Dynamic Modelling of Diesel Engine Emissions Using Parametric Volterra Series

Burke R.D.[1]*, Baumann, W.[2], Akehurst, S.[1] and Brace C.J.[1]

1. Department of Mechanical Engineering, University of Bath, Bath, UK
2. IAV GmbH, Berlin, Germany

* Contact Author: Dept. of Mechanical Engineering, University of Bath, Bath, B2 7AY, UK; email: R.D.Burke@bath.ac.uk

ABSTRACT

Design of powertrain controllers relies on the availability of data driven models of the emissions formation from internal combustion engines. Typically these are in the form of tables or statistical regression models based on data obtained from stabilised experiments. However, as the complexity of engine systems increase, the number of experiments required to capture the effects of each actuator becomes large. In addition, the models are only valid under stable operating condition, and do not give any information as to dynamic behaviour. In this paper, the authors present the use of Volterra Series (dynamic polynomial models) calculated from dynamic measurements as an alternative to the steady state models. Dynamic measurements of gaseous exhaust emissions were taken for a 2.0L automotive Diesel engine installed on a transient engine dynamometer. Sinusoidal based excitations were used to vary

engine speed, load, main injection timing, Exhaust gas Recirculation (EGR) valve position and fuel injection pressure. Volterra models calculated for NOx and CO_2 emissions presented high levels of fit with R^2 and normalised RMSE values of 0.85, 0.91 and 6.8% and 6.6% respectively for cold start NEDC. Models for CO and THC emissions presented poorer levels of fit (normalised RMSE 26% and 17% respectively), with difficulties in capturing the high nonlinearities of the measured data, notably for very high emissions levels.

Keywords: diesel engine emissions, engine control systems, engine dynamics, engine testing, diesel engine design/ development

1 INTRODUCTION

Engine manufacturers in the automotive sector are coming under increasing pressure to deliver lower harmful emissions and fuel consumption. In Europe, Oxides of Nitrogen (NOx) emissions requirements for Diesel passenger cars are currently 180g/km with a planned reduction to 80g/km in 2014. CO_2 emissions will be regulated on all vehicles form 2015 in the form of a fleet average, initially set at 130g/km, but reducing to 95g/km by 2020. In areas similar limitations are appearing on fuel consumption. In addition to this, the drive cycles used for the certification of vehicles are becoming much more dynamic in order to better represent in-service behaviour. Fairly steady cycles like the New European Drive Cycle (NEDC) or Japan Mode 10-15 are being replaced by more dynamic cycles such as the World harmonised cycle [1].

To meet these stringent targets, engine manufacturers are using an increasing number of sub-systems such as low and high pressure Exhaust gas Recirculation (EGR), multi ignition combustion systems, advanced thermal management systems, multistage turbocharging and variable valve timing [2, 3]. Each of these gives the engine further flexibility but makes the system controller more complex. The current industry standard for controller optimisation is to run experiments under stationary engine operating conditions, using design of experiments (DoE). Statistical response models are then used to represent the measured engine behaviour in optimisation routines [4]. However, as engine systems increase in complexity through more control actuators, the experimental effort also increases and more operating points are required to explore the system interactions and calculate the response surface.

A shift to dynamic engine characterisation offers the opportunity to significantly reduce experimental work during model identification for optimisation. This avoids the long thermal settling times between each operating point which can require as much as 5 minutes [5]. The dynamic operating points are much closer to real world use in automotive applications where steady state conditions rarely occur. With a move to hybrid powertrains, engine use will become much more intermittent meaning fully warm operation will be less frequent, replaced by a series of thermal transients [6]. A move to dynamic experiments and models to capture cold start, dynamic behaviour will allow both reduced experimental effort and focus on conditions closer to real world applications.

This paper describes the application of Volterra Series Models, an extended polynomial approach to the modelling of Diesel engine exhaust gaseous emissions. Dynamic experiments have been conducted on an automotive Diesel engine installed on a transient dynamometer. The measured data was used to calculate models for various gaseous emissions to assess the suitability of this modelling approach.

2 BACKGROUND

2.1 Dynamic modelling of engine emissions

Recent interest in the dynamic modelling of engine emissions using data driven models has been driven by the need to improve the engine development process. Research in this field can be roughly split into the following two areas:

1. Design of dynamic experiments;
2. Dynamic model type and training

Different types of dynamic experiments have been considered for use with internal combustion engines. Amplitude modulated Pseudo Random Binary Signals (APRBS) [7, 8] and varying frequency sinusoidal signals (Chirps) [9] have received the most interest because of their ability to cover broad frequency ranges. Baumann et al [10] presented a comparison of Triangular, sinusoidal, APRBS and Chirp signals and showed that sinusoidal based signals presented significant advantages for engine development. These signals are less problematic with regards to safe engine operation because of their continuous nature rather than step disturbances. Although APRBS signals are, from a theoretical perspective, superior for

identification purposes because they cover a broader frequency range, the harsh nature of the step changes are not suited to all engine systems. In contrast, Chirp signals are a continuous signal however, the frequency range is limited.

A range of common mathematical models are suitable dynamic applications if augmented to capture the measured dynamics. Generally this is achieved by including additional model inputs relating to the previous states of both the input (independent) variables and the output (dependent) variable. Guehmann and Riedel [11] compared ten different dynamic modelling approaches for NOx and CO emissions. Volterra series performed best for NOx emissions whilst a neural network approach was recommended for CO. The modelling work was conducted on measured data from an engine subjected to chirp excitations for engine speed, load, main injection timing, variable geometry turbocharger (VGT) position, EGR valve opening and fuel injection pressure. The predictive power of the models was assessed based on a portion of the chirp experiments not used during the training. Consequently, these findings do not give an insight into the prediction of drive cycle behaviour. A further study by the present authors [12] compared the performance of recurrent neural networks and Volterra series for the modelling of NOx emissions. Again the models were calculated based on chirp experiments, but validated over an NEDC. This resulted in similar performance for both model types, but highlighted the added complication of neural network model training.

Baumann et al. [9] used sinusoidal chirp input signals to develop a parametric Volterra model for NOx emissions based on engine speed, engine load, injection timing, fuel injection

pressure, EGR rate and boost pressure. The model fitted measured data with a normalised RMSE of 6% over the NEDC. However the model did not include any way of accounting for the temperature variation during warm-up; consequently cold start NOx emissions prediction was overestimated. This paper aims to build on the previous works to assess the performance of the Volterra series for other gaseous emissions types.

2.2 Emissions formation

This paper will assess the performance of Volterra model to capture NOx, CO_2, carbon monoxide (CO) and unburned hydrocarbons (THC). The proportions of each species in the exhaust gas results from a number of factors controlling the combustion process in the cylinder.

NOx emissions comprise NO and NO_2, but are most commonly grouped together as NOx. They are mostly formed through the reaction of atmospheric nitrogen and oxygen under conditions of high temperature (above 1900-2000K) [13, 14]. There is therefore a link between NOx emissions levels, combustion temperatures and availability of oxygen [15-17]. Hence systems that reduce combustion temperature (such as intake air cooling or main injection retard) or reduce the availability of oxygen (such as EGR) allow reductions in NOx emissions. Under cold-start conditions, the engine and combustion chamber walls are significantly colder and the combustion air enters the cylinder at a colder temperature. Subsequently during combustion, more heat is lost to the walls meaning combustion occurs at colder temperatures making conditions less favourable for NOx formation. [18, 19].

Carbon dioxide (CO_2) results from the complete combustion of fuel with atmospheric oxygen; reductions in fuel consumption will lead to reductions in CO_2 emissions. Consequently CO_2 emissions are linked to the efficiency of torque production and therefore to the phasing of combustion with cylinder volume changes as well as parasitic losses (friction losses, accessory systems). With respect to the combustion process, measures used to reduce NOx emissions are typically detrimental to CO_2 as the thermal efficiency is reduced. Under cold start conditions, engine friction is higher due to the higher viscosity of the lubricant; this results in higher fuel consumption to deliver similar brake torque output as under hot conditions. Consequently CO_2 emissions are higher by 3-10% during warm up [18].

Carbon monoxide (CO) and total un-burnt hydrocarbons (THC) are formed throughout the combustion phase within the Diesel jet [20]. They are the products of the rich premixed flame where initial combustion occurs with insufficient oxygen and along with soot precursors serve as the reactants for the diffusion flame. The majority of products are transformed into water and carbon dioxide, however as the piston moves down expanding the gases in the cylinder, the temperature drops and the chemical reactions freeze. The remaining CO and HC are then included in the exhaust gases.

This paper aims to use the dynamic modelling approaches to capture the formation of Diesel engine gaseous emissions. This approach offers an alternative to conventional engine controller design that relies heavily on measurements taken under stable engine operating conditions. The application of these models to engine development process could result in

shorter development times and allow the controller design to be more suited to the dynamic duty cycles encountered in real driving situations.

3 THEORY

3.1 Dynamic Modelling

Figure 1 illustrates the need for dynamic modelling and shows the effect of applying a static model to a dynamic training data-set for a single-input-single-output system. The measured system response is shown following a step change in input and, before the system is allowed to settle, the input is returned to its original value. If a static model is used (a), then only the current settings of the actuators can be used in the mathematical representation of the data. Consequently the model estimates the average response over the measurement period and underestimates the actual settling value. If a dynamic model is used (b), the additional parameters from previous input settings and output response capture the dynamics whilst the static model will represent the system behaviour if it were allowed to settle.

3.2 Volterra Series

Polynomial models are widely used in engine applications because of their simplicity, ease of training using least squares regression and explicit formula. A practical extension of polynomial models to the dynamic range is the parametric Volterra series [21]. The static model is augmented using previous states of model inputs and feedback of model output as described by equation 1; this is the general form of the Volterra series. Because of the increased number

of parameters and the presence of output feedback, the regression process also becomes more complex, but crucially still relies on least squares.

$$\hat{y}(t) = C + \sum X_{static}^{n} + \sum X_{delays} + \sum X_{int} + Y_{fback} \quad \textbf{1}$$

Where

$$X_{static}^{n} = \beta_{n,1} x_1^n(t) + \beta_{n,2} x_2^n(t) + \cdots + \beta_{n,k} x_k^n(t)$$

$$X_{delays,k} = \gamma_{k,1} x_k(t-1) + \gamma_{k,2} x_k(t-2) + \cdots + \gamma_{k,j} x_k(t-j)$$

$$X_{int} = \varepsilon_1 x_1(t) x_2(t) + \varepsilon_2 x_1(t) x_3(t) + \cdots + \varepsilon_k x_1(t) x_k(t)$$

$$Y_{fback} = \delta_1 \hat{y}(t-1) + \delta_2 \hat{y}(t-2) + \cdots + \delta_l \hat{y}(t-l)$$

The different terms of the Volterra series in equation 1 are:

- **Model order**: this defines the highest exponent order for the static model (value of n in X_{static}).
- **Delay order**: This defines the number of previous input events that are used in the model (largest value of j in X_{delays}). The delay terms can also have higher order exponents and typically this was allowed to the same order as the model order.
- **Interaction order**: this defines the number of inputs that are grouped together for interaction terms. Interactions can also occur between delay terms.
- **Feedback order**: this defines the number of previous output terms included in the model (value of l in Y_{fback}). This was maintained as 1 for all modelling work presented.

3.3 Temperature Dependant Model

For the modelling of engine behaviour during warm-up, the operating temperature is the main descriptor. In this paper it will be assumed that whilst the effect caused by temperature may be non-linear, it acts globally on the effects from the other inputs. This is illustrated in equation 2 and means the model could not capture individual interactions between temperature and other inputs.

$$\hat{y}_{cold}(t) = f(T) \times \left(\hat{y}_{hot}(t)\right) \quad \textbf{2}$$

Where $f(T) = (\alpha_0 + \alpha_1 T + \alpha_2 T^2 + \cdots \alpha_n T^n)$

3.4 Evaluation of Model quality

The quality of each model was assessed using the fit statistics detailed in equations 3 to 5. The statistics indicate the level of fit if applied to the prediction of training data whereas if applied to validation data (not used in the training) they give a measure of predictive performance. Ideally the fit should be similar for both training and validation data; if a model tends to predict training data better than validation data, this is a sign the model is over-fitted.

$$R^2 = 1 - \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2} \quad \textbf{3}$$

$$RMSE = \sqrt{\frac{\sum(\hat{y}-y)^2}{n}} \quad \textbf{4}$$

$$nRMSE = \frac{RMSE}{\bar{y}} \quad (5)$$

4 Experimental Equipment and Data acquisition

4.1 Experimental Apparatus and Input factors

A 2.0L Diesel engine meeting EURO IV emissions specifications was used in this work. The engine included a variable geometry turbocharger, cooled high pressure EGR and common rail fuel injection. The typical application of the engine was a light commercial vehicle. The engine was installed on a 215kW transient AC dynamometer with all but vital auxiliary systems removed (cooling and lubrication). Three key combustion control parameters were chosen as the basis of this study; these are summarised in table 1. Also given in the table are the ranges of variation for each parameter and the upper and lower frequencies used in constructing the dynamic experiments. The upper frequencies were determined from frequency analysis of the NEDC whilst the lower frequencies were defined to give least correlation between input variables. In principle, the lower frequencies should be zero to cover steady state operation, but to remain with dynamic experiments these were defined at least an order of magnitude lower than the upper frequency.

Variable	Excitation method	Excitation range	Frequency Range
Engine speed	Direct control of set-point from host system	1000-2500rpm	0.003-0.1Hz
Engine load	PID control of electronic pedal position based on torque set-point within host system	20-250Nm	0.01-0.1Hz
Main Injection timing	'adder' function resident in ECU; offset to production calibration was set directly in ^{o}CA	+/-2^{o}CA (Absolute: -6 to 8^{o}BTDC)	0.006-0.06Hz
Common Rail fuel pressure	'adder' function resident in ECU; offset to production calibration was set directly in bar	+/-100bar (Absolute: 200 to 1400bar)	0.005-0.15Hz
EGR valve position	Indirect control through Mass air flow set point using 'multiplier' function resident in ECU. This was necessary as the EGR valve position itself is controlled closed loop according to a MAF demand.	+/-10% Absolute (6 to 60% EGR by CO2 ratio)	0.001-0.06Hz

Table 1: Variable descriptions, excitation methods and excitation ranges

The engine facility was controlled by the host system and the communications layout is summarised in figure 2. Communication with the engine control unit (ECU) was ensured using a calibration tool that was linked to the host system via ASAP3 link. Horiba MEXA 7000 emissions analysers were used to measure pre-catalyst emissions.

The control of engine behaviour according to the dynamic test designs was conducted by the host system (CP Engineering Cadet). Engine speed was controlled by the AC dynamometer and engine torque was controlled using a PID controller acting on the pedal position. Other actuators were controlled via the calibration tool and ECU (Accurate Technologies ATi Vision). The implementation of this control was through embedded "adder" and "multiplier" functions resident within the engine strategy. As the names suggest, an adder function allows an offset to be added manually to the ECU generated demand for a specific actuator and the 'multiplier' scales that signal. These capabilities have different applications during the calibration process; however it is fortuitous that these features can be addressed rapidly by an appropriately

configured host system to assume transient control over engine actuators without needing to resort to modified code or prototype hardware. Full details of the control methods for each variable are summarised in table 1.

4.2 Model Training Data

To capture the dynamic engine behaviour of the various control variables, chirp signals were used. These are sine wave functions with a frequency that varies as a function of time. The basic equation for a normalised chirp signal is given in equation 6. This is defined to vary between an upper and lower frequencies f_1 and f_0 occurring at times t_{tot} and t_0 respectively. Other parameters in equation 6 relate to the total length of the chirp between the upper and lower frequencies and a phase shift. These parameters can be used as variables to optimise the phasing of multiple input excitations. To ensure that the total length of each excitation signal is the same, these can be assembled back to back as shown in figure 3.

$$x_n = \cos\left((f_0 + K^t)t + \varphi_n\right) \quad \textbf{6}$$

Where $K = (f_1 - f_0)/t_1^2$, and $t_1 = \frac{t_{tot}}{k}$

An individual chirp signal was calculated for each of the input signals and the phasing of each signal calculated to optimise the design space coverage for all variables. However, this will result in operating points that are not achievable in practice because of limitations either in terms of mechanical integrity of the engine or because of operation in unstable conditions. For example, the injection timing window for sensible engine operating will be a function of engine

speed. To account for this, the static operating limits were determined and used to scale the affected input variables. Figure 4 shows the example for torque as a function of engine speed; this was achieved through scaling rather than a saturation limit.

The experiments were conducted in two phases: the first to train the Volterra series representing the engine under fully warm conditions and the second to capture the warm-up behaviour. The excitation signals for the fully warm conditions for each of the five input variables are shown in in figure 5. The test is split into two phases: a loaded condition covering 60 minutes and an idle phase lasting 15 minutes.

The engine speed varies consistently between the lower and upper limits (1000-2500rpm). For engine torque, the signal has been scaled according to the region of interest which has been defined as a function of speed. This is due to the higher torque the engine can achieve at higher speeds. The three other signals do not require scaling in this case because they have been implemented through “adder” and “multiplier” functions and consequently the scaling is already an inherent part of the engine strategy. In other applications where direct set-points are defined, the scaling of these functions would also be required.

The resulting coverage of the design space is shown in figure 6 as pairwise projections of the multidimensional design space. In each of the ten plots the chirp loaded phase, chirp idle phase and NEDC are distinguished. This representation gives a view of the overall operating points that are covered by the experimental data. It is important to bear in mind though that although the experiments are represented as a cluster of points, in fact they are a continuous

signal. This is best shown by considering figure 7 which shows the plot of engine speed against engine torque for the first 25 seconds of the experiment. There is a marked start point and the design space is swept according to the test plan.

To capture the temperature dependent behaviour, a separate experiment was conducted. The aim was to record engine emissions showing the difference between cold and hot conditions. This was based on a chirp signal for engine speed and load as shown in figure 8; this was constructed using the same principles as previously described. The experiment was performed twice: once from ambient start (overnight soak at 20^{o}C) and once from hot start (40min thermal soak at 1500rpm/100Nm). In each case, care was taken to ensure that other engine control variables were set to the same levels throughout both experiments.

4.3 Validation Data

To provide an independent data set for the validation, the NEDC was used. The speed and torque traces for the engine used in this study are shown in figure 9; these have been defined based on the application in light commercial vehicle Validation data was recorded both for cold and hot starting engine.

5 Volterra Model Identification

5.1 Overview

The fitting algorithm of the model needs to be defined to account for the large number of factors of the Volterra model listed in equation 1. This is split into six phases as summarised in

figure 10. The many possibilities in terms of model orders and delay terms very quickly increases the number of possible model terms. This can cause problems for the parameter selection algorithms that fail to identify key effects. To avoid this, a pre-selection of delay terms and model order was performed (phase 1) before the automated fitting routine could be implemented (phase 2).

The fitting of the input model corresponds to the initial identification of all terms except Y_{fback} (see equation 1) and is achieved through least squares regression using orthogonal least squares parameter selection [21]. The calculation of the Y_{fback} terms must be performed separately from the other coefficients as the output delay terms are highly correlated with the output itself. This is achieved by allowing the parameters identified from regression to float with the output feedback coefficients whilst the fit RMSE is minimised. This is performed using an unconstrained optimisation routine (phase 3). The best models are then chosen based on predictive performance (phase 4) Calculation of the temperature dependent function is then performed using the dedicated experimental data using conventional least squares regression (phase 5). Finally the results of phases 4 and 5 are combined to produce the complete model.

5.2 Illustration of fitting process

5.2.1 Phase 1: Pre-selection of Model terms

The pre-selection of modelling terms define the overall Volterra series structure. In this work these were defined according to four parameters listed in table 2. As the initial phase of the model training relies on least squares, the calculation of the model is fast and therefore many

variations in model structure can be evaluated. This was performed for around 100 different cases.

Value	Description	Range
n_{poly}	Maximum order of single input terms	2-4
$n_{interactions}$	Maximum number of cross terms or interaction order	1-2
n_{delay}	Maximum number of delay terms per input	0-2
Δt_{delay}	Time interval between delay terms	1 / 4 / 7

Table 2: Pre-selection of model terms defining structure of Volterra series

5.2.2 Phase 2: Least squares fit of input terms

For each emission species, all possible combinations of pre-selection terms were calculated based on the training data. This resulted in fitted models that included all terms except Y_{fback} (see equation 1). The quality of the fit of these models was assessed using nRMSE values for both training and validation datasets (equation 3). The relative quality of the various models can be assessed by comparing these two values as shown for each species in figure 11. For all emission species, a range of model qualities appear. With the exception of CO_2 emissions, a trade-off appears between the training and validation fits. Along the pareto front, models with lowest training nRMSE tend to have high validation nRMSE and vice-versa. This is due to the over fitting of models to the training data. The training data fit can always be improved by increasing the number of terms in the model as this allows more mathematical flexibility, however at the same time, this increases the tendency of the model to capture noise in the data set.

An ideal model would present low but similar values of RMSE for the training and validation

data, meaning the model fits equally well both data sets. For the models calculated here, this is only the case for NOx emissions with all other model having poorer fit with the validation data. This is evidence of over-fitting of the models. The selection of best models was based on achieving a suitable balance between training and fit RMSE. For each emission species, the best models according to these rules are circled. In the case of NOx emissions, these are the models with similar levels of RMSE. For other species, a judgement has been made with an emphasis on minimum validation RMSE.

5.2.3 Phase 3: Output feedback fitting

Following the selection from phase 2, the output feedback term was added to the model. This was determined by fixing the model equation according to that defined in phase 2, and allowing the coefficients to float in an optimisation routine, minimising the RMSE value for the training data. Only the training data was used for the fitting of the model at this phase. Figure 12 shows an example of the fitting results before and after the inclusion and identification of the feedback term.

This training process was conducted for a small number of models for each emission species as identified in figure 11. The improvements in fit quality through the inclusion and training of output feedback are given in table 3. This shows a typical improvement of R^2 of around 0.01 to 0.05 and in nRMSE of around 0.5%. From these results it is clear the models produced for CO and THC have a poor level of fit with R^2 values of 0.35 and 0.56 respectively.

Emission spices	Without Output Feedback (Phase 2 Models)			With Output Feedback (Phase 3 Models)		
	R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
NOx	0.88	124	8.9	0.91	107	7.7
CO_2	0.92	0.52	5.1	0.93	0.47	4.7
CO	0.29	543	5.8	0.35	519	5.6
THC	0.52	104	3.5	0.56	100	3.4

Table 3: Average of fit statistics with and without input feedback for various emissions species

5.2.4 Phase 4: Model Selection

As for phase 2, the best models were selected based on the fit statistics for the validation NEDC. A summary of the best model structures and fit statistics is given in table 4.For each of the models considered in this phase of the work, the evolution of nRMSE has been plotted on the trade-off plots in figure 13. For the majority of models there is an improvement both in training and validation fit, however even in the small number of models tested here, in some cases there are large improvements in training fit accompanied by deterioration in validation fit. This is a sign of over fitting and highlights the need to consider multiple models at this stage.

Emission	n_{order}	$n_{Interaction}$	n_{delays}	Δt_{delay}	N_{terms}	R^2	RMSE	nRMSE
NOx	2	2	2	4	20	0.84	67	6.3
CO_2	3	2	2	7	20	0.90	0.88	8.2
CO	3	1	1	1	13	0.50	315	8.1
THC	3	1	2	4	13	0.05	71	7.6

Table 4: Model structures and validation fit statistics for best models for each emission species

5.2.5 Phase 5: Temperature function identification

The ratio of cold to hot emissions is plotted for each spiecies against oil tempeature in figure 14. In each case the regression fit is plotted through the data. This captures the simple temperature dependent trend in the data but clearly for each emission spiecies there are more complex phenomenon.

NOx emissions at 20^{o}C oil temperature were 50% of those at 105^{o}C (fully warm) with an approximatly linear trend. CO_2 emissions are slightly higher at lower oil temperatures, around 4% at 20^{o}C; this is consistant with studies of engine warm-up behaviour. For CO and THC, the effect of engine temperature is much larger with emissions 3 and 10 times larger at 20^{o}C respectivly. Model fit statistics and orders are given in table 5.

Emission	α_2 (**T^2**)	α_1 (**T**)	α_0 (Cnst)	**R^2**
NOx	N/A	4.96e-3	0.42	0.38
CO_2	N/A	4.04e-4	1.05	0.25
CO	1.39e-4	-3.80e-2	3.49	0.32
THC	2.02e-3	-3.24e-1	14.7	0.49

Table 5: Fit statistics for temperature dependent scaling factors

6 MODELLING RESULTS

The final model predictions for the cold start NEDC drive cycle are illustrated in figure 15 and the fit is quantified by the statistics summarised in table 6. The fit for NOx emissions is reasonable throughout the drive cycle. The fit for CO_2 emissions is excellent, notably towards the end of the cycle. For CO and THC emissions the predictions are poor, notably the very large values (above 2000ppm) are not captured by the model. During the early part of the drive cycle when the engine is cold, these could be attributed to the cold starting behaviour however these continue to occur even in the later phase of the drive cycle as shown by the plot for CO emissions.

Emission	Hot start NEDC			Cold Start NEDC		
	R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
NOx	0.84	67	6.3	0.85	59	6.8
CO_2	0.90	0.88	8.2	0.91	0.80	6.6
CO	0.50	315	8.1	0.33	1225	17
THC	0.05	71	7.6	0.13	2114	26

Table 6: Summary fit statistics for NOx, CO_2, CO and THC over hot and cold start NEDC

Equations 7 to 8 summarise the dynamic models of NOx and CO_2 emissions. Due to the poor level of fit, the models for CO and THC are not given in this publication. For the NOx model, the application of the Volterra series is presented graphically in figure 16. The time variation of each input is shown over a certain time period. Equally, the predicted emissions are shown until point *t-1*. The current point to be predicted, *y(t)*, is shown as a star point whilst each of

the current and previous input and output variables that are used in the calculation of that point are emphasised in the respective time series.

$$\begin{aligned} NOx(t) = \{&\alpha_1 N_t^2 + \alpha_2 N_t + \alpha_3 N_{t-8}^2 + \alpha_4 N_{t-8} + \alpha_5 \tau_{t-4} + \alpha_6 \tau_{t-8}^2 + \alpha_7 \tau_{t-8} + \alpha_8 \theta_t \\ &+ \alpha_9 \dot{m}_t + \alpha_{10} \dot{m}_{t-8}^2 + \alpha_{11} P_t + \alpha_{12} P_{t-8} + \alpha_{13} (N_t\, \tau_t) + \alpha_{14} (N_{t-8}\, \tau_{t-8}) \\ &+ \alpha_{15} (N_t\, P_t) + \alpha_{16} (\tau_t\, \theta_t) + \alpha_{17} (\tau_t\, \dot{m}_t) + \alpha_{18} (\theta_t\, \dot{m}_t) + \alpha_{19} (\dot{m}_t\, P_t) \\ &+ \alpha_{20}\} \end{aligned} \quad 7$$

$$\begin{aligned} CO2(t) = \{&\alpha_1 N_t + \alpha_2 N_{t-7}^2 + \alpha_3 \tau_t^3 + \alpha_4 \tau_t + \alpha_5 \tau_{t-7}^3 + \alpha_6 \tau_{t-7}^2 + \alpha_7 \dot{m}_t + \alpha_8 \dot{m}_{t-7}^3 \\ &+ \alpha_9 \dot{m}_{t-7}^2 + \alpha_{10} \dot{m}_{t-7} + \alpha_{11} \dot{m}_{t-14}^3 + \alpha_{12} \dot{m}_{t-14}^2 + \alpha_{12} \dot{m}_{t-14} + \alpha_{13} P_{t-7} \\ &+ \alpha_{14} (N_{t-7}\, \tau_{t-7}) + \alpha_{15} (N_{t-14}\, \tau_{t-14}) + \alpha_{16} (N_{t-7}\, \dot{m}_{t-7}) \\ &+ \alpha_{17} (\tau_{t-7}\, P_{t-7}) + \alpha_{18} (\tau_{t-7}\, \dot{m}_{t-7}) + \alpha_{19} (\dot{m}_t\, P_t) + \alpha_{20}\} \end{aligned} \quad 8$$

7 DISCUSSION

7.1 Volterra modelling process and input data

The results show the potential of using dynamic models based on dynamic training data to capture the NOx and CO_2 emissions of Diesel engines in response to a number of control variables. In contrast, the approach has not given satisfactory results for CO and THC emissions.

The modelling approach proposed in this paper is based on least squares regression and as such allows a repeatable training method unlike the fitting of non-explicit models such as

Neural Networks. However, such is the large number of possible terms described by the Volterra series, it is still necessary to calculate a considerable number of models to obtain suitable results. Even in the case of the CO_2 model which achieved a high level of fit, a large number of models during the first phase of the identification process presented poor levels of fit. The multiple stage approach presented in this paper illustrates a method to avoid poor model qualities.

7.2 Model structure analysis

Detailed analysis of the terms apparent in the NOx emissions (equation 7) shows that the response was significantly influenced by all of the input parameters. There are a large number of terms relating to engine speed and torque showing both the dominance but also the complexity of the response to the overall engine operating point. In this approach, a global model has been produced where engine speed and torque are given the same stature as other control variables. However, an approach taken by other researchers is to produce dynamic models at a range of local engine speed and torque points. The present approach was justified as it allows capturing of dynamics related to the change in speed and torque condition, however evidently requires the modelling of a more complex function. Main injection timing only appears as a static term in the model (terms 8 and 18) suggesting the impact on NOx emissions does not present significant dynamics in the 10Hz time frame of the model. In contrast, the other inputs present a number of delay terms which are required to capture the

dynamics associated with the engine turbocharger inertia, EGR and intake path volumes and thermal inertia.

Considering the CO_2 equation (8); this is dominated by terms relating to engine torque and air flow. A combination of static and dynamic terms are included in the model which, as for the NOx emissions, will be required to capture the dynamics of the various engine components. Also in this case, the mechanical inertia of the engine may be significant. It is interesting that the injection timing term does not feature in the model. One explanation for this is that the effect is hidden under the effects of other variables, i.e. this effect is small and can't be extracted from the data. This is not necessarily an issue with the modelling approach, but rather with the individual test design and resulting training data. In this study a relatively small window of injection timing was considered (4^{o}Crank Angle) which may be significant for the formation of other emissions in the cylinder, but may not have a significant effect on fuel consumption and CO_2.

7.3 Temperature scaling function

Simple first and second order equations were derived for the temperature behaviour of emissions formation. This was based on the assumption that the effect of temperature was independent of other input parameters. This approach required only a simple experiment for identification and this appears sufficient for NOx and CO_2, evidenced by the similar levels of fit in table 6 for hot and cold NEDC. However, if this assumption were strictly true, then the data

used for the fitting in figure 14 would lie considerably closer to the fitted lines. The large excursions from this may result from two effects:

1. Measurement issues resulting in misalignements in the hot and cold start data sets
2. Real effects dependent on the engine operating point

To consider first of these effects, the hot and cold emissions measurements for NOx are plotted in figure 17. This shows that whilst the aligment is reasonable, there are cases where this is poor. The issues arise from variable time delays associated with the transport of emissions gases from the engine exhaust to the sensors in the analysers.

Further investigations into the temperature dependent scaling function were conducted by using additional factors in the fitted model. Two further iterations were included, the first using brake power as an input and second using engine speed and torque. The fit statistics are shown in table 7. These show an improvement in fit from the additional parameters but this would always be the case as the flexibility of the function is increased.

Emission	**R^2**		
species	$y=f(T_{oil})$	$y=f(T_{oil},W_{brake})$	$y=f(T_{oil}, N_{eng},\tau_{eng})$
NOx	0.38	0.39	0.42
CO_2	0.25	0.29	0.32
CO	0.32	0.38	0.39
THC	0.49	0.49	0.51

Table 7: Fit statistics for augmented temperature scaling function

Each of the new scaling functions was used in combination with the Volterra model for NOx and CO_2 to predict cold start NEDC emissions; the resulting fit statistics are given in table 8. This shows that for NOx emissions there is a small improvement in nRMSE from 6.8% to 6.3% and 6.5% using the augmented temperature scaling functions. For CO_2 on the other hand the augmented scaling factor causes deterioration in predicted cold start statistics with nRMSE increasing from 6.6% to 9.8%. Coupling this to the low levels of fit from (table 7) suggests a significant level of noise and a tendency to over-fitting for the augmented scaling factor models.

Emission	$y=f(T_{oil})$			$y=f(T_{oil}, W_{brake})$			$y=f(T_{oil}, N_{eng}, \tau_{eng})$		
	R^2	RMSE	nRMSE	R^2	RMSE	nRMSE	R^2	RMSE	nRMSE
NOx	0.85	59	6.8	0.88	55	6.3	0.86	56	6.5
CO_2	0.91	0.80	6.6	0.91	0.81	6.8	0.88	1.2	9.8

Table 8: Fit statistics for cold start NEDC NOx and CO_2 using a range of temperature scaling functions

8 CONCLUSIONS

Dynamic polynomial models (Volterra Series) have been used to model various gaseous emissions species from a multi-cylinder Diesel engine. Dynamic experiments were conducted, varying five control parameters according to swept frequency sinusoidal excitations. The measured data was used to calculate Volterra models and their predictive performance was assessed over the NEDC drive cycle. A simple approach for capturing temperature dependent

behaviour during engine warm-up was also presented that uses a simple temperature dependent scaling factor. Based on this work the following conclusions have been drawn:

1. Modelling of NOx and CO_2 emissions over the NEDC resulted in predicted nRMSE values of 6.8% and 6.6% respectively using conventional measurement equipment. This high level of fit can provide useful models for engine simulation work.
2. The modelling of CO and THC emissions is more problematic suggesting significant levels of random variation that was not controlled by the test procedure in this work. Normalised RMSE levels were 26% and 17% for cold start NEDC respectively.
3. A simple approach to capture temperature dependant behaviour using a scaling function independent of other inputs provides a reasonable prediction of cold start behaviour. The inclusion of additional terms in this model such as engine power can improve the prediction but care should be taken to avoid capturing measurement noise.

The models developed in this work offer the possibility to replace steady state approaches for engine calibration by offering reduced experimental effort and also yielding additional information relating to transient response. The resulting models could be used in optimisation procedures or integrated into the engine ECU to aid the controller algorithms which are becoming more physically based. During vehicle development, the models could also be used as a replacement for map based engine models or as an enhancement to mean values models. An example application is the development of hybrid vehicles which typically rely on models of

stationary engine behaviour. These simple mathematical models could improve the products by allowing early estimation of transient emissions.

9 FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

10 REFERENCES

1. Delphi. Worldwide Emissions Standards: Passenger Cars and Light Duty Vehicles. 2010.
2. Owen N. and Jackson N. A New Look at the Low Carbon Roadmap. *Low-Carbon Vehicles 2009*. London (UK): Chandos Publishing Ltd., 2009, p. pp. 3-14.
3. Taylor A.M.K.P. Science Review of Internal Combustion Engines. *Energy Policy*. 2008; 36: 4657-67.
4. Watanabe W., Ehara M., Ohata A., Kaneko S., Yoshida S., Baumann W. and Kohler B.-U. Direction of Model Based Engine Calibration. In: Ropke K., (ed.). *Design of Experiments (DoE) in Engine Development*. Berlin, Germany: expert verlag, Renningen, 2009.
5. Sugita M., Harada S. and Arakawa H. Transient Measurement for Steady State Calibration. In: Ropke K., (ed.). *Design of Experiments (DoE) in Engine Development*. Berlin, Germany: expert verlag, Renningen, 2009, p. pp. 78-91.

6. Bent E., Shayler P.J. and La Rocca A. Investigation of Temperature Fields During Stop-Start Operation of a Spark Ignition Engine and Hte Effect on Engine Behaviour after Restart. *VTMS 11*. Gaydon, UK2011.
7. Deflorian M. and Klopper F. Design of Dynamic Experiments. In: Ropke K., (ed.). *Design of Experiments (DoE) in Engine Development*. Berlin, Germany: expert verlag, Renningen, 2009, p. pp. 31-40.
8. Ezzeddinne M., Castro E. and Lengelle R. Dynamic Design of Experiments for Engine Pollutants Emissions Modelling and Optimization. *SAE Powertrains, Fuels and Lubricants Meeting*. Rosemont, Illinois: SAE International Warrendale Pennsylvania USA, 2008.
9. Baumann W., Klug K., Kohler B.-U. and Ropke K. Modelling of Transient Diesel Engine Emissions. In: Ropke K., (ed.). *Design of Experiments (DoE) in Engine Development*. Berlin, Germany: expert verlag, Renningen, 2009, p. pp.41-53.
10. Baumann W., Schaum S., Roepke K. and Knaak M. Excitation Signals for Nonlinear Dynamic Modelling of Combustion Engines. *Proceedings of the 17th World Congress The international Federation of Automatic control*. Seoul, Korea2008.
11. Guhmann C. and Riedel J.-M. Comparaison of Identification Methods for Nonlinear Dynamic Systems. In: Ropke K., (ed.). *6th Design of Experiments (DoE) in Engine Development*. Berlin, Germany: expert verlag, Renningen, 2011, p. pp.41-53.
12. Burke R.D., Zhang Q. and Brace C.J. Dynamic Models for Diesel Nox and CO2. *Powertrain Modelling and Control 2012*. Bradford, UK2012.

13. Dec J.E. and Canaan R.E. Plif Imaging of No Formation in a Di Diesel Engine, Sae Paper Number 980147. 1998.
14. Flynn P.F., Hunter G.L., Durrett R.P., Farrell L.A. and Akinyemi W.C. Minimum Engine Flame Temperature Impacts on Diesel and Spark-Ignition Engine Nox Production. 2000.
15. Ahmad T. and Plee S.L. Application of Flame Temperature Correlations to Emissions from a Direct-Injection Diesel Engine. SAE International, 1983.
16. Pierpont D.A. and Reitw R.D. Effects of Injection Pressure and Nozzle Geometry on D.I. Diesel Emissions and Performance, Sae Paper Number 950604. 1995.
17. Kidoguchi Y., Yang C. and Miwa K. Effects of Fuel Properties on Combustion and Emission Characteristics of a Direct-Injection Diesel Engine. SAE International, 2000.
18. Burke R.D., Brace C.J., Hawley J.G. and Pegg I. Review of the Systems Analysis of the Interactions of Thermal, Lubricant and Combustion Processes of Diesel Engines. *Proc Inst Mech Eng Part D-J Automob Eng*. 2010; 224: 681-704.
19. Burke R.D., Brace C.J., Cox A., Lewis A., Hawley J.G., Pegg I. and Stark R. Systems Approach to the Improvement of Engine Warm-up Behaviour. *Proc Inst Mech Eng Part D-J Automob Eng*. 2011; 225(2): 190-205.
20. Dec J.E. A Conceptual Model of D.I. Diesel Combustion Based on Laser-Sheet Imaging, Sae Paper Number 970873. 1997.
21. Nelles O. *Non Linear System Identification*. Berlin, Germany: Springer Verlag, 2001.

11 APPENDIX: Notation

11.1 Abbreviations

APRBS	Amplitude Modulated Pseudo Random Binary Signal
ASAP3	Communication Protocol
CO	Carbon Monoxide
CO_2	Carbon Dioxide
DoE	Design of Experiments
ECU	Engine Control Unit
EGR	Exhaust gas Recirculation
NEDC	New European Drive Cycle
NOx	Oxides of Nitrogen
nRMSE	Normalised RMSE
PID	Proportional, Integral, Derivative (controller)
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
SOI	Start of main injection
THC	Total Unburnt Hydrocarbons
VGT	Variable Geometry Turbocharger

11.2 Symbols

f	frequency	rad/s
j	Input delay offset	
l	Output delay offset	
N	Rotational speed	rev/min
n	Number of measurements	
$\dot{m}$	Mass flow rate	kg/hr
P	Fuel injection Pressure	Bar
t	time	S
T	Temperature	^{o}C
W_{brake}	Brake engine power	W
x	Regression Input	
y	Measured dependent variable	
$\hat{y}$	Output predictor	
$\bar{y}$	Mean measured dependent variable	

Greek Letters

φ	Phase Shift	rad
α	Regression parameter	
β	Regression parameter	
γ	Regression parameter	

Δ	Regression Coefficient	
ε	Regression parameter	
θ	Main Injection Timing	^{o} CA BTDC
τ	Torque	Nm

Subscripts

n	Input number
k	Parameter number

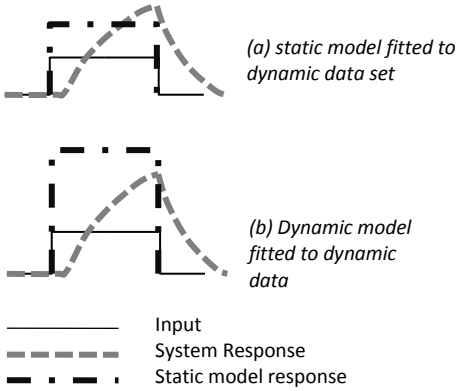

Figure 1: Illustration of (a) static model prediction and (b) dynamic model fitted to dynamic measurement data

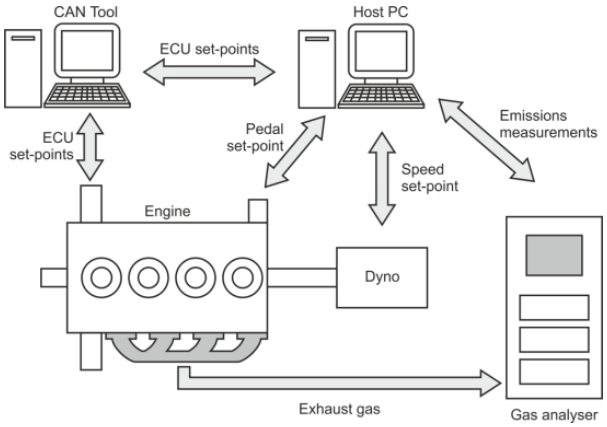

Figure 2: Experimental facility communications layout

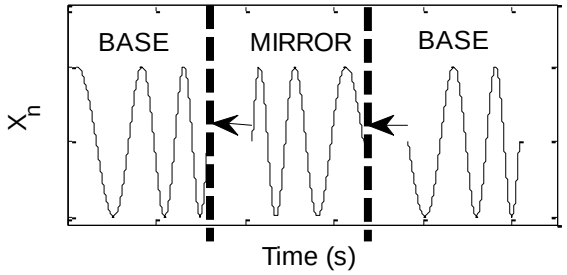

Figure 3: Assembly of full excitation signals from base chirps

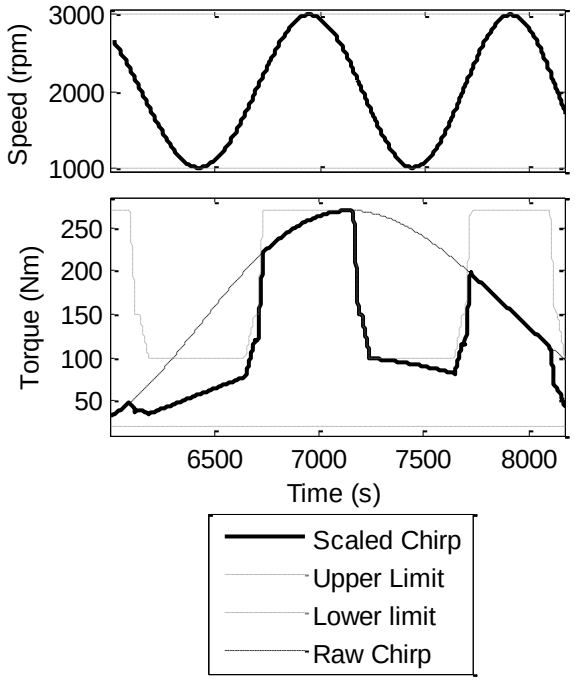

Figure 4: Scaling of torque chirp signal according to speed

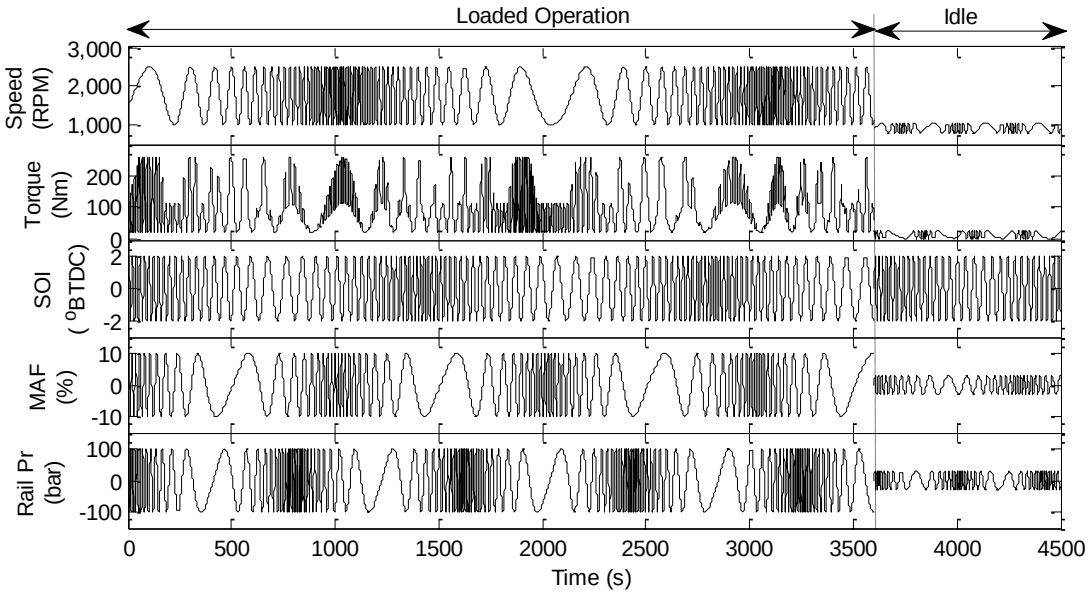

Figure 5: Torque based Chirp excitation signals for main and idling test

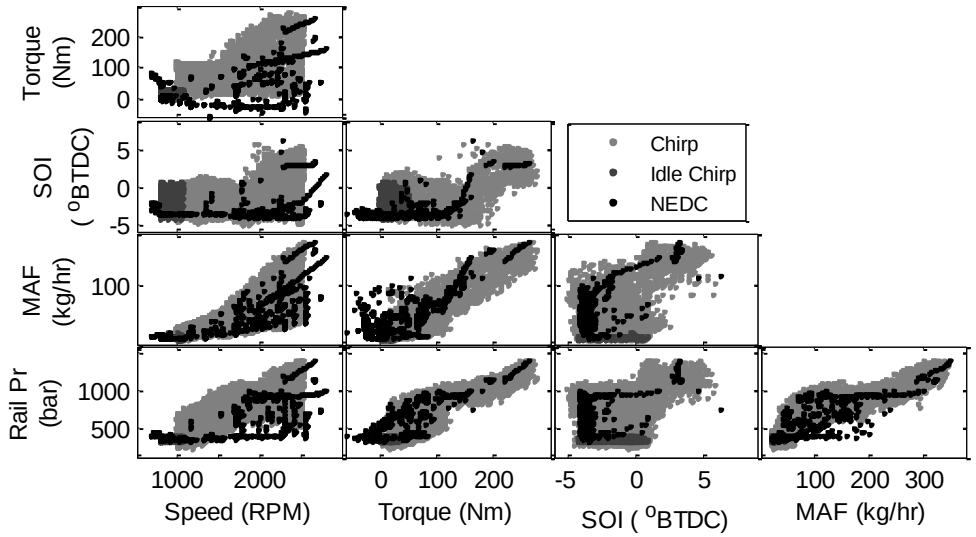

Figure 6: 2dimensional projections of 5 dimensional design space for the dynamic chirp experiment, idle chirp and NEDC

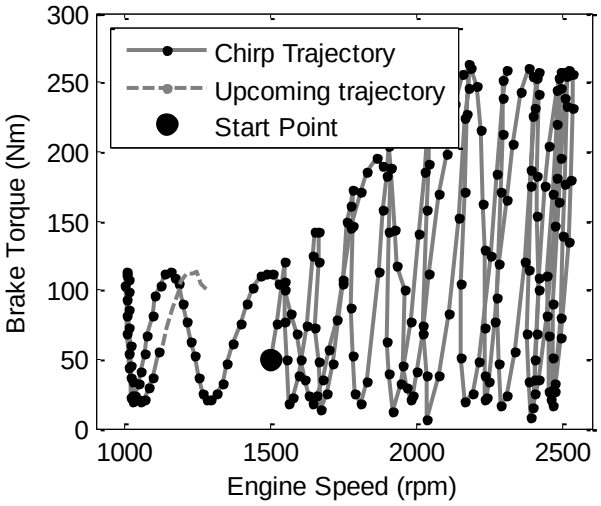

Figure 7: Brake Torque and engine speed evolution over the first 25 seconds of the Chirp experiment

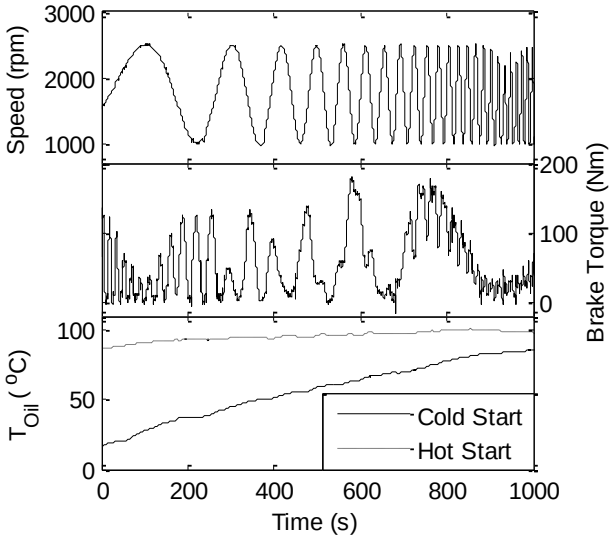

Figure 8: Engine speed, brake torque and oil temperature for the temperature function identification test

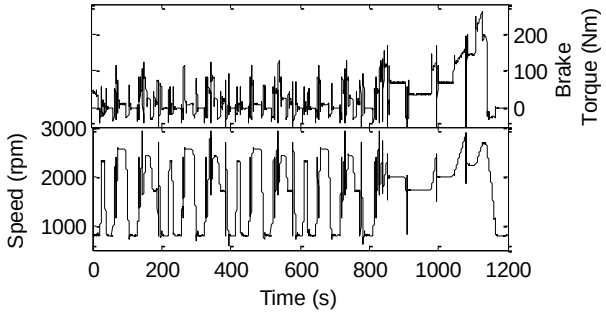

Figure 9: NEDC speed and torque traces used for temperature function training and model validation

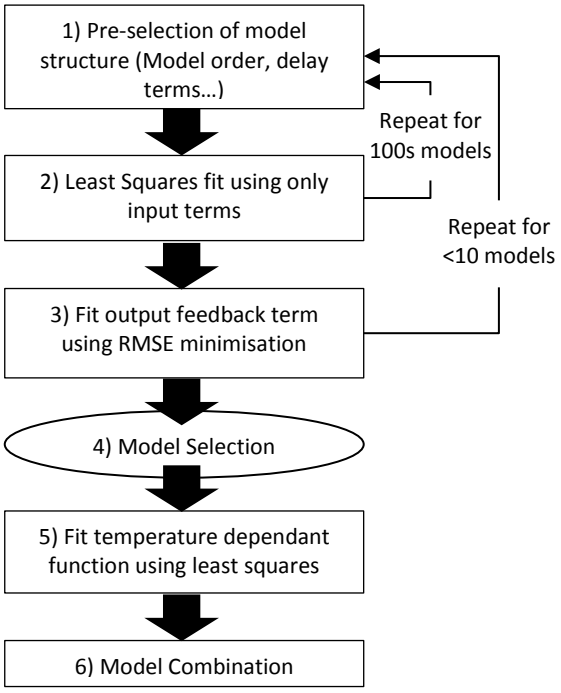

Figure 10: Model Training process

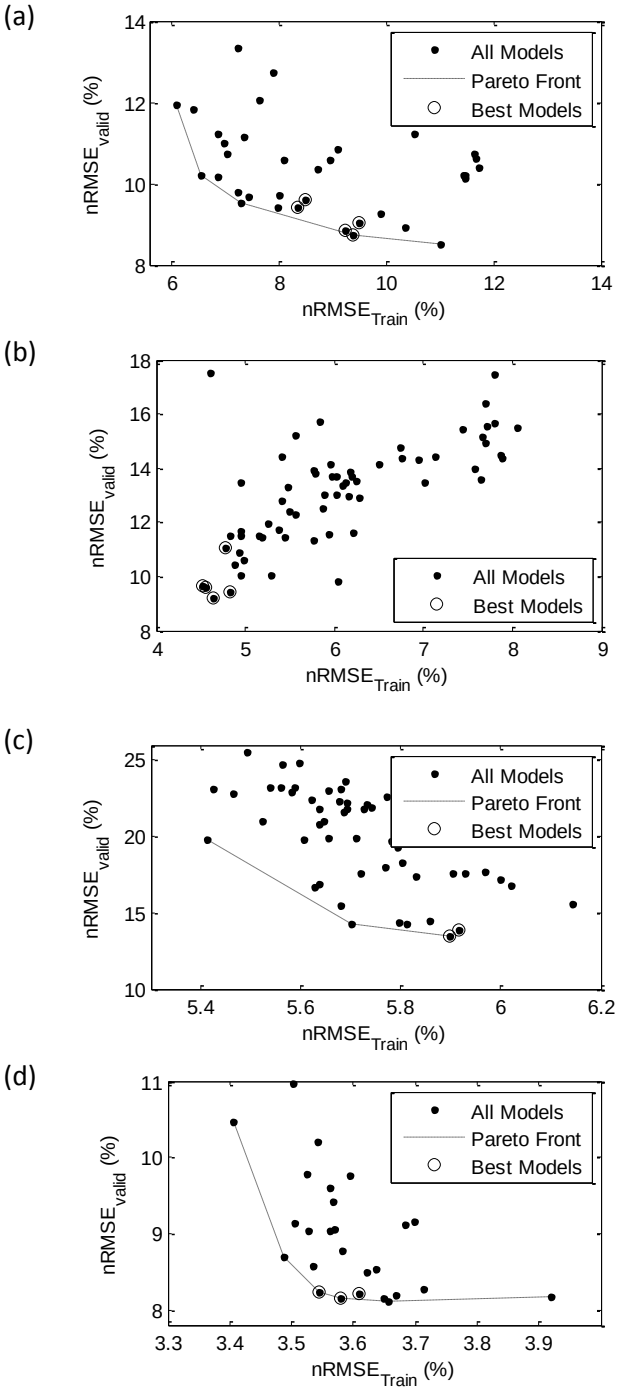

Figure 11: Normalised RMSE trade-off for training and validation data for (a) NOx, (b) CO_2, (c) CO and (d) THC emissions

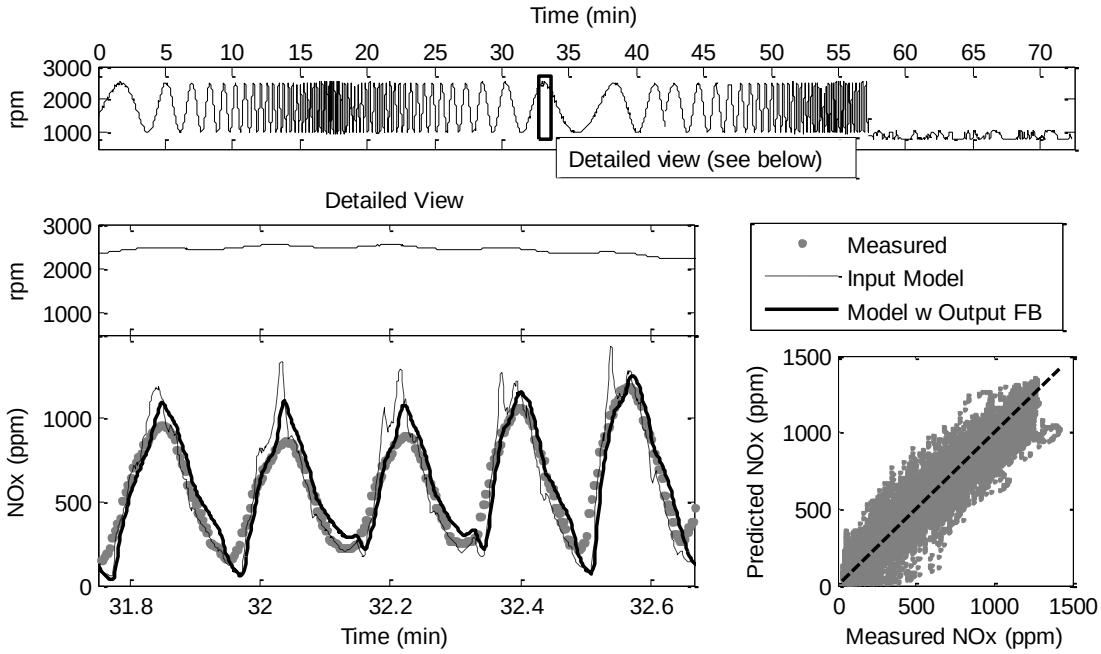

Figure 12: Fitted NO_x emissions for training data both with and without feedback term

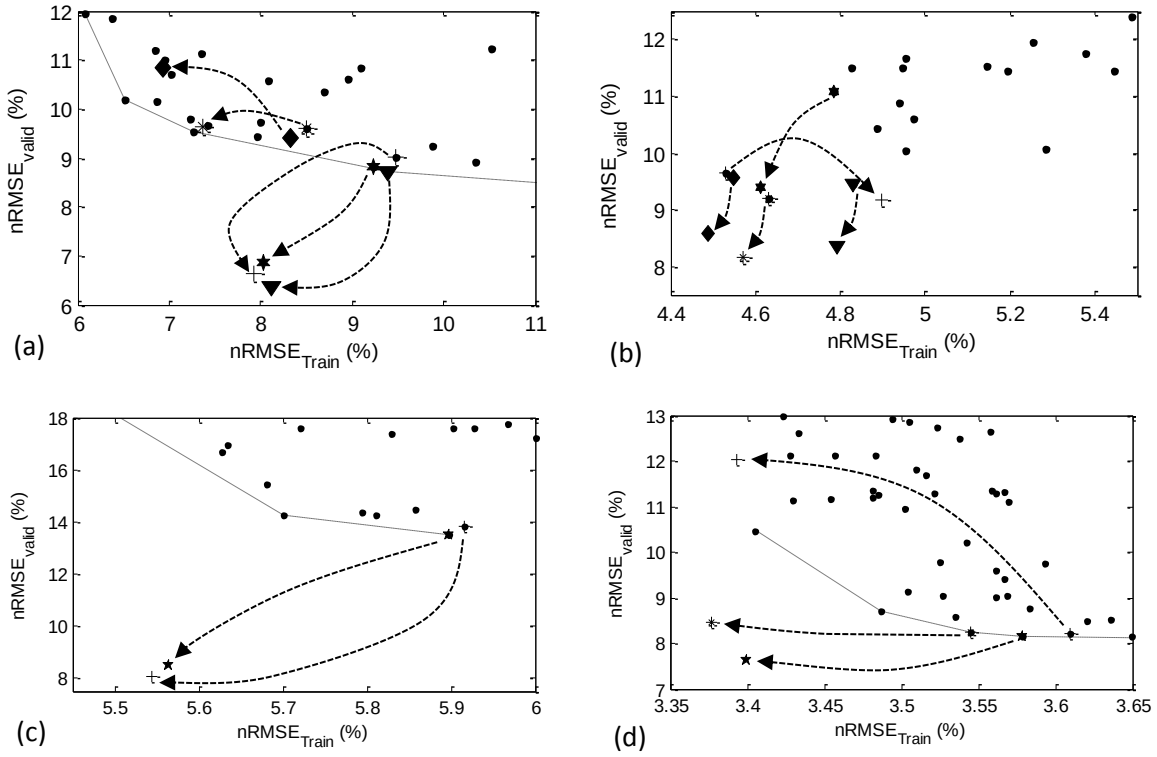

Figure 13: Variation of model fit through inclusion of output feedback for (a) NOx, (b) CO_2, (c) CO and (d) THC emissions. Arrows indicate evolution of nRMSE through inclusion and training out output feedback term

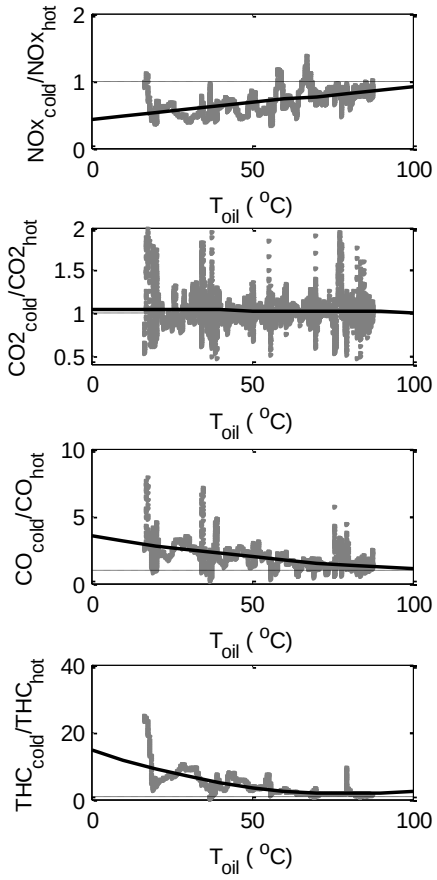

Figure 14: Ratio of hot to cold emissions for NOx, CO_2, CO and THC as a function of oil temperature

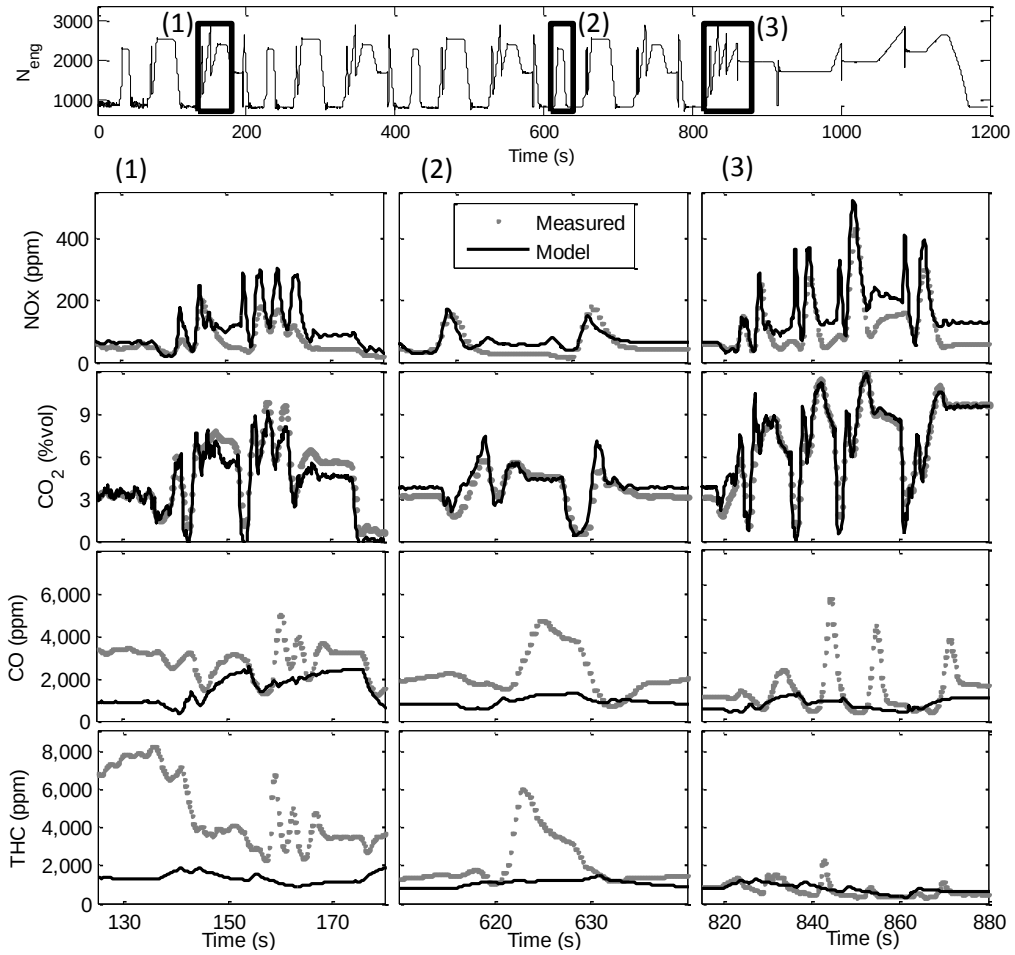

Figure 15: Model prediction of cold start NEDC NOx, CO_2, CO and THC at time windows

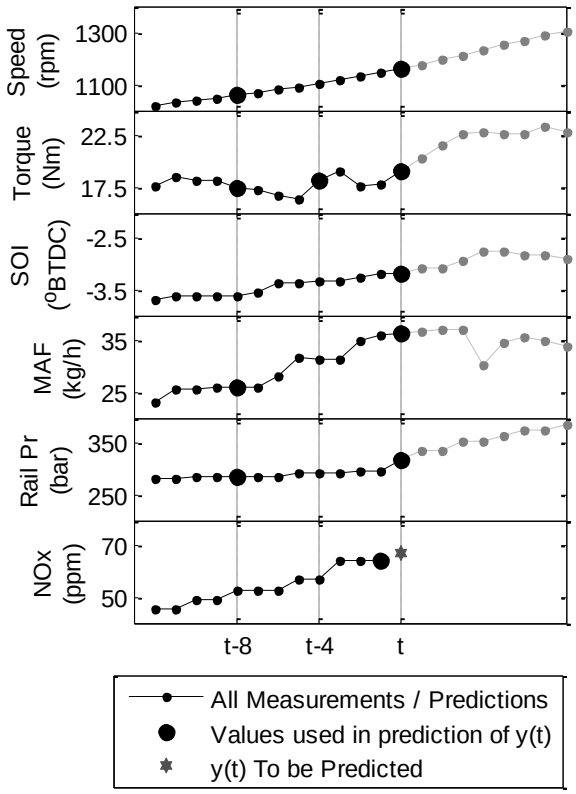

Figure 16: Illustration of significant terms of the Volterra model for NOx emissions

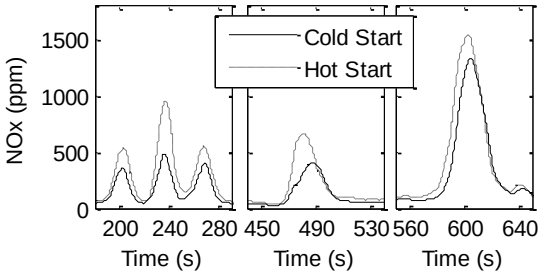

Figure 17: NOx emissions measurement during hot and cold tests for temperature function identification