SAE INTERNATIONAL

2017-01-0601

Published 03/28/2017

Emissions Modeling of a Light-Duty Diesel Engine for Model-Based Control Design Using Multi-Layer Perceptron Neural Networks

Huayi Li, Kenneth Butts, and Kevin Zaseck
Toyota Motor North America

Dominic Liao-McPherson and Ilya Kolmanovsky
University of Michigan

CITATION: Li, H., Butts, K., Zaseck, K., Liao-McPherson, D. et al., "Emissions Modeling of a Light-Duty Diesel Engine for Model-Based Control Design Using Multi-Layer Perceptron Neural Networks," SAE Technical Paper 2017-01-0601, 2017, doi:10.4271/2017-01-0601.

Abstract

The development of advanced model-based engine control strategies, such as economic model predictive control (eMPC) for diesel engine fuel economy and emission optimization, requires accurate and low-complexity models for controller design validation. This paper presents the NOx and smoke emissions modeling of a light duty diesel engine equipped with a variable geometry turbocharger (VGT) and a high pressure exhaust gas recirculation (EGR) system. Such emission models can be integrated with an existing air path model into a complete engine mean value model (MVM), which can predict engine behavior at different operating conditions for controller design and validation before physical engine tests. The NOx and smoke emission models adopt an artificial neural network (ANN) approach with Multi-Layer Perceptron (MLP) architectures. The networks are trained and validated using experimental data collected from engine bench tests. Model inputs (including input delays) are selected based on physics-based analyses supplemented with data-driven cross-covariance studies. Special care is taken during the training process to avoid overfitting and ensure strong generalization performance. Various neural network architectures, including static networks, dynamic networks, and classifiers, are compared in terms of model complexity and accuracy. Simulation results indicate that MLP networks are capable of capturing the highly nonlinear engine NOx and smoke emissions at both steady state and transient conditions.

Introduction

The need to meet increasingly stringent emissions regulations and simultaneously increasing vehicle fuel economy standards are driving the development of more advanced and complex engine systems [1]. As a result, engine control problems become substantially more challenging. During the preliminary controller design phase, accurate computer models are needed to efficiently characterize engine behaviors, verify control methodologies, and facilitate controller tuning before physical engine testing. Motivated by the development of economic model predictive control (eMPC) strategies to seek the direct optimization of fuel economy and engine-out emissions for a light-duty diesel engine, control-oriented modeling of NOx and smoke emissions for off-line validation has been investigated.

Various emissions modeling methods have been proposed in the literature. Map-based quasi-steady models are often used due to their low complexity and good steady-state accuracy. However, real-world engines rarely operate at steady state, and quasi-steady models may fail to accurately predict transient effects such as smoke spikes during tip-in conditions [2]. Grey-box models are interpretable and have the ability to capture both steady state and transient effects. However, the structure of the physical equations can introduce bias in the model, limiting the performance. In [3], physics-based grey-box NOx models were developed based on the extended Zeldovich mechanism described in [4]. The Zeldovich model is a zero-dimensional model that considers kinetics of NOx formation, so it is able to capture transient effects while retaining relatively fast calculation speed. The drawback is that it requires in-cylinder temperature (or pressure) measurements, which are rarely available on production engines, or an accurate combustion model with low computational cost, which is challenging to create.

An alternative to grey or white box models are data-driven statistical machine learning based methods, which have been proposed in many publications for not only emissions modeling, but also engine control in general [5]. In [6] a multiple-zone polynomial model was developed to model NOx emissions. More sophisticated methods, such as Volterra series in [7], have been applied to model transient

soot emissions. Furthermore, a significant number of studies on virtual emissions sensing technologies have adopted Artificial Neural Network (ANN) based approaches [8][9][10]. ANNs are able to represent very general non-linear functions but must be carefully trained to prevent overfitting. The complexity and accuracy of these black-box models are highly dependent on the nature of the model architecture as well as methods used to train the parameters.

This paper presents the development of NOx and smoke models using Multi-Layer Perceptron (MLP) neural networks to support and validate the eMPC design of a turbocharged light-duty diesel engine. Instead of directly using all available measurements, domain knowledge is used to generate features which, when used as predictors, reduce the size of the network needed to achieve the required accuracy. In smoke models, for example, neural networks have been combined with a physics based gas composition model where effective features of several signals are extracted to become a single neural network input. This approach reduces the cost of neural network training and potentially improves the model accuracy. For time delay neural networks, a cross-covariance based approach is used for feature selection, considering discrete delays as features. In addition, for smoke modeling, the classification method is proposed and compared with the commonly used regression approach, as the classification result meets the need as a constrain for the controller, in addition that the classifier is easier to be retrained with big data for better model accuracy. While the results are specific to the aforementioned application, the detailed discussions in this paper on suitable neural network architectures, input selection, and identification procedures are relevant to and can provide guidance for a large number of potential applications.

The paper is organized as follows. First, the MLP neural networks are introduced, followed by the discussion of data preparation for network parameter training, including experimental setup and inputs selection. Then in the emissions modeling section, NOx and smoke models are presented explicitly, and different neural network architectures are compared based on model accuracy and complexity. Finally, the prediction quality of emission models is validated by running drive cycle simulations, and conclusions are drawn in the last section.

Multi-Layer Perceptron Neural Network

Architectures

To begin the discussion of MLP neural networks, Figure 1 shows a Single-Layer Perceptron (SLP) neural network where x and y represent network inputs and output, w is the weight of each input, and θ is the bias of the neuron [11]. Each neuron has a function $\phi(\cdot)$ called the activation function. This network output can be written as

$$y = \phi(\boldsymbol{w}^T\boldsymbol{x} - \theta) \tag{1}$$

By using a nonlinear function as the activation function, such as logistic and hyperbolic tangent sigmoid functions, the network is able to capture the nonlinearity of the system.

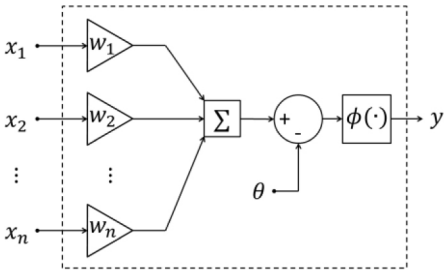

Figure 1. Single-Layer Perceptron neural network architecture.

However, for more difficult problems, a single activation function may not have sufficient representation ability to capture the input-to-output relations. Furthermore, from the perspective of engine emissions modeling, the availability of input signals that have direct impact on emissions may be limited. In this case, a deeper neural network with multiple layers can provide a better approximation of the underlying process. Figure 2 shows the architecture of a four-layer MLP network. Each grey circle corresponds to a similar operation as the one inside the dashed box in Figure 1. These four layers include an input layer, two hidden layers, and an output layer. The outputs of the hidden layers, such as $a_j^{(2)}$ and $a_k^{(3)}$ (where j and k are respectively the neuron number of the first and second hidden layer), are called hidden features, because they represent a transformation of the inputs into a form that can more directly correlate the output. In addition, the increased number of activation functions between inputs and the output enables the network to better capture the nonlinear behavior of the system.

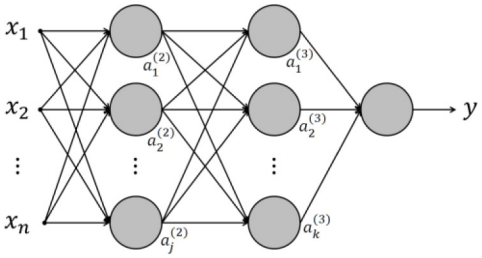

Figure 2. Multi-Layer Perceptron neural network architecture.

There are several extensions of the basic MLP that can be used to model engine emissions with higher accuracy. The network shown in Figure 2 is a static network because it does not have any memory at the level of inputs or outputs, while in Figure 3, both input and input delays are included in the input layer. This type of neural network, commonly called time delay neural network (TDNN), is suitable for modeling the target that depends on current and previous input values. Later in the NOx modeling section, the performance of the static model will be compared against the TDNN model. It is worth mentioning another major type of dynamic neural network, the recurrent neural network (RNN), which has output and output delays fed back as inputs. RNNs can provide a very good prediction accuracy on the training dataset but can become unstable during validation because it depends on the prediction from previous step(s). Although there are some potential methods to solve this problem, RNNs are not investigated for emissions modeling in this paper. Last but not least, to model the highly transient smoke emissions for the purpose of eMPC development, a classification neural network (or a classifier) could be

used because it may be easier to categorize smoke as 'acceptable' or 'bad' than to develop a quantitatively accurate nonlinear regression model. In the later section on smoke modeling, both TDNN and classification neural network will be developed and compared.

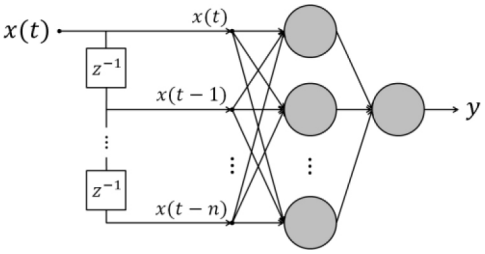

Figure 3. Time delay neural network architecture.

Neural Network Training

The selection of the training algorithm for neural network parameters is critical because it has a significant impact on model accuracy, generalization performance, and computational footprint. For supervised neural network training, backpropagation is widely used in conjunction with gradient descent (a first-order optimization method) or Newton's method (a second-order optimization method). In this paper, MATLAB Neural Network Toolbox is used and the training algorithm adopted is called Levenberg-Marquardt backpropagation. It is a relatively fast training method that requires little memory because it uses a quasi-Newton method, where the Hessian matrix is approximated instead of directly calculated [11]. The early stopping method is integrated to prevent overfitting and improve generalization performance. In this method, the training dataset is randomly divided into three groups called training, validation and test dataset. Training data is for parameters (weights and bias) training. Validation data is used to monitor the network performance during the training process. Test data is not involved in the training process and thus can be used to validate the model performance after the training is finished. During the training process, the mean square error (MSE) between predictions and output targets of the validation dataset is calculated at every epoch. Typically, this MSE will decrease at the beginning and then start to rise as the model begins to overfit the data. Once the validation MSE continues increasing for a certain number of iterations (the default is 6), the training will automatically stop.

Due to the characteristics of MLP, deeper neural networks are able to represent more sophisticated nonlinear functions. However, increasing the network depth may also increase the risk of falling into local minima during network training. On the other hand, three-layer MLPs with sigmoid activation functions has been mathematically proven to be able to approximate any sufficiently continuous multivariate function to any accuracy over a compact domain [11]. Thus, only two-hidden-layer MLPs are considered in this paper, while the number of neurons in each hidden layer varies. Consequently, the total number of network parameters will differ, which affects the model accuracy as well as the possibility of overfitting.

Engine Data Preparation and Analysis

Experimental Setup

The experimental data for model identification and validation was gathered through bench tests on a 2.8L Toyota GD engine. As shown in Figure 4, this four-cylinder diesel engine is equipped with a variable geometry turbocharger (VGT) and a high pressure exhaust gas recirculation (EGR) system, in addition to a diesel throttle which is used to adjust EGR rate at low engine speed conditions. The GD engine also contains an intercooler after the compressor and an EGR cooler before the EGR valve. High pressure common rail and multiple injection strategy are used by the fuel injection system to improve in-cylinder combustion and reduce engine-out emission.

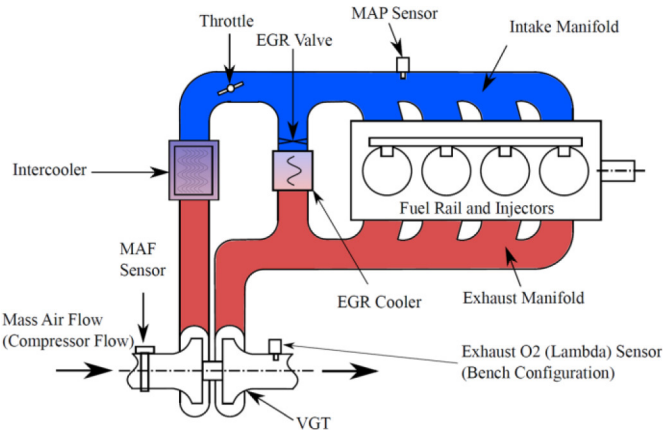

Figure 4. Toyota GD engine air path diagram

During the engine bench test, the engine was run with commands of engine speed, fuel (injection quantity, timing and pressure), and actuators positions, either open-loop or with a controller. For drive cycle simulation, in addition to the engine controller there is a driver model with simple PID control logics to govern the engine operating conditions.

The test signals measured by sensors include mass air flow (MAF), manifold air pressure (MAP), intake temperature, and oxygen concentration of the exhaust gas at the turbine outlet. NOx is measured by the gas analyzer in the unit of parts per million (ppm). The gas analyzer also measures exhaust oxygen concentration. Due to the distance between the exhaust manifold and measurement locations, transport delay and slower dynamics have been considered for both signals of exhaust oxygen concentration and the signal of NOx. For smoke emissions, exhaust gas is sampled at the turbine outlet and then delivered to an opacity meter. The measurement unit is percentage opacity. In contrast with NOx, the smoke measurement is relatively instantaneous.

To acquire sufficient data for neural network training and validation, more than 100 test hours have been spent on the bench. Both steady state and transient data has been acquired. Considering the variety of the data coverage, only a portion of the data has been selected for training. Figure 5 shows the normalized fuel vs. engine speed map with open-loop steady state data as well as transient operating points from the closed-loop pattern test and the NEDC test for network

training. The black curve indicates maximum fuel at different engine speeds. As shown, the test data is widely spread and covers most of the engine operating conditions.

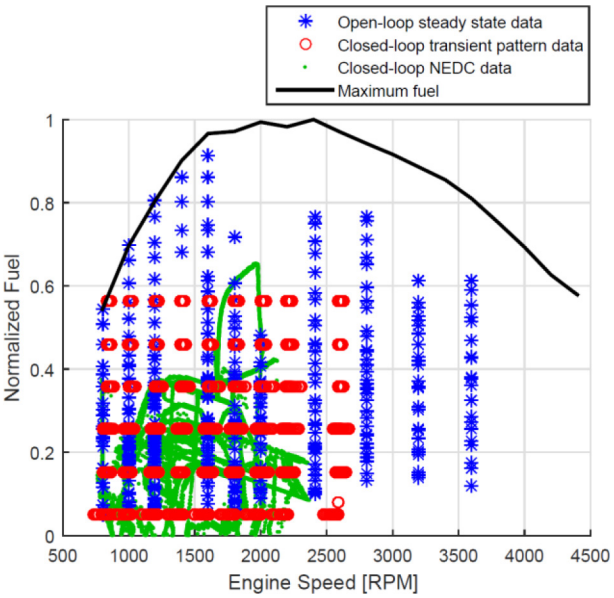

Figure 5. Engine operating conditions coverage

Model Input Selection

The selection of model inputs (including delays) has a large impact on the performance of black-box models. Direct relationships between inputs and outputs can increase the predictive power of the resulting neural networks. However, for engine emissions modeling, available measurements are very limited. As a result, the inputs to emission models were chosen carefully based on physical insight into the pollutant formation mechanisms supplemented with cross-covariance studies applied to experimental data.

During the closed-loop operation of the engine, the operating condition is decided by the engine speed and fuel injection quantity. The boost pressure and EGR rate will be controlled by varying actuators positions to track the targets from a 2-D lookup map [12]. As a result, engine speed and fuel injection quantity are two dominant factors for the GD engine-out emissions. Generally speaking, for diesel engines, EGR has a large impact on NOx and smoke formation. For NOx emissions, increasing EGR rate leads to higher heat capacity of the in-cylinder gas, which as a result reduces the flame temperature and resists the NOx formation. However, if there is excessive amount of EGR in the cylinder, it could cause incomplete combustion and thus increase formation of smoke. Since there is no EGR flow measurement in the GD engine, intake manifold gas temperature is taken as the indicator of intake gas EGR rate, because fresh air temperature is typically much lower and more stable than EGR temperature. Engine-out emissions are also highly affected by pre-combustion in-cylinder air-to-fuel ratio (AFR). MAF and MAP can be used to indicate fresh air induction quantity [13]. However, due to the effect of EGR, in-cylinder gas would contain air from both the compressor flow and the recirculated exhaust gas. In this case, oxygen concentration or air fraction of the intake manifold should be used to represent in-cylinder air. For smoke emissions, the change of fuel injection strategies, including injection timing, quantity and common rail pressure, could cause the sudden increase of pollutant formation. Thus, the history of the injection parameters should be considered as model inputs.

To visualize and quantify the relationships between different variables and emissions, Matlab function *xcov* has been used to perform the cross-covariance study. The equation to calculate the cross-covariance is given by [14]

$$c_{xy}(m) = \begin{cases} \sum_{n=0}^{N-m-1} \left(x_{n+m} - \frac{1}{N}\sum_{i=0}^{N-1} x_i \right)\left(y_n^* - \frac{1}{N}\sum_{i=0}^{N-1} y_i^* \right), & m \geq 0 \\ c_{yx}^*(-m), & m < 0 \end{cases} \quad (2)$$

where x and y represent engine variables and emissions in time series, c_{xy} is the cross-covariance between x and y, N is the length of the data array, and m is delays in timestep. To perform dimensionless comparison for different variables, the cross-covariance can be normalized by

$$r_{xy} = \frac{c_{xy}(m)}{\sqrt{c_{xx}(0) \cdot c_{yy}(0)}} \quad (3)$$

where c_{xx} and c_{yy} are the autocorrelations of engine variables and emissions.

Figure 6 and Figure 7 are example results of the cross-covariance studies. The two numbers at the top right side indicate the time lag and the maximum cross-covariance value when the maximum cross-covariance occurs. For the NEDC drive cycle dataset, fuel quantity and MAF show strong relationships with smoke emission at approximately 0.6 second time lag. This can be considered as feature selection, with the delayed signals as features.

Table 1 summarizes selected inputs for each emission neural network. Common inputs for both NOx and smoke emissions are engine speed, fuel injection quantity, MAF, MAP and intake temperature. For static NOx model and smoke models, the main injection timing and injection pressure have been included. Smoke models also have intake oxygen concentration as an input. Except for engine speed, all inputs have input delays of 0.2, 0.4, 0.6, 0.8 and 1 second.

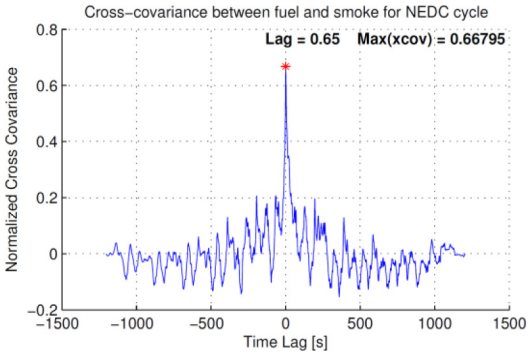

Figure 6. Cross-covariance between fuel and smoke for NEDC cycle

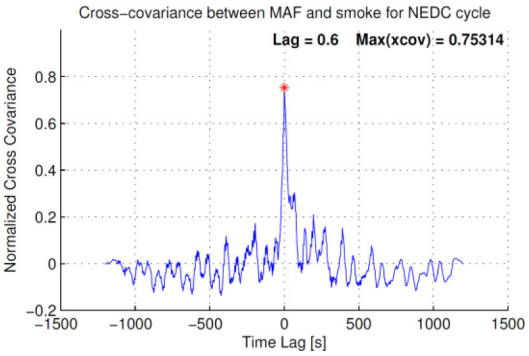

Figure 7. Cross-covariance between MAF and smoke for NEDC cycle

Table 1. Emission NNs input selection summary(Y: selected; -N-: not selected)

		NOx (static)	NOx (dynamics)	Smoke (dynamics)	Smoke (classification)
Engine speed		Y	Y	Y	Y
Fuel		Y	Y	Y	Y
Main injection timing		Y	-N-	Y	Y
Injection pressure		Y	-N-	Y	Y
MAP		Y	Y	Y	Y
Intake temperature		Y	Y	Y	Y
MAF		Y	Y	Y	Y
Intake O_2 concentration		-N-	-N-	Y	Y
Delays	Fuel	-N-	Y	Y	Y
	Main injection timing	-N-	-N-	Y	Y
	Injection pressure	-N-	-N-	Y	Y
	MAP	-N-	Y	Y	Y
	Intake temperature	-N-	Y	Y	Y
	MAF	-N-	Y	Y	Y
	Intake O_2 concentration	-N-	-N-	Y	Y
Total number of Neural Network inputs		**7**	**25**	**43**	**43**

Engine-Out Emission Models

NOx Models

A NOx model was first built with the static neural network architecture and trained using the steady-state data. The steady-state dataset contains 567 points which cover most of the operating region of the engine as shown in Figure 5. The network has been trained 100 times with the dataset randomly divided into three parts before training. As mentioned above, these three parts are the training, validation and test datasets, where the training data is used for parameters training and the validation data is used to monitor the network performance to avoid overfitting. Consequently, the overall performance of each trained network is different. The neural network with minimum MSE or maximum correlation coefficient (the value of R) between model results and measured targets within the complete steady-state dataset is selected to be the static NOx model. Considering the data coverage and the relatively small data size, 70% of the steady-state data is assigned for training, while 15% is reserved for validation and 15% for test. Similar to the architecture shown in Figure 2, this neural network has two hidden layers with 7 neurons in each layer. Thus, there are 120 parameters in total, which is about 1/5 of the steady-state data points. Figure 8 shows the fitting result of the neural network with best performance after 100 trainings. As shown, this static model is able to predict steady state NOx emission with reasonable accuracy.

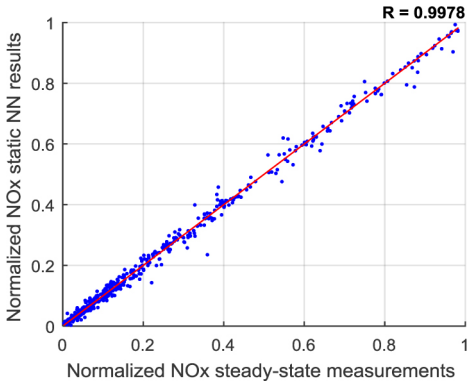

Figure 8. Static NOx neural network training result

To evaluate the transient performance of this static NOx model, the neural network has been run with inputs from a transient test. As shown in Figure 9 and Figure 10, such a transient test is referred to as a pattern test so that it can be distinguished from the drive cycle test. A fixed 1000 RPM sequence was selected because at low engine speed NOx emissions can actively vary within a wide range depending on inputs that are time-varying. If one ignores the extreme case where fuel exceeds the limit (near 130 to 140 seconds), the predicted value of NOx emissions follows the overall trend of the measurements but does not capture the dynamics very well during transient conditions.

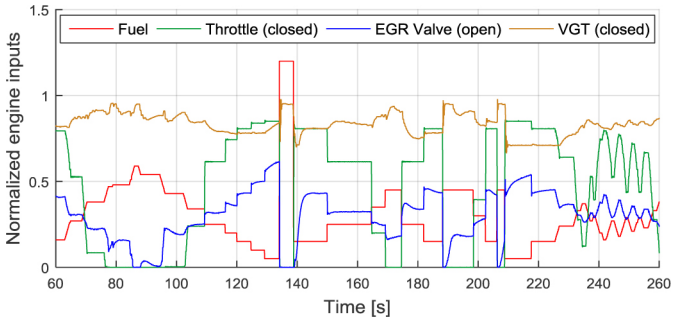

Figure 9. Transient simulation inputs for NOx models at 1000 RPM

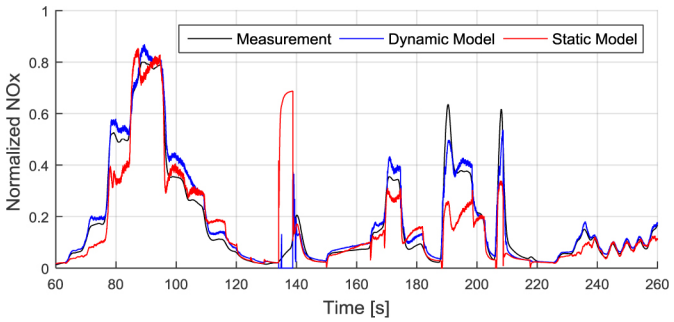

Figure 10. Transient simulation results with NOx models at 1000 RPM

In order to make the NOx model better predict the transient dynamics, a dynamic neural network with the TDNN architecture was developed and trained using transient data (where the transient pattern test in Figure 9 is not included). The total number of this transient dataset contains more than 5×10^5 points, which permits an increase of the neuron numbers in the two hidden layers to improve the neural network accuracy without the risk of overfitting. In this dynamic NOx model, the number of neurons in each hidden layer was set to be 20, making the total number of parameters equal to 961. However, the large number of training data points as well as network

parameters significantly increases the training time. In order to train the model multiple times within reasonable intervals, only 20% of the data (randomly selected) has been used, with 70% allocated to training and 30% to validation.

The dynamic neural network with best performance out of 8 trainings, with RMSE of 28.12 ppm and the correlation coefficient of 0.9973 within the complete transient dataset, has been selected for comparison with the static model. As compared in Figure 10, the dynamic neural network predicts NOx values with overall smaller error than the static neural network due to its ability to capture the NOx dynamics. In addition, the dynamic model is more accurate during transient input changes such as the ones at approximately 170 seconds and 210 seconds, where the static model predicts NOx undershoots not present in the test data. For these reasons, the dynamic TDNN is considered to be better at modeling NOx emissions and was selected to support the subsequent eMPC design and validation process.

It should be noted that an engine in a vehicle is constantly exposed to transients and steady-state conditions are often not reached. As a result, emissions models that are developed from steady-state data may not always provide accurate prediction of vehicle emissions. In particular, the cylinder wall temperatures, which have relatively slow dynamics, are known to affect NOx emissions. The increased accuracy of the dynamic neural network trained from transient data may also be indicative of an important ability of the dynamic neural network to inherently capture some of the natural patterns present during driving. This hypothesis will be further investigated in future work.

Smoke Models

In this section, a gas composition model is first created to calculate the intake O_2 concentration since it is one of the inputs to the smoke model. This gas composition model can be considered an engineered feature, used to incorporate domain knowledge into the black-box model. Then a TDNN dynamic neural network is trained to model the smoke emissions. Later, to deal with the highly transient behavior of smoke emissions, a classification neural network is built and compared against the dynamic model using TDNN.

A physical manifold gas composition model is built based on [15], with inputs including pressure and temperature in intake and exhaust manifolds, fuel injection quantity, MAF, and estimations of EGR flow and cylinder-in flow. The physical model can be represented by the following two dynamic equations:

$$\dot{F}_1 = \frac{W_{21}(F_2 - F_1) - W_{Dth}F_1}{\frac{p_1 V_1}{RT_1}} \tag{4}$$

$$\dot{F}_2 = \frac{W_{e2}(F_{e2} - F_2)}{\frac{p_2 V_2}{RT_2}} \tag{5}$$

where the symbol F is used for burned gas fraction, W is used to designate gas mass flow rates, p designates manifold pressures, T designates manifold temperatures, V is the manifold volume, and R is the specific heat constant. The subscript 1 denotes the intake manifold, 2 denotes the exhaust manifold, e denotes cylinders, and Dth denotes the diesel throttle. Once the burned gas fraction of the intake manifold is computed, the intake oxygen mass fraction can be calculated as an algebraic output using

$$O_{2_{intake}}\% = (1 - F_1) \times O_{2_{FreshAir}}\% \tag{6}$$

where the mass concentration of oxygen in the fresh air is assumed to be a constant value of 23%.

To gather intake O_2 concentration data for smoke neural networks training, a gas composition model was built in Simulink and simulated with inputs from transient tests. Figure 11 and Figure 12 show an example of the gas composition model results, where the exhaust O_2 concentration prediction matches well with the measurements at steady-state, and the prediction of intake O_2 concentration shows strong correlation with transient spikes of the smoke measurement, which further confirm the conclusion that the intake O_2 concentration and its delays are strong indicators of smoke production.

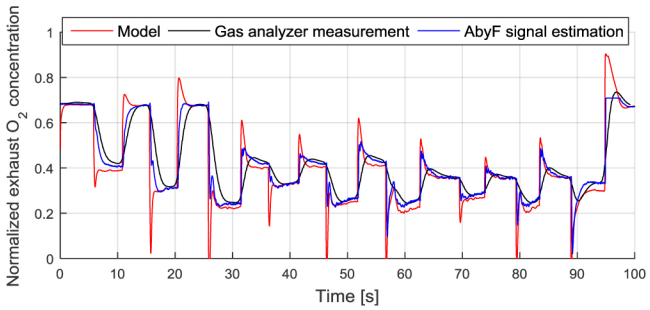

Figure 11. Exhaust manifolds gas composition model results

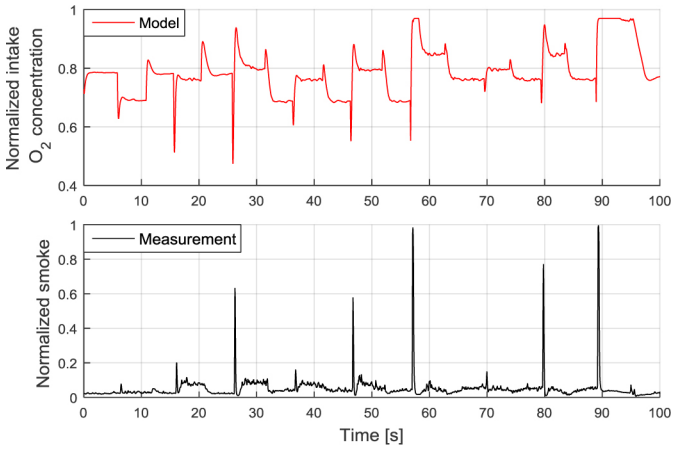

Figure 12. Intake manifold gas composition model results and correlation with smoke measurement

Using the same architecture as the dynamic NOx model, the smoke TDNN with inputs listed in Table 1 was trained with experimental data and simulation data from the gas composition model. More than 2×10^5 transient data points have been gathered, while only half of the data was randomly used. 35% of the total data was assigned for training and 15% of the total data was assigned for validation. The neural network has been trained 8 times and the one with the best performance, for which RMSE equals to 0.79% and the correlation

coefficient equals to 0.9887 considering the complete transient dataset, was combined with the gas composition model to form a complete dynamic smoke model.

Figure 13 and Figure 14 shows the simulation results of this complete dynamic smoke model using input data from a transient test which has not been used for the network training. Again, a fixed 1000 RPM sequence was selected since large smoke spikes are more likely to present at low engine speeds. The predicted smoke generally follows the trend of the smoke meter measurement. However, although the model is able to predict smoke spikes in transient operating conditions, the predicted value could be more than 50% off from the measurement, such as at 57 seconds.

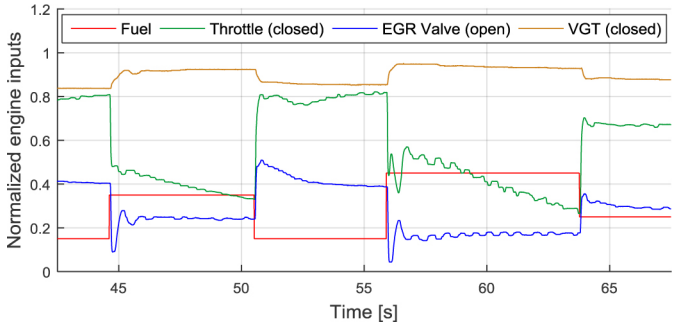

Figure 13. Transient simulation inputs for smoke models at 1000 RPM

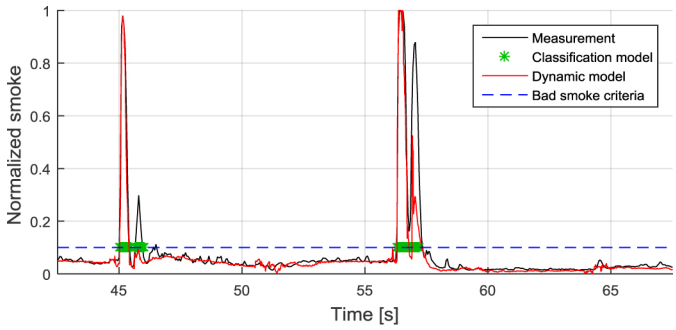

Figure 14. Transient simulation results of smoke models at 1000 RPM

To treat this problem, the chosen method is to avoid the direct calculation of smoke opacity value and use a classification neural network to classify the smoke emission as either 'acceptable' or 'bad'. Figure 15 shows the workflow of building and validating a smoke classification model. The test measurements of smoke are preprocessed before network training. In the eMPC application of this paper, keeping smoke level acceptable is a priority and the 'bad' smoke corresponds to the value exceeding 10% of normalized smoke, but to provide some robustness margin, a lower value corresponding to 'bad' smoke was assumed during training. Specifically, if the smoke opacity is higher than or equal to 8% of normalized smoke, it is labeled as 'bad' and the target output is set to 1. In contrast, the target output is set to 0 if the smoke opacity is less than 8% of normalized smoke. The output of the classifier, is the probability of having 'bad' smoke, with the value varying between 0 and 1 in a continuous range. The prediction is categorized as 'bad' if the neural network output is greater than or equals to 0.5, and otherwise as 'acceptable'. During model performance validation, the 'bad' smoke corresponds to 10% of normalized smoke and above, while 'acceptable' smoke corresponds to values below 10% of normalized smoke.

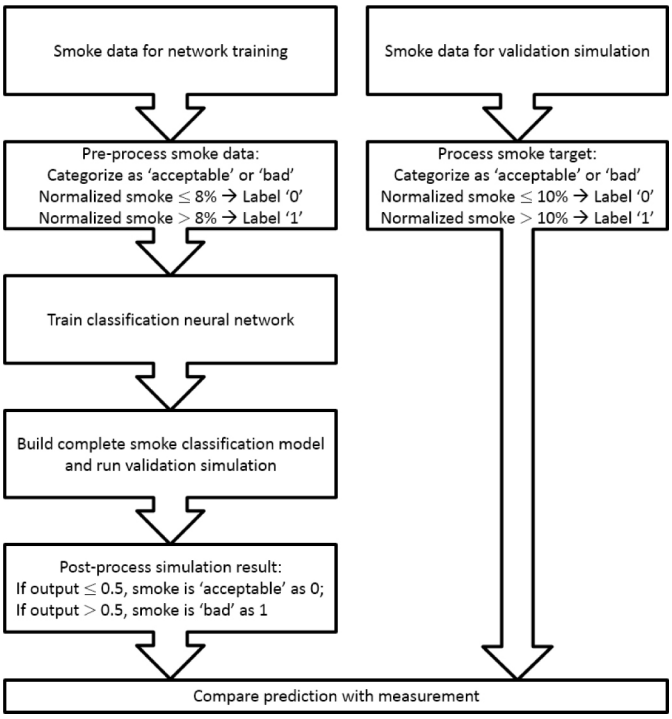

Figure 15. Smoke classifier workflow

Similar to the dynamic smoke TDNN, the smoke classification neural network has two hidden layers with 20 neurons in each layer. It was also trained with only half of the data randomly selected and combined with the gas composition model after training. The classifier with the best performance out of 8 trained neural networks was selected. Its total prediction accuracy considering the complete transient dataset is 94.08%, with true positives rate of 97.18% and true negatives rate of 93.78%.

The simulation result of the selected smoke classifier is shown in Figure 15, where green dots indicate the time when the classifier predicts 'bad' smoke. Note that the classification neural network is able to predict smoke spikes with similar and sometimes better accuracy as the previous developed TDNN model in transients without directly computing the smoke opacity value. Note also that the only major difference between these two types of neural networks is the activation function in the output layer. However, it was observed that the training of the smoke classifier takes much shorter time compared with the training of the smoke TDNN. Since in eMPC design the smoke is constrained below a prescribed level, the classification neural network represents a suitable modeling solution to represent this constraint, with an advantage of being more easily re-trainable if more data become available. On the other hand, such a re-training will, in general, be needed if the designer decides to change the threshold separating 'bad' and 'acceptable' smoke.

Drive Cycle Simulation Results

To evaluate the performance of NOx and smoke TDNN models, drive cycle simulations have been run using input data from NEDC and WLTC engine test. Table 2 summarizes the neural networks performance in the two drive cycle simulations. Since some of the NEDC data (after randomization) has been used for the neural network training, both NOx and smoke models have very small

cumulative error in the NEDC simulation. On the contrary, none of the WLTC data was involved for training, and thus the WLTC simulation could better validate the neural networks performance.

Table 2. Emission models performance in drive cycle simulations

		NOx	Smoke
NEDC	Data involved in training	20%	50%
	Cumulative error	-0.005%	0.079%
WLTC	Data involved in training	0%	0%
	Cumulative error	6.250%	-22.207%

Figure 16, Figure 17 and Figure 18 shows a portion of the WLTC simulation results for both emission models. As can be seen, the NOx prediction stays very close to the measurement and it is able to capture the transient behavior with reasonable accuracy. The smoke model predicts correct classifications most of the time, and false positive result (where the model predicts no spike but the measurement shows spikes, which is not good for the controller to minimize emissions) is not seen over the entire 200 seconds in the figure. It occasionally over predicts the transient spike, causing the increase of cumulative error over the complete drive cycle. Potential solutions are to include more training data, or increase the number of model inputs. Some oscillations are seen in both emissions results. One reason may be that the engine inputs of most training data are less transient than the ones for drive cycles, as can be compared among Figure 9, Figure 13 and Figure 16. Another possible reason is that all training and validation data is directly from measurements which contain a lot of noise, and sensor dynamics are not considered in this work. Thus, in addition to increasing training data and model inputs, filtering and pre-processing the training and validation data could also help smoothing the results. While further improvements can be made, some over-prediction of transient emissions may not be entirely undesirable, as it can provide some robustness margin in the control design. Further, for the purposes of controller validation, it is desirable that the emissions models mimic the behavior of physical sensors, including noise. Readers can decide what improvements are necessary based on their applications.

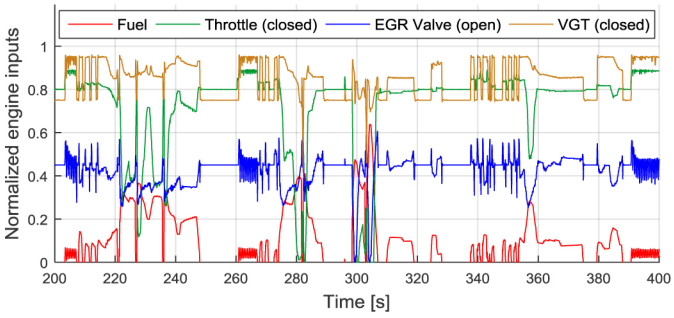

Figure 16. WLTC drive cycle simulation inputs

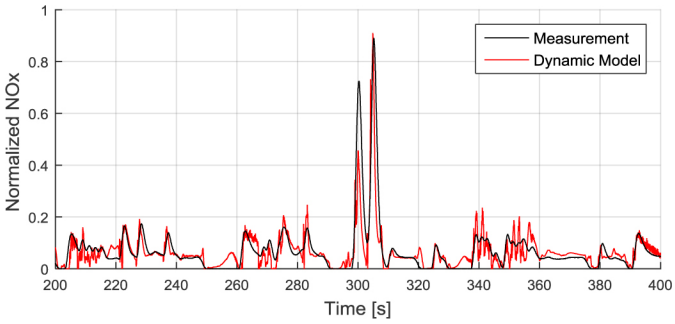

Figure 17. WLTC drive cycle simulation NOx results

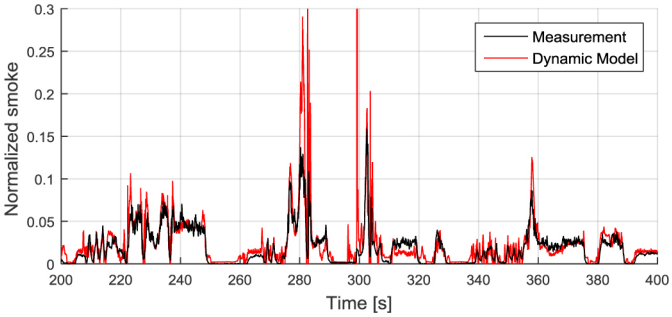

Figure 18. WLTC drive cycle simulation smoke results

Conclusions

In this article, neural networks with differing architectures have been introduced, developed and compared for the modeling of NOx and smoke emissions. Neural network training data was collected through engine bench tests and data covering most of the engine operating range has been gathered. Emission model inputs were selected by analyzing the engine physical operating principles in addition to results from the cross-covariance analysis. The Matlab Neural Network Toolbox was used for network parameter training, and the early stop method has been applied to avoid overfitting.

For NOx emissions, a static neural network and a dynamic neural network were built and compared against each other. The dynamic NOx model with TDNN architecture has been shown to be able to achieve better model accuracy in transients. To model smoke emissions, a gas composition model was first built to provide intake $\mathbf{O_2}$ concentration data for network training and then combined with the neural networks to form complete smoke dynamic models. A TDNN model and a classifier were developed and compared. Both models are able to capture transient smoke spikes with good accuracy, while the classification neural network presents significantly faster training speed. Last but not least, NOx and smoke TDNN models were simulated over NEDC and WLTC drive cycles using experimental data as inputs. The simulation results, including the drive cycle cumulative errors, show that both models are capable of capturing the highly nonlinear engine-out emissions high accuracy and strong generalization performance.

Overall this work confirm that MLP neural networks are a powerful tool for diesel engine-out emissions modelling, while neural network architectures, training algorithms, input selection, and data preprocessing were all found to have a very strong effect on both model accuracy and generalization performance. Future work may include investigating the emissions modeling with neural networks under more transient and real-life driving conditions, in addition to the use of big data and more sophisticated artificial neural network techniques such as deep learning to improve the model accuracy.

References

1. Chan, M., Das, S., and Reitz, R., "Modeling Multiple Injection and EGR Effects on Diesel Engine Emissions," SAE Technical Paper 972864, 1997, doi:10.4271/972864.
2. Hagena, J., Filipi, Z., and Assanis, D., "Transient Diesel Emissions: Analysis of Engine Operation During a Tip-In," SAE Technical Paper 2006-01-1151, 2006, doi:10.4271/2006-01-1151.

3. Ericson, C., Westerberg, B., Andersson, M., and Egnell, R., "Modelling Diesel Engine Combustion and NOx Formation for Model Based Control and Simulation of Engine and Exhaust Aftertreatment Systems," SAE Technical Paper 2006-01-0687, 2006, doi:10.4271/2006-01-0687.
4. Heywood, John B. "Internal Combustion Engine Fundamentals." Vol. 930. New York: Mcgraw-hill, 1988.
5. Bloch, Gérard, Lauer Fabien, and Colin Guillaume. "On learning machines for engine control." In Computational Intelligence in Automotive Applications, pp. 125-144. Springer Berlin Heidelberg, 2008.
6. Winkler-Ebner, B., Hirsch, M., Del Re, L., Klinger, H. et al., "Comparison of Virtual and Physical NOx-Sensors for Heavy Duty Diesel Engine Application," *SAE Int. J. Engines* 3(1):1124-1139, 2010, doi:10.4271/2010-01-1296.
7. Ahlawat, Rahul, Hagena Jonathan R., Filipi Zoran S., Stein Jeffrey L., and Fathy Hosam K.. "Volterra Series Estimation of Transient Soot Emissions from a Diesel Engine." In 2010 IEEE vehicle power and propulsion conference, pp. 1-7. IEEE, 2010.
8. Prokhorov, Danil. "Neural networks in automotive applications." In Computational Intelligence in Automotive Applications, pp. 101-123. Springer Berlin Heidelberg, 2008.
9. Johri, Rajit, Salvi Ashwin, and Filipi Zoran. "Real-time Transient Soot and NOx Virtual Sensors for Diesel Engine Using Neuro-Fuzzy Model Tree and Orthogonal Least Squares." Journal of Engineering for Gas Turbines and Power 134, no. 9 (2012): 092806.
10. De Cesare, M. and Covassin, F., "Neural Network Based Models for Virtual NOx Sensing of Compression Ignition Engines," SAE Technical Paper 2011-24-0157, 2011, doi:10.4271/2011-24-0157.
11. Hagan, Martin T., Demuth Howard B., Beale Mark H., and De Jesús Orlando. "Neural Network Design." Vol. 20. Boston: PWS publishing company, 1996.
12. Huang, Mike, Nakada Hayato, Butts Ken, and Kolmanovsky Ilya. "Nonlinear Model Predictive Control of a Diesel Engine Air Path: A Comparison of Constraint Handling and Computational Strategies." Proceedings of IFAC Symposium on Nonlinear Model Predictive Control, IFAC-PapersOnLine 48, no. 23 (2015): 372-379.
13. Wang, Z., Zhu, Q., and Prucka, R., "A Review of Spark-Ignition Engine Air Charge Estimation Methods," SAE Technical Paper 2016-01-0620, 2016, doi:10.4271/2016-01-0620.
14. MathWorks, Signal Processing Toolbox Functions: xcov, http://www.mathworks.com/help/signal/ref/xcov.html, Sep 2016.
15. Kolmanovsky, Ilya, Morall P., Nieuwstadt Michiel. Van, and Stefanopoulou Anna. "Issues in Modelling and Control of Intake Flow in Variable Geometry Turbocharged Engines." Chapman and Hall CRC research notes in mathematics (1999): 436-445.

Contact Information

huayi.li@toyota.com

Acknowledgments

The authors would like to appreciate the support and critical comments on this work by Hayato Nakada and Masanori Shimada from Toyota Motor Corporation. The authors would also like to acknowledge the support on bench test data acquisition by Mike Huang and Tony Kim from Toyota Technical Center.

ISSN 0148-7191

http://papers.sae.org/2017-01-0601