Estimating Soot Emission in Diesel Engines Using Gated Recurrent Unit Networks

Gokhan Alcan*∗, Emre Yilmaz*, Mustafa Unel*, Volkan Aran*, Metin Yilmaz**, Cetin Gurel***, Kerem Koprubasi***

* Faculty of Engineering and Natural Sciences, Sabanci University
Istanbul, Turkey
** Integrated Manufacturing Technologies Research and Application Center, Sabanci University, Istanbul, Turkey
*** Product Development, Ford Otosan, Istanbul, Turkey

Abstract: In this paper, a new data-driven modeling of a diesel engine soot emission formation using gated recurrent unit (GRU) networks is proposed. Different from the traditional time series prediction methods such as nonlinear autoregressive with exogenous input (NARX) approach, GRU structure does not require the determination of the pure time delay between the inputs and the output, and the number of regressors does not have to be chosen beforehand. Gates in a GRU network enable to capture such dependencies on the past input values without any prior knowledge. As a design of experiment, 30 different points in engine speed - injected fuel quantity plane are determined and the rest of the input channels, i.e., rail pressure, main start of injection, equivalence ratio, and intake oxygen concentration are excited with chirp signals in the intended regions of operation. Experimental results show that the prediction performances of GRU based soot models are quite satisfactory with 77% training and 57% validation fit accuracies and normalized root mean square error (NRMSE) values are less than 0.038 and 0.069, respectively. GRU soot models surpass the traditional NARX based soot models in both steady-state and transient cycles.

Keywords: Diesel Engine, Combustion Process, Soot Emission, Experiment Design, Gated Recurrent Unit.

1. INTRODUCTION

Diesel engines are widely preferred due to their high thermal efficiency, endurance, reliability, and low operating costs, especially for heavy-duty vehicles in the market (Resitoglu et al. 2014). However, they emit high exhaust emissions particularly soot particles which is responsible for serious environmental and health problems. Regarding this, ever-mounting stringent emission regulations enforce the engine manufacturers to design innovative engines by introducing after-treatment systems (Hsieh and Wang, 2011) or in-cylinder combustion control techniques (Sindhu et al., 2017) to reduce exhaust emissions to the environment. Moreover, increase in fuel prices and the demand for more powerful engines reveal the need for searching the optimum engine conditions in both steady-state and transient operations for engine manufacturers. Therefore, obtaining a sufficiently accurate model of engine combustion process is required to employ it in powertrain development for the optimization of engine components, testing and model-based calibration of combustion and after-treatment control (Bertram et al., 2014).

Modeling exhaust emissions on both steady-state and transient operating cycles of diesel engines can significantly reduce the experimentation time and cost for tuning of the parameters. Pfeifer et al. (2003) stated that measurements need to be taken not only on the after-treatment system but also on the engine raw emissions in order to meet the actual and upcoming stringent emission legislation. In literature, some studies focus on the phenomenological nature of the soot formation and present physics or chemistry-based models. On the other hand, some existing methods tackle the soot emission modeling problem by exploiting the potential of empirical data and try to find a global model or multiple local models which explain the relation between measured input and output signals. Tauzia et al. (2017) presented a semi-physical sub-models to predict NOx and soot emissions for a compression ignition (CI) diesel engine. Soot formation is described by a global equation relates the formation with O2 concentration, in-cylinder pressure, temperatures, heat release rate duration and turbulence intensity. Walke et al. (2016) proposed sub-models for cylinder pressure, two zone temperature, NOx and soot emissions. In the study, Hiroyasu model (Hiroyasu and Kadota, 1976) is taken as base soot model and the model is corrected by experiments. Tanelli and Maranta (2015) compared three semi-empirical soot models for internal combustion engine simulations called Moss, Lindttest-Leung and Wen. They also extended the representative interactive flamelet combustion model to predict soot emissions. It should be noted that the phenomenological models include several local physics and chemistry based relations, therefore generalization perfor-
manances of such models are not promising due to their complex structures.

Benz et. al. (2010) proposed a nonlinear extended quasi-static model for raw emissions of heavy-duty and light-duty diesel engines derived by a symbolic regression algorithm. In order to choose the input signals, they employ an input variable selection algorithm based on genetic programming and artificial neural network (ANN). Prediction performances of their NOx models are quite satisfactory, but the soot emission predictions are not quite well due to relatively high measurement errors. Tschanz et. al. (2010) presented a novel model for particulate matter (PM) emissions of diesel engines that meets the requirements of being control oriented, easily identifiable and portable. They also assumed that the engine-out PM emissions are quasi-statically influenced by the conditions inside the cylinder at intake valve close and injection parameters. PM emissions are modeled as relative deviations of stationary base maps and a polynomial model is employed to estimate the influence of each input on PM emissions. However, employing a polynomial approach in such method brings certain disadvantages such that the number of parameters for the presented method is relatively high and polynomial approach has diminished extrapolation ability. Ericson et. al. (2005) presented another quasi-stationary modeling approach for fuel consumption, CO, HC, NOx and PM emissions. They claimed that transient correction methods are required to obtain well-performing models, and torque and speed must be employed as inputs to those correction models for better generalization. Their approaches are primarily useful for predicting the emissions and are not offered to be used for engine control or optimization. In order to obtain more consistent results and employ that approach to engine control, the time delay estimation must be substantially improved.

Sequenz et. al. (2010) presented a global model structure composed of adaptive local polynomial models. In order to reduce the number of parameters to be estimated and the variance error in estimation, a regressor selection method based on Mallows Cp-statistics (Mallows, 1973) is described. It is seen that the error between soot measurements and simulated output increases after a constant level of air-fuel ratio, and so R² performance of the soot models are not high enough like NOx emissions. Mrosek et. al., (2010) modeled the simplified combustion process and the emission formation as a stationary batch process. Diesel engine emissions are modeled locally at discrete operations points and these points are defined by engine speed and the desired injection quantity. Local polynomial model (LPM) (Sequenz et. al., 2010) structure takes the measures and the combustion characteristic (MBF50) as inputs. Raw emission outputs are approximated by polynomials of equal order 3 and considering cross-terms of two inputs. Performances of NOx emission models were quite well but the soot models still need to be improved. Hafner et. al., (2000) compared two neural models for the stationary soot formation with different input signals. The first model is based on the engine control settings, injected fuel, injection angle and engine speed, while in the second one the characteristics of the measured cylinder pressure signal are utilized. They employed a special local linear model tree (LOLIMOT) with radial basis function (RBF) introduced by Nelles and Isermann (Nelles and Isermann, 1996). Static soot models with cylinder pressure characteristics show comparable good performances with the model based on engine actuator signals. Atkinson and Mott (2005) proposed a neural network based transient engine modeling where engine operating parameters such as engine speed, engine temperature, MAP, manifold air temperature and the engine control inputs such as fueling rate, injection timing, injection pressure, VGT setting and EGR rate are considered as inputs, and emissions such as NOx, HC, CO, PM, and CO2 are taken as outputs. Their work covers a very narrow range of operating points and the PM results are not satisfactory.

In this paper, data-driven modeling of a diesel engine soot emission formation using gated recurrent unit (GRU) networks is proposed for the first time. Different from the traditional time series prediction methods, GRU structure does not necessitate to determine the pure time delay between inputs and output, and the number of regressors does not need to be chosen beforehand. As a design of experiment, 30 different operating points in engine speed - injected fuel quantity plane are selected and the rest of the input channels (rail pressure, main start of injection, equivalence ratio, and intake oxygen concentration) are excited with chirp signals in the intended regions of operation. To reduce the amplitude of sudden spikes and encapsulate the nonlinearities between inputs and the output, soot measurements are logarithmically normalized. Experimental results show that the prediction performances of GRU based soot models are quite satisfactory and surpass the traditional NARX based soot models.

Organization of this paper is as follows: Diesel engine combustion is briefly described in Section 2. Design of experiment is presented in Section 3. Gated Recurrent Unit Network structure is explained in Section 4. Experimental results are provided in Section 5. Finally, the paper is concluded with some remarks and future directions in Section 6.

2. DIESEL ENGINE

The experimental measurements are performed on Ford Otosan’s Ecotorq 13L Euro 6 diesel engine. The diesel engine combustion composes of two fundamental paths named air and fuel paths. On its air-path, the engine has high-pressure Exhaust Gas Recirculation (EGR) routing and swing vane Variable Geometry Turbine (VGT) turbocharger as shown in Fig. 1.

The common pressurized fuel rail, solenoid injectors and high-pressure fuel pump are components of the fuel system. The ambient fresh air which flows through air filter is measured before entering to the compressor. This air flow measurement is stated as Mass Air Flow (MAF). After the compressor, its pressure is increased to a regulated level, and in order to boost the volumetric efficiency of the engine, the compressed air flow is forwarded to the charge air cooler which reduces the flow temperature to a regulated level. Before the EGR mixing point, the temperature and the pressure of the cooled charger air are measured. The cooled charger air flow which is mixed with EGR flow is injected into the cooled charger air through the pressurized intake manifold. The pressure of
the mixed air flow in the intake manifold is named as Manifold Absolute Pressure (MAP). After the combustion, through the exhaust manifold the exhaust gases exit the chamber and certain portion of them are recirculated back via EGR valve.

Fuel is injected into a pressurized rail. The fuel amount and the pressure in the rail are measured as Fuel Quantity (QNT) and Rail Pressure (RailP), respectively. The starting angle of the main fuel injection quantity is defined as Start of Injection (SOI). Impact angle of the exhaust gases to turbine blades is regulated by a valve called Variable Geometry Turbine (VGT) turbochargers. The exhaust gas expanded in the turbine is sent to the exhaust line and its soot content is measured by the AVL 483 Micro Soot Sensor. Micro soot sensor is a photoacoustic system designed for the measurement of black carbon aerosols. It is a suitable test bench with high sensitivity down to $\mu g/m^3$ level, and good linearity and reproducibility (Schindler et al. 2004).

3. DESIGN OF EXPERIMENT

Design of experiment includes the selection of model inputs, the determination of the excitation signals and the selection of the validation test. As aforementioned in the former section, the signals in both air-path and fuel-path play an important role in the combustion process and exhaust emissions. Within this context, engine speed (SPD), total injected fuel quantity (QNT), main Start of Injection (miSOI), rail pressure (railP) and intake oxygen ratio (intO2) are chosen as input channels for soot emission models. Furthermore, the equivalence ratio (EQVR) (Ramadhas et. al., 2014), which is the ratio of actual fuel-air (QNT/MAF) ratio to the theoretical fuel-air ratio, is also included in the input set.

Exciting the dynamical system with adequate signals plays a decisive role in the parameter estimation process. In a model-free design of experiment approaches for dynamic systems, covering the whole intended input space by using space filling methods (Santner et. al. 2013) and encapsulating the frequency range of underlying process (Ljung, 1999) are very crucial. Therefore, chirp signals are commonly preferred due to their persistent excitation capabilities. In this study, a diesel engine's intended region of operation was determined by 30 different points in engine speed - injected fuel quantity plane, where the engine speed and the injected fuel quantity are selected in the ranges of 800-2200 rpm and 4-30 mg/stroke, respectively. The rest of the input channels were excited by chirp signals, which have a sinusoidal waveform with changing frequencies over time given by

$$y(t) = Asin(2\pi f(t))$$

where the frequency of the chirp signal can be a linear, quadratic or an exponential function of time. In this work, a quadratic function is employed to excite the system sufficiently in both high and low frequencies as follows

$$f(t) = f_{min} + \left(\frac{f_{max} - f_{min}}{t_{max}} \right) t$$

where $f_{max}$ is the maximum frequency (0.25 Hz), $f_{min}$ is the minimum frequency (0.05 Hz) and $t_{max}$ is the time that the system operates at maximum frequency.

Duration of each experiment was 5 minutes and the signals were collected in 10 Hz. 10 of 30 operation points were chosen for validation purposes in two different ways (Fig. 2). In Case 1, the validation points were randomly selected. In Case 2, the validation points were selected from the boundary of the intended region of operations.

For reasons of confidentiality, input channels that were excited by chirp signals are presented as percentages in Fig. 3.
In order to reduce sensitivity to sudden spikes in soot measurements, i.e. increasing numerical stability of the obtained models, logarithmic normalization was applied to soot measurements as follows

\[
\text{lognorm}(x) = \begin{cases} 
\log(x) & x > 1 \\
0 & |x| \leq 1 \\
-\log(-x) & x < -1 
\end{cases} 
\]  

It should be noted that all input signals and logarithmically normalized soot measurements were scaled to a range between 0 and 1 to increase numerical stability.

4. GATED RECURRENT UNIT NETWORKS

In order to capture long-term dependencies without the need of delay information for time series problems, Long-Short Term Memory (LSTM) networks were first introduced by Hochreiter and Schmidhuber (1997). Recent developments in CPU and GPU technologies enabled to train such complex structures with high dimensional data. Although the NARX soot models are trained at most 59% and the models are validated with at most 47% (NRMSE) given by

\[
\text{NRMSE} = \frac{1}{\max(y) - \min(y)} \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} 
\]  

where \(y_i\) and \(\hat{y}_i\) denote the actual and predicted values. The performances of obtained models are assessed by the fit metric and normalized root mean square error (NRMSE) given by

\[
\text{Fit} = \left(1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|}\right) \times 100 
\]  

5. EXPERIMENTAL RESULTS

Diesel engine combustion process results in various exhaust emissions including NOx and soot. Several studies exist regarding exhaust emissions modeling particularly NOx emissions by utilizing Nonlinear Autoregressive with Exogenous Input (NARX) network structure (Boz et. al., 2015, Alcan et. al., 2018, Alcan et. al. 2019). In order to compare the prediction performances of the proposed GRU soot modeling approach, NARX network based soot modeling is also performed.

The NARX structure has several hyperparameters including the number of output regressors, the number of input regressors (nb) and the amount of delays between inputs and output. Since the proposed GRU modeling structure does not include feedback from the estimated output and captures the relations between time delayed input signals automatically, the number of output regressors in NARX models is set to zero and the number of input regressors is selected in the range of 2 to 5. Moreover, the amount of delay between inputs and output is decided as 3.5 seconds by considering the physical conditions of the experimental setup. Levenberg-Marquardt optimizer is employed in the parameter estimation process of NARX models.

The performances of obtained models are assessed by the fit metric and normalized root mean square error (NRMSE) given by

\[
\text{Fit} = \left(1 - \frac{\|y - \hat{y}\|}{\|y - \bar{y}\|}\right) \times 100 
\]

\[
\text{NRMSE} = \frac{1}{\max(y) - \min(y)} \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} 
\]  

Training and validation performances of some representative models obtained by NARX and GRU structures are tabulated in Table 1 and Table 2.

Although the NARX soot models are trained at most 59% and the models are validated with at most 47% fit accuracy, training performances of GRU models are
Table 1. Performances of NARX models

<table>
<thead>
<tr>
<th>Case</th>
<th>nb</th>
<th>Unit</th>
<th>Training</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fit (%)</td>
<td>Fit (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NRMSE</td>
<td>NRMSE</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>46.89</td>
<td>38.95</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>15</td>
<td>49.70</td>
<td>47.59</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>20</td>
<td>48.87</td>
<td>45.30</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>25</td>
<td>55.45</td>
<td>44.83</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>30</td>
<td>54.48</td>
<td>40.40</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>50</td>
<td>52.79</td>
<td>41.00</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>10</td>
<td>46.60</td>
<td>38.69</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>15</td>
<td>45.46</td>
<td>43.94</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>20</td>
<td>51.99</td>
<td>40.46</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>25</td>
<td>54.75</td>
<td>40.83</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>30</td>
<td>55.07</td>
<td>43.03</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>50</td>
<td>59.48</td>
<td>49.93</td>
</tr>
</tbody>
</table>

Table 2. Performances of GRU models

<table>
<thead>
<tr>
<th>Case</th>
<th>Unit</th>
<th>Training</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fit (%)</td>
<td>Fit (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NRMSE</td>
<td>NRMSE</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>76.19</td>
<td>53.28</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>77.17</td>
<td>57.37</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>68.53</td>
<td>54.86</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>63.95</td>
<td>50.37</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>74.74</td>
<td>53.72</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>73.03</td>
<td>53.59</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>59.72</td>
<td>47.75</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>76.99</td>
<td>52.53</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>49.23</td>
<td>42.97</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>75.67</td>
<td>49.06</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>54.17</td>
<td>47.08</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>57.42</td>
<td>45.58</td>
</tr>
</tbody>
</table>

able to reach up to 77%, corresponding validation performances exceeds 55%. Furthermore, training and validation NRMSE values are around 0.037 and 0.069 in the selected GRU model, however, NRMSE values are always above 0.65 for training and validation of NARX models. The best results of GRU structure for both cases are achieved when the number of units is selected as 15.

Time plots of NARX and GRU models with 15 units are depicted in Fig. 5, 6, 7 and 8. Since the number of input regressors and the pure time delay between inputs and output are pre-defined and can not be adjusted during parameter estimation in a NARX structure, the predictions of NARX models are highly oscillatory and their accuracies are not good at steady-state operating points compared to transient sections. Furthermore, it is not easy to determine the model parameters that result in a smooth response by fixing these values, particularly in soot emissions modeling. However, the predictions of GRU models are smoother thanks to the gated structure which enables the model to adjust the effects of former input signals adaptively with update and reset gates. Moreover, the prediction accuracies of GRU models are better on both steady-state and transient sections of test cycles. In Fig. 7, it is seen that the prediction performance of GRU model in the first 10 minutes is inadequate because the engine works in its idle state with speed less than or equal to 1000 rpm, and soot emission dynamics differ in this speed range. In order to increase the prediction performances, separate models for different ranges of engine speed can be trained.

It should be noted that this work explores the possibility of employing a GRU network in soot emissions modeling for an extensive range of engine speeds. We do not claim any optimality regarding structure of the GRU network used in this work. Therefore obtained results are not optimal in any sense, however they are satisfactory when compared to the existing results in the literature.

6. CONCLUSION

We have now presented gated recurrent unit (GRU) networks to estimate soot emissions in diesel engines. Due to having both reset and update gates in GRU networks, dynamical relations between past inputs and the output are automatically captured. Thanks to this important memory capability, GRU structure does not require the
determination of the pure time delay and the number of regressors beforehand.

Prediction performances of GRU based soot models are quite satisfactory with 77% training and 57% validation fit accuracies and NRMSE values are less than 0.038 and 0.069, respectively. GRU soot models surpass the traditional NARX based models in both steady-state and transient test cycles. Utilization of a logarithmic normalization on soot measurements in the parameter estimation process reduces model sensitivity to sudden variations in soot measurements and increases the modeling accuracies.

As a future work, prediction errors in steady-state regions of the test cycles will be investigated and the obtained models will be validated with different cycles such as the New European Driving Cycle (NEDC) and the Worldwide Light-duty Test Cycle (WLTC).

Acknowledgements

The funding provided by Ford Otosan is gratefully acknowledged.

References


Tanelli, C., and Maranta, S. Modeling soot emissions in Diesel engines by means of semi-empirical approaches and a combustion model based on detailed chemistry, Politecnico di Milano.

