Engine Fuel Consumption Modelling using Prediction Error Identification and On-road Data

Anil K. Madhusudhanan, Xiaoxiang Na, Daniel Ainalis and David Cebon

***Abstract*—Engine modelling is an important step in predicting the fuel consumption of a vehicle. Existing methods in the literature require dedicated tests on a test track or on a chassis dynamometer or they require measurements from several days of vehicle operation. This article proposes a new method to model fuel flow rate of a diesel engine and a compressed gas engine using prediction error identification and on-road data collection. The model inputs are the engine torque and speed. The on-road vehicle data was collected during normal transport operations. The identification data set was approximately 99% shorter than the baseline method. The proposed method is applicable for other types of vehicles, including electric vehicles. The identified engine models have less than 1.3% mean error and 2.5% RMS error.**

***Index Terms*—Engine model, prediction error identification, vehicle fuel consumption.**

I. Introduction

ASSESSING fuel consumption benefits of different interventions, e.g. aerodynamic features, low rolling resistance tyres and cruise control systems, is a necessary decarbonisation activity in the transport sector. For example, evaluating the effects of semi-trailer modifications on fuel consumption and carbon emissions of heavy goods vehicles [1], and analysing the effects of biomethane as a truck fuel on carbon emissions and lifetime vehicle cost [2]. In the recent years, there have also been several work on exploiting autonomous driving technologies to reduce fuel consumption and carbon emissions [3]–[5]. A vehicle fuel consumption model is often required for such analyses. To develop fuel consumption models, there are several methods in the literature [2], [6]–[16].

Most of these methods do not model the engine in detail using empirical methods [6], [7], [9]–[11], [13], [14], [16]. In [6], second order polynomials were proposed to predict fuel consumption. The effect of different driving patterns on fuel consumption was studied in [7] using a polynomial based model. In [9], a vehicle mass dependent fuel consumption model was developed using statistical data. Fuel consumption factors such as engine speed and torque, road elevation, aerodynamic drag and rolling resistance were ignored in this work. A model with the fuel consumption rate as a function of vehicle speed and acceleration was proposed in [10]. The model development involved fuel consumption estimation using data from a portable emission measurement system. In [11], a fuel consumption map, as a function of vehicle speed and acceleration, was developed using fuel consumption and emission measurements. This fuel consumption map development required laboratory testing using a chassis dynamometer. In [13], data from a vehicle's on-board diagnostic port, which can be measured on-road, was used to develop a fuel consumption model as a function of the vehicle speed, acceleration and gas pedal position. In [14], a fuel consumption model using vehicle speed and altitude data from a GPS receiver was proposed. This model development assumed constant engine efficiency. In [16], a framework to develop computationally efficient vehicle energy consumption models was proposed. It considers vehicle parameters, drive cycles in which the vehicle operates and vehicle mass. But it does not use an engine or motor model.

Some of the methods in the literature do consider engine models [2], [8], [12], [15]. These engine models are mainly in the form of steady state engine maps, which are empirical look-up tables of fuel consumption as a function of engine speed and torque. In [8], a fuel consumption model, including an engine map, was developed using dedicated tests, which required the vehicle to be taken out of business operations. In addition to the vehicle speed and acceleration, the model proposed in [12] uses a steady state engine map, created using test data from a chassis dynamometer. In [15], a commercial power train simulation software, AVL Cruise, was used to develop a fuel consumption model using the longitudinal equations of motion and an engine map. This work used test data from a chassis dynamometer to create the engine map. To reduce the experimental cost, the number of data points in the engine map was limited to less than 40. In [2], fuel consumption models of Compressed Gas (CG) and diesel trucks were developed using on-road data, which was collected from the vehicles' Fleet Management System (FMS) interface. Steady-state engine maps, using data from several days of transport operations, were developed in this work.

Having an engine model can improve the accuracy of a fuel consumption model. However, most of the engine modelling methods in the literature require dedicated tests [8], [12], [15], either on a test track or on a chassis dynanometer, which requires significant time and cost. The method proposed in [2] does not require such dedicated tests as on-road data, collected when the vehicle was performing routine transport operations, was used. However, it required several days of data and the engine model only captured the steady-state charactersitics, i.e. it ignored the dynamic characteristics. This article proposes an

Anil is with the Mechatronics Research Group, Department of Mechanical Engineering, University of Southampton, Southampton SO17 1BJ, UK. Correspondance email: anil@soton.ac.uk

Xiaoxiang, Daniel and David are with the Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK.

This research was partly supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant EP/R035199/1: Centre for Sustainable Road Freight 2018-2023.

Manuscript accepted on April 12, 2022.

engine modelling method to address these problems.

The main contributions of this article are:

- A method to model fuel flow rates of a diesel engine and a Compressed Gas (CG) engine using Prediction Error Identification (PEI) and on-road data collection is proposed. The method is also applicable to other vehicle types, including electric vehicles.
- Dynamic models of a diesel engine and a CG engine were identified and validated using different data sets.
- The identified diesel engine model and CG engine model were compared against the baseline models in [2], which were also developed using on-road data.

This article is structured as follows. Section II describes how the vehicles were instrumented and how the on-road data was collected. Data preprocessing before identifying the engine models is explained in Section III. In Section IV, the model identification method is described. The identification and validation results are shared in Section V. It also contains comparisons between the identified engine models and the baseline models in [2]. Finally, the conclusions and future work are mentioned in Section VI. The identified diesel engine model and CG engine model are given in the article Appendix.

II. Instrumentation and Data Collection

The propsed method was used to model two different articulated heavy goods vehicles with maximum gross vehicle masses of 44 t. One had a tractor unit with a Euro 6 diesel engine, while the other had a spark ignition engine that ran on Compressed Gas (CG). The vehicles were measured in-service with a comprehensive data logging system.

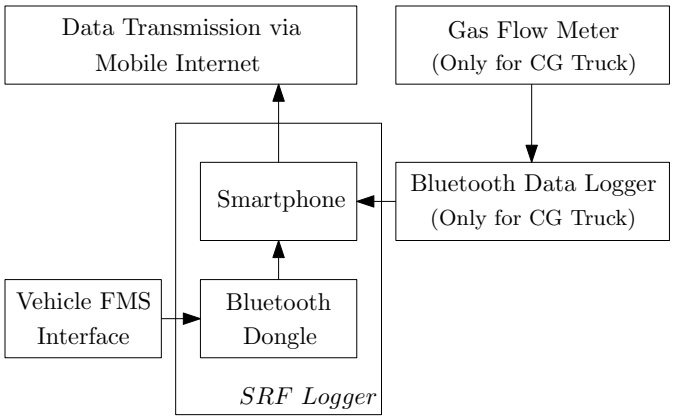


Fig. 1: A schematic diagram of the vehicle instrumentation.

Fig. 1 shows a schematic diagram of the instrumentation and Fig. 2 shows one of the instrumented vehicles. The instrumentation uses an SRF Logger, developed at the Centre for Sustainable Road Freight (SRF) at the University of Cambridge. It consists of a smartphone, a software app written for the smartphone and a VIACONT Bluetooth dongle. The Bluetooth dongle connects to the vehicles Fleet Management System (FMS) interface and transmits the FMS data to the smartphone. The FMS data includes the accelerator pedal position, brake pedal position, vehicle speed, engine speed, engine torque and fuel flow rate. The smartphone obtains GPS coordinates of the vehicle using its internal GPS sensor. For the vehicles used in this work, the FMS interface reported the engine torque as a percentage of the maximum torque. Therefore, it was multiplied by the maximum torque, which was extracted from the On-Board Diagnostic (OBD) port.

Fig. 2: One of the instrumented vehicles, a Scania P340 Gas.

As the FMS data from the CG truck (a Scania P340 Gas) did not include the fuel flow rate, it was instrumented with a MASS-STREAM D-6300 gas flow meter and a Bluetooth data logger (see Fig. 1). The gas flow meter uses a constant temperature anemometer principle. It can measure in the range $[2.5, 125]$ kg/h within 2% full scale accuracy and comes with a factory calibration certificate. Its output signal is in the range $[0, 10]$ V. The fuel flow meter was connected to an analog input of the Bluetooth data logger, BTH-1208LS, which communicated via Bluetooth with the SRF Logger. For the diesel truck (a Scania P320 Diesel), the fuel flow rate was available from the vehicle's FMS interface.

The vehicles were manufactured by Scania AB and are part of a fleet operated by the John Lewis Partnership Plc, UK. The data collection was performed during normal transport operations and did not require any intervention from the driver. Once the vehicles began to move, the SRF Loggers acquired and automatically uploaded the data via 3G or 4G mobile data connection and sent the data to a database server at the University of Cambridge.

III. Data Preprocessing

The engine models have two inputs, the engine speed and torque. The model output is the fuel flow rate. As mentioned in the introduction, many researchers have used vehicle speed, longitudinal acceleration, vehicle mass or throttle pedal position as an input for the fuel consumption model. But the model's complexity, i.e. the order, non-linearities and uncertainties, may increase if these inputs are used instead of the engine torque and speed. This increase in complexity is caused by the additional vehicle dynamics and effects of external time varying factors such as road slope and wind speed. As a consequence, using the vehicle speed, acceleration, mass or throttle pedal position as an input, it will be difficult to identify a model that meets the cross-correlation and auto-correlation criteria, which are discussed in the next section. On the other hand, considering the engine speed and torque

as the model inputs, it is feasible to identify a dynamic model that meets these criteria.

The input and output measurements were preprocessed before identifying a model. The identified models are Linear Time Invariant (LTI) and the engine idling state with an engine speed of 600 rpm was chosen as the linearisation point. Therefore, average values of the input and output measurements, when the engine was idling, were subtracted from the respective measurements. The selected identification data set was approximately 1.5 hrs long with a sampling interval of 0.3 s. The diesel truck's identification data set with the preprocessed input and output measurements is shown in Fig. 3. Fig. 4 shows the validation data set for the diesel truck, which is clearly different from the identification data set in Fig. 3.

Similar to Fig. 3 and Fig. 4, identification and validation data sets for the CG truck were also preprocessed before identifying a dynamic model for the CG engine.

IV. Model Identification

This section describes the model identification method. The following system model is considered:

$$Y(z) = G_0(z)U(z) + H_0(z)E(z). \tag{1}$$

Here $Y(z)$ is the z-transform of the system output (fuel flow rate), $G_0(z)$ is the true discrete-time transfer function between the system inputs (engine speed and torque) and output, $U(z)$ is the z-transform of the system inputs, $H_0(z)$ is the true discrete-time transfer function to shape white noise, and $E(z)$ is the z-transform of white noise. In (1), the second term on the right hand side is to consider the effect of noise in the measurements so that an accurate transfer function between the system inputs and output can be identified.

In the identification method used, i.e. in Prediction Error Identification (PEI), candidate parametric transfer functions, $G(z,\theta)$ and $H(z,\theta)$, are chosen for the true transfer functions, $G_0(z)$ and $H_0(z)$. Here θ is a vector, containing the model parameters. Using the parametric transfer functions, the following prediction model is used:

$$\begin{aligned}\hat{Y}(z,\theta) =& H(z,\theta)^{-1}G(z,\theta)U(z) \\ &+ (1 - H(z,\theta)^{-1})Y(z).\end{aligned} \tag{2}$$

Here $\hat{Y}(z,\theta)$ is the z-transform of the prediction model output.

Using the true model in (1) and prediction model in (2), a prediction error model, i.e. a model for the error between measured and predicted outputs, is shown in Fig. 5 and can be calculated as:

$$\begin{aligned}\mathcal{E}(z,\theta) &= Y(z) - \hat{Y}(z,\theta) &(3)\\ &= H(z,\theta)^{-1}\left[Y(z) - G(z,\theta)U(z)\right]. &(4)\end{aligned}$$

Here $\mathcal{E}(z,\theta)$ is the z-transform of the prediction error.

The time domain equivalent of (3) is:

$$\epsilon(k,\theta) = y(k) - \hat{y}(k,\theta). \tag{5}$$

Here $\epsilon(k,\theta)$ is the prediction error at sample k, $y(k)$ is the system output at sample k and $\hat{y}(k,\theta)$ is the predicted system output at sample k.

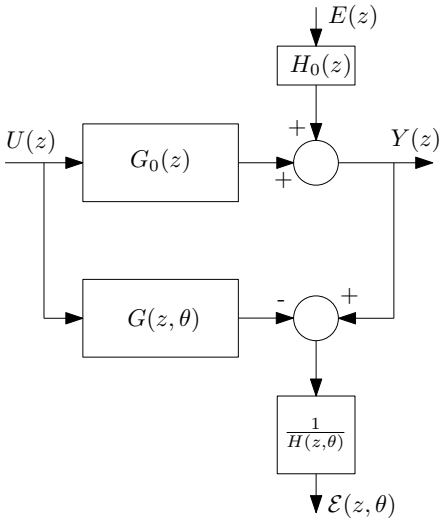


Fig. 5: A schematic diagram of the prediction error model.

An estimate of the parameter vector with N data samples, $\hat{\theta}_N$, is found by solving the following optimisation problem to minimise the sum of squares of prediction error values:

$$\begin{aligned}\hat{\theta}_N &= \underset{\theta}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N}\epsilon(k,\theta)^2 &(6)\\ &= \underset{\theta}{\operatorname{argmin}} \frac{1}{N}\sum_{k=1}^{N}\left[y(k) - \hat{y}(k,\theta)\right]^2. &(7)\end{aligned}$$

To identify an engine model, the following Box-Jenkins model structures were used for the parametric transfer functions, $G(z,\theta)$ and $H(z,\theta)$:

$$\begin{aligned}G(z,\theta) &= \frac{z^{-n_k}B(z,\theta)}{F(z,\theta)} \text{ and} &(8)\\ H(z,\theta) &= \frac{C(z,\theta)}{D(z,\theta)}, \text{ where} &(9)\end{aligned}$$

$$\theta = [b_0 \ ... \ b_{n_b-1} \ f_1 \ ... \ f_{n_f} \ c_1 \ ... \ c_{n_c} \ d_1 \ ... \ d_{n_d}], \tag{10}$$

$$\begin{aligned}B(z,\theta) &= b_0 + b_1 z^{-1} + \ ... \ + b_{n_b-1}z^{-n_b+1}, &(11)\\ F(z,\theta) &= 1 + f_1 z^{-1} + \ ... \ + f_{n_f}z^{-n_f}, &(12)\end{aligned}$$

$$\begin{aligned}C(z,\theta) &= 1 + c_1 z^{-1} + \ ... \ + c_{n_c}z^{-n_c} \text{ and} &(13)\\ D(z,\theta) &= 1 + d_1 z^{-1} + \ ... \ + d_{n_d}z^{-n_d}. &(14)\end{aligned}$$

As the engine model has two inputs and one output, the parametric transfer function for $G_0(z)$ is a matrix transfer function:

$$\begin{aligned}G(z,\theta) &= \begin{bmatrix}G_w(z,\theta) & G_T(z,\theta)\end{bmatrix} &(15)\\ &= \begin{bmatrix}\frac{z^{-n_w}B_w(z,\theta)}{F_w(z,\theta)} & \frac{z^{-n_T}B_T(z,\theta)}{F_T(z,\theta)}\end{bmatrix}. &(16)\end{aligned}$$

Here $G_w(z,\theta)$ is the transfer function between the engine speed and fuel flow rate, and $G_T(z,\theta)$ is the transfer function between the engine torque and fuel flow rate. These two transfer functions include the system dynamics. $G_w(z,\theta)$ was parameterised so that B_w has 4 parameters, F_w has 5 parameters and $n_w = 3$. $G_T(z,\theta)$ was parameterised so

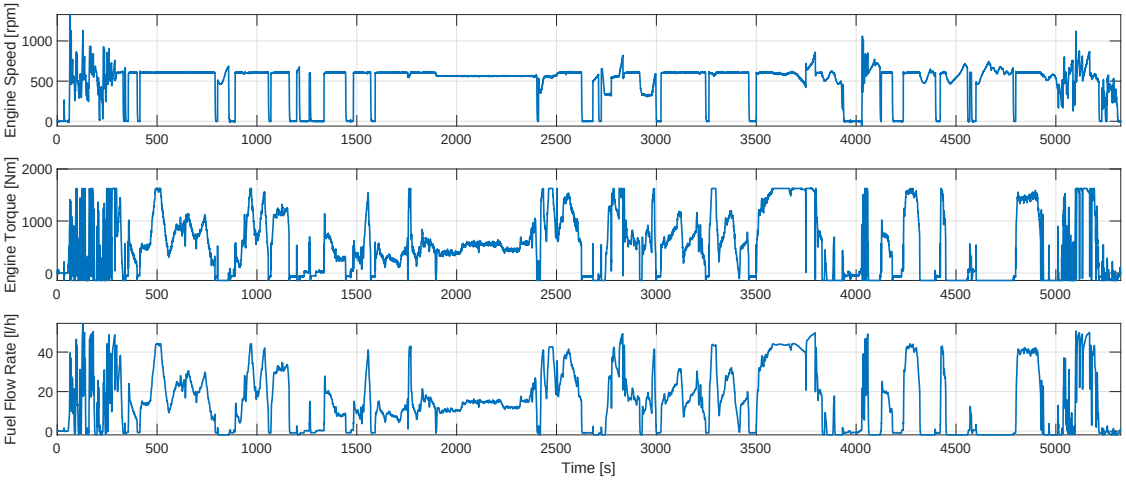


Fig. 3: The diesel truck's model identification data set.

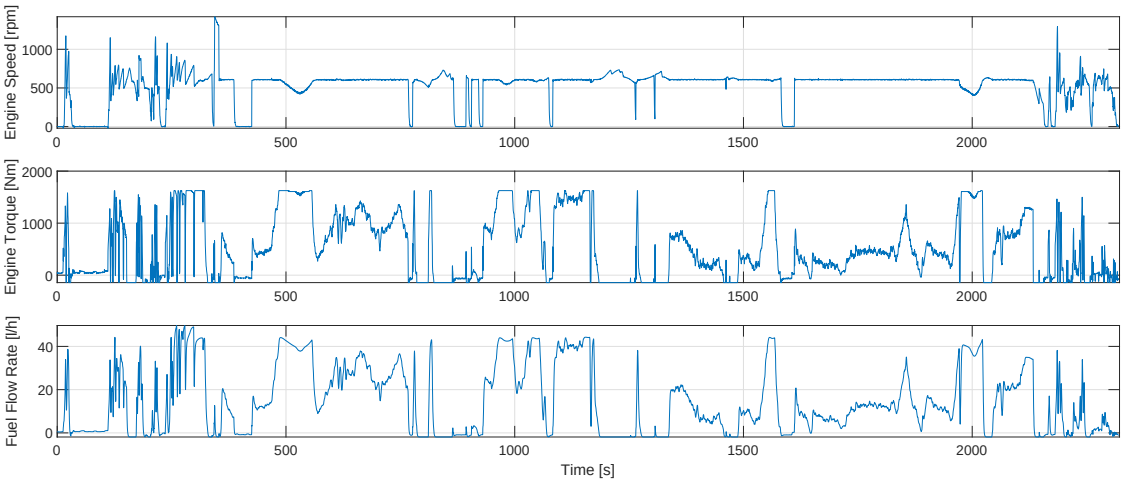


Fig. 4: The diesel truck's model validation data set.

that B_T and F_T have 5 parameters each, and $n_T = 0$. The candidate engine model is of the following form:

$$\hat{F}(z,\theta) = \begin{bmatrix} G_w(z,\theta) & G_T(z,\theta) \end{bmatrix} \begin{bmatrix} W(z) \\ T(z) \end{bmatrix} + H(z,\theta)E(z). \tag{17}$$

Here $\hat{F}(z,\theta)$ is the z-transform of the predicted fuel flow rate, $W(z)$ is the z-transform of the engine speed and $T(z)$ is the z-trasnform of the engine torque. $H(z,\theta)$ was parameterised so that C and D have 11 parameters each. The model form is shown in Fig. 6.

The optimisation problem in (7) was solved with the identification data set using the *polyest* command in MATLAB. The identified transfer functions, $G_w(z,\theta)$, $G_T(z,\theta)$ and $H(z,\theta)$, have the same model structure as the true system transfer functions if the auto-correlation of the prediction error, and the cross-correlation between the prediction error and the system inputs, are in their 99% confidence region. These conditions

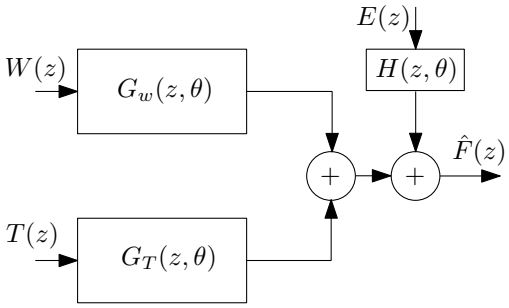


Fig. 6: The candidate model form.

are met if the following two inequalities are true [17], [18]:

$$\left|\hat{R}_\epsilon(h)\right| < 2.58\sqrt{\frac{\hat{R}_\epsilon(0)}{N}} \text{ and} \tag{18}$$

$$\left|\hat{R}_{\epsilon u}(h)\right| < 2.58\sqrt{\frac{\sum_{i=-\infty}^{\infty}\hat{R}_\epsilon(i)R_u(i)}{N}}. \tag{19}$$

Here $\hat{R}_\epsilon(h)$ is the auto-correlation of the prediction error with

a sample lag of h, $\hat{R}_{\epsilon u}(h)$ is the cross-correlation between the prediction error and the system inputs with a sample lag of h, and $R_u(i)$ is the auto-correlation of the system inputs with a sample lag of i. The numbers of parameters in the numerator and denominator of G_w, G_T and H were tuned so that these two conditions are met.

V. Results

This section shares the model identification and validation results for the diesel and gas engines.

A. Diesel Engine

The identified dynamic model of the Scania P320 diesel engine is given in the Appendix. For the identified model, Fig. 7 shows the auto-correlation of the prediction error and cross-correlation between the prediction error and the system inputs.

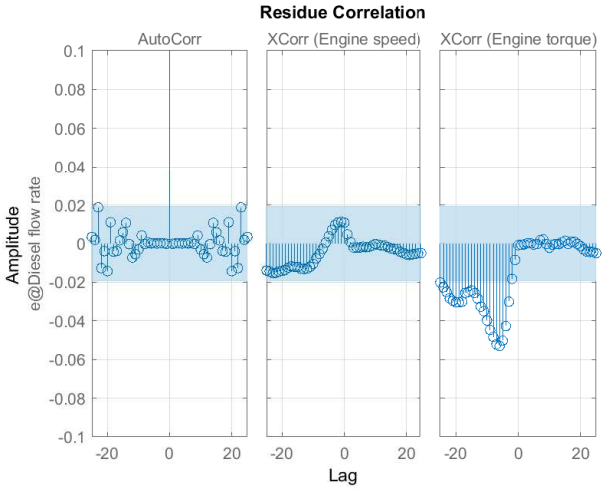


Fig. 7: Auto-correlation of the prediction error and cross-correlation between the prediction error and the model inputs, for the diesel engine model.

In Fig. 7, the 99% confidence regions, described by (18) and (19), are highlighted in blue. From the auto-correlation plot on the left, it is clear that the prediction error's auto-correlation values for all non-zero lag are within the confidence region, i.e. the criterion in (18) is met.

In Fig. 7, in the cross-correlation plot in the middle for engine speed, the cross-correlation values between the prediction error and engine speed are within the confidence region for all lag values, i.e. the criterion in (19) is met. In the cross-correlation plot on the right for engine torque, the cross-correlation values between the prediction error and engine torque are within the confidence region for all non-negative lag values. This is indeed acceptable as the cross-correlation values between the prediction error and a system input can be outside the confidence region for negative lag values if there is feedback in the system [19], i.e. if the current diesel flow rate (system output) can affect future engine torque (one of the system inputs), which indeed is the case for an internal combustion engine.

Fig. 8 compares the measured and modelled diesel flow rates for the identification data set in Fig. 3. Fig. 9 shows two zoomed in versions of Fig. 8. As mentioned in Fig. 8, for the identification data set, there is a 91% fit (normalised root mean square fitness) between the identified LTI model and measurements, which is quite good given the internal combustion engine is a non-linear system. In Fig. 9a, the model output deviates from the measurements at multiple peaks around $t = 130$ s and $t = 170$ s. These deviations are probably due to the engine non-linearities associated with gear shifts. In Fig. 9b, the model output correlates well with the measurements. When the fuel flow rate is close to zero or negative, the model output deviates slightly from the measurements. This behaviour is mostly caused by the engine non-linearities that are not captured by the identified LTI model.

In theory, choosing more model inputs will facilitate inclusion of the engines fuel consumption dependencies on other factors such as air temperature, air pressure, vehicle speed, vehicle acceleration, etc. However, as mentioned in Section III, the models order and uncertainties may increase if these dependencies are considered. In system identification theory, the aim usually is to use the lowest number of inputs and lowest model order to obtain a model that meets the cross-correlation and auto-correlation criteria in (18) and (19). If with the chosen number of inputs and model order, these criteria are not met, then the next step is to consider additional or other model inputs to consider their dependencies. But given these criteria are met with the chosen model inputs and model structure, and the time domain comparison has a 91% fit, from a system identification point of view, the model is sufficiently accurate. However, considering more model inputs may reduce the deviations around $t = 130$ s and $t = 170$ s in Fig. 9a. But it may also require a higher order linear model structure or require a non-linear model structure to capture the additional dynamics. But given the cross-correlation and auto-correlation criteria are met, and the time domain comparison has a 91% fit, it is an unnecessary step for the scope of this article. However, considering more model inputs is an interesting step for future research.

Fig. 10 compares the measured and identified model's diesel flow rates for the validation data set in Fig. 4. Fig. 11 shows two zoomed in versions of Fig. 10. As mentioned in Fig. 10, for the validation dataset, there is a 91.2% fit between the identified LTI model and measurements, which is similar to the fit in Fig. 8 for the identification data set. In the zoomed in versions in Fig. 11, similar to Fig. 9, the LTI model output deviates from the measurements around gear shifts and when the fuel flow rate is close to zero or negative, probably caused by the engine non-linearities.

The identified LTI diesel engine model was compared against the baseline diesel engine model in [2]. The baseline engine model was developed using on-road data from the same diesel truck, which was collected as described in section II. The baseline model takes the form of a two dimensional lookup table (engine map) with 'steady-state' values of the engine speed and engine torque as the inputs, and 'steady-state' values of the diesel flow rate as the output.

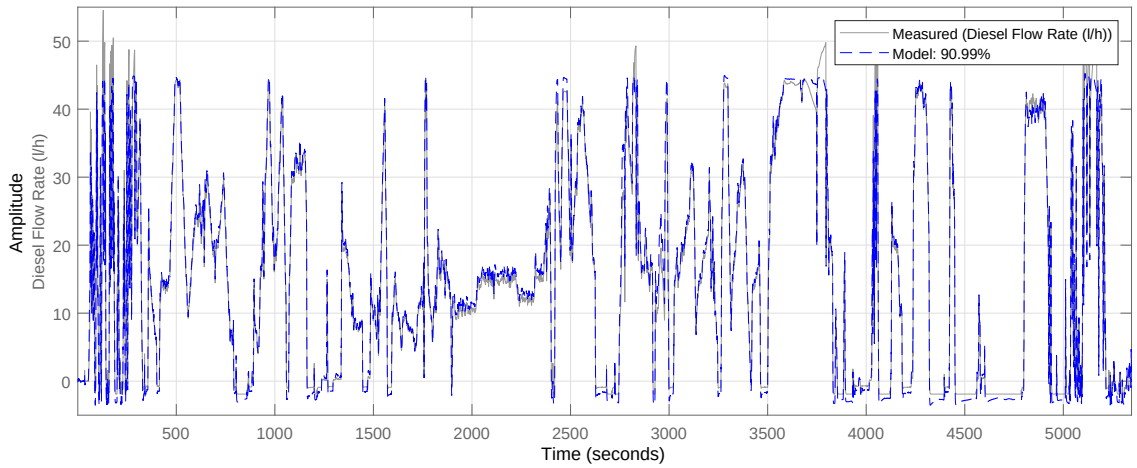
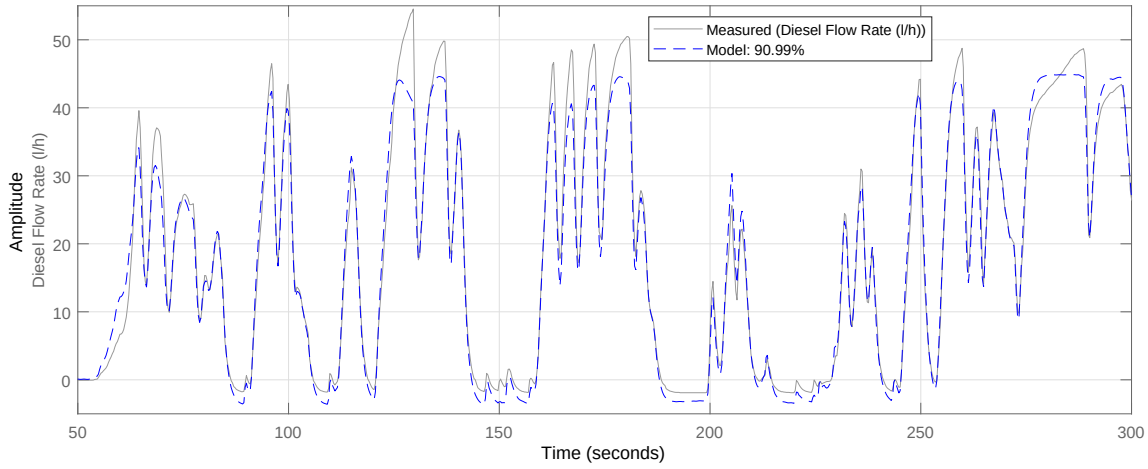
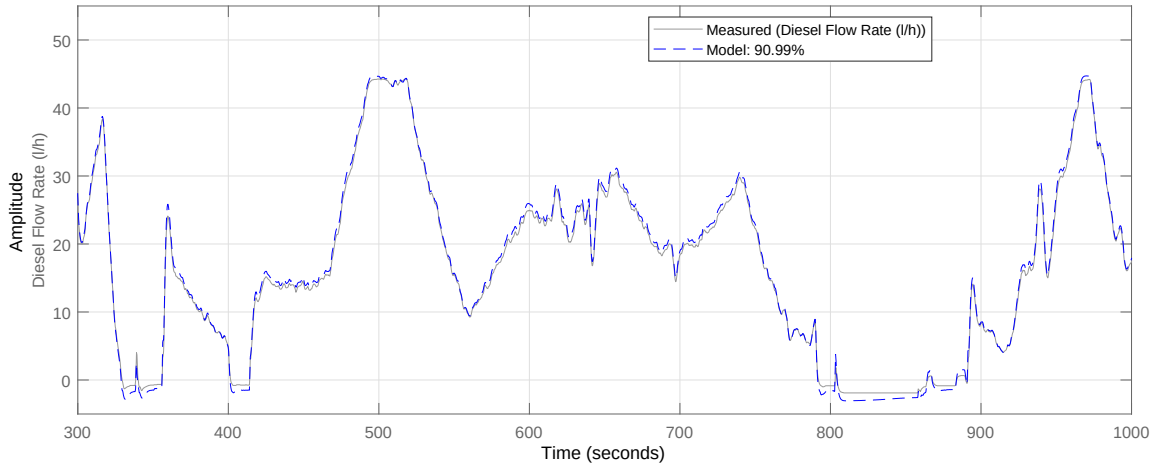


Fig. 8: The measured diesel flow rate and the identified model's fuel flow rate for the identification data set.



(a) From 50 s to 300 s.



(b) From 300 s to 1000 s.

Fig. 9: Zoomed in versions of Fig. 8, showing the measured and modelled diesel flow rates for the identification data set.

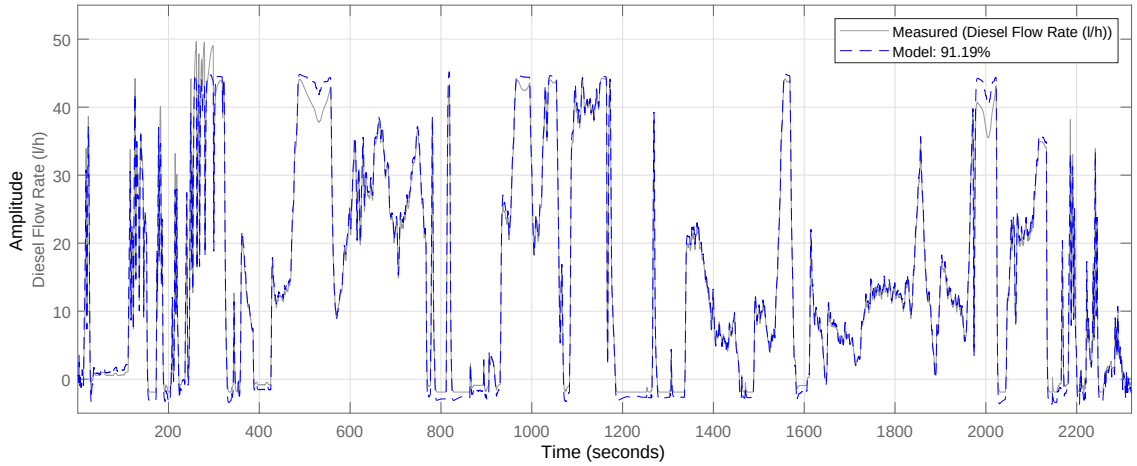
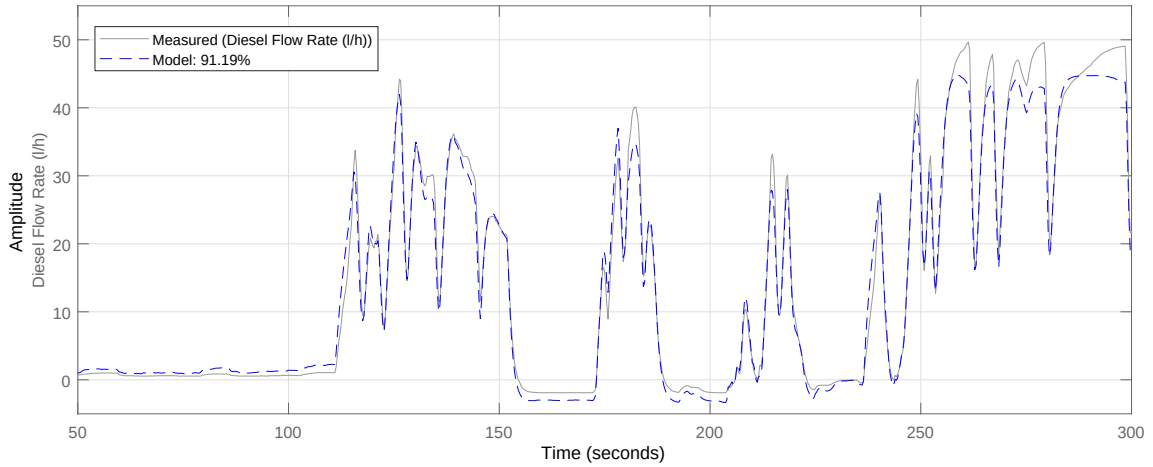
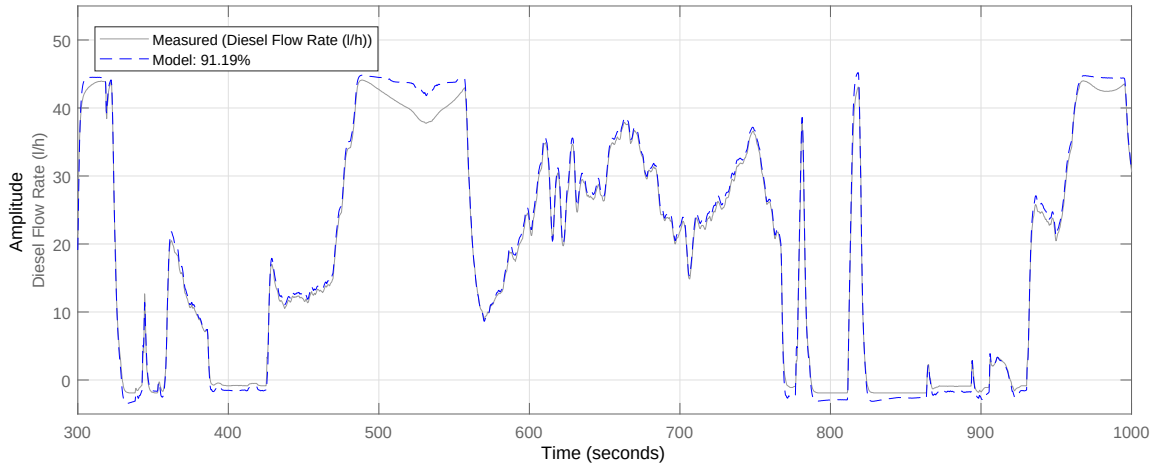


Fig. 10: The measured diesel flow rate and the identified model's diesel flow rate for the validation data set.



(a) From 50 s to 300 s.



(b) From 300 s to 1000 s.

Fig. 11: Zoomed in versions of Fig. 10, showing the measured and modelled diesel flow rates for the validation data set.

The lookup table data points were obtained from on-road measurements by locating data segments of three seconds or longer for which the vehicle speed was essentially constant. Whenever such an operating condition was found, the corresponding engine speed, engine torque and diesel flow rate were saved as a data point in the lookup table. This procedure was performed over all data collected for a period of one week. As mentioned in Section III, it is worth noting that the proposed modelling method used an identification data set of 1.5 hrs, which is approximately 99% shorter than the one week of data used by the baseline method. For intermediate input data points, the baseline model interpolates the diesel flow rate. For more details about the baseline model, see [2]. This baseline model was chosen as its development, similar to the proposed method, did not require chassis dynamometer tests or dedicated tests on a test track, which prevented extra costs for the truck operator. Table I shows the comparison between diesel consumption error of the baseline model and the identified model for 11 on-road drive cycles from multiple days. Each drive cycle lasts 30 minutes. The diesel consumption was calculated by integrating the model output, i.e. the diesel flow rate, with respect to time. For each drive cycle, the error was calculated using the following equation.

$$Error = \frac{\int_0^t f_{model}(t)dt - \int_0^t f_{measured}(t)dt}{\int_0^t f_{measured}(t)dt} \times 100\% \quad (20)$$

Here, f_{model} is the model's fuel flow rate, $f_{measured}$ is the measured fuel flow rate and t is the drive cycle duration. As shown in Table I, the identified model's mean error is close to zero, whereas the baseline model's mean error is 7.7%. The identified model's RMS error is approximately 70% lower than that of the baseline model. The identified model is more accurate than the baseline model because it was identified considering dynamic characteristics of the diesel engine using optimisation, whereas the baseline model was developed only using steady state values.

B. Gas Truck

The proposed methods for on-road data collection, data preprocessing and model identification are applicable for other vehicle types. To demonstrate this, a spark ignition gas engine model was identified. The model identification was done using the Compressed Gas (CG) truck data, collected and preprocessed as described in Section II and III. A dynamic LTI model of the Scania P340 gas engine was identified using the same method, described in Section IV.

Fig. 13 shows auto-correlation of the prediction error, and cross-correlation between the prediction error and system inputs, for the identified gas engine model. The 99% confidence regions are highlighted in blue. From the auto-correlation plot on the left, it is clear that the prediction error's auto-correlation values for all non-zero lag are within the confidence region, i.e. criterion (18) is met. In the cross-correlation plots in the middle and right, the cross-correlation values between the prediction error and system inputs are within the confidence region for all lag values, i.e. criterion (19) is met.

TABLE I: Comparison of the identified and baseline diesel engine models.

Drive Cycle	Baseline Model Error [%]	Identified Model Error [%]
1	-4.4	2.4
2	-6.1	-4.1
3	-6.8	-1.7
4	-7.5	-1.9
5	-5.6	2.5
6	-8.9	-0.5
7	-7.7	-0.7
8	-6.8	3.5
9	-8.9	-1.6
10	-15.2	0.0
11	-6.2	-3.6
Mean	7.7	-0.5
RMS	8.1	2.4

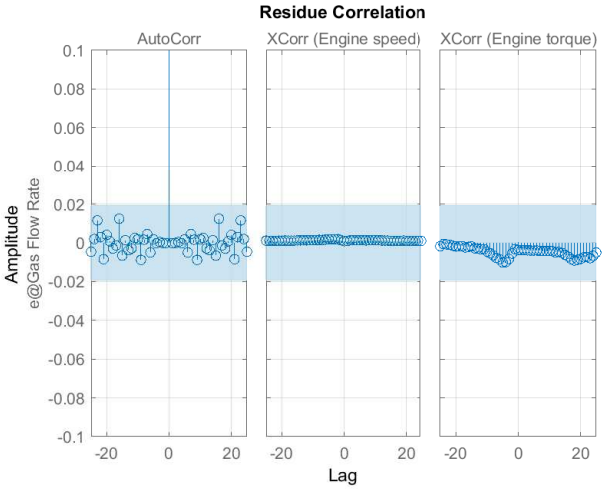


Fig. 13: Auto-correlation of the prediction error, and cross-correlation between the prediction error and model inputs, for the gas engine model.

Fig. 12 compares the measured and modelled gas flow rates for a sample data set. There is an 85% fit between the identified gas engine model and measurements, which is quite good given the gas engine is a non-linear system. Around $t = 2560$ s, the model output deviates from the measurements. This is probably caused by the engine non-linearities that are not captured by the identified LTI model.

The identified gas engine model was compared against the baseline gas engine model in [2]. The baseline engine model was developed using on-road data from the same CG truck, which was collected as described in section II. Similar to the baseline diesel engine model, the baseline gas engine model is in the form of a two dimensional lookup table (engine map) with 'steady-state' values of the engine speed and engine torque as the inputs, and 'steady-state' values of the fuel (gas) flow rate as the output. For more details about the baseline model, see [2].

Table II shows the comparison between gas consumption error of the baseline model and the identified model for 10 on-road drive cycles. For each drive cycle, the error was calculated using the same equation, which was used for the diesel engine and is shown in (20). Similar to the diesel engine

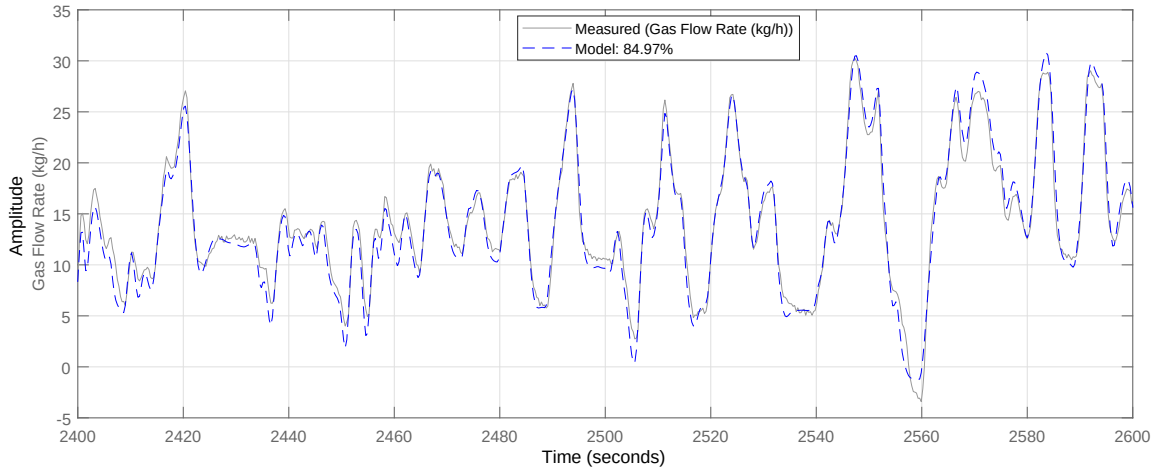


Fig. 12: The measured and identified model's gas flow rates for a zoomed-in version of the identification data set.

TABLE II: Comparison of the identified and baseline gas engine models.

Drive Cycle	Baseline Model Error [%]	Identified Model Error [%]
1	-4.4	-4.3
2	-1.3	-0.7
3	-4.6	-2.0
4	2.1	-0.3
5	-1.8	-1.1
6	-2.3	0.8
7	-1.2	1.1
8	-5.5	-1.3
9	-0.5	-0.3
10	-5.7	-3.8
Mean	-2.5	-1.2
RMS	3.5	2.1

case, these drive cycles are from multiple days and each of them lasts 30 minutes. The gas consumption was calculated by integrating the model output, i.e. the gas flow rate, with respect to time. As shown in Table II, the identified model's mean and RMS errors are lower than those of the baseline model. The identified model is more accurate than the baseline model as it was identified considering dynamic characteristics of the gas engine using optimisation, whereas the baseline model was developed only using steady state values.

The results for the diesel and gas engines demonstrate the applicability of the proposed method for on-road data collection, data processing and model identification to multiple vehicle types. The identified dynamic models of the diesel and gas engines are given in the Appendix. For another vehicle with the same engine and fuel injection system, the identified engine model should be valid. But for a vehicle with a different engine or fuel injection system, a new model should be identified using on-road data, which can be collected with an SRF Logger. It is thought that the same modelling method can be used to identify models of other vehicle types, e.g. to model the electric motor of an electric vehicle. For electric vehicles, instead of fuel flow rate, electric power will be used in the modelling process. As electric motors have faster dynamics than internal combustion engines, a lower data sampling time may be needed.

VI. Conclusions and Future Work

A method to model fuel flow rate of an internal combustion engine using on-road data collection and Prediction Error Identification is proposed in this article. The model inputs are engine speed and torque. Unlike the existing methods, the proposed method does not require dedicated tests on a proving ground or chassis dynamometer, which are expensive and time consuming.

In this work, the on-road data from a diesel truck was collected during normal vehicle operation, without affecting the vehicle operator. The identification data set was only 1.5 hours long. On the other hand, it took a full week of data collection to develop an engine map for the same vehicle [2]. The identified diesel engine model is Linear Time Invariant (LTI). The prediction error's auto-correlation values are within the 99% confidence bounds for all non-zero lags. The prediction error's cross-correlation values with the model inputs are within the 99% confidence bounds for all non-negative lags. In the time-domain, the LTI model output has a 91% fit with the output measurements in the identification data set. The identified diesel engine model was validated using a validation data set, different from the identification data set. The LTI model output has a 91.2% fit with the output measurements in the validation data set. For the 11 on-road drive cycles, the diesel consumption of the identified model (i.e. time integral of the model output) has an RMS error of 2.4%, compared to 8.1% for the baseline model [2]. For these drive cycles, the diesel consumption's mean error is -0.5%.

A gas engine model was also identified using the proposed method for on-road data collection, data processing and model identification. The identified gas engine model's mean and RMS errors, -1.2% and 2.1% respectively, are lower than those of the baseline model [2].

It is thought the the proposed modelling method is applicable to model power train components in other types of vehicles, e.g. to model inverters and motors of electric

vehicles. Using the proposed method for on-road data collection, data processing and model identification, identifying and validating a dynamic model for an electric vehicle's motor is a next step. The next steps also include design of an online engine/motor modelling tool at https://data.csrf.ac.uk/ using the proposed method. Such a tool will facilitate online fuel/electricity consumption modelling for different types of vehicles using on-road data, collected and uploaded to a central server.

APPENDIX

The identified LTI model of the P320 Scania diesel engine is:

$$\hat{F}(z)=\begin{bmatrix}G_w(z) & G_T(z)\end{bmatrix}\begin{bmatrix}W(z)\\T(z)\end{bmatrix}+H(z)E(z), \text{ where} \tag{21}$$

$$G_w(z)=\frac{0.001171z^{-3}-0.001928z^{-4}+0.001929z^{-5}-0.001095z^{-6}}{1-2.042z^{-1}+2.415z^{-2}-1.78z^{-3}+0.551z^{-4}-0.1025z^{-5}}, \tag{22}$$

$$G_T(z)=\frac{0.08494+0.05512z^{-1}-0.03854z^{-2}-0.06077z^{-3}-0.03114z^{-4}}{1-0.5971z^{-1}-0.5782z^{-2}-0.2039z^{-3}+0.3611z^{-4}+0.03823z^{-5}}, \tag{23}$$

$$H(z)=\frac{H_n(z)}{H_d(z)}, \tag{24}$$

$$\begin{aligned}H_n(z)=&1-3.455z^{-1}+4.74z^{-2}-3.948z^{-3}+3.758z^{-4}\\&-3.864z^{-5}+2.36z^{-6}-0.6951z^{-7}+0.2516z^{-8}\\&-0.2643z^{-9}+0.1683z^{-10}-0.04875z^{-11} \text{ and}\end{aligned} \tag{25}$$

$$\begin{aligned}H_d(z)=&1-4.547z^{-1}+8.628z^{-2}-9.613z^{-3}+8.999z^{-4}\\&-9.039z^{-5}+7.537z^{-6}-4.104z^{-7}+1.584z^{-8}\\&-0.6304z^{-9}+0.213z^{-10}-0.02728z^{-11}.\end{aligned} \tag{26}$$

The identified LTI model of the P340 Scania gas engine is:

$$\hat{F}(z)=\begin{bmatrix}G_w(z) & G_T(z)\end{bmatrix}\begin{bmatrix}W(z)\\T(z)\end{bmatrix}+H(z)E(z), \text{ where} \tag{27}$$

$$G_w(z)=\frac{0.001887z^{-3}-0.001204z^{-4}+0.001189z^{-5}-0.0003614z^{-6}}{1-1.591z^{-1}+0.9563z^{-2}+0.5942z^{-3}-0.9717z^{-4}+0.5044z^{-5}}, \tag{28}$$

$$G_T(z)=\frac{0.05099+0.04245z^{-1}-0.01672z^{-2}+0.006218z^{-3}+0.02182z^{-4}}{1-1.038z^{-1}-0.02435z^{-2}+0.8159z^{-3}-0.6473z^{-4}+0.1803z^{-5}}, \tag{29}$$

$$H(z)=\frac{H_n(z)}{H_d(z)}, \tag{30}$$

$$\begin{aligned}H_n(z)=&1-0.1898z^{-1}-0.3046z^{-2}-0.3604z^{-3}-0.3639z^{-4}\\&+0.1212z^{-5}+0.4104z^{-6}-0.5819z^{-7}+0.2884z^{-8}\\&+0.05158z^{-9}+0.06619z^{-10} \text{ and}\end{aligned} \tag{31}$$

$$\begin{aligned}H_d(z)=&1-1.379z^{-1}+0.1144z^{-2}+0.1047z^{-3}-0.06714z^{-4}\\&+0.442z^{-5}+0.1539z^{-6}-1.087z^{-7}+1.1z^{-8}\\&-0.3445z^{-9}-0.0315z^{-10}.\end{aligned} \tag{32}$$

For both models, the sampling time is 0.3s.

REFERENCES

[1] A. K. Madhusudhanan, D. Ainalis, X. Na, I. V. Garcia, M. Sutcliffe, and D. Cebon, "Effects of semi-trailer modifications on HGV fuel consumption," *Transportation Research Part D: Transport and Environment*, vol. 92, p. 102717, 2021.

[2] A. K. Madhusudhanan, X. Na, A. Boies, and D. Cebon, "Modelling and evaluation of a biomethane truck for transport performance and cost," *Transportation Research Part D: Transport and Environment*, vol. 87, p. 102530, 2020.

[3] J. Rios-Torres and A. A. Malikopoulos, "Impact of partial penetrations of connected and automated vehicles on fuel consumption and traffic flow," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 4, pp. 453–462, 2018.

[4] G. Li, D. Gorges, and M. Wang, "Online optimization of gear shift and velocity for eco-driving using adaptive dynamic programming," *IEEE Transactions on Intelligent Vehicles*, pp. 1–1, 2021.

[5] S. E. Li, Q. Guo, S. Xu, J. Duan, S. Li, C. Li, and K. Su, "Performance enhanced predictive control for adaptive cruise control system considering road elevation information," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 150–160, 2017.

[6] H. A. Rakha, K. Ahn, K. Moran, B. Saerens, and E. Van den Bulck, "Virginia tech comprehensive power-based fuel consumption model: model development and testing," *Transportation Research Part D: Transport and Environment*, vol. 16, no. 7, pp. 492–503, 2011.

[7] H. Wang, L. Fu, Y. Zhou, and H. Li, "Modelling of the fuel consumption for passenger cars regarding driving characteristics," *Transportation Research Part D: Transport and Environment*, vol. 13, no. 7, pp. 479–482, 2008.

[8] S. Hunt, A. Odhams, R. Roebuck, and D. Cebon, "Parameter measurement for heavy-vehicle fuel consumption modelling," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 225, no. 5, pp. 567–589, 2011.

[9] Y. Xiao, Q. Zhao, I. Kaku, and Y. Xu, "Development of a fuel consumption optimization model for the capacitated vehicle routing problem," *Computers & operations research*, vol. 39, no. 7, pp. 1419–1431, 2012.

[10] G. Song, L. Yu, and Z. Wang, "Aggregate fuel consumption model of light-duty vehicles for evaluating effectiveness of traffic management strategies on fuels," *Journal of Transportation Engineering*, vol. 135, no. 9, pp. 611–618, 2009.

[11] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *Journal of transportation engineering*, vol. 128, no. 2, pp. 182–190, 2002.

[12] M. Zhou and H. Jin, "Development of a transient fuel consumption model," *Transportation Research Part D: Transport and Environment*, vol. 51, pp. 82–93, 2017.

[13] G. N. Bifulco, F. Galante, L. Pariota, and M. R. Spena, "A linear model for the estimation of fuel consumption and the impact evaluation of advanced driving assistance systems," *Sustainability*, vol. 7, no. 10, pp. 14 326–14 343, 2015.

[14] I. Skog and P. Händel, "Indirect instantaneous car-fuel consumption measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 12, pp. 3190–3198, 2014.

[15] Y. Oh, J. Park, J. Lee, M. Do Eom, and S. Park, "Modeling effects of vehicle specifications on fuel economy based on engine fuel consumption map and vehicle dynamics," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 287–302, 2014.

[16] A. K. Madhusudhanan, X. Na, and D. Cebon, "A computationally efficient framework for modelling energy consumption of ICE and electric vehicles," *Energies*, vol. 14, no. 7, 2021.

[17] L. Lennart, "System Identification: Theory for the user," *PTR Prentice Hall, Upper Saddle River, NJ*, pp. 1–14, 1999.

[18] P. Van den Hof, "System identification," *Lecture Notes SC4110, Delft Center for Systems and Control, TU Delft, The Netherlands*, January 2006.

[19] L. Ljung, "System identification toolbox for use with MATLAB," *MathWorks*, 2004.

Anil K. Madhusudhanan is a Lecturer in Future Mobility Systems at the University of Southampton, UK. Before moving to Southampton, he was a Senior Research Associate (2020-21) and Research Associate (2017-20) at the University of Cambridge, UK. During 2015-17, he worked at the Netherlands Organisation for Applied Scientific Research (TNO), the Netherlands. He received PhD, and Master of Science *Cum Laude* in Systems and Control from Delft University of Technology, the Netherlands, in 2016 and 2011 respectively.

During 2006-09, he worked at STMicroelectronics (Greater Noida, India), Indian Institute of Science (Bengaluru, India) and Cranes Software (Bengaluru, India) for one year each. He received Bachelor of Technology in Electronics and Communication Engineering from National Institute of Technology Allahabad, India, in 2006.

Xiaoxiang Na received the B.Sc. and M.Sc. degrees in automotive engineering from the College of Automotive Engineering, Jilin University, China, in 2007 and 2009, respectively. He received the Ph.D. degree in driver-vehicle dynamics from the Department of Engineering, University of Cambridge, U.K. in 2014.

He is currently a Senior Research Associate with the Centre for Sustainable Road Freight, University of Cambridge, and a Borysiewicz Interdisciplinary Fellow. His main research interests include driver-vehicle dynamics, operations monitoring of road freight vehicles, and vehicle energy performance assessment.

Daniel Ainalis received the B.Eng. and Ph.D. degrees in mechanical engineering from Victoria University, Australia in 2010 and 2015, respectively.

He is currently a Senior Research Associate with the Engineering Department, University of Cambridge, U.K. and the Research Manager of the Centre for Sustainable Road Freight. He was a previously a postdoctoral researcher with the University of Mons, Belgium, and the Imperial College London, U.K. on a variety of topics related to sustainable transport.

David Cebon received the B.E. degree in mechanical engineering from the University of Melbourne, Australia in 1980, and the Ph.D. degree in engineering from the University of Cambridge, U.K. in 1985.

Since 1985, he has been with the Department of Engineering at the University of Cambridge, where he is currently a Professor of Mechanical Engineering. His research covers the mechanical, civil, materials, and energy aspects of road transport engineering.

Prof. Cebon is a Fellow of the Royal Academy of Engineering. He is the Director of the Centre for Sustainable Road Freight and the Cambridge Vehicle Dynamics Consortium. He also leads the research theme Energy, Transport, and Urban Infrastructure in the Department of Engineering at the University of Cambridge.