figure 5.

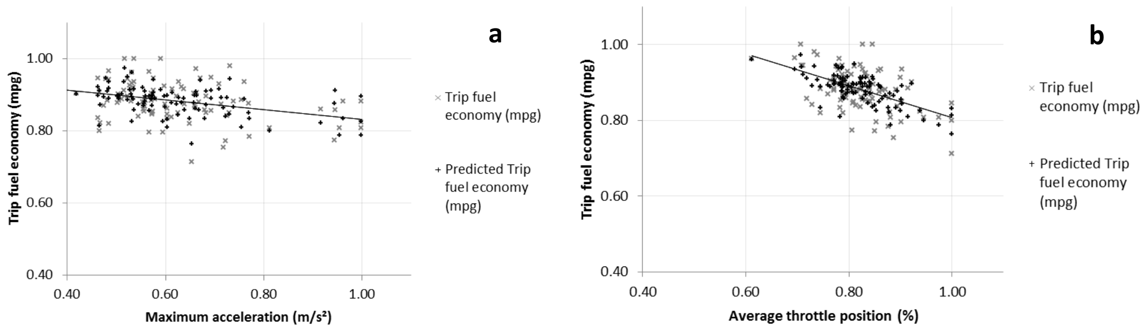

Figure 5: Normalised fuel economy for maximum acceleration (a) and average throttle position (b).

The R^2 value for MLR performed on the normalised data set is reduced to 0.518, but inspection shows the model is now being applied less as a classifier and more as a predictor; the predicted data points no longer directly match the raw data, and the data is visually moving towards a linear trend instead of data groups. The correlation co-efficients (P-value) give a measure of the contribution of each parameter to the model as shown in Table 6.

Table 6: Correlation co-efficients (P-value) determined from the MLR analysis.

Parameter	**P-Value**
Standard deviation of speed	0.865
Average engine speed (RPM)	0.768
Maximum deceleration (m/s^2)	0.762
Maximum throttle position (%)	0.335
Maximum lateral deceleration (m/s^2)	0.216
Maximum engine speed (RPM)	0.210
Maximum speed (mph)	0.155
Maximum lateral acceleration (m/s^2)	0.086
Maximum acceleration (m/s^2)	0.066
Standard deviation of throttle	0.047
Average speed (mph)	0.005
Average throttle position (%)	0.001

Three factors have significantly higher P-values, indicating them as the major contributing factors to the model:

- Standard deviation of speed
- Average engine speed
- Maximum deceleration

4.1 General Linear Regression

At least one parameter (maximum engine speed – figure 3d) was identified as possibly having an exponential or polynomial response, suggesting the assumption of linear response of fuel economy to the changes in parameters, required for MLR, may not apply to all parameters. An alternative approach to regression analysis, general linear regression (GLR), removes this dependency from the regression model, removing any underlying assumptions in regard to the data, such as a normally disturbed or linear response of the prediction variable to a change in any of the parameters (Montgomery, 2013). This model also more readily reduces to a combined parameter matrix solution, allowing a single combined prediction variable from the model instead of a series of regression equations as used in MLR.

The model using the combined prediction variable has a correlation of 0.714. This value is a measure of how well the predicted data represents the raw data.

Table 7: Correlation co-efficients (P-value) determined from the GLR analysis.

Parameter	**P-Value**
Standard deviation of speed	0.896
Average engine speed (RPM)	0.773
Maximum deceleration (m/s^2)	0.605
Maximum throttle position (%)	0.388
Maximum lateral deceleration (m/s^2)	0.296
Maximum speed (mph)	0.169
Standard deviation of throttle position	0.102
Maximum engine speed (RPM)	0.088
Maximum lateral acceleration (m/s^2)	0.060
Maximum acceleration (m/s^2)	0.053
Average throttle position (%)	0.001
Average speed (mph)	0.001

The R^2 value for this GLR is 0.509, utilising 12 parameters, removing time to complete from the variable set. The top three parameters, having produced significantly higher P-values, selected here are:

- Standard deviation of speed
- Average engine speed
- Maximum deceleration.

Whilst a linear model is suggested by inspecting the data, the noisy nature of the available data makes goodness-of-fit less than desired, and possibly may not be the optimal model for this data. The correlation value is sufficient for the model to be representative of the data, however the R^2 value is low (values of 0.9 or greater indicate a good fit to the linear model). Other common models for prediction were considered for comparison.

4.2.3 Neural Network

Neural networks are a non-parametric method of analysing data, inspired by biological processing carried out in nature's nervous systems (Haykin, 2009). Parameter data is input in a weighted manner to processing 'neurons' (known as the hidden layer), which process the data and combine their outputs to produce a result. Multiple output neurones can be used to construct a classifier, while a single output neuron is used for prediction. The weightings of the data paths within the neural network are defined through training the network using test data, and using the error of the result as feedback to adjust the weights.

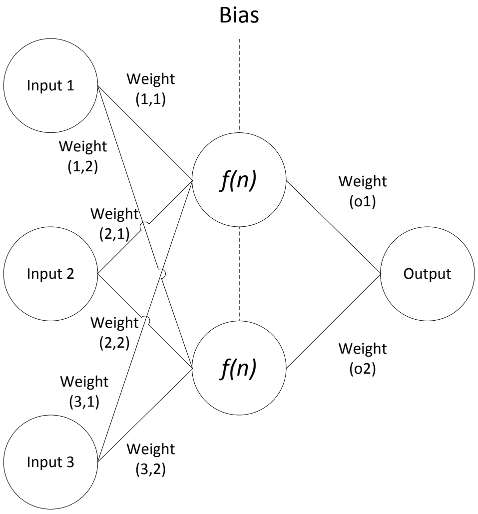

Figure 6: A simple multi-layer perceptron neural network with 3 input neurons, 2 neurons in the hidden layer and a single output neuron

A neural network was constructed, using a multi-layer perceptron (structurally illustrated in Figure 8) with 5 neurons in the hidden layer, 10 parameters as input neurons and one output neuron. Due to the limited availability of data for each study, the neural network was trained using a bootstrap training approach (Ananda Rao and Srinivas, 2003), splitting the data into randomised subgroups, and training the network using each of the subgroups as the testing group in turn, and resetting the network between each group. The MLP was able to achieve a prediction accuracy of 65% (the per cent of test data points that have 90% confidence at a 5% level of matching the source data) on these testing groups, and a model correlation of 0.775. The performance of this non-parametric approach is roughly comparable to the MLR and GLR approaches, with the noise affecting the accuracy, and a model with similar correlation strength.

Table 8: Predictor importance as determined from the Neural Network analysis.

Parameter	**Predictor Importance**
Average Throttle Position	0.24
Maximum Acceleration	0.13
Standard Deviation of Speed	0.13
Maximum Engine Speed	0.08
Maximum Lateral Acceleration	0.07
Maximum Throttle Position	0.07
Average Engine Speed	0.07
Maximum Speed	0.05
Standard Deviation of Throttle Position	0.05
Minimum Lateral Deceleration	0.04
Maximum Deceleration	0.04
Average Speed	0.03

Three parameters have a clear significance in the predictor:

- Average throttle position
- Maximum acceleration
- Standard deviation of speed

4.2.4 CHAID Tree

A CHAID (Chi-squared Automatic Interaction Detector) tree (Tuffery, 2011) differs from the other approaches presented here, as it builds up the parameter set by the addition of effective parameters, rather than starting with the entire parameter set and removing ineffective parameters. At each step of the tree, a parameter is added and the accuracy calculated using an F-test (due to the continuous nature of the data, this is used over the definition chi-squared test), retaining the parameter with the best accuracy. The data are split into accuracy groups, and the process repeated across available groups. This is repeated until all parameters are applied or the accuracy is no longer improved.

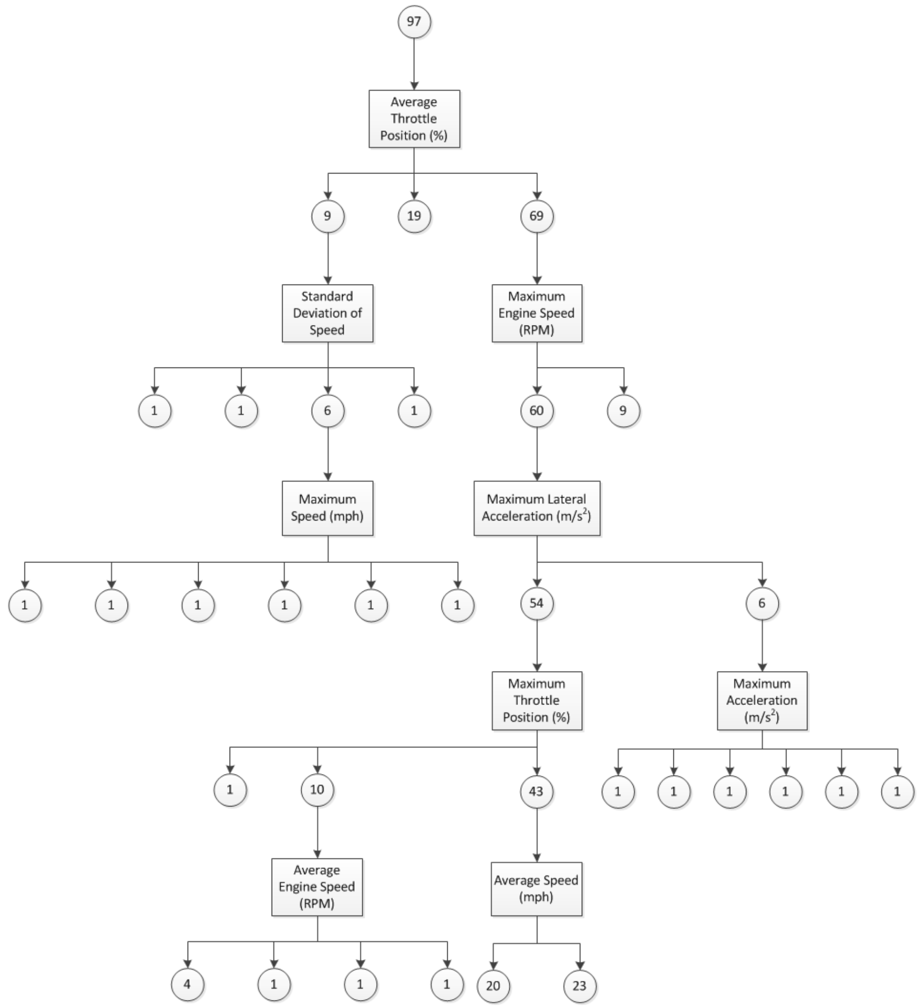

Figure 7: The CHAID tree output from the three independent studies, with the most important parameters located at the top of the tree.

The tree starts with all the data points and performs the F-test against each parameter. The parameter that produces the most distinction is chosen, and the data points split into statistically significant groups. The F-test is repeated across each of the groups at this level using the remaining parameters, choosing the parameters with best distinction. If a threshold level of distinction is not reached, that branch is closed and no more tests will be performed on it. This process repeats until single classifications are reached, no more groups reach the distinction threshold or all parameters have been used within the tree.

The complete CHAID tree (figure 7) uses 9 parameters, omitting maximum deceleration, maximum lateral deceleration and standard deviation of throttle position, and achieves a model correlation of 0.808, and the first three parameters applied were:

- Average throttle
- Standard deviation of speed
- Maximum engine speed

5. Discussion

The expert knowledge analysis performed by the authors identified high level, visual, relationships to be determined from normalised data. By pairing this information with their expert knowledge, they were able to sort the parameters into groups. The combination of analysis and knowledge made it possible to select a parameter from each of the main parameter categories (table 4) to represent the dependence of fuel economy on these parameters. However, the trends and data observed at this level gave no ability to judge whether one factor exerted more influence on the fuel economy than another.

The MLR and GLR approaches selected identified linear trends when comparing several parameters to fuel economy. Even though MLR considered the relationship on a parameter-by-parameter basis and GLR considered the data as a whole, both selected the same variables; standard deviation of speed was chosen as the most important parameter, with average engine speed and maximum deceleration being second and third parameter.

The non-parametric models (neural network and CHAID) approaches identified different parameters which ranked highly in terms of their influence on fuel economy. Both of these methods identified average throttle as the factor with the most influence on fuel economy, and both placed standard deviation of speed in the top three factors. However, the neural net chose maximum acceleration for its third factor while the CHAID tree selected maximum engine speed. Table 9 summarises the results of the analysis for the five methods employed.

Table 9: Comparison of parameters for the different analytical methods

Category	Driving Parameter	Expert	MLR	GLR	NN	CHAID
Acceleration	Max acceleration	✓			2	
	Max deceleration		3	3		
	Max lateral accel (L)					
	Max lateral accel (R)					
Engine Speed	Ave engine speed		2	2		
	Max engine speed	✓				3
Vehicle Speed	Ave speed					
	Max speed					

	SD of speed	✓	1	1	3	2
Throttle	Ave throttle position Max throttle position SD of throttle position	✓			1	1

It is clear from the above table that in particular the data driven approach supports the expert knowledge of the researchers for selection of the Standard Deviation of Speed and Average Throttle parameters. With respect to engine speed the stats could be interpreted to support the selection of average engine speed over maximum engine speed, wither either maximum acceleration or deceleration being an appropriate parameter to be selected.

When the data driven approach is used, interestingly the top 3 parameters for each approach all come from different parameter categories. This indicates the importance of considering a range of parameters for analysis of the effect of driver behaviour on fuel economy. These parameters match the groupings as identified by expert analysis. It should be noted the fourth ranked parameter in each of the statistical analyses, while not as strong as the three chosen factors, was still significantly stronger than the rest of the parameters and belonged to the remaining grouping. Ranked 4th for the regression models were max throttle, and for the non-parametric tests was maximum engine speed (neural network) and maximum speed (CHAID tree).

By traditional quantifiable measures, the models did not perform well, although certain levels of acceptable linear trends could be generalised from them. However, the data comes from 3 real-world studies carried out independently, in a non-standardised manner to investigate the same dependent factor, fuel economy. The levels of noise and uncertainty in the data are very high, so high goodness-of-fit should not be expected (and require further investigation when they do occur, such as during the first stage of analysis). In light of this, the analysis performed acceptably, drawing conclusions from the data by means of linearity and the relative importance of factors, even if a solid predictor model could not be established.

One danger that needs to be considered during the regression analysis is producing a model which over-fits the data, as seen during the non-normalised analysis. The model fits the data, but acting in a classification mode instead of the desired predictor mode. For the purpose of predicting fuel economy based on vehicle factors, this was undesirable. However, remembering that the same technique can act as both as predictor and classifier may lead to additional observations. Figure 4 shows the data forming clusters, showing that another distinct factor may be at work (for instance, driver aggressiveness).

The ease of visual analysis of regression was important to determining the properties of the model, as well as offsetting the issue of over-fitting models. While the abstract non-parametric models (neural network and CHAID tree) performed marginally better in quantifiably ranking the variables, this was at the expense of identifying the underlying trend within the data, in addition to the more complex approach to analysis. While the MLR showed the individual contribution of each parameter, the GLR allows a more unified appreciation of the data, giving an overall picture of a linear, if noisy, relationship between the parameters and fuel economy existing across all the data sets.

The outcomes from this study have demonstrated that there are two principle parameters which were identified by both the expert and data driven approach to have a significant impact on fuel economy during real-world driving; these were standard deviation of speed and average throttle position. Both methods of analysis also identified two further parameters of importance from the two remaining categories of engine speed and acceleration. These relationships indicate that it might be possible for fuel economy of vehicles to be positively impacted by driver behaviour through reducing maximum engine speed and average throttle position, but also by reducing acceleration and deceleration rates and standard deviation of speed. These findings are supported by the literature presented at the beginning of this paper.

Of particular interest was the impact that standard deviation of speed had on real-world fuel economy. McGordon and colleagues (2014) have shown that over different simulated drive cycles with the same average speed, fuel economy can differ by a factor of two, and they suggest that the addition of standard deviation of speed could improve the accuracy of traffic emissions models. In addition, within the eco driving literature it is commonly stated that planning ahead to maintain a consistent speed profile is beneficial for fuel economy, however finding a measure to represent this has proved a challenge. Ericsson (2001) identified speed oscillations and stop factor, whereas Gonder et al (2011) categorised stop/start cycles. This paper suggests that standard deviation of speed could also be used as a reliable factor for categorising fuel economy in real-world driving scenarios.

6. Conclusions

The expert analysis of the data made observations of the data to arrive at a set of conclusions, based on visual analysis of the data and interpreting observed features through expert knowledge, categorising the variables into groups and selecting the most favourable feature from each groups as the best indicator of fuel efficiency. The data-driven analysis applied regression to the data, reducing the data set to a number of most significant variables based on their contribution to the regression. The data-driven analysis supports the expert’s opinions, and, importantly, provides a means to quantitatively classify their observations.

The analysis presented also shows how three differing independent studies looking at the same factor can be brought together for analysis, and, despite their notable differences in environment, route, vehicle and drivers, compared on broadly similar terms. While standardised methodologies will provide a better quality of analysis, this considerably limits the scale and scope of the analysis that can be performed, and what historic data can be used to support a study. The different factors present permit the construction of a more generalised model that can be built on for more specific circumstances, allowing future work be targeted more easily at an early stage of planning.

The various models and experts agreed that a combination of at least one factor from the each of the categories of vehicle speed, engine speed, acceleration and throttle position were required to predict the impact on fuel economy. While the model could not generate a clear predictor model, enough evidence is available to show the trends that arise from the impact factors, and the general effect of changing a parameter could be predicted.

With this diverse non-standardised data, the most informative approach to analysis is to use a multiple linear regression model or general linear regression model. While other models were able to produce slightly more statistically accurate models from the data, the visual analysis of the regression model provides a more easily presented connection between factors and fuel efficiency, and more clarity for identifying trends within the data.

Finally, this study has illustrated how various seemingly independent studies can be brought together, analysed as a whole and meaningful conclusions extracted from the combined data set. The identification of standard deviation of speed as the primary contributing factor to fuel economy is also an important finding.

References

Ananda Rao M and Srinivas J, 2003, Neural Networks: Algorithms and applications, Alpha Science International

Azzalini A and Scarpa B, 2012, Data analysis and data mining: an Introduction, Oxford University Press

De Vlieger, I., 1997. On-board emission and fuel consumption measurement campaign on petrol-driven passenger cars. Atmospheric Environment, 31(22), 3753-3761.

De Vlieger, I., De Keukeleere, D. and Kretzschmar, J., 2000. Environmental effects of driving behaviour and congestion related to passenger cars. Atmospheric Environment, 34, 4649-4655.

EC, 2009. Impact of Information and Communication Technologies on Energy Efficiency in Road Transport. TNO report for the EC (pp. 1-126). Available at: http://ec.europa.eu/information_society/activities/esafety/doc/studies/energy/energy_eff_study_final.pdf

El-Shawarby, I., Ahn, K., Rakha, H., 2005. Comparative field evaluation of vehicle cruise speed and acceleration level impacts on hot stabilized emissions. Transportation Research Part D 10, 13-30.

Ericsson, E., 2001. Independent driving pattern factors and their influence on fuel use and exhaust emission factor. Transportation Research Part D 6, 325-345.

Gonder, J., Earleywine, M. and Sparks, W., 2011. Final report on the fuel saving effectiveness of various driver feedback approaches. National Renewable Energy Laboratory, U.S. Department of Energy, March 2011, 1-31.

Haykin S, 2009, Neural Networks and Learning Machines, Pearson

Lim PC, Shepard E, Crundal, D. 2013. Cross-cultural effects on drivers' hazard perception. Transportation Research Part F, 21, 194-206

McGordon A, Birrell S, Poxon J, Jennings P. 2014. Comparison of Fuel Economy over Different Drive Cycles Each Having the Same Average Speed. Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014, Kraków, Poland 19-23 July 2014.

Montgomery DC, 2013, Introduction to Linear Regression Analysis, Wiley

Ozkan T, Lajunen T, Chiliaoutakis JE, Parker D, Summala H. 2006. Cross-cultural differences in driving skill: A comparison of six countries. Accident Analysis and Prevention, 38 1011-1018.

Tuffery S, 2011, Data mining and statistics for decision making, Wiley

van de Burgwal, H., and N. Gense. 2002. Application of 'New Style Driving' and Effects on Passenger Car Fuel and Emissions, TNO Automotive, TNO Report 02.OR.VM.004.1/HVDB, 1–59.

Waters, M. and Laker, I., 1980. Research on fuel conservation for cars. Crowthorne: Transport and Road Research Laboratory (pp. 1-38).

Young MS, Birrell SA and Stanton NA., 2011. Safe driving in a green world: A review of driver performance benchmarks and technologies to support 'smart' driving. Applied Ergonomics, 42, 533-539.

Acknowledgements

The authors acknowledge the support given for this research. The ‘WMG’ trails were supported by the Technology Strategy Board through the WMG centre High Value Manufacturing Catapult and EPSRC through Warwick Innovative Manufacturing Research Centre. The TeleFOT project, for which the ‘MIRA’ trials were conducted is an EU sponsored project under the Seventh Framework Programme. The ‘DIGST’ trials were supported in part by the DGIST Research Program of the Ministry of Science, ICT and Future Planning, and Establishment Program of Industrial Original Technological Base of the Ministry of Trade, Industry and Energy (MOTIE).