DERIVATION OF PREDICTIVE MODELS OF VEHICLE EXHAUST EMISSIONS FROM INSTRUMENTED VEHICLE MEASUREMENTS: ANALYTICAL AND STATISTICAL CHALLENGES

Glyn Rhys-Tyler

Postgraduate Researcher

Transport Operations Research Group, Newcastle University

Professor Margaret C. Bell

Science City Chair of Transport and the Environment

Transport Operations Research Group, Newcastle University

Abstract

This paper explores some of the analytical and statistical challenges encountered when seeking to derive predictive models of vehicle exhaust emissions from instrumented vehicle measurements. Data at a microscopic temporal resolution (1Hz) from forty drivers on a suburban route is utilised as a case study. A range of analytical issues are highlighted, including temporal synchronisation of dependent and independent variables, determining the most appropriate temporal sampling resolution, multi-modality in the frequency distribution of exhaust emissions data relating to vehicle operating modes, variability in driver behaviour, and specification of predictive models. Illustrative examples of each issue are presented, utilising the case study as a test case.

1. Preamble

The development of predictive models of road vehicle (internal combustion engine) exhaust emissions of sufficient robustness for use in local air quality assessment requires an adequate understanding of the relevant generative system processes encompassing:

- Automotive control systems (engine management, emissions control, sensors and actuators);
- Fuel delivery, combustion, and exhaust system operation;
- Driver behaviour (nature, range and variability of system inputs);
- Combustion chemistry;
- Catalyst / emissions control system chemistry;
- Atmospheric chemistry.

Further, the nature, operation, and limitations of the system monitoring (data collection) methodology and associated instrumentation need to be fully understood to allow correct processing and interpretation of data gathered from monitoring campaigns. This encompasses issues such as:

- Type of sensor technology used to measure variation in relevant system parameters;
- Type of sensor technology employed to measure exhaust gas composition;
- Sensor sensitivity, specificity, cross-sensitivity, resolution, and response times;
- Physical location of sensor measurements within the system;
- Quantification of time lags between system inputs (potential independent variables), system outputs, and sensor outputs, including variations in components of exhaust emissions;
- Time resolution of the sensor measurements themselves;
- Range of operational conditions under which measurements were taken.

Clearly, the scope of this problem is wide, and the potential complexity challenging. This paper does not attempt to provide a comprehensive treatment of the issues, but provides some illustrative examples of the type of problem encountered when analysing data obtained from a vehicle emissions monitoring campaign, with a view to the development of predictive models. A more comprehensive treatment will be the subject of later publication. Initially, this paper provides brief details of the measurement campaign which provided the data for this research. This is followed by a description of the issues relating to temporal synchronisation and resolution. The frequency distributions of measured vehicle characteristics paves the way for an elaboration of the driver characteristics based on a cluster analysis. Finally, before conclusions are drawn, the reader is introduced to preliminary ideas of the proposed modelling framework expected to evolve from this work.

2. Background to the dataset

The RETEMM project (Real-world Traffic Emissions Monitoring and Modelling), funded by the UK Engineering and Physical Sciences Research Council, and completed in 2008, investigated the impact of driver behaviour on emissions (Bell et al 2008). Simultaneous measurements of real-world traffic conditions, driver behaviour, and instantaneous vehicle emissions were undertaken. The analysis presented in this paper utilises some of the data generated in the RETEMM project to illustrate various aspects of the data analysis process being developed to improve the tailpipe emission prediction capability in traffic micro-simulation models. A short (0.6km) circular route was defined in a suburban residential area characterised by having priority junctions only. The route comprised four reasonably straight sections varying in length between 140 and 165 metres. Forty drivers were recruited, comprising 20 male and 20 female. The sample of drivers spanned a wide range of driving experience, and ages from 21 to 63. Drivers drove ten laps of the circuit. The analysis presented here utilises data collected from a Euro 4 standard petrol 1800cc SI passenger car with manual transmission fitted with an HORIBA OBS-1300 emissions measurement system, combined with GPS vehicle positioning (Daham 2006). Data were collected at 1 Hz. Further information can be found in Bell M.C. and Rhys-Tyler G.A. (2008) and Rhys-Tyler G.A and Bell M.C. (2009a & 2009b).

3. Temporal synchronisation

According to INRETS(2006), there are a number of potential systematic problems associated with the measurement of instantaneous emissions. The emissions signals recorded by exhaust gas analysers are delayed in time and smoothed compared to the emission events at the location of formation due to:

1. The transport of the exhaust gas to the analysers;
2. The mixing of exhaust gas, especially in the vehicle silencer and measurement equipment;
3. The response time of the analysers.

Research published by Le Anh et al (2006) presents mathematical techniques to attempt to address these issues by explaining the change of the emission value from its location of formation to the exhaust gas analyser signal by formula, and then by inverting these formulae to obtain equations which transform the analyser signal into the engine out (or catalyst-out) emission value. This technique has been used within the EU 5th Framework ARTEMIS project by both the Technical University of Graz and by the EMPA research institute in Switzerland (INRETS 2006). Figure 1 is adapted from the ARTEMIS work, and illustrates both the emissions signal time offset and smoothing effects, and the signal reconstruction using the equation inversion approach. In this example, an oxygen signal at the catalyst outlet (measured at location ‘C’) is reconstructed from the analyser signal (measured at location ‘F’). One of the conclusions drawn from this work is that the use of data from sensor measurements which have not been corrected for time alignment and signal smoothing will potentially lead to errors in the allocation of emissions to corresponding engine operating conditions (INRETS 2006). Static time realignment may be satisfactory in some limited circumstances where engine operation is constant in terms of engine speed, load, and resultant gas transit times in the vehicle exhaust and measurement system. However, in the real world, engine speed and load have the potential to be highly variable, due to:

- The designed operational range of the vehicle’s engine;

- Fuel type (diesel engines tend to operate over a narrower and lower engine speed range than petrol engines);
- Variability of behaviour (system inputs) within individual drivers;
- Variability of behaviour (system inputs) across drivers within the wider population.

The RETEMM data set utilised within this case study was originally processed by colleagues at the University of Leeds employing techniques described in Ropkins et al 2007. A constant time offset was assumed which accounted for the delay time due to sample measurement and sensor response time i.e. the delay between points E and F in Figure 1. However, it was recognised by the researchers that this approach did not explicitly take account of variation due to the dynamic nature of gas transit times within the vehicle exhaust system due to changes in engine speed and load. Ropkins et al (2009) have subsequently suggested alternative approaches using correlation optimised warping techniques.

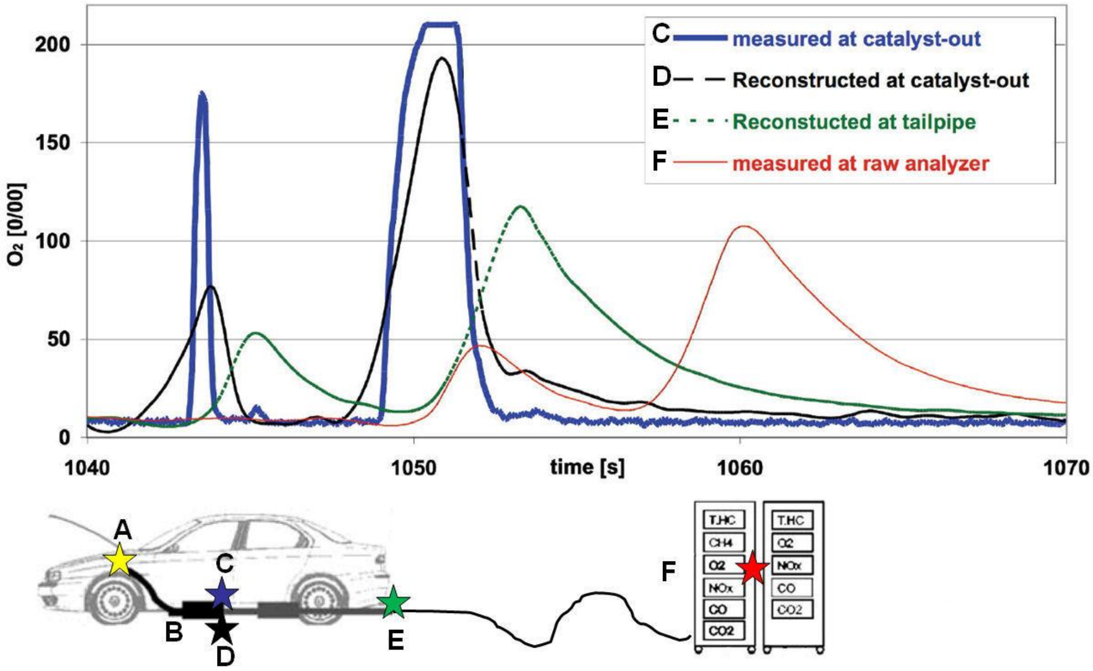

Figure 1: Variability of time alignment between engine operation and measured emission signal, and reconstruction of source emissions using sensor signal inversion techniques. Adapted from INRETS (2006), and Ajtay et al (2004).

The constrained nature of the route used to collect the data in this case study would (arguably) limit the range of variability one would expect to observe in the dataset in terms of temporal offset between emission generation and emission measurement. However, variability may be observed within the sample of 40 drivers. To investigate this phenomena further, a simple linear correlation was carried out between emissions (g/sec), and vehicle acceleration (-/+m/s^2) and throttle position (0-100%) with varying time offsets. Similar techniques have been used by North et al (2006). Intuitively (and assuming that a relationship exists between the prospective dependent and independent variables), one would expect to observe different time offsets for throttle position and acceleration since throttle application (and consequent combustion) occurs before acceleration takes place. A time offset range of -5 seconds (before the emissions event) to +5 seconds (after the emissions event) was explored in 1 second steps. Figure 2 presents the results graphically for the 40 drivers. This analysis has the potential both to indicate the appropriate static time lag between prospective dependent and independent variables, and to highlight variability between drivers, but also to indicate where linear relationships (positive or negative) might exist. Carbon dioxide emissions display the strongest positive correlation with both acceleration and throttle position. All drivers displayed the strongest correlation for CO_2 with throttle position at $\Delta t=0$ time offset. The majority of drivers (32) displayed the strongest correlation with acceleration at $\Delta t=-1$ time offset, although 8 drivers were observed to have the strongest correlation at $\Delta t=-2$.

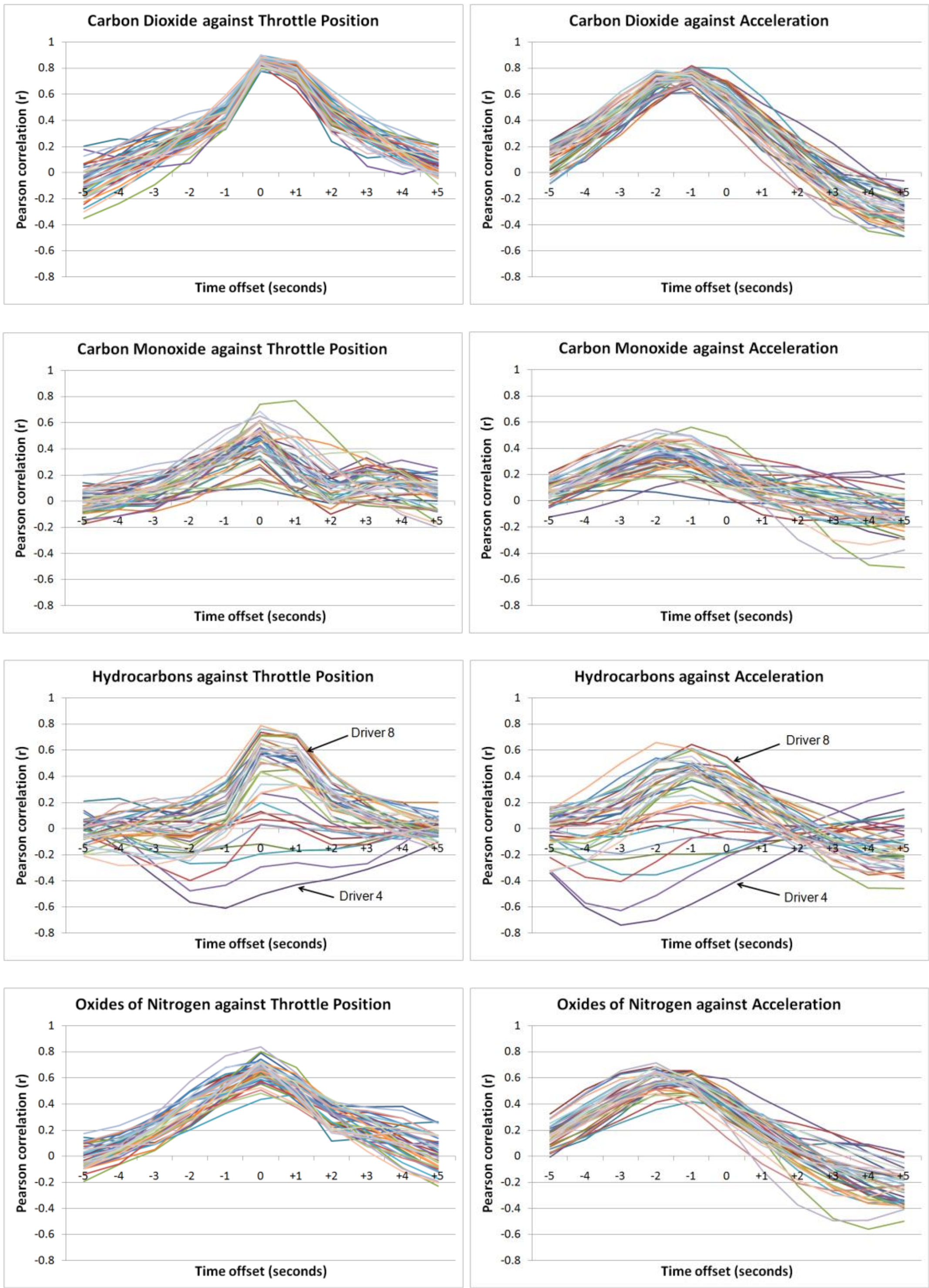

Figure 2: Investigation into variation in static time offsets using linear correlation

Oxides of nitrogen were observed to display the next strongest positive correlation with both throttle position and acceleration[1]. The majority of drivers (39) displayed the strongest positive correlation for NO_x with throttle position at Δt=0 time offset, although 1 driver was observed to have the strongest correlation at Δt=+1. The majority of drivers (24) displayed

[1] The NO_x emissions values should be interpreted with caution. It has been demonstrated that the sensor utilised in the experiment is cross-sensitive to ammonia (NH_3), rendering measurements sometimes unreliable, especially under rich engine operating conditions (Ropkins et al 2008).

the strongest correlation with acceleration at Δt=-2 time offset, although 16 drivers were observed to have the strongest correlation at Δt=-1.

Table 1: Summary of correlation coefficients derived from static time offset analysis

Throttle Position		CO_2	CO	HC	NO_x	Acceleration		CO_2	CO	HC	NO_x
Δt= -1	Min	-	0.34	-	-	Δt= -3	Min	-	0.08	-	-
	Max	-	0.42	-	-		Max	-	0.47	-	-
	Mean	-	0.38	-	-		Mean	-	0.32	-	-
	Median	-	0.38	-	-		Median	-	0.42	-	-
	n	-	2	-	-		n	-	3	-	-
Δt= 0	Min	0.78	0.09	0.12	0.48	Δt= -2	Min	0.66	0.18	0.12	0.47
	Max	0.90	0.69	0.79	0.84		Max	0.78	0.55	0.66	0.72
	Mean	0.86	0.44	0.54	0.66		Mean	0.74	0.37	0.44	0.60
	Median	0.85	0.46	0.59	0.67		Median	0.76	0.37	0.54	0.62
	n	39	36	26	39		n	8	25	3	25
Δt= +1	Min	0.79	0.50	0.33	0.48	Δt= -1	Min	0.62	0.16	0.17	0.42
	Max	0.79	0.77	0.72	0.48		Max	0.82	0.56	0.64	0.66
	Mean	0.79	0.63	0.51	0.48		Mean	0.74	0.37	0.43	0.57
	Median	0.79	0.63	0.51	0.48		Median	0.75	0.38	0.44	0.58
	n	1	2	7	1		n	32	12	28	15

N.B. Negative and indeterminate coefficients relating to hydrocarbons have been excluded, resulting in sum of n<40.

The correlation between carbon monoxide and throttle position/acceleration was less strong than for CO_2 or NO_x, but a positive correlation can be observed (throttle position Δt in the range -1 to +1, and acceleration Δt in the range -1 to -3 with the majority of drivers at Δt=-2).

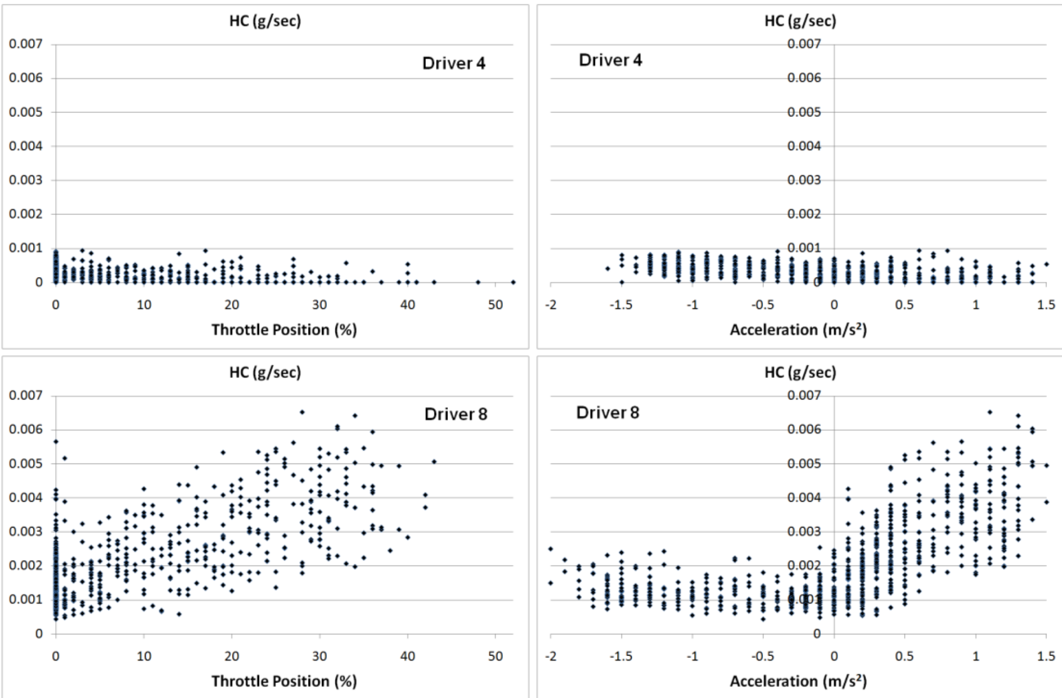

Figure 3: Scatter plots of Hydrocarbons (g/sec) against Throttle Position (%) and Acceleration (m/s^2) respectively for Drivers 4 and 8.

The results for hydrocarbons are interesting and warrant further investigation to better understand the relationships and formation processes. Whilst a positive correlation can be observed for many drivers (typically throttle position Δt=0, acceleration Δt=-1), the strength of the correlation is highly variable. In addition, a small number of drivers display a negative correlation together with a longer time lag. This can be partly explained by the fact that the nature of Pearson 'r' is such that it imparts no information about the magnitude of the variables being analysed, only the extent to which 'x' increases or decreases with changes in 'y' (and of course implies no causality). In the hydrocarbon plots in Figure 2, two drivers have been highlighted; Driver 8 displaying a more typical positive correlation, and Driver 4 displaying a less typical negative correlation. Figure 3 presents scatter plots of hydrocarbons against throttle position and acceleration respectively for these two drivers. It can be seen that the absolute levels of hydrocarbon emissions for Driver 4 are significantly lower than Driver 8, and that the higher levels generated by Driver 8 result in a more defined positive relationship with the independent variables, whereas the relationships for Driver 4 are far more ambiguous. Thus it can be seen that whilst correlation can be used successfully to derive static time alignment of dependent and independent variables, it should be used cautiously and with a knowledge of the distribution and magnitude of the underlying data.

4. Temporal sampling resolution

The data gathered in this case study was collected at a sampling frequency of 1Hz across all variables. However, there is no reason to assume that independent variables relevant to instantaneous modelling of emissions have a maximum frequency of the order of 1Hz, or indeed that they all have the same frequency of operation. Sensor response time (which may vary for different pollutants) is often a practical constraint, particularly for measurement of specific gases in the exhaust emissions.

The Shannon (or Nyquist) Sampling Theorem provides a theoretical basis for determining the appropriate sampling rate of a variable. Essentially, "the sampling frequency should be at least twice the highest frequency contained in the signal" (Olshausen 2000). Mathematically;

$$f_s \geq 2f_c \tag{1}$$

where f_s is the sampling frequency, and f_c is the highest frequency contained in the signal. If $f_s < 2f_c$, then 'aliasing' can occur where a misleading/false understanding of the form of the original signal can be arrived at.

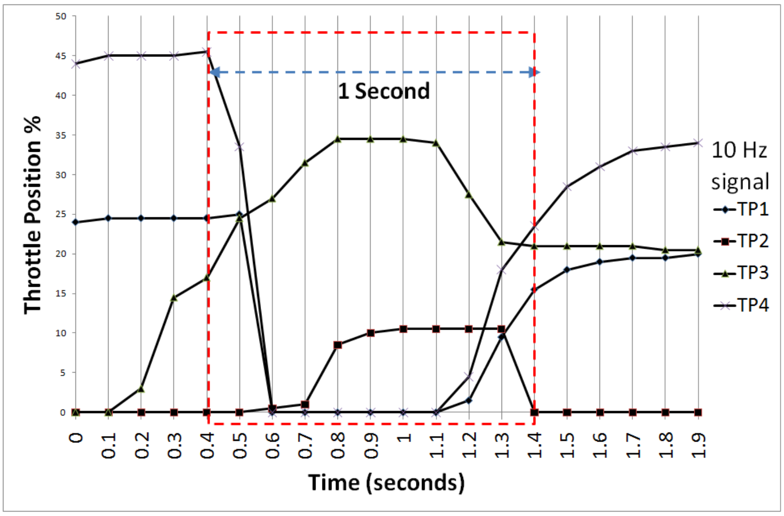

Figure 4: Changes in vehicle throttle position (%), measured at 10Hz.

Determining the highest frequency contained in a signal can be determined by measurement. A simple, but relevant, illustrative example is presented in Figure 4. Changes in throttle (accelerator) position were measured at a sampling frequency (f_s) of 10Hz on an instrumented vehicle during normal driving. Figure 4 presents four short (2 second duration) samples of 10Hz data (TP1 – TP4) from these measurements. By observation, it can be seen that significant transitions in throttle position (+/- 20 to 40%) occur within time periods of 0.3 to 0.1 seconds (3Hz - 10Hz). Based on this data set, it cannot be stated with certainty that the maximum signal frequencies (f_c) are not higher than 10Hz (this can only be determined by sample measurement at a higher frequency than 10Hz). From this simple example, it can be seen that if throttle position was a significant independent variable in our analysis, it should be sampled at a rate f_s = 20Hz to capture transitions taking place at f_c = 10Hz. However, this raises practical concerns for experimental design about achieving consistency of sampling frequency if, for example, sensor response times for some variables are constrained by limitations of technology.

5. Frequency distribution of data

When the frequency distribution of emissions data (CO_2, HC, NO_x, CO etc) is examined, it is often seen to exhibit more than one local maxima, in addition to significant skew and kurtosis. Multi-modality in the distribution (which can be a function of inherent system characteristics, and variability in driver behaviour), can inhibit the efficacy of 'standard' data transformations used traditionally to address issues of non-normality. De-convolution of the frequency distribution into vehicle operating modes has been observed to result in uni-modal distributions which are more amenable to statistical analysis and modelling. There is a practical rationale for such de-convolution in a modelling context since the independent variables/emissions precursors may be different in each operating mode. This issue is explored further.

Figures 5 and 6 respectively illustrate the multi-component nature of the frequency distributions for CO_2 and NO_x for four vehicle operating modes, acceleration, cruise, deceleration, and stop;

Acceleration	–	where acceleration (m/s^2) > 0
Cruise	–	where velocity (m/s) < 0, and acceleration (m/s^2) = 0
Deceleration	–	where acceleration (m/s^2) < 0
Stop	–	where velocity (m/s) = 0, and acceleration (m/s^2) = 0

It is of course possible to define vehicle operating modes in greater detail.

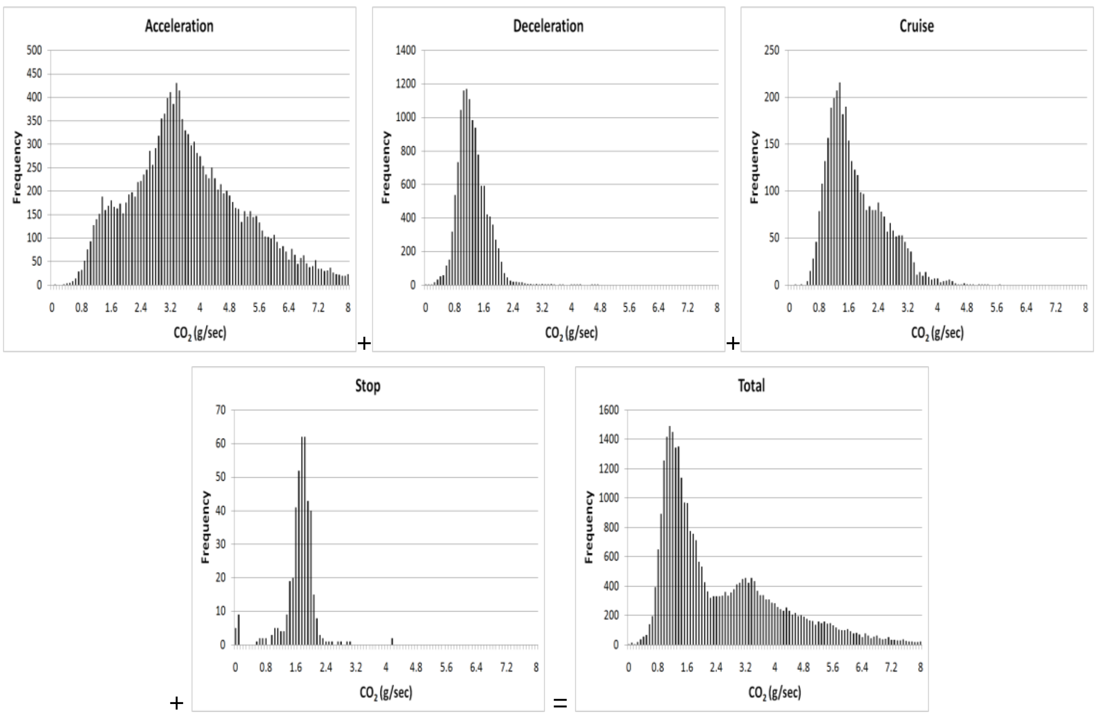

Figure 5: Modal components of CO_2 (g/sec) exhaust emissions frequency distribution

It can be seen clearly in Figure 5 that the frequency distribution of total CO_2 (g/sec) (sum of acceleration, deceleration, cruise, and stop) has two local maxima, one at approximately 1.12 g/sec, and a second at approximately 3.2 g/sec. However, it is clear that the peak at 3.2 g/sec is generated largely by the 'acceleration' component of vehicle operation, whereas the peak at 1.2 g/sec is dominated by the 'deceleration' component.

On the other hand, the frequency distribution of total NO_x g/sec emissions in Figure 6 is uni-modal and highly skewed. However, it can be seen that the individual operating mode components are quite distinct. As is perhaps to be expected, 'acceleration' dominates the right hand tail of the distribution, and whilst still skewed, the 'acceleration' distribution is more 'normal' (for want of a better descriptor) than the other components. 'Deceleration' and 'stop' modes have similar distributions, whereas the 'cruise' mode has a denser right hand tail.

This type of analysis is useful, not only because it provides guidance on the most appropriate statistical techniques to adopt (and possible data transformations), but also because it can provide indications regarding which operating modes are most dominant, and therefore where the analyst should be focusing attention when specifying predictive models of instantaneous tailpipe emissions.

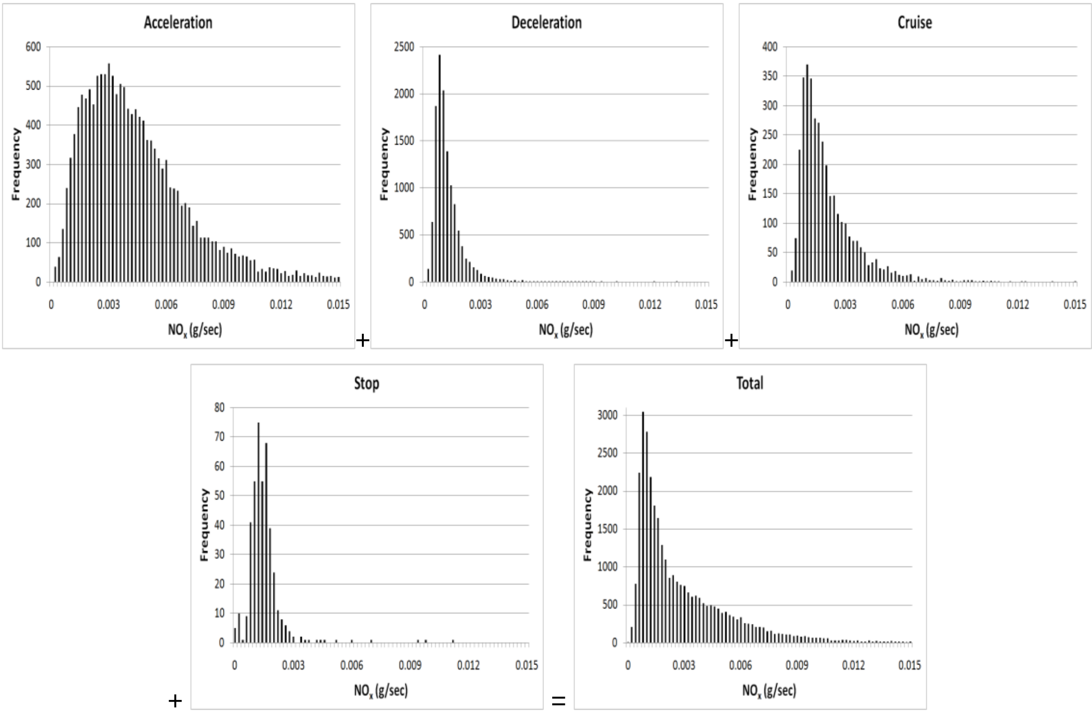

Figure 6: Modal components of NO_x (g/sec) exhaust emissions frequency distribution

6. Variability in driver behaviour

Whilst emissions models can be derived from data collected from instrumented vehicle surveys and laboratory campaigns, with current vehicle technology, variability in driver behaviour (i.e. how drivers operate the primary vehicle controls such as the accelerator, gears, brake, clutch etc.) will have a significant influence on exhaust emissions, both green house gases and local pollutants. A knowledge of such variability, its nature and extent, is necessary if emissions models are to be used as part of a wider modelling framework to, for example, predict local air quality. Relatively little research data is currently available regarding the nature and extent of such variability in the 'amateur' (non-professional) driver population.

Cluster analysis techniques can be used to analyse data collected as part of the case study, grouping driver cases with similar attributes or characteristics. In this illustrative example, hierarchical cluster analysis was applied to the driver behaviour data to investigate how engine speed (rpm), throttle position (%), and vehicle acceleration (+ve or –ve m/s^2) could be used to group or 'cluster' drivers. The environmental efficiency of these clusters of drivers could then be investigated to provide an insight into the relationship between driver behaviour and environmental performance.

The distributions of the behavioural data for each variable by driver were standardised by generating percentile values at 5 percentile intervals. These percentile values were then utilised in the cluster analysis. The average (between-group) linkage cluster method was adopted, with squared Euclidean distance being used as the measure of similarity. Separate cluster analyses were carried out for each variable, engine speed, throttle position, and acceleration respectively (it is of course also possible to cluster all three variables at the same time). Data from 37 of the 40 drivers were included in the analysis, 3 drivers being discarded due to instrumentation reliability issues. Hierarchical cluster analysis (unlike other methods such as k-means clustering) makes no prior assumptions about the number of clusters to be generated. The number of clusters is determined by the analyst using metrics from the analysis such as the measure of proximity between clusters. In this case, four clusters of drivers were identified for each variable respectively.

Table 2 – Clustering of drivers by variable

	Cluster (R1)	**Cluster (R2)**	**Cluster (R3)**	**Cluster (R4)**
(R) Engine speed	10, 15, 17, 36	7, 8, 9, 12, 13, 16, 18, 19, 20, 21, 22, 23, 26, 29, 30, 33, 34, 35, 37, 38, 40	1, 5, 6, 11, 24, 25, 31, 32	2, 14, 27, 28
	Cluster (T1)	**Cluster (T2)**	**Cluster (T3)**	**Cluster (T4)**
(T) Throttle Position	15*	20, 22, 24, 25, 27	2, 8, 10, 11, 14, 16, 17, 18, 19, 21, 23, 26, 28, 30, 33, 37	1, 5, 6, 7, 9, 12, 13, 29, 31, 32, 34, 35, 36, 38, 40
	Cluster (A1)	**Cluster (A2)**	**Cluster (A3)**	**Cluster (A4)**
(A) Vehicle acceleration	15, 20, 24	2, 10, 18, 21, 22, 23, 25, 27, 30, 33, 37	1, 7, 8, 11, 12, 14, 16, 17, 19, 26, 28, 29, 31, 32, 36, 40	5, 6, 9, 13, 34, 35, 38

**N.B. Driver 15 was an outlier for the Throttle Position variable, and was allocated to its own cluster.*

Investigation of the clusters generated by the analysis determined that they could be characterised by the attributes in Table 3.

Table 3 – Behavioural attributes of clustered drivers by variable

		Cluster (R1)	**Cluster (R2)**	**Cluster (R3)**	**Cluster (R4)**
(R) Engine speed (RPM)	Mean 25th %tile	1561 rpm	1343 rpm	1184 rpm	825 rpm
	Mean 50th %tile	2094 rpm	1754 rpm	1493 rpm	1311 rpm
	Mean 75th %tile	2500 rpm	1970 rpm	1706 rpm	1573 rpm
	Mean 95th %tile	2896 rpm	2359 rpm	2021 rpm	1927 rpm
		Cluster (T1)	**Cluster (T2)**	**Cluster (T3)**	**Cluster (T4)**
(T) Throttle Position (%)	Mean 65th %tile	21%	19%	12%	6%
	Mean 75th %tile	35%	28%	20%	11%
	Mean 85th %tile	73%	36%	28%	15%
	Mean 95th %tile	98%	47%	38%	22%
		Cluster (A1)	**Cluster (A2)**	**Cluster (A3)**	**Cluster (A4)**
(A) Vehicle acceleration (m/s^2)	Mean 5th %tile	-1.97 m/s^2	-1.82 m/s^2	-1.45 m/s^2	-1.10 m/s^2
	Mean 25th %tile	-0.86 m/s^2	-0.66 m/s^2	-0.40 m/s^2	-0.30 m/s^2
	Mean 75th %tile	0.85 m/s^2	0.72 m/s^2	0.52 m/s^2	0.36 m/s^2
	Mean 95th %tile	1.60 m/s^2	1.34 m/s^2	1.09 m/s^2	0.92 m/s^2

In principle, the production of 4 clusters of drivers for each of the 3 variables produces 4^3 (64) potential cluster combinations (R1 to 4 by T1 to 4 by A1 to 4). However, with a relatively small sample of 37 drivers, the three dimensional cluster matrix could not be fully populated. In addition, some combinations of driver behaviour clusters may have a greater probability than others, and others may not be feasible in practice. It transpired that when the clusters of drivers by variable are cross-tabulated, 18 of the possible 64 matrix cells are populated, 11 of these by individual drivers.

The environmental 'performance' of these 18 cluster cells is presented in Table 4 in terms of CO_2 emissions, fuel consumption, and pollutant emissions HC, CO, and NO_x.

Table 4 – Environmental performance of clustered drivers

	Carbon dioxide		Fuel Consumption		Hydrocarbons		Carbon monoxide		Oxides of nitrogen	
Cluster cell	Mean rate (g/sec)	Mean rate (g/km)	Mean rate (g/sec)	Mean rate (g/km)	Mean rate (g/sec)	Mean rate (g/km)	Mean rate (g/sec)	Mean rate (g/km)	Mean rate (g/sec)	Mean rate (g/km)
R1_T1_A1 *(n=1)*	3.67	385	1.10	115	0.00411	0.43	0.1219	12.77	0.1217	1.27
R1_T3_A2 *(n=1)*	2.72	344	0.83	106	0.00261	0.33	0.0277	3.51	0.0059	0.75
R1_T3_A3 *(n=1)*	3.02	365	0.87	105	0.00020	0.02	0.0056	0.67	0.0020	0.25
R1_T4_A3 *(n=1)*	2.41	340	0.72	102	0.00077	0.11	0.0033	0.46	0.0014	0.20
R2_T2_A1 *(n=1)*	3.11	346	0.91	101	0.00049	0.05	0.0126	1.40	0.0026	0.30
R2_T2_A2 *(n=1)*	2.87	346	0.84	101	0.00026	0.03	0.0083	1.01	0.0023	0.28
R2_T3_A2 *(n=6)*	2.79	348	0.83	104	0.00114	0.14	0.0104	1.29	0.0030	0.38
R2_T3_A3 *(n=4)*	2.60	349	0.76	102	0.00133	0.18	0.0114	1.54	0.0030	0.41
R2_T4_A3 *(n=4)*	2.22	325	0.66	97	0.00131	0.19	0.0049	0.72	0.0025	0.37
R2_T4_A4 *(n=5)*	2.14	330	0.63	98	0.00094	0.14	0.0024	0.38	0.0017	0.26
R3_T2_A1 *(n=1)*	3.21	369	0.99	113	0.00069	0.08	0.0233	2.68	0.0041	0.47
R3_T2_A2 *(n=1)*	3.22	375	0.98	115	0.00079	0.09	0.0131	1.53	0.0035	0.41
R3_T3_A3 *(n=1)*	2.51	328	0.75	98	0.00161	0.21	0.0107	1.40	0.0033	0.44
R3_T4_A3 *(n=3)*	2.11	312	0.63	93	0.00088	0.13	0.0047	0.70	0.0027	0.41
R3_T4_A4 *(n=2)*	1.93	307	0.58	93	0.00110	0.18	0.0032	0.51	0.0030	0.48
R4_T2_A2 *(n=1)*	2.89	343	0.93	111	0.00184	0.22	0.0219	2.61	0.0043	0.51
R4_T3_A2 *(n=1)*	2.77	350	0.84	106	0.00038	0.05	0.0112	1.41	0.0041	0.52
R4_T3_A3 *(n=2)*	2.34	358	0.72	110	0.00154	0.24	0.0133	2.02	0.0036	0.56

When interpreting the data, it should be remembered that the context of the measurements was a low speed suburban route with short links connected by left-hand turns at priority junctions. Drivers were generally either accelerating or decelerating between corners with little opportunity to ‘cruise’. Gear selection was dominated by 2nd and 3rd gears. Hence, measured CO_2 emissions and fuel consumption would be expected to be significantly higher than ‘typical’ rates for mixed driving conditions. The main objective of the analysis is to assess the degree of variability displayed by drivers when presented with these constrained driving conditions, the vehicle and the highway geometry being held constant (ambient traffic conditions were extremely light with very little interaction with other traffic).

It is clear that the throttle application behaviour of Driver 15 was extreme relative to the other drivers. This behaviour tended to result in very high levels of fuel consumption and emissions for all pollutants. Generally, lower rates of fuel consumption and CO_2 emissions are associated with lighter throttle applications and lower rates of acceleration (a degree of symmetry was observed between positive (+ve) and negative (–ve) acceleration; the drivers who accelerated heavily also tended to brake heavily). However, it was observed that very low engine speeds (for example, associated with the R4 cluster) are not always desirable, perhaps because they are associated with engine ‘labouring’. The engine appeared to operate more efficiently in the R2 and R3 clusters when combined with light throttle application and low levels of acceleration, although some of the lowest emissions results for HC, CO, and NO_x were associated with the R1 cluster when combined with light throttle application (T3, T4) and low levels of acceleration (A3). It should also be noted that there is sometimes a trade-off between the rate of emissions in g/sec and the rate of emissions in g/km, where average speed is a factor. Total emissions for a journey can be high, even with a low rate g/sec, if average speed is very low. This implies that over-cautious, hesitant driving can increase emissions in g/km for a total journey, relative to a more competent driver who maintains a reasonable g/sec emissions rate, and completes the journey expeditiously, resulting in a lower g/km.

Clearly, such a small sample of drivers is not necessarily representative of the whole driver population, but it is a subset of the UK driver population. The highway network used for the measurements is also only a subset of the total network, and only one vehicle was utilised in the measurements. Nevertheless, the research has provided an insight into the nature and potential scale of driver behaviour variability in the population, and can be used to inform future experimental design.

7. Predictive models

Taking into account issues discussed above such as static time synchronisation and modality of frequency distributions, predictive models of instantaneous exhaust emissions can be explored. As an illustrative example only, a simple linear regression model of CO_2 emissions during the positive acceleration mode only (see Figure 5) is presented below. Vehicle power (kW) and throttle position (%) are used as independent variables (Vehicle power is defined as the power required to propel the vehicle (encompassing rolling resistance, aerodynamic resistance, road gradient, and acceleration).

Model Summary [b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
	.843[a]	.710	.710	.87700682163	2.017

a Predictors: (Constant), Throttle Position (%), Vehicle Power (kW)
b Dependent Variable: CO_2 (g/sec)

Coefficients [a]

Model		Unstandardised Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	1.842	.012	155.392	.000
	Vehicle Power (kW)	.088	.001	80.142	.000
	Throttle Position (%)	.052	.001	83.307	.000

a Dependent Variable: CO_2 (g/sec)

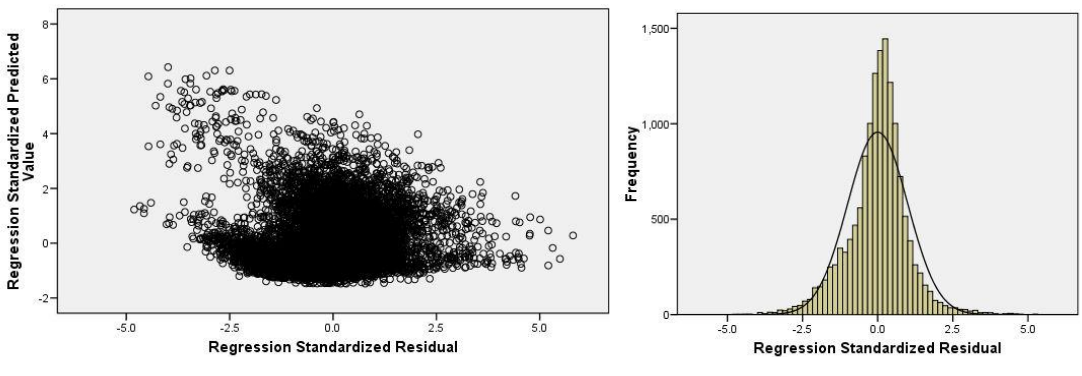

Figure 7: Residual plots for prototype CO_2 emissions (positive acceleration) model

Whilst the adjusted R^2 value of 0.71 is reasonable (for an initial prototype model), the distribution of residuals are not normally distributed. In addition, the constant is relatively dominant, suggesting that further investigation, perhaps of model form or independent variables, is required. Clearly, such an investigation of the relationships between dependent and independent variables relating to instantaneous vehicle exhaust emissions will be wide ranging, and the subject of further research and publication.

8. Conclusions

This paper has explored some of the analytical and statistical challenges encountered when seeking to derive predictive models of instantaneous vehicle exhaust emissions from instrumented vehicle measurements. Time synchronisation of dependent and independent variables using static alignment procedures may prove to be increasingly limiting for model development, especially as sampling rates increase in future experimental design. Model development by mode of vehicle operation appears to be a promising avenue of future research, but explicit recognition of the significance of variability in driver behaviour will be required in future models if instantaneous vehicle emissions are to be more robustly represented within the wider integrated modelling framework for local air quality assessment.

9. Acknowledgements

The authors wish to acknowledge EPSRC and Newcastle University for funding the PhD studentship to enable this research to be carried out, the instrumented City Facility and the RETEMM project (GR/S31136/01) which provided the data for the research presented here. The authors are indebted to colleagues at the Institute for Transport Studies at the University of Leeds for their valuable assistance.

10. References

Ajtay D., Weilenmann M., and Soltic P. (2004). Towards accurate instantaneous emission models. Atmospheric Environment, Volume 39, Issue 13, Pages 2443-2449

Bell M.C., Tate J.E., Li H., Ropkins K. and Chen H. (2008). RETEMM: Real-world Traffic Emissions Monitoring and Modelling. Final Report to the EPSRC, University of Leeds.

Bell M.C. and Rhys-Tyler G.A (2008). Understanding the effects on tailpipe emissions of integrated vehicle and systems technologies. Proceedings of the 15th ITS World Congress, New York 2008.

Daham B (2006). Measuring real-world emissions from spark ignition engines. PhD thesis, University of Leeds.

Felstead T., McDonald M. & Fowkes M. Extremes in driving style and their potential effect on vehicle emissions. Proceedings of the 40th UTSG Conference, Southampton 2008.

INRETS (2006). Accuracy of exhaust emissions measurements on vehicle bench (Artemis deliverable 2), Report No. LTE 0522, December 2006

Le Anh T., Hausberger S., and Zallinger M. (2006). Correction for accurate instantaneous emission measurements of passenger cars, in WIT Transactions on Ecology and the Environment 2006, Vol. 86, pp 703-712. ISBN: 1-84564-165-5.

North R.J., Noland R.B., Ochieng W.Y., and Polak J.W. (2006). Modelling of particulate matter mass emissions from a light-duty diesel vehicle. Transportation Research Part D 11 (2006), pp344-357.

Olshausen B.A. (2000). Aliasing. PSC 129 – Sensory Processes. https://redwood.berkeley.edu/bruno/npb261/aliasing.pdf, accessed on 1/12/2009.

Rhys-Tyler G. A. and Bell M.C. (2009a). Understanding the significance of variability in driver behaviour on tailpipe emissions of a Euro 4 spark ignition passenger car. Proceedings of the 41st UTSG Conference, London 2009.

Rhys-Tyler G.A. and Bell M.C. (2009b). Influencing driver behaviour for environmental benefit: The role of ITS technologies. Proceedings of the 16th ITS World Congress, Stockholm 2009.

Ropkins K., Quinn R., Beebe J., Li H., Daham B., Tate J., Bell M., and Andrews G. (2007). Real-world comparison of probe vehicle emissions and fuel consumption using diesel and 5% biodiesel (B5) blend. Science of the Total Environment 376 (2007) pp267-284.

Ropkins, K., Tate, J., Li, H., Andrews, G., Hawley, J. G., Bell, M. (2008). Chassis Dynamometer Evaluation of On-Board Exhaust Emission Measurement System Performance in SI Car under Transient Operating Conditions. In: SAE International Powertrains, Fuels and Lubricants Congress, 23-25 June 2008, Shanghai, China. SAE Paper No. 2008-01-1826.

Ropkins K., Carslaw D.C., Goodman P.S., Tate J.E. (2009). Application of non-linear time-alignment and integration methods to environmental time series. Trends in Analytical Chemistry, Vol. 28, No. 3, 2009