Dynamic Models for Diesel NOx and CO2

RD. Burke [1], Q. Zhang [1] and CJ. Brace [1]

[1] Department of Mechanical Engineering, University of Bath, UK

Abstract: Dynamic modelling of engine emissions is important because it promises significant improvements over static modelling for engine calibration through reduced testing and development times. Volterra series and neural network dynamic model structures were trained using transient data from an engine test stand under sinusoidal excitations and the predictive power over the New European Drive Cycle (NEDC) cycle was assessed for a EURO IV specification Diesel engine. Models were identified for oxides of nitrogen (NOx) and carbon dioxide (CO_2) emissions concentrations based on engine speed, torque, injection timing, EGR rate and fuel injection pressure. The fit R^2 values for CO_2 emissions for Volterra series and Neural networks were 0.92 and 0.99 respectively; for NOx emissions these were 0.92and 0.998. Although this suggests better flexibility from the Neural network to represent the nonlinearity there were large variations in predictive power resulting from the partially random nature of model training. Also, close observation of the model predictions suggested higher accuracy for the Volterra series for most of the cycle, but that this model suffered from fewer, large prediction errors compared to smaller but more frequent errors from the neural network.

Keywords— Design of Experiments, Dynamic Models, Neural Networks, Volterra Series.

1-Introduction

The current standard approach to engine calibration is to capture a high fidelity mathematical representation of the engine behaviour through high fidelity statistical models. Although physical engine models can avoid the need for experimental data in the early stages of control system design, the thermodynamic and chemical processes in engine combustion are often too complex to give an accurate representation without significant resources and model tuning [1]. In this process, the data driven models are fitted to experimental results measured from the engine operating under steady state conditions. Because of the large number of control variables on modern engines, design of experiments (DoE) has become the standard approach to reduce

experimental effort by optimising the combinations of variables at each testing point [2]. However, the current approach presents two major limitations:

- Testing time remains long because of the need to settle the engine at each steady state point between measurements.
- The model obtained from the experimental data is only valid under steady conditions and does not capture the engine behaviour during transient events. This imposes a limitation on the optimisation procedure which can only account for steady state operating conditions.

These limitations can be overcome with a move to a dynamic calibration process using dynamic engine modelling. A dynamic training sequence in the form of a transient test sequence is used rather than a steady state test plan to obtain the model training data. In this paper different mathematical structures for dynamic models are compared: extended Volterra Series (polynomial) and Neural Network. Gühmann and Riedel [3] compared a number of different modelling approaches on a single training data set and identified neural networks and Volterra series having best performance. In their study, both the training and validation data sets were sinusoidal in nature however in this work the performance of their two best performing models are assessed over the NEDC.

2-Methodology

2.1-Experimental approach

A EURO IV specification, 2.0L Turbocharged Diesel engine was used in this study for which the usual vehicle application was a light commercial vehicle. The engine was installed in a transient engine dynamometer facility allowing dynamic excitation and full emulation of the NEDC. The engine was controlled from a *CP engineering Cadet* host system with an interface to the engine control unit (ECU) using *Accurate Technologies Ati Vision* system. Engine emissions were sampled between the turbocharger and the close coupled catalyst and concentrations were measured using *Horiba MEXA 7000* analysers. The test cell and communications networks are shown in figure 1.

The region of interest for the models represented the operating region during the NEDC: this defined the speed and torque operating regions for the training data. Three calibration variables

were identified: injection timing, fuel injection pressure and EGR valve position. As the engine was supplied with a production calibration, the region of interest was defined relative to the "production" operating points and implemented through "adder" and "multiplier" functions within the ECU software. These functions are a core part of the engine strategy and have various uses during the calibration process. In this case they offer a simple method for applying an offset to the production calibration for the modelling purposes. The control of these variables was specified in the host system and passed to the ECU via the calibration tool using an *ASAP3 link*. The excitation ranges and methods are summarised in table 1.

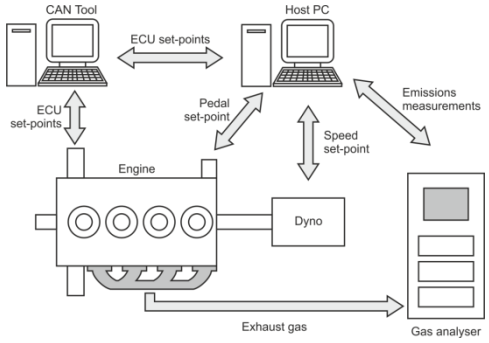


Figure 1: Test facility and communication network

Variable	Excitation method	Full Range	Idle Range
Engine speed	Direct control from host system	1000-2500rpm	800-1000rpm
Engine load	PID control of electronic pedal position from host system	20-250Nm	0-20Nm
Injection timing	'adder' function resident in ECU;	+/-2^{o}CA	+/-2^{o}CA
Common Rail fuel pressure	'adder' function resident in ECU;	+/-100bar	+/-25bar
EGR valve position	Indirect control through 'multiplier' function for Mass air flow set point.	+/-10%	+/-5%

Table 1: Excitation variables, ranges and implementation

Sinusoidal chirp signals were used to excite the system as they offer a good compromise between static and dynamic space coverage and are more suited to engine operation than the harsh changes experienced in step based signals [4]. Although random binary signals are theoretically better for dynamic system identification [5], their harsh nature can cause problems and even damage engine hardware. The training data sequence is shown in figure 2 and comprises of a full load and an idle phase; the torque signal has been scaled as a function of engine speed to represent the engine limitations and the region of interest defined by the NEDC.

2.2-Model Structures

Two dynamic model structures were considered for this work and will be described in the following section. The dynamic models mathematical functions that depend on the current input settings, the previous states of these inputs and the previous states of the system output, as defined by equation 1; in each case the dynamic models were trained for operation at 10Hz.

$$y(t) = f\left(y(t-1), y(t-2), \dots, y(t-n_y), x(t), x(t-1), x(t-2), \dots, x(t-n_x)\right) \quad (1)$$

<u>Extended Volterra Series</u>: This polynomial based modelling structure is a reduced version of the Kolmogorov-Gabor polynomial (equation 2) which models the system by assuming that the non-linearity remains only between the inputs and not the output feedback (i.e. exclusion of terms θ_8, θ_9 and θ_{12} to θ_{15}). The Volterra model is represented schematically form in figure 3. The identification of such a model is achieved by initially by regression of parameter *u* onto the inputs followed by a post optimisation phase to include the output feedback term. The model parameterisation is primarily dependent on regression and therefore a stable and repeatable process. The Volterra model is defined by the following parameters:

- Model order: the highest exponent for the static model; typically 4th order.
- Delay order: the number of previous input events; typically 1 or 2.
- Interaction order: the number of grouped inputs; typically only 2-way.
- Feedback order: the number of previous output terms included in the model; typically 1.

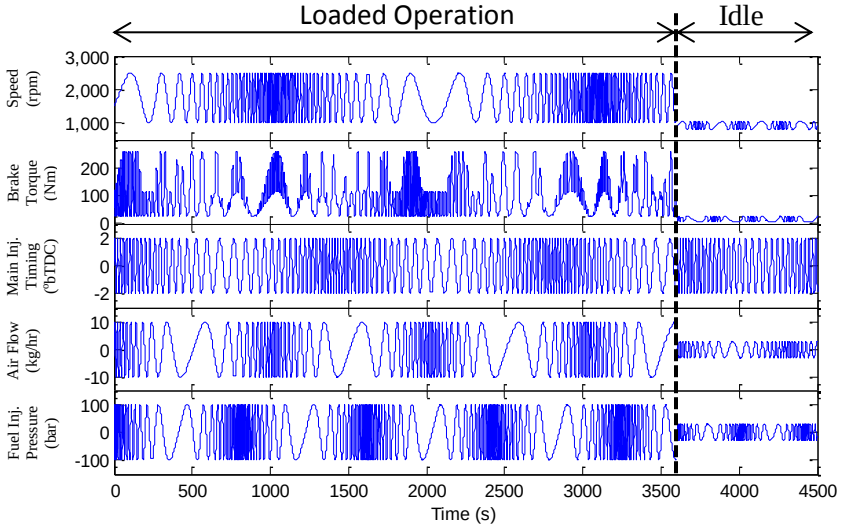


Figure 2: Training data test plan

$$y(t) = \theta_1 + \theta_2 u(t-1) + \theta_3 u(t-2) + \theta_4 y(t-1) + \theta_5 y(t-2) + \theta_6 u^2(t-1) + \theta_7 u^2(t-2) + \theta_8 y^2(t-1) + \theta_9 y^2(t-2) + \theta_{10} u(t-1)u(t-2) + \theta_{11} u(t-1)y(t-1) + \theta_{12} u(t-1)y(t-2) + \theta_{13} u(t-2)y(t-1) + \theta_{14} u(t-2)y(t-2) + \theta_{15} y(t-1)y(t-2) \quad (2)$$

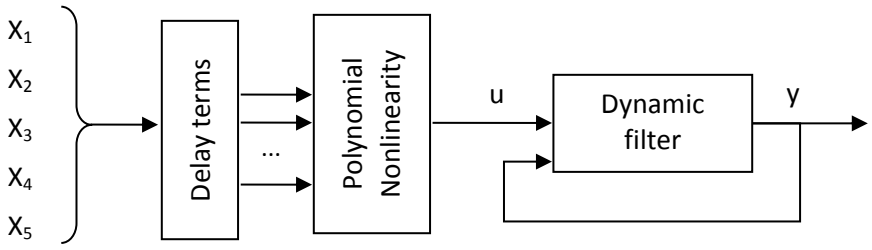


Figure 3: Schematic representation of the Volterra series model

Neural Networks: The neural network models are interconnections of neurons which can perform calculations independently. The model imitates the behaviour of simplified biological neural network by storing highly complex and nonlinear information through varied weights (W) and biases (b) of the inter-connections. A typical dynamic neural network structure is shown in figure 4. Training of the neural network models is done in an iterative way aiming to minimize the fit error. For a dynamic neural network, the feedback loop is disconnected and measured output terms are used in their place during training. After training, the feedback loop is reconnected so that the simulated output is used thus providing a predictive tool. The initialisation of the training algorithm is a random selection of weights and biases and this can result in variations in fit quality on repeated training runs. Usually several rounds of training are conducted to give an idea of the spread and allow a good model to be chosen. For each neuron, the input and output relationship is defined by equation 3, the grouping of the neurons results in a function represented by equation 1.

$$a = tansig(Wp + b) \quad (3)$$

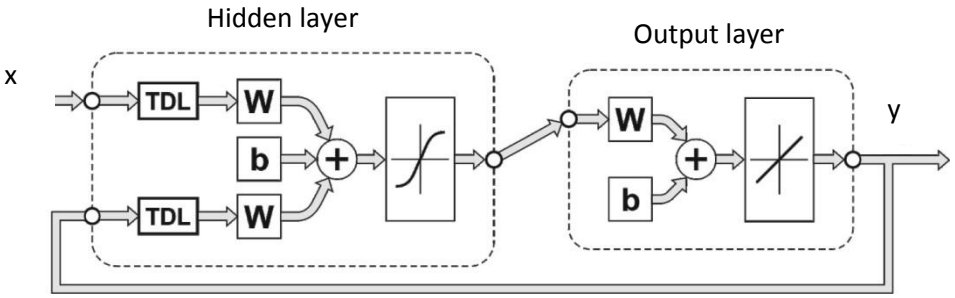


Figure 4: Dynamic neural network model

The neural network model is defined by the following parameters:

- Layer number: the number of hidden layers defines the complexity of the whole neural network. Usually, one hidden layer composed of around ten neurons proves to be sufficient.
- Delay order: the number of previous input and output events to be fed into the hidden layer.
- Neuron number: since the weights and biases were separately defined for each inter-connection between neurons, neuron number decides how much information can be stored in the model. Given that the inter-connection number is the factorial of neuron number, small neuron numbers are usually chosen to prevent over fitting.
- Transfer function: the nonlinear behaviour of the system is represented using the nonlinear transfer function with each neuron.

2.3-Model Assessment

The models were trained paying attention to avoid over-fitting by considering reasonable model complexities in each case. With these statistical models it is always possible to reach a perfect model fit if enough terms or neurons are included; therefore it is important to assess the predictive power of the models on an independent validation data set. The trained models were validated over separate tests conducted over the NEDC cycle. The model predictions were compared to the measured behaviour and the quality of fit was assessed using coefficient of determination (R^2), Root mean square error (RMSE), normalised RMSE and signal to noise ratio (SNR) as defined by equations 4 to 7. Before calculating the fit statistics for the validation sequence, points lying outside the multi-dimensional training region were removed from the data set to avoid significant use of the model in extrapolation.

$$R^2 = 1 - \frac{\sum(\hat{y}-y)^2}{\sum(y-\bar{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum(\hat{y}-y)^2}{n}} \tag{5}$$

$$nRMSE = \frac{RMSE}{y_{max}-y_{min}} \tag{6}$$

$$SNR = 10log_{10}\left(\frac{(\sum\hat{y}^2)/n}{(\sum(y-\hat{y})^2)/n}\right) \tag{7}$$

3-Results

An example of the Volterra series fit for NOx emissions is shown in figure 4: a detailed portion of the training data is shown with a *fitted vs. measured* plot for the complete data set.

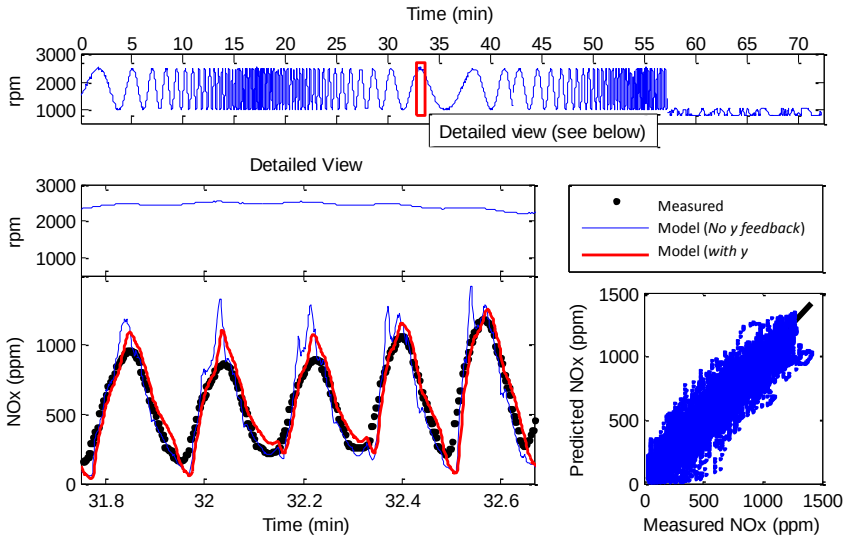


Figure 4: Fitted NOx emissions for training data for Volterra Series model.

The model structures for best fitting NOx and CO_2 emissions are detailed in tables 2 and 4 for Volterra series and neural network models respectively; the corresponding fit statistics for both model types are show in tables 3 and 5. For the Volterra model, the fit quality is broken down into three stages of the regression process; for the neural network modelling only the 1st and 3rd stages were performed due to the time required for model training.

1. Static polynomial: referring to the model only with current time step inputs.
2. Dynamic polynomial: referring to a model with current and previous time step inputs.
3. Autoregressive model: referring to the complete model with output feedback.

For the Volterra models, NOx emissions fitted best without input delays (the 5-8s transport delay in the measurement process was accounted for independently by appropriate time alignment [6]); the autoregressive model improved R^2 from 0.9 to 0.92. For CO_2, there was a significant improvement from the inclusion of delay terms, with the best fit resulting from a delay term at *t-0.9s* (R^2 improving from 0.82 to 0.93). The inclusion of an autoregressive term improved the fit most noticeable through a reduction in RMSE and an increase in signal to noise ratio.

	NO_x	CO_2
Model order*	4^{th}	4^{th}
Interaction order#	2^{nd}	2^{nd}
Delay terms	None	1 (0.9s)
Output transform&	0.25	0.75
Total Number of terms after parameter selection	17	27

Table 2: Volterra model structures for NOx and CO_2 emissions.

	NOx			CO_2		
Fit Statistic	*Static polynomial*	*Dynamic Polynomial*	*Autoregressive polynomial*	*Static polynomial*	*Dynamic Polynomial*	*Autoregressive polynomial*
R^2	0.9		0.92	0.82	0.93	0.93
RMSE	114		96	0.77	0.48	0.46
nRMSE	8.2		6.9	7.6	4.8	4.5
SNR	13.7		15.1	19.6	23.7	24.1

Table 3: Fit statistics for Volterra series models

Due to the instability of training results of neural network model each candidate structure was repeatedly trained and the models giving best validation results were chosen. The dynamic network shows significant improvement on the fitting statistics (Table 5). Both NOx and CO_2 models achieved virtually perfect fit. For the NOx emission model, a dynamic autoregressive network with 3 neurons and 2 delay terms was found with lowest validation Normalised RMSE. For the CO_2 model, the optimal model was found with 6 neurons and 3 delay terms. The fitting capability of the neural network models exceeds that of the Volterra series model because it allows for more flexibility in representing the nonlinearities present within the data. This is reflected in the fit quality for all models such as the Dynamic autoregressive NOx function which achieves an R^2 of 0.99 for neural net compared to 0.93 for Volterra series.

	NO_x	CO_2
Layer number	1	1
Delay terms	2	3
Neurons number	3	6
Transfer Function	tansig	tansig

Table 4: Neural network model structures for NOx and CO_2 emissions.

	NOx		CO_2	
Fit Statistic	*Static network*	*Dynamic Autoregressive Network*	*Static network*	*Dynamic Autoregressive Network*
R^2	0.95	0.998	0.84	0.99
RMSE	77.9	15.85	0.72	0.48
nRMSE	5.6	1.14	7.1	4.8
SNR	17	30.8	20.2	23.7

Table 5: Fit statistics for Neural network models

Table 6 shows the modelling statistics for the prediction for the NEDC validation tests. Extrapolation was allowed for some points close to the periphery of the hull as specified by the tolerance limit (defined as a percentage of hull size): these were set to 3% for the NOx and 1% CO_2 and were based on model stability in extrapolation. The NEDC validation results showed a statistical advantage of Neural network model over the Volterra series model, with similar NOx predictive results (R^2=0.84 of Neural net against R^2=0.81 of Volterra series) and better CO_2 predictive results (0.92 against 0.82).

Model Structure	**Volterra**		**Neural Network**	
Emissions	**NO_x**	**CO_2**	**NOx**	**CO2**
Hull limit tolerance	3%	1%	3%	1%
% points included using hull	71%	65%	71%	65%
Predicted R^2	0.81	0.82	0.84	0.92
RMSE (ppm or %)	65	1.22	60.6	0.84
Normalised RMSE (%)	6.2	11.3	7	7.8
Signal to Error ratio (dB)	10.3	14	9.2	16.4

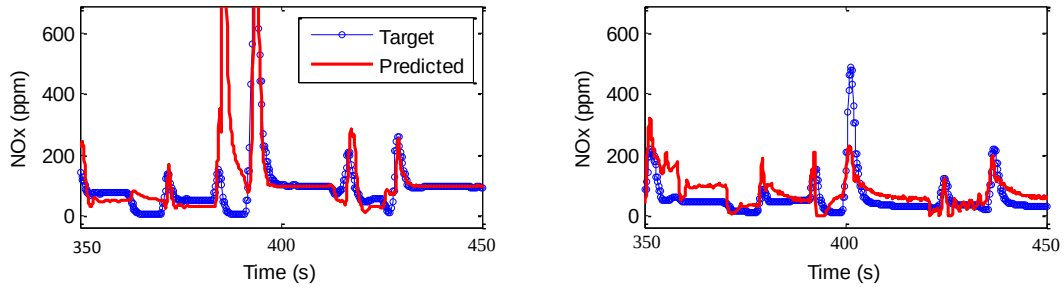
Table 6: Volterra series and Neural Network fit statistics for predicted NEDC

4-Discussion

From the Volterra modelling results the obvious difference between emissions species is the requirement for delay term in the CO_2 model. The NOx model sees no significant increase from the inclusion of these terms however the fit R^2 increases from 0.82 to 0.93 for the inclusion of a 0.9s delay on input terms. This suggests that the NOx emissions are primarily a static event whereas CO_2 is more dependent on the time history of engine operation. A similar trend is seen for the neural network fit statistics. The Volterra series offers an advantage for this analysis as the explicit mathematical formula can be analysed which is not possible with the black box neural network approach.

Although the neural network performs better than the Volterra series in terms of fit statistics, when observing the detailed simulation predictions over the NEDC the results appear somewhat different. The model prediction and measured validation for a portion of the NEDC is shown in figure 5 for both models. In this representation, the Volterra series appears to have more accurate prediction of engine behaviour, notably during the steady state periods. The observed differences

in the validation fit statistics may result from the less frequent but larger errors in the Volterra model. In contrast, the neural network gives more persistence but smaller errors.



(a) *(b)*

Figure 5: (a) Volterra and (b) neural network NEDC prediction

Another key difference between the two modelling approaches is the stability of the parameterisation. The Volterra series is predominantly based on least squares regression and therefore gives an identical model if the process is repeated: the limited reliance on optimisation for the autoregressive aspects of the function is well controlled. In contrast, the neural network identification is based on an initial random model parameterisation that is subsequently adjusted using optimisation to minimise mean square error. The resulting models can be significantly different depending on the initial parameterisation. For each of the neural network configurations, a number of repeat identifications were performed (100 repeats for static networks; 50 repeats for dynamic networks) and the variation in fit normalised RMSE is presented in figure 6. It is obvious from these graphs that the static neural network is most stable giving a tight range of fit qualities. The inclusion of dynamic terms can improve the model fit quality, but at the expense of more variation and larger risk of a lower quality model.

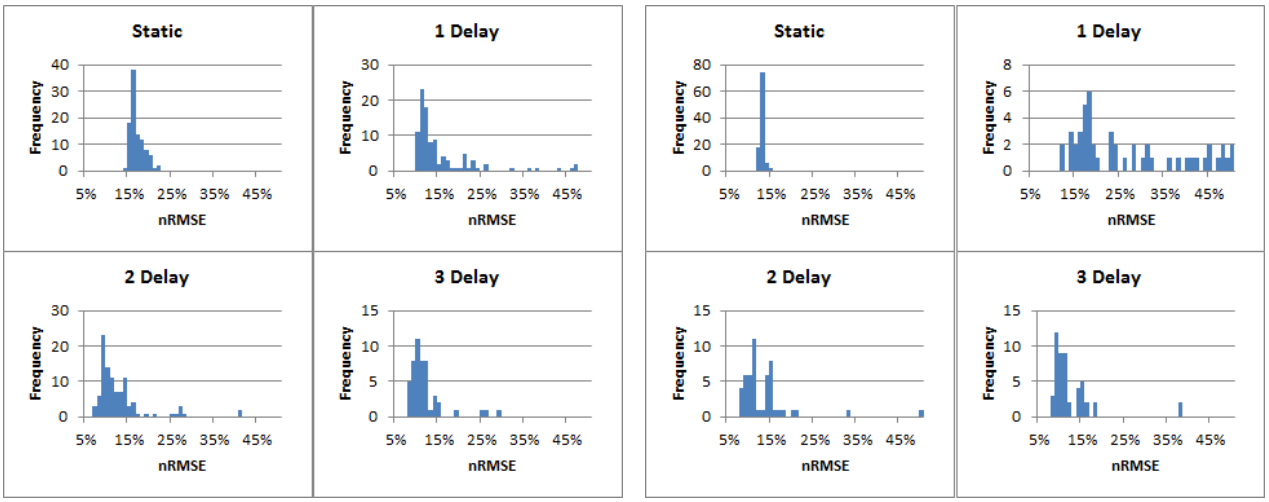
5-Conclusions

The major conclusions from this work are listed as follows:

- Dynamic model structures are required for accurate representation of transient data sets. For NOx emissions the dynamics do not seem significantly dependent on the historical states of input actuators however this is not the case for CO2 emissions.
- Fit statistics for neural networks and Volterra series suggested higher performance of the neural network models in predicting NEDC performance. However, visualisation of the

model fits suggests that the Volterra series can provide high fidelity, but statistics are skewed by a small number of larger deviations causing large errors. The neural networks appeared more flexible in fitting, but appeared to suffer from lower predictive accuracy suggesting a higher degree of over fitting.

- The variation of model fit quality of neural networks was worse for dynamic models than static models although a higher quality of fit could be achieved. This results in a more time consuming model parameterisation procedure as a larger number of repeated training routines are required to ensure a high quality model is achieved. Volterra series models are based on least squares and do not suffer from this issue.



(a) *(b)*

Figure 6: Variation in Normalised RMSE for static and dynamic neural networks for (a) NOx and (b) CO_2 emissions

References:

1. Ropke, K., Nessler, A., Haukap, C., Baumann, W., Kohler, B-U. and Schaum, S., 2009, Model-based Methods for Engine Calibration - Quo Vadis, in 3. Internationales Symposium fur Entwicklungsmethodik. Herausforderungen im Spannungsfeld neuer Antriebskonzepte, Kurhaus Wisebaden, Germany
2. Baumann, W., Klug, K., Kohler, B-U. and Ropke, K., 2009, Modelling of Transient Diesel Engine Emissions, in Design of Experiments (DoE) in Engine Development, K. Ropke, Editor. 2009, expert verlag, Renningen: Berlin, Germany. p.41-53

3. Guhmann, C. and J.-M. Riedel, 2011, Comparaison of Identification Methods for Nonlinear Dynamic Systems, in 6th Design of Experiments (DoE) in Engine Development, K. Ropke, Editor., expert verlag, Renningen: Berlin, Germany. P.41-53.
4. Baumann, W., Schaum, S., Roepke, K. and Knaak, M., 2008, Excitation Signals for Nonlinear Dynamic Modelling of Combustion Engines, in Proceedings of the 17th World Congress The international Federation of Automatic control, Seoul, Korea
5. Nelles, O., 2001, Non Linear System Identification. Springer Verlag, Berlin, Germany
6. Bannister, C.D., Hawley , J.G., Brace, C.J., Cox, A., Ketcher, D. and Stark, R., 2004, Further Investigations on Time Alignment, SAE Paper Number 2004-01-1441

Notation:

Abbreviations

CO2	Carbon Dioxide	NOx	Oxides of nitrogen
DoE	Design of Experiments	nRMSE	Normalised RMSE
ECU	Engine Control Unit	R^2	Coefficient of Determination
EGR	Exhaust gas Recirculation	RMSE	Residual Mean Square Error
NEDC	New European Drive Cycle	SNR	Signal to Noise Ratio

Variables

a	Output Vector	x	Model input
b	Neuron bias	y	Measured output
p	Input Vector	$\bar{y}$	Mean measured output
t	Time	$\hat{y}$	Model Output
u	Modelling variable	Θ	Polynomial coefficient
w	Neuron weight		