Heavy duty vehicle fuel consumption modelling using artificial neural networks

Oskar Wysocki
Department of Energy and Industrial Apparatus
Gdansk University of Technology
Gdansk, Poland
oskwys@gmail.com

Lipika Deka
Department of Computer Science and Informatics
De Montfort University
Leicester, UK
lipika.deka@dmu.ac.uk

David Elizondo
Department of Computer Science and Informatics
De Montfort University
Leicester, UK
elizondo@dmu.ac.uk

Abstract – In this paper an artificial neural network (ANN) approach to modelling fuel consumption of heavy duty vehicles is presented. The proposed method uses easy accessible data collected via CAN bus of the truck. As a benchmark a conventional method, which is based on polynomial regression model, is used. The fuel consumption is measured in two different tests, performed by using a unique test bench to apply the load to the engine. Firstly, a transient state test was performed, in order to evaluate the polynomial regression and 25 ANN models with different parameters. Based on the results, the best ANN model was chosen. Then, validation test was conducted using real duty cycle loads for model comparison. The neural network model outperformed the conventional method and represents fuel consumption of the engine operating in transient states significantly better. The presented method can be applied in order to reduce fuel consumption in utility vehicles delivering accurate fuel economy model of truck engines, in particular in low engine speed and torque range.

Keywords – neural network, combustion engine, heavy duty truck, fuel economy

I. INTRODUCTION

Trucks, whether used for freight transportation or as utility vehicles, play an important role in a countries economy and improving their fuel efficiency can undoubtedly prove highly beneficial. There is a large volume of published studies describing the operational efficiency of diesel engines in vehicles [1],[2]. However vast majority of work focuses on fuel consumption during transportation [3][4], and there is scarce research on duty cycle fuel efficiency of utility trucks i.e. garbage trucks, dump trucks, concrete pump and concrete mixer trucks, trucks with crane etc. Such vehicles use Power Take-Off (PTO) units to draw power from chassis combustion engine to perform other functions such as waste compaction, lifting bins, crane operation, drum rotation in concrete mixers and dump lifting. PTO is an additional shaft located on the engine or the gearbox, where a power receiving device can be mounted (e.g. hydraulic pump). One of the most popular examples of a utility truck is a Refuse Collection Vehicle (RCV), which is equipped with a hydraulic system. Annual production of these vehicles exceeds 6500 units, and about 95% of them are powered by diesel engines [5]. RCV’s efficiency and energetic model are introduced in [6]-[8]. The important fact is that utility vehicles during PTO operation, work at low engine speed and low torque range, in comparison to the values obtained while driving. It was shown in [6] and [9], that proper drivetrain configuration can lead to a significant fuel consumption reduction of up to 35% in extreme cases. Thus, an engine fuel map, commonly also referred to as a general engine characteristic (GEC), contains information about the most efficient engine operation points. This map is a requirement for the optimal design of a new utility truck drivetrain or the calibration of an existing one. The map allows not only to set the engine to the optimal engine operation point, but also to compare different chassis in terms of their fuel consumption while performing similar duty cycles. Such knowledge may lead to optimal vehicle choice or route planning leading to significant fuel and cost reduction for utility truck fleets.

Normally, combustion engine characteristics are obtained in laboratories using test benches or chassis [10],[11]. However, truck body manufacturers and truck users usually have little access to such sophisticated equipment, which usage is time consuming and associated to high costs. An alternative solution to this problem is the use of the vehicle’s On-Board Diagnostics system (OBD) to log the engine operation data and acquire it via the
CAN bus. An example of this for passenger cars is presented in [12]. Since the Fleet Management System (FMS) standard to log chassis data [13] is widely used in trucks manufactured after 2002, data acquisition is relatively easy and does not require interfering with the vehicle's onboard systems. This standard provides, among other parameters: engine speed \( n \), torque \( T \), instantaneous fuel consumption or fuel rate \( G_c \).

The purpose of this paper is to propose an Artificial Neural Network (ANN) based approach for fuel consumption prediction from real operational data in order to accurately determine diesel engine performance. Conventional methods require tests to be conducted in quasi-steady states. This assumes that changes to the engine parameters (such as increase/decrease in engine speed or torque) do not exceed previously assumed limits [14]-[17] in given time. However, such methods prove infeasible as was shown in [18],[19] that in real world, combustion engines rarely operate in steady states. Thus, to overcome the challenges of conventional methods, effective alternative methods must be developed whereby logged CAN bus data is used instead.

In [12], GEC obtained based on dynamometer tests was used to predict fuel consumption and then compared to real values logged in via On-Board Diagnostics (OBD). Such application resulted in prediction errors of up to 10%. This was caused by differences in steady and transient engine operations as described in [11]. An alternative method using artificial networks can be used to deal with the problem of transient state prediction inaccuracy. Proposed in 2000 [20], the model includes a history window of a few seconds of the same data i.e. values from few preceding observations before the actual observation. This model allows the network to learn non-linear dependencies between engine working parameters and fuel consumption and exhaust gas emissions. Neural networks are also widely used in combustion engines tests [21][22] and can achieve high accuracy. However, the majority of this research focuses on fuel consumption and exhaust gas emissions measured using sophisticated apparatus and hence costly equipment. Thus its application for truck body manufacturers or truck users is limited.

A key aspect of the method presented in this paper is based on the development of an ANN model built by using CAN data collected from the truck, with no need for dynamometer tests. Therefore, the proposed method can be easily used without special equipment, examining the data collected during real operation of the truck, both during transportation and using PTO. However, for the purpose of this study, in order to simulate arbitrary load with no need to drive the vehicle, a test bench was used. The test bench was connected to the chassis engine via the PTO. In [23] the authors describe the design principles of this equipment. It uses a hydraulic system to simulate the duty cycle load or any arbitrary load defined by the test needs. Data is then collected using the CAN bus, as it can be done during real daily operation of the vehicle.

II. DESCRIPTION OF EXAMINED CHASSIS AND TEST BENCH

The main concept of tests presented in this study, was to set the engine in desired operation point and measure its fuel consumption. This was done both in quasi-steady and transient states, and in the entire \( n \cdot F_{PTO} \cdot T_{max} \) domain. Then, the collected data were used to calibrate different models and assess their accuracy.

Tests were performed using a test bench mounted on a Scania P320 chassis, equipped with 9.3 dm³ displacement engine (max power: 235 kW and max torque: 1600 Nm) [24]. The Test Bench (TB) concept is based on the load application using the PTO as a controllable engine load generator. A pump attached to the PTO drives the hydraulic system. The load of the system can be adjusted due to expected load history. The engine power is determined by its torque \( T \) and rotational speed \( n \). This is transmitted to the hydraulic system. The power of this system is defined by the product of the oil flow \( Q \) and the pressure \( p \):

\[
\eta \cdot T \cdot n = Q \cdot p
\]

where \( \eta \) represents the efficiency of the pump. Using a test bench, it is possible to obtain an arbitrary dense grid of measuring points within the range of the PTO maximum power. The TB can be mounted on any truck chassis as long as it is equipped with a PTO. In case there is not enough space on the frame, only the pump is mounted on the vehicle and the oil is delivered to the hydraulic unit using hydraulic hoses.

For the purpose of this research, the TB was attached to the chassis frame with screws and the hydraulic pump was connected to the engine’s PTO by means of a homokinetic shaft. The final step was to connect the TB control unit to the chassis CAN bus. The entire mounting process took less than -an hour for 2 workers. The final test stand is presented in Fig. 1. Three types of tests were performed: (1) quasi-steady state, (2) transient state and (3) validation test. They are described in detail in the following sections.

![Figure 1. Test bench during tests, mounted on Scania chassis](image)

III. REFERENCE MODEL – POLYNOMIAL REGRESSION

A conventional approach to determine general characteristics of internal combustion engine (GEC) is to apply a constant load at a constant speed at different operating points and to measure the fuel consumption.
Due to the nature of the combustion engine, the above conditions are considered as quasi-steady. This method is widely used to assess engine performance and fuel consumption [25]. Then, a 3rd degree polynomial regression is used as an approximating function. As the GEC is considered as the function of two variables, the result of the approximation corresponds to a surface, with the coefficients of the polynomial calculated by using the least squares method. It is worth noting, that in this method all measuring points are used to build the model, leaving no unused points for testing purposes. Quasi-steady states characteristics are often used to evaluate fuel consumption in transient states with acceptable levels of error [1], [26].

IV. TRANSIENT STATE TEST

During the RCV operation, as well as in most utility vehicles, the engine speed is usually set to a constant value. Due to power demand generated by the hydraulic system, the engine indicated torque keeps changing in order to maintain a constant speed. This process is managed by the Electronic Engine Control Unit (ECCU), which also allows for data to be logged via the CAN bus. A transient state test was performed based on this settings: the engine speed was set constant for 50 seconds and the load was changed every 3, 4 or 5 seconds at random. Then the engine speed was set stepwise on the next value from range: [700, 750, …, 1150, 1200]. Then, 50 seconds of load changes were applied again. This process was repeated, resulting in collected measurements for 11 different engine speeds. An excerpt of this is presented in Fig. 2. A delay (lag) Δt between the G, and the T increase or decrease can be easily visualized with values ranging from 0.5 to 1.5 seconds. Thus, in order to obtain the most accurate model, it seems that engine parameters from preceding observations should be included in the model in addition to actual values, as suggested by [20], [27].

![Figure 2. Engine speed n, torque T and fuel rate G, during transient state tests. Δt = delay between local torque and G, extremum](image)

Transient state test resulted in 6000 observations within the entire examined engine speed range. For further analysis, the dataset was randomly sampled to the size of 600 and split into train and test set in proportion of 80:20 respectively. Training points in the n-T domain are presented in Fig. 4.

V. VALIDATION LOAD TEST

To validate the models a second set of measurements was collected. Similarly to the transient state test, for each constant engine speed, the load was applied 6 times with 10 min intervals inbetween tests. This time, the focus was on fuel consumption in 33 s load cycles, which are similar to the real RCV duty cycle [6], [7]. Examples of this are shown in Fig. 3. The validation test resulted in 65 load cycles, 6 per engine speed (1 cycle was discarded, due to logging error). The test summary is presented in Table 1.

![Figure 3. Load cycle in Nm for n=900 (blue), n=1000 (red) and n=1200 (green) rpm in validation tests](image)

<table>
<thead>
<tr>
<th>Test</th>
<th>Transient states</th>
<th>Validation load</th>
</tr>
</thead>
<tbody>
<tr>
<td>N range</td>
<td>650 ÷ 1250 rpm</td>
<td>6000 cycles</td>
</tr>
<tr>
<td>T range</td>
<td>50 ÷ 500 Nm</td>
<td>12000 cycles</td>
</tr>
<tr>
<td>Nb of observations</td>
<td>6000</td>
<td>2670 g</td>
</tr>
<tr>
<td>Duration time</td>
<td>10 min</td>
<td>20 min</td>
</tr>
<tr>
<td>Total fuel used</td>
<td>1110 g</td>
<td>2670 g</td>
</tr>
</tbody>
</table>

VI. POLYNOMIAL REGRESSION FOR TRANSIENT STATES

Before presenting our ANN based approach it is useful/necessary to investigate the result of the application of a more conventional method to transient state data. Training set points were approximated using a 3rd degree polynomial surface (Fig. 4). The train scores for this model are: RMSE = 1.33, $R^2 = 0.811$ and MAE = 1.02. In the following chapters the polynomial regression model trained on transient state data is referred to as the TSM (Transient State Model). This model can also be regarded as the GEC obtained from non steady states using a steady state method. Based on the Least Squared Criterion, the conventional approach does not seem to be robust enough when it comes to outliers or high leverage points. The polynomial surface fit is highly sensitive to the observation distribution in the n-T domain and it may cause overfitting and an incorrect surface shape. A solution to this problem may be the application of more sophisticated regression models, e.g. LASSO regression, Elastic Net. However, this is beyond the scope of the presented work.
Figure 4. Transient state model (TSM): 3rd degree polynomial surface approximation of transient states observations (red points).

VII. NEURAL NETWORK MODEL

In the analysis of transient states of combustion engines, the inclusion of information related to speed and torque changes in the model, generally results in higher accuracy of the model. In [27] as additional input variables, derivatives of \( n \) and \( T \) were used. In [20] a number of values from a time window before observation were considered. However, the examined engines differ from the one presented in this paper in terms of size, ignition type and even emission regulations which they need to comply to. There is no direct suggestion as to which additional values preceding observations should be included in order to obtain an accurate model. Moreover, it seems that this issue depends on the inertia of the engine, which is directly related to the engine displacement. In this research it was assumed, that changes should be consider up to 1.5 s before the measuring point.

The neural network model used in this study is the ANN. This model is based on a multi-layer feedforward artificial neural network, that is trained with stochastic gradient descent using back-propagation. A rectifier was used as activation function. In order to produce an optimal ANN, 5 sets of predictors and 5 numbers of neurons in one hidden layers were analyzed resulting in 25 models. These are summarized in Table 2. A subscript in predictors notation indicates the time in milliseconds before the observation. Additionally, \( \Delta \) indicates that the variable corresponds to the difference between the actual value and the value x milliseconds before; e.g. \( T_{300} \) corresponds to torque value 500 ms before the observation, and \( \Delta T_{1500} \) corresponds to difference between \( n \) and \( \theta_{1500} \).

In ANN 1, only 2 input variables (\( n \) and \( T \)) are included and this is the model with closest similarity to the polynomial reference model in terms of predictors. As the polynomial regression is replaced by a neural network algorithm, nonlinear dependencies in the data are expected to be explained better. ANN 2 and ANN 4 consist of actual and preceding \( n \) and \( T \) values. Thus, they can be considered as models averaging actual observation values with preceding ones. However, thanks to the deep learning capabilities, nonlinear dependencies can be unveiled. In ANN 3 and ANN 5 from \( n \) and \( T \), differences between actual and preceding values are included. This idea seems to focus more on the actual observation and the changes for preceding \( \Delta t \).

Each neural network model was computed using the training set with a 10-fold cross-validation; the epochs were set to a 100. To measure the performance of the model RMSE was used (Table 3). As expected, ANN 1 performed significantly worse than other models, with the 3 times higher values of RMSE. Higher number of predictors led to better scores, what is shown by lower RMSE for ANN 4 and 5 than for ANN 2 and 3. The number of neurons in hidden layer does not have significant impact on the models’ performance. RMSE scores for each \( N_{\text{hidden}} \) differ no more than one standard deviation from each other. Thus, it was assumed that there is no point in using the model with \( N_{\text{hidden}} \) higher than 5. ANN 5 was chosen to be applied to the validation dataset, as it performed best with RMSE equal to 0.097, and it is referred to the ANN model later in the text.

VIII. VALIDATION DATASET – RESULTS

The level of generalization the polynomial regression model and ANN model was compared using a validation dataset consisting of 65 load cycles, described in section VII. For each load cycle the RMSE, \( R^2 \) and MAE were calculated. Additionally an absolute relative error score was introduced, as it focuses on the engineering application of the model. This is given by the equation:

\[
\Delta e_{\text{abs,rel}} = \left| \frac{\sum_{i=1}^{\Delta T} (G_{e_{i}} \cdot \Delta t) - \sum_{i=1}^{\Delta T} (G_{e_{i,\text{pred}}} \cdot \Delta t)}{\sum_{i=1}^{\Delta T} (G_{e_{i}} \cdot \Delta t)} \right|
\]

Where, \( G_{e_{i}} \) – corresponds to the measured fuel rate, \( G_{e_{i,\text{pred}}} \) – to the predicted fuel rate, \( \Delta t \) – to the time step between \( i \)-th and \( i+1 \) to the actual observation. The level of accuracy of the models was based on the total fuel consumption prediction for the load cycle with no consideration for goodness of fit on a particular observation. Mean scores from 65 load cycles for each model are presented in Table 4 together with test scores from the training process. Examples of real and predicted fuel rates for each model and 4 engine speeds are presented in Fig. 5. Engine speeds 750, 850, 1000 and 1150 rpm were chosen to be shown, as they correspond to the different areas of the engine performance map.

In terms of statistical scores the ANN model outperforms the TSM model. Both the RMSE and the MAE scores are almost three times and more than two times smaller, accordingly for the ANN model. The \( R^2 \) values of the ANN model are close to 1 compared to TSM model which are equal to 0.83. The ANN model exhibits also much improved fit to the \( G_e \), revealing an almost overlaying of the real and predicted \( G_e \) values (As seen in Fig. 5). On the contrary, the TSM model fails to predict accurately the \( G_e \), when fast changes of \( G_e \) appear. This is
not a crucial problem as long as the total fuel consumption is predicted well. The mean absolute relative error is 3.6±2.8 % and 2.4±1.2 % for the TSM and ANN models respectively. These are reasonable values for engineering applications. However, considering the maximal $e_{abs,rel}$ in some cases the TSM can be inaccurate by as much as 10%. It is worth noting, based on the presented results, that the ANN model scores are better in all assumed criteria.

TABLE II. NEURAL NETWORK MODELS: 5 SETS OF PREDICTORS, 5 SETS OF HIDDEN LAYERS EACH

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>$N_{\text{input}}$</th>
<th>$N_{\text{hidden}}$</th>
<th>$N_{\text{output}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN 1</td>
<td>$n, M$</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>ANN 2</td>
<td>$n, M, n_{1000}, M_{1000}$</td>
<td>4</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>ANN 3</td>
<td>$n, M, \Delta n_{1000}, \Delta M_{1000}$</td>
<td>4</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>ANN 4</td>
<td>$n, M, n_{500}, n_{1000}, n_{1500}, M_{500}, M_{1000}, M_{1500}$</td>
<td>8</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>ANN 5</td>
<td>$n, M, \Delta n_{500}, \Delta n_{1000}, \Delta M_{500}, \Delta M_{1000}, \Delta M_{1500}$</td>
<td>8</td>
<td>30</td>
<td>1</td>
</tr>
</tbody>
</table>

TABLE III. RMSE$_{\text{abs}}$ SCORES FOR 25 NEURAL NETWORK MODELS

<table>
<thead>
<tr>
<th>$N_{\text{hidden}}$</th>
<th>ANN 1</th>
<th>ANN 2</th>
<th>ANN 3</th>
<th>ANN 4</th>
<th>ANN 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.306 (0.026)</td>
<td>0.116 (0.023)</td>
<td>0.118 (0.031)</td>
<td>0.108 (0.09)</td>
<td>0.097 (0.024)</td>
</tr>
<tr>
<td>10</td>
<td>0.307 (0.028)</td>
<td>0.112 (0.023)</td>
<td>0.111 (0.024)</td>
<td>0.099 (0.045)</td>
<td>0.092 (0.026)</td>
</tr>
<tr>
<td>15</td>
<td>0.306 (0.032)</td>
<td>0.111 (0.022)</td>
<td>0.109 (0.028)</td>
<td>0.095 (0.043)</td>
<td>0.091 (0.025)</td>
</tr>
<tr>
<td>20</td>
<td>0.307 (0.03)</td>
<td>0.111 (0.022)</td>
<td>0.107 (0.028)</td>
<td>0.096 (0.06)</td>
<td>0.087 (0.021)</td>
</tr>
<tr>
<td>30</td>
<td>0.307 (0.031)</td>
<td>0.109 (0.024)</td>
<td>0.106 (0.027)</td>
<td>0.087 (0.029)</td>
<td>0.087 (0.022)</td>
</tr>
</tbody>
</table>

Figure 5. Real fuel rate $G_f$ (black) and predicted (red) in validation test for $n=750, 850, 1000$ and $1150$ rpm using TSM and ANN

TABLE IV. MODEL SCORES COMPARISON

<table>
<thead>
<tr>
<th>Validation load score</th>
<th>RMSE$_{\text{cv}}$ (M)</th>
<th>$R^2$ (M)</th>
<th>$\sigma_{\text{train}}$ (M)</th>
<th>$\sigma_{\text{test}}$ (M)</th>
<th>$\sigma_{\text{error}}$ (M)</th>
<th>Max $\sigma_{\text{error}}$ (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSM</td>
<td>1.33</td>
<td>0.96 (0.14)</td>
<td>0.83 (0.04)</td>
<td>-1.5% (4.3)</td>
<td>3.6% (2.6)</td>
<td>10.2%</td>
</tr>
<tr>
<td>ANN</td>
<td>0.39</td>
<td>0.32 (0.12)</td>
<td>0.99 (0.01)</td>
<td>-0.6% (2.3)</td>
<td>2.4% (1.3)</td>
<td>4.4%</td>
</tr>
</tbody>
</table>

IX. DISCUSSION

The proposed ANN method should be validated in other diesel engines to confirm its level of generalisation. No more tests have been done so far on the test bench, however using CAN data registered during real operation of two MAN trucks, some conclusions can be drawn. In both trucks, equipped with diesel engines (12.4 dm$^3$ (309 kW) and 10.5 dm$^3$ of disp. (235 kW) respectively) the delay $\Delta t$ does not exceed 1.5 s. Thus, it can be concluded that the model ANN 5 (including 8 predictors) should perform similarly to the Scania engine, examined in previous sections. The neural network should adjust its parameters properly to the magnitude of $\Delta t$ for a particular engine.

The ANN model allows for good fit in transient states without need for a deep understanding of dependencies between $n, T, \Delta n, \Delta T$ and $G_f$, for a particular engine. Expanding the conventional model, by formulating additional equations manually appears to be ineffective and it is unlikely that such model will result in a better fit. ANN model with 8 predictors and 5 neurons in hidden layer consist of much larger set of equations. Its weights and biases are automatically optimized by means of large number of iterations in backpropagation. It allows for the detection of the nonlinear dependencies more efficiently. Thus, it can be concluded that exploiting modern Machine Learning methods in order to obtain GEC for fuel prediction purposes has at least two main benefits:

- results in a better fit than conventional models
- does not require analytical/formulated explanation of analyzed objects as long as the model formulates optimal equations automatically by means of the learning process

The neural network model was trained using only 510 observations off 6000 total observations. Even with this reduced data set, the model performed better than the equivalent polynomial regression model. This means that the theoretically 60 s (including 15% for the test data set) transient state test is sufficient to obtain good a ANN model. However, in reality in such short time it may be difficult to examine the engine in the whole desired range due to its inertia. Nonetheless, the results shows the potential of using ANN to solve real world problems. Data collected during regular operation of the vehicle usually has a non-uniform distribution in the n-T domain. As a uniform distribution is desired for building both an ANN and a polynomial regression models, a manual observation selection or clustering algorithm can be applied. Obtaining uniform distribution leads to the reduction of sample size. However, as shown in this paper even a relatively small
dataset can result in an ANN model with sufficient accuracy in fuel consumption prediction, which performs better than a polynomial regression. Further research needs to be done to investigate the quantitative influence of non-uniform distribution in sample.

X. CONCLUSIONS

A neural network model used to predict fuel consumption of combustion engine presented in this paper resulted in higher accuracy and better fit compared to a conventional polynomial regression model. The model was trained based on transient state data and validated on real load cycles, scoring mean RMSE=0.32 and mean $\varepsilon_{\text{abs. rel}}=2.4\%$. The presented method can be applied to fuel prediction in trucks, where the engine parameters can be logged using a CAN bus during daily operation. An engine performance model then can be used to calibrate a power receiver mounted via PTO, e.g. by choosing optimal engine speeds. The PTO can be used throughout a large part of the work day of the vehicle. In [28] it was shown, that 15\% of the time the PTO was engaged and the engine speed was set to constant 900 rpm. 50\% of the time the engine was idling with PTO engaged. As a result, the power receiver mounted via PTO was active during 65\% of the daily operation time, consuming approximately 30\% of the total fuel used. Thus, an application of the method presented in this paper may lead to reduction of the fuel consumption and exhaust gases emission during significant part of daily operation time.

The same duties can be performed by using similar vehicles, but equipped with various engines. Along with the simplicity of the applied method, it is possible to obtain engine models from different chassis in an easy and cost effective manner in order to compare their efficiency in terms of operation in low torque range. This is an important and useful information for truck users, fleet management and utility truck body manufacturers.

REFERENCES

[14] ISO 15550:2016, Internal combustion engines — Determination and method for the measurement of engine power — General requirements