2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

Diesel Engine Emissions Prediction Using Parallel Neural Networks

Bastian Maaß Prof. Richard Stobart Dr. Jiamei Deng
R.K.Stobart@lboro.ac.uk
Department of Aeronautical and Automotive Engineering
Loughborough University, LE11 3TU, UK

Abstract— **Emission legislation has forced the pace of development of engine management functions. Legislation that will be applied to diesel engines during the period 2010-2020 continue to put great emphasis on both nitrogen oxides NO_x and particulate matter (PM). With the increasing effort to reduce emissions and maintain fuel economy manufacturers are focussing on engine control. Engine control requires data acquisition and acquisition requires sensors, but hardware in the form of sensors adds further cost to the production. As a result, so called virtual sensors are introduced. These are estimators that predict the required data, which is costly to measure or simply incapable of measurement.**

In this paper a parallel neural network structure is built. It consists of three Non-linear autoregressive exogenous input (NLARX) neural network models used to predict the smoke emissions of a diesel engine operated in a Non-Road-Transient Cycle. Existing resources from Matlab toolboxes are used in order to monitor both the cost and computational expenses of analysis. The data is re-ordered into training and validation sets and processed. To overcome the weakness of the neural network approach in respect of high frequency signals, the data is divided into layers to split up the frequencies and cut high amplitudes. Three horizontal layers of the signal are processed in parallel through independent NLARX-models and their performances are added to give an overall result.

Keywords: Diesel Engine, Emission, Smoke, Neural Networks, NLARX, Modelling

I. INTRODUCTION

In all parts of the world, emission standards for diesel engines are becoming increasingly stringent. In particular, the new legislation is focussed on severe reduction in nitrogen oxides NO_x and particulate matter (PM) [1].

The new goals require a more comprehensive understanding of input parameters to the combustion process. However, monitoring and controlling of these parameters brings growing complexity and increasing costs [2]. Costs and computational demand are especially added by sensor systems.

Instead of installing additional hardware and increasing the delay of control systems by measuring and processing data, estimators can replace the actual sensor. This approach is called virtual sensing where an estimator predicts costly or immeasurable data from available sensor signals. It is used in different areas such as engine and emission monitoring or vehicle dynamics. So called observers are more based on physical equations or data-maps such as those used in the work of Stéphant et al. [3] in a vehicle simulator. The modern term virtual sensors is marked by their less required understanding of physical processes. Black-Box modelling is often used and enables to manage complex and uncertain physics. Prokhorov [4] describes the use of neural networks for different automotive applications such as misfire detection, torque monitoring or tyre pressure change detection. In the work of Hanzevack et al. [5] a neural network is chosen in on-board diagnostics and emission estimations for a SI engine. Atkinson et al. [6] describe an approach to use a neural network for the prediction of smoke in diesel engines.

Neural Networks not only satisfy with their ability to manage system's complexity, but also with their capability of real-time calculations and adaptiveness toward new situations. This enables them to improve control strategies by decreasing delay time. In addition, they can be used for fault detection and the consequent error recovering.

Artificial neural networks are used in different areas of automotive applications. For the case of real-time estimation they are mainly built for control design purposes. The area of engine development shows neural network models for common control problems for instance such as fuel injection, injection timing, output power and speed [7], [8]. Also new control features like variable turbine geometry (VTG), exhaust gas recirculation (EGR) or variable valve timing (VVT) are in the focus of artificial network modelling [9]. Combustion and emission modelling promises to be another satisfying area since these are the fields of focus. Prediction of single combustion parameters as well as considering the whole combustion process have been tried by different authors [10], [11]. Parameters such as the Air-to-Fuel Ratio (AFR) or pressure and temperature that are known for their direct and big impact on combustion quality are estimated by neural network models from Potenza et al. [10]. He et al. [11] built a model that considers several engine parameters such as boost pressure and EGR and it generates several outputs amongst other things also soot emissions.

In this paper three parallel neural networks based on a non-linear autoregressive exogenous input (NLARX) model are used. They are trained and validated to predict the smoke output of a heavy duty diesel engine. The smoke is assumed as a good indicator for the description of particulate matter. The engine is operated in the Non-Road-Transient Cycle (NRTC) [12], shown in figure 2, as it is used for standards certification in the EU and USA. An easily accessible model of the neural network toolbox in Matlab is implemented in order to minimise costs. Three data layers are build from the training and validation set respectively. These three layers are the base for the parallel structure of three neural networks. This new approach tries to cope with the weakness of each of the network's slow response to high frequency patches in the signal.

In Section 2 a short introduction and discussion toward soot formation and the correlation with the NRTC signal is given. Section 3 explains the implemented model architecture. The experiment, the recorded signals and their processing is described in Section 4. Section 5 deals with the newly implemented approach of dividing the signals into three horizontal layers, which is followed by the presentation of the estimation results, discussion and conclusion.

II. Smoke Signal Response to NRTC Signal

Despite the fact that the approach of black-box modelling obscures physical processes it is necessary to identify the influencing parameters. Initial parameters that can be investigated are the torque and speed signals of the NRTC. Closer examination of the cycle characteristics and emission signal have brought about the assumption that the cycle curve has an observable impact on the smoke signal characteristics. In this paper the smoke signal refers as an indicator for the PM or soot amount.

The NRTC speed curve at the top of figure 2 has a fast oscillating curve in the first half of the cycle. The characteristics of the smoke signal in figure 3 show a visible correspondence. A slower rise of the speed at about 100 s with a following flat part till about 150 s lead to a flat part in the smoke output signal. Comparable behaviour can be determined in the second half of the curve where the NRTC shows a fairly slow change in speed ranges. The emission signal is marked by much less oscillations. Hence, it can be assumed that fluctuations in the curve are a response to a fast speed change during the cycle.

The behaviour identified can be traced back to the fact that with a rapid change of speed the combustion conditions also change. Soot formation is regulated by a number of different parameters, which are indirectly influenced by the change of loading conditions of the engine. On the one hand the amount of oxygen that is available for building organic compounds by oxidation reactions is critical. On the other hand the formation of the spray is crucial. High injection pressures ensure that a sufficient atomisation of the fuel can take place because smaller droplets are less likely to build soot formations. A third feature is a high combustion temperature that leads to complete combustion and less in-cylinder soot formation by breaking up fuel droplets through oxidisation [13].

Taking these thoughts into account the smoke signal (Figure 3) can be explained as follows. The first half of the smoke signal is a result of rapid changes in engine speeds. After such a change the engine stays in a transient condition for a period of time. During this phase the amount of oxygen that flows into the cylinder settles towards a steady state. Whereas the fuel injected may rise due to a load increase. This initial excessive fuel may coincide with a reduction in oxygen flow and consequently soot is more likely to be formed. In addition, the duration of combustion is dependent on the amount of oxygen present and the engine speed. Consequently, a shorter period of combustion with a decrease in required oxygen can lead to incomplete combustion. The second half of the signal is dominated by steady speed resulting in a flat output signal. Small visible peaks breaking this signal are due to the sporadic fast speed changes. Due to fairly high speeds around 80% of the rated speed, the temperature can rise and kept at a high level. Hence, the conditions during the combustion process are more likely to break up the fuel droplets. At the same time less fluctuations mean less transient states with varying conditions favouring the soot generation.

Other parameters such as fuel injection timing and duration or in-cylinder pressures and temperatures have also an impact on observed engine behaviour. In this case however, it was related principally to the controlled parameters of the NRTC test cycle.

III. NLARX Model

Neural networks can be split into the following three categories:

- Single-Layer Feedforward Networks (SLFN)
- Multi-layer Feedforward Networks (MLFN)
- Recurrent Neural Network (RNN)

In this paper a NLARX model is applied. It represents a recurrent neural network (RNN), which fits the purpose of non-linearity of the problem. This network is chosen for its simplicity and the availability in the Matlab neural network toolbox.

The NLARX model is an Input-Output recurrent model and can be implemented with multiple inputs and outputs [14]. Its feature of recurrence enables the model to take into account precedent states and hence model dynamical behavior.

The present multiple input and single output structure of the NLARX model is illustrated in figure 1. The inputs are represented by $u_k(t)$ and the output is described by $y'(t)$. The dynamical behaviour of this NLARX model can be described by the formulation:

$$y(t+1) = F(y(t), ..., y(t-p+1), u_k(t-m_k), ..., u_k(t-q_k+1)) \quad (1)$$

Where $y(t)$ describes the output and $u_k(t)$ the input k at time-step t. Consequently, $k = 1, ..., 12$ for the presented model that has twelve inputs. The Matlab toolbox enables to manipulate the recurrent output and inputs that are fed back with the delay steps p and q_k respectively. The parameter m_k and q_k define the inputs that are used to predict the next time step. The multilayer perceptron part of the NLARX model is realised through a binary-tree structure that estimates the non-linear behaviour. Its training is known as a nonlinear unconstrained optimisation problem of the form:

$$\min_{\Theta} F_M(\Theta, Z_M) = \frac{1}{2M} \sum_{t=1}^{M} \|y(t) - y'(t|\Theta)\|^2 \quad (2)$$

where $Z_M = [y(t), u_k(t)]_{t=1,...,M}$ is a data set that is split into training and validation parts. The $y(t)$ and $u_k(t)$ represent the measured outputs whereas $y'(t\|\Theta)$ is the generated output at time step t from the NLARX model that is dependent the weight vector Θ. This optimisation problem minimises the averaged distance between the NLARX's outputs $y'(t|\Theta)$ and the target values of the training samples $y(t)$. In this paper, this training is realised by using the MATLAB Neural Network Toolbox.

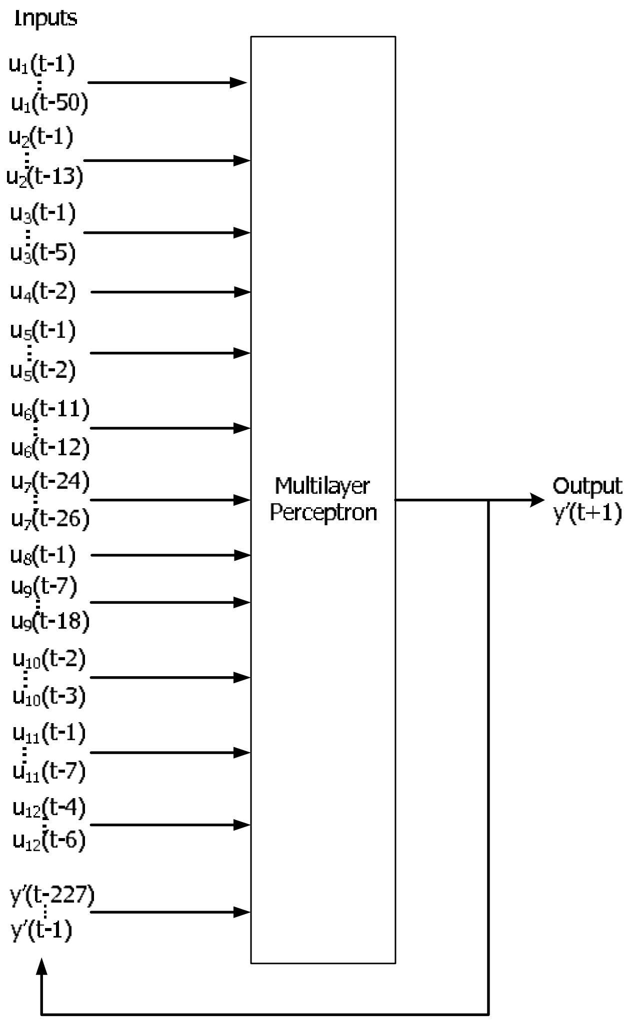

Fig. 1. General NLARX - model [14]

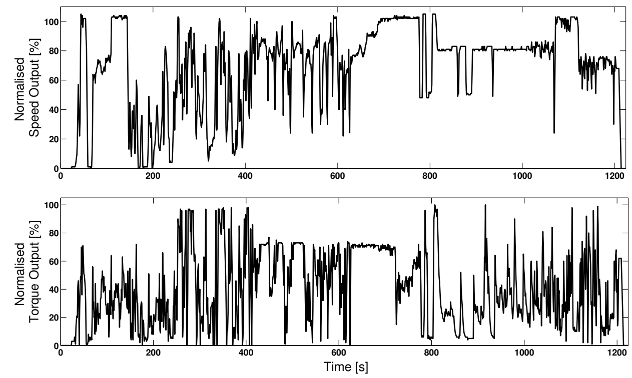

Fig. 2. Non-Road Transient Cycle [12]

parameters were used from the beginning and investigated and revised for their impact on the model.

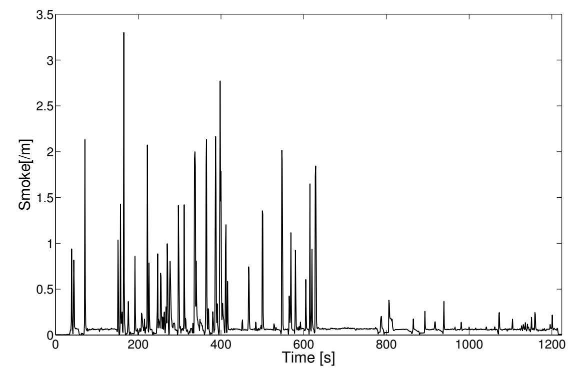

Fig. 3. Smoke output signal

IV. Experiment

The experiment was conducted on a heavy duty diesel engine in the Perkins powertrain labs that is run under the conditions of a Non-Road-Transient cycle (NRTC). This cycle is applied to test and certificate non-road engines for EPA- and EU-standards. It represents common scenarios of load and speed changes for an engine. It is applied to generate emission values for comparison with specified values of the regulations. The graph in figure 2 shows the normalised engine speed and torque over the 1200s test period. The resulting smoke signal can be seen in figure 3.

A. Data Set

Different engine parameters are sampled at 1Hz under the NRTC conditions. The whole data set consists of 12 inputs and 5 outputs. Among NO_x, HC CO and CO_2 in this paper the focus lies in the smoke output. Its signal is predicted on the basis of 12 inputs such as torque, boost pressure, engine speed, liquid pilot fuel quantity, final fuel injection, back pressure, intake manifold temperature, exhaust temperature, intake depression and coolant temperatures in and out. All parameters were used from the beginning and investigated and revised for their impact on the model.

B. Data Partitioning

The initial output signal shows two characteristic halves. In the first half strong fluctuations and high peaks are typical, whereas the second half displays a much flatter characteristic with small oscillations. The approach of modelling and estimating the signal requires a training and validation data set. Therefore the signal is bisected. However, a training set requires preferably a broad spectrum of features provided by the signal. The signal is first divided into quarters accordingly and then newly-arranged. As a result the training and validation set cover a high oscillating part with high peaks and a flat, low oscillating part - figure 4. Every set consists of a correspondingly split smoke output and twelve inputs. As well as the data partitioning a normalisation process is applied to the inputs and output.

V. Approach of Data Filtering

The NLARX model is known for its weakness towards signals with high-frequency components. In an initial approach of modelling the signal with an single NLARX network it was recognised that noise is introduced by the model. This occurs especially then, if the signal contains large amplitudes

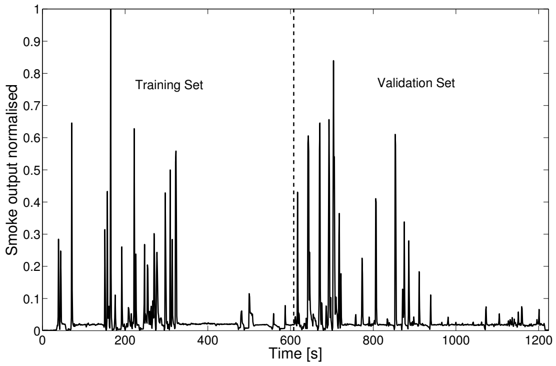

Fig. 4. Processed smoke output signal

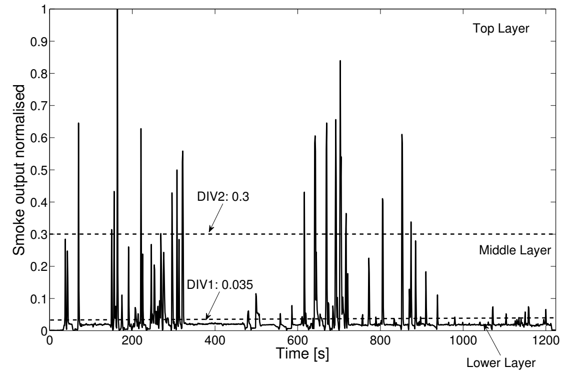

Fig. 6. Layer approach with correlating divisions

and high-frequencies. In Figure 5 the modelling results of a single NLARX model are plotted over the measured signal. The early phase of the signal is well predicted. However, in the second phase of the characteristic the prediction data starts oscillating in high-frequencies as well as an underlying lower frequency. The model becomes unstable. This is assumed to be forced by the training on high amplitudes in the first stage and hence the development of a hypersensitive behaviour. Other approaches are known to overcome those issues such as fuzzy logic and wavelet networks [15]. They offer a much better response to highly fluctuating signals.

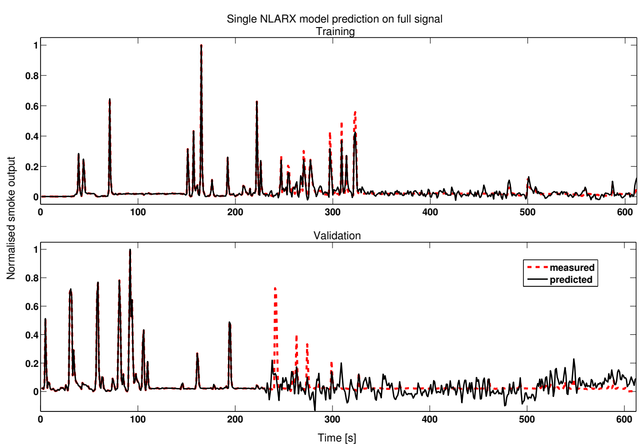

Fig. 5. Single NLARX model: Measured output signal correlated versus predicted output signal

Among those approaches, Guoyin et al. [16] introduced three classes of parallel network systems. Here, a parallel network system with multiple tasks is chosen. Lee [17] states that due to the approach of more than one network the risk to get stuck in a local minimum decreases. Additionally, the performance increases due to the fact that particular networks handle a specific subspace instead of dealing with the whole problem.

In the current work the signal is divided into different vertical layers. Consequently the amplitudes are cut and the frequency of the residual signal part is decreased. With lower frequencies the NLARX model promises satisfying results regarding performance and cost.

By trial and error three layers are determined as a reasonable amount of divisions. The first layer called lower layer (LL) contains the signal noise and low frequencies. The remaining part is split into a mid layer (ML) and a top layer (TL). The ML covers a part of the signal with a medium density of oscillations and peaks of a smoke value up to y=0.3. The residual signal peaks are covered by the TL. Its characteristic is marked by only a few single peaks, the occurrence of which is not distracted by noise or smaller peaks. The division borders in this approach are chosen as outlined in table I and illustrated in figure 6.

TABLE I
DIVISION BORDERS OF THE APPROACH

$0 < \text{LL} < 0.035$	$\Rightarrow$	$\Delta y_{LL} = 0.035$
$0.035 < \text{ML} < 0.3$	$\Rightarrow$	$\Delta y_{ML} = 0.265$
$0.3 < \text{TL} < 1$	$\Rightarrow$	$\Delta y_{TL} = 0.7$

Each division is processed and estimated independently. This leads to a parallel processing model structure as shown in figure 7. The input vector U with its twelve input signals is used for all three independent layers whereas the predicted output is split into the three divisions, *top*, *mid* and *low*. After estimation they are combined to $y'_{overall}$ and compared against the overall measured output.

VI. Estimation Results

An estimation with an artificial neural network (ANN) is processed by initially training and then validating it with the corresponding signals. Every layer is estimated independently. The NLARX-model are initialised with an arbitrarily state and taught with the corresponding training data set. Based on this data the NLARX-model is designed to estimate the desired output signal. The designing process consists of changing the design parameters in Matlab by teacher-forced learning until a satisfactory result is achieved. Design parameters are the input/output delays.

A. Estimation Results for the Lower Layer

The lower part is marked by (1) the lowest values of the higher oscillations of the signal and (2) small oscillations

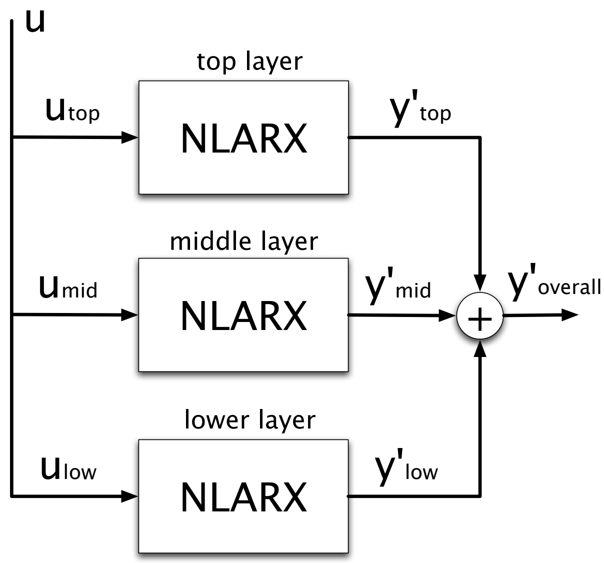

Fig. 7. Scheme of applied model structure

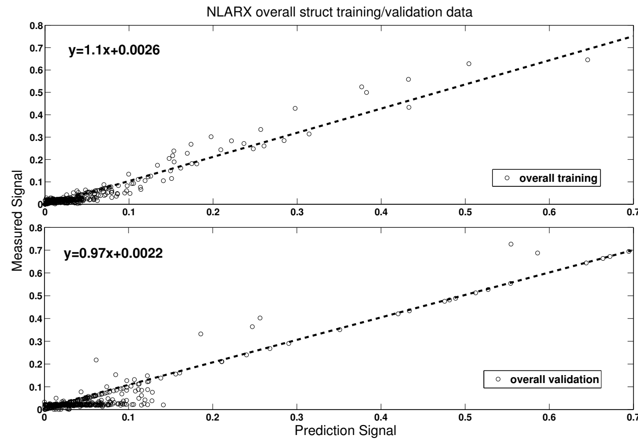

Fig. 8. Overall Performance: Training and validation sets estimation in correlation to measured data

TABLE II

USING RMS AS PERFORMANCE INDICATOR OF MODEL OUTPUT

Layer	NARX-Model	
	train	valid
TL	0.8330	0.9163
ML	0.9328	0.9006
LL	0.9743	0.9164
Total	0.9706	0.9616

that are introduced by noise. By cutting off a lower part of the signal a more homogeneous distribution of the height of oscillations is generated. This enables a better estimation with the chosen NLARX approach.

The training of the network generates a correlation between the measured and estimated signal of $R^2 = 97\%$. The coefficient of determination R^2 is expressed by equation 3,

$$R^2 = 1 - \frac{\sum_{i=1}^{t} (y_i - y_i')^2}{\sum_{i=1}^{t} (y_i - \bar{y}_i)^2}, \tag{3}$$

where y_i describes the measured data and y_i' the prediction respectively.

Validating the network leads to a performance of $R^2 = 92\%$, which demonstrates the practicability of the chosen design. However, the model introduces additional noise to the signal. This effect is discussed in more detail in the following sections.

B. Estimation Results for the Middle Layer

This middle layer represents the central section of the high peaks and the medium peaks. The lowest values of the large signal excursions are included in the lower layer. Through training the NLARX model achieves a correlation of $R^2 = 93\%$ with the measured signal. The model's quality is confirmed by the validation set, which achieves a performance of $R^2 = 90\%$. The performance is predictably lower than in the first layer due to the higher frequencies. Higher frequencies occur because of an expanded range of y-values.

The characteristic of the graph is marked by noise in the second, low oscillating part of the signal. It is assumed that this noise is introduced as a result of the network design. There is a fast response identified by the network when managing high oscillating signals. In consequence, this leads to an oscillating estimation signal.

C. Estimation Results for the Top Layer

The top layer covers the high peaks of the signal. Consequently high frequencies are introduced and a lower correlation performance is expected. The design process achieves a result of $R^2 = 83\%$ compared to a $R^2 = 92\%$ for the validation data. Validation shows a better result because the main peaks of the validation signal are predicted well, whereas the training signal shows some missing details in the middle part for three consecutive spikes.

D. Overall Estimation

The overall estimation is achieved by adding the three estimated signals together and correlating it with the measured output. The comparison of the measured and predicted signal shows a distribution around the linear correlation in figure 8. The reason that a cluster of points forms close to the origin is due to the fact that the most of the data samples are measured in the lower data scope. However, the results of an overall correlations of the smoke output signal are about $R^2 = 97\%$ and $R^2 = 96\%$ for training and validation set respectively as illustrated in figure 9. It can be seen that parts intially classified as difficult due to big amplitude differences and high frequencies are modelled well. The patterns of high peaks and high density of oscillations show appropriate correlations. However, the flatter parts are marked by the introduction of noise through the model design as mentioned earlier in connection with the single NLARX attempt.

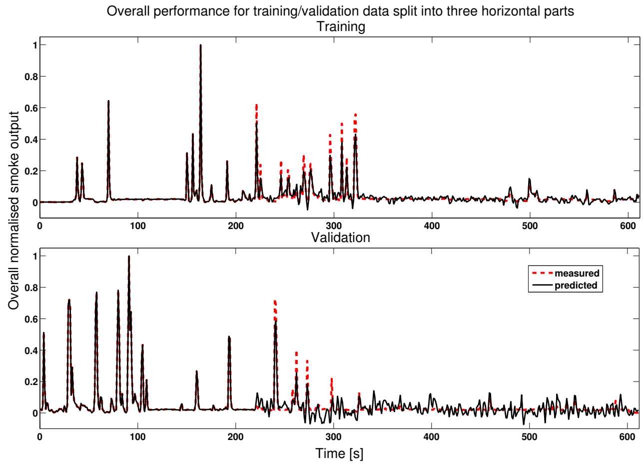

Fig. 9. Overall Performance: Training and validation sets estimation of measured and predicted signal

VII. Conclusion

This paper uses a parallel neural network structure to predict the smoke output of a diesel engine based on NLARX-models. The model is chosen due to its good generalisation. Its weakness of not being capable of high-frequency signals which is shown with a single NLARX model for comparison, is overcome by a new approach of frequency filtering.

The NLARX architecture is cut into different layers to reduce the frequency bands so that it can cope with the present signal. A lower layer for the signal scope that covers noise and the base of higher peaks, followed by a middle layer for medium density of oscillations and a top layer for the peak tops. Through this approach the frequencies are cut into smaller bits the model can handle.

A parallel model structure is built, based on this approach, which is used to process and estimate the layers independently. It generates satisfactory results for the available data. The overall performance lies at 97% and 96% for training and validation data, respectively. It proves to be a well designed model which is developed with a convenient and simple approach.

The performance could probably be improved by introducing additional layers and decompose frequencies further. In addition, the practicability of the approach needs to be tested on further data sets. A further 10Hz smoke data set is tested and suggests similar good results. It also shows the generalisation capability by using the developed model on 26 different calibrated engine data runs. Indeed, the performance decreases but settles at an acceptable level.

Acknowledgement

The reported work has been conducted during a project that is co-funded by the U.K. Technology Strategy Board's Collaborative Research and Development programme, following an open competition (Grant Reference,TP/3/DSM/6/I/152). The Technology Strategy Board is an executive body established by the U.K Government to drive innovation. It promotes and invests in research, development and the exploitation of science, technology and new ideas for the benefit of business - increasing sustainable economic growth in the UK and improving quality of life. The authors wish to thank the Technology Strategy Board and its representative, Mr Ian Massey for support throughout the project.

References

[1] E. Parliament and C. of the European Union, "Regulation (ec) no 715/2007," Internet, June 2007.

[2] W. Knecht, "Diesel engine development in view of reduced emission standards," *Energy*, vol. 33, no. 2, pp. 264–271, February 2008.

[3] J. Stéphant, A. Charara, and D. Meizel, "Virtual sensors: Application to vehicle sideslip angle and transversal forces," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 2, pp. 278–289, 2004.

[4] D. Prokhorov, "Virtual sensors and their automotive applications," *ISSNIP*, pp. 411–416, 2005.

[5] E. L. Hanzevack, T. W. Long, C. M. Atkinson, and M. L. Traver, "Virtual sensors for spark ignition engines using neural networks," *Proceedings of the American Control Conference*, pp. 669–673, June 1997.

[6] C. Atkinson, M. Traver, T. Long, and E. Hanzevack, "Predicting smoke," Internet, 2002.

[7] M. Hafner, M. Schüler, O. Nelles, and R. Isermann, "Fast neural networks for diesel engine control design," *Control Engineering Practice 8*, vol. 8, pp. 1211–1221, 2000.

[8] M. Ouladsine, G. Bloch, and X. Dovifaaz, "Neural modelling and control of a diesel engine with pollution constraints," *J. of Intel. and Robot. Sys.*, vol. 41, pp. 151–171, 2004.

[9] G. Thompson, C. Atkinson, N. Clark, T. Long, and E. Hanzevack, "Neural network modelling of the emissions and performance of a heavy-duty diesel engine," *Proc Instn Mech Engrs*, 2000.

[10] R. Potenza, J. F. Dunne, S. Vulli, D. Richardson, and P. King, "Multicylinder engine pressure reconstruction using narx neural networks and crank kinematics," *Int. J. Engine Res.*, vol. 8, pp. 499–518, 2007.

[11] Y. He and C. J. Rutland, "Application of artificial neural networks in engine modelling," *Int. J. Engine Res.*, vol. 5, no. 4, pp. 281–296, 2004.

[12] Dieselnet, "Nrtc data set," Internet, June 2007.

[13] J. Heywood, *Internal Combustion Engine Fundamentals*, 2nd ed. McGraw-Hill International Editions, 1988.

[14] S. Haykin, *Neural Networks - A Comprehensive Foundation*, 2nd ed. Prentice Hall, Inc., 1999.

[15] K. Parasuraman and A. Elshorbagy, "Wavelet networks: An alternative to classical neural networks," *Proc. of Intern. Joint Conf. NN, IEEE*, pp. 2674–2679, August 2005.

[16] W. Guoyin and S. Hongbao, "Parallel neural network architectures," *IEEE Proc. of Intern. Joint Conf. NN*, pp. 1234–1239, 1995.

[17] B. Lee, "Parallel neural networks for speech recognition," *IEEE, Proc. of Intern. Joint Conf. NN*, pp. 2093–2097, 1997.