

A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles

Alexander Schoen*, Andy Byerly†, Brent Hendrix†, Rishikesh Mahesh Bagwe*, Euzeli Cipriano dos Santos Jr.*, and Zina Ben Miled*, *Department of Electrical and Computer Engineering (IUPUI) Indianapolis, IN 46202

†Allison Transmission, Inc., One Allison Way, Indianapolis, IN 46222

Abstract—This paper advocates a data summarization approach based on distance rather than the traditional time period when developing individualized machine learning models for fuel consumption. This approach is used in conjunction with seven predictors derived from vehicle speed and road grade to produce a highly predictive neural network model for average fuel consumption in heavy vehicles. The proposed model can easily be developed and deployed for each individual vehicle in a fleet in order to optimize fuel consumption over the entire fleet. The predictors of the model are aggregated over fixed window sizes of distance traveled. Different window sizes are evaluated and the results show that a 1 km window is able to predict fuel consumption with a 0.91 coefficient of determination and mean absolute peak-to-peak percent error less than 4% for routes that include both city and highway duty cycle segments.

Index Terms—vehicle modeling, neural networks, average fuel consumption, data summarization, fleet management.

I. INTRODUCTION

Fuel consumption models for vehicles are of interest to manufacturers, regulators, and consumers. They are needed across all the phases of the vehicle life-cycle. In this paper, we focus on modeling average fuel consumption for heavy vehicles during the operation and maintenance phase. In general, techniques used to develop models for fuel consumption fall under three main categories:

- Physics-based models, which are derived from an in-depth understanding of the physical system. These models describe the dynamics of the components of the vehicle at each time step using detailed mathematical equations [1], [2].
- Machine learning models, which are data-driven and represent an abstract mapping from an input space consisting of a selected set of predictors to an output space that represents the target output, in this case average fuel consumption [3], [4].
- Statistical models, which are also data-driven and establish a mapping between the probability distribution of a selected set of predictors and the target outcome [5], [6].

Trade-offs among the above techniques are primarily with respect to cost and accuracy as per the requirements of the intended application.

In this paper, a model that can be easily developed for individual heavy vehicles in a large fleet is proposed. Relying

on accurate models of all of the vehicles in a fleet, a fleet manager can optimize the route planning for all of the vehicles based on each unique vehicle predicted fuel consumption thereby ensuring the route assignments are aligned to minimize overall fleet fuel consumption. These types of fleets exist in various sectors including, road transportation of goods [7], public transportation [3], construction trucks [8] and refuse trucks [9]. For each fleet, the methodology must apply and adapt to many different vehicle technologies (including future ones) and configurations without detailed knowledge of the vehicles specific physical characteristics and measurements. These requirements make machine learning the technique of choice when taking into consideration the desired accuracy versus the cost of the development and adaptation of an individualized model for each vehicle in the fleet.

Several previous models for both instantaneous and average fuel consumption have been proposed. Physics-based models are best suited for predicting instantaneous fuel consumption [1], [2] because they can capture the dynamics of the behavior of the system at different time steps. Machine learning models are not able to predict instantaneous fuel consumption [3] with a high level of accuracy because of the difficulty associated with identifying patterns in instantaneous data. However, these models are able to identify and learn trends in average fuel consumption with an adequate level of accuracy [4].

Previously proposed machine learning models for average fuel consumption use a set of predictors that are collected over a time period to predict the corresponding fuel consumption in terms of either gallons per mile or liters per kilometer. While still focusing on average fuel consumption, our proposed approach differs from that used in previous models because the input space of the predictors is quantized with respect to a fixed distance as opposed to a fixed time period. In the proposed model, all the predictors are aggregated with respect to a fixed window that represents the distance traveled by the vehicle thereby providing a better mapping from the input space to the output space of the model. In contrast, previous machine learning models must not only learn the patterns in the input data but also perform a conversion from the time-based scale of the input domain to the distance-based scale of the output domain (i.e., average fuel consumption). Using the same scale for both the input and output spaces of the model offers several benefits:

- Data is collected at a rate that is proportional to its impact on the outcome. When the input space is sampled with respect to time, the amount of data collected from

a vehicle at a stop is the same as the amount of data collected when the vehicle is moving.

- The predictors in the model are able to capture the impact of both the duty cycle and the environment on the average fuel consumption of the vehicle (e.g., the number of stops in an urban traffic over a given distance).
- Data from raw sensors can be aggregated on-board into few predictors with lower storage and transmission bandwidth requirements. Given the increase in computational capabilities of new vehicles, data summarization is best performed on-board near the source of the data.
- New technologies such as V2I and dynamic traffic management [10]–[12] can be leveraged for additional fuel efficiency optimization at the level of each specific vehicle, route and time of day.

The remainder of the paper is organized as follows: Section II includes a review of previous related work, Section III introduces the proposed machine learning model, Section IV describes the process used for data collection and data summarization, Section V presents the results of applying the proposed model under different configurations, and Section VI summarizes the principle findings of this study and offers direction for future work.

II. RELATED WORK

As mentioned above, physics-based, machine learning, and statistical models have all been used to model average fuel consumption. The EPA and the European Commission developed physics-based, full vehicle simulation models for heavy duty vehicles [1], [2]. These models are capable of predicting average fuel consumption with an accuracy of $\pm 3\%$ compared to real measurements obtained from a flowmeter [2]. This level of accuracy comes at the cost of a substantial development effort.

At the other end of the modeling spectrum are statistical procedures which are applied under strict testing conditions to ensure that the reported results are standardized and repeatable. For example, the model proposed by the Code of Federal Regulation (CFR) [5] estimates fuel consumption for new vehicles by using well defined statistical methods for specific duty cycles created from segments of real world trips. Similarly, the SAE J1321 [6] standard is used to estimate fuel consumption after market modifications or under varying operating conditions for trucks and buses. This standard compares similar vehicles following the same route under similar operating conditions using real data collected from the field. For example, the standard was used in [13] to compare the fuel consumption of a control vehicle to that of two test vehicles after changing lubrication fluids in the engine, transmission and axle. The standard was also used in [8] to measure the performance of three fuel technologies in two vehicles operating in coal mines.

The generalizable characteristics of machine learning models to different vehicles and different operating conditions made this modeling methodology attractive for fuel consumption prediction in many studies. In the remainder of this section we discuss these models with respect to the underlying

machine learning technique, the representation of the input space and the representation of the output space.

Different types of machine learning techniques have been used and compared for the purpose of modeling fuel consumption. For instance, gradient boosting, neural networks and random forest are compared in [3]; neural networks and multivariate regression splines are compared in [4]; and support vector machine, neural networks and random forest are compared in [7]. Based on the results, these studies identify a technique of choice. However the differences between these techniques are mostly marginal and as stated in [7] and [14], the techniques are comparable. We believe that the differences are primarily due to different data collection and data summarization methodologies. In this paper, we opted to use neural networks because this technique is best suited for models with continuous input and output variables. Moreover neural networks are less susceptible to noisy data [15].

The input of previously proposed fuel consumption models also varies considerably. A holistic model might attempt to capture driver behavior, vehicle dynamics and the impact of the environment on the vehicle. For example, the models introduced in [4] use combinations of first, second, third and fourth orders of vehicle acceleration and speed as predictors. In [3], the predictors include vehicle speed, distance traveled, elevation, longitude, latitude and day of the week. Predictors related to the road condition (e.g., grade, curvature and roughness) and the vehicle's operating conditions (e.g., vehicle speed, acceleration, gear, and % torque) are used in [7]. The most important predictors in this previous study were found to be acceleration, % torque, and gradient. Vehicle speed was not important because it was maintained nearly constant during data collection. In excess of 30 predictors were investigated in [15] including wind speed, platooning, engine strength and breaking rate and the most important predictors were found to be road grade, vehicle speed and vehicle weight. Vehicle weight is not typically available as a standard sensor and the weight in [15] was estimated using the suspension. In this paper, we also use vehicle speed and road grade to derive the predictors of the proposed model. These variables can be directly obtained from non-invasive, affordable and widely available telematics devices.

Typically, the predictors of the models are derived from different sensor values that are sampled at fixed time intervals [3], [4]. In [15], the author compares the accuracy of the proposed fuel consumption models with respect to input data collected at 1 minute and 10 minute intervals and concludes that the 10 minute interval yields more accurate models. In [7], measurements are collected each 1 minute or 1 mile, whichever is the smallest. Given that the vehicles were traveling at constant speed in this study, this amounts to collecting input data over a fixed distance of ≈ 1 mile. Both [7] and [15] seem to hint that collecting input data over distance traveled is more suited for fuel consumption modeling. In this paper, we aggregate the values of the predictors over fixed windows of traveled distance. We also investigate the impact of the length of the window on the accuracy of the model for real world duty cycles with varying vehicle speeds.

The output of the fuel consumption models can be ei-

ther fuel rate (liters/hour) or average fuel consumption (liters/100km). The objective of fuel rate models is to predict instantaneous fuel consumption. These models tend to suffer from a low point-wise accuracy (i.e., at the level of each sample). However, by averaging the predicted fuel rates over an extended time period or distance, the models are able to deliver relatively accurate average fuel consumption. This approach was used in [4] with a resulting accuracy in the range of 3.73% to 6.83% from the measured average fuel rate over an entire real-world duty cycle. The approach was also used in [3] and after the predicted fuel rates are translated into fuel consumption, the model is able to predict total fuel consumption with an accuracy of 2% over an entire trip covering a distance of 365 km. However, an analysis of how well point-wise predicted fuel rates in this study track measured fuel rates showed a low coefficient of determination (< 0.3). The input and output of the models in both of the aforementioned studies are in the time domain.

The models introduced in [7] use the ratio of consumed fuel over the distance traveled as an output. As previously mentioned, the speed of the vehicles in this study was constant, essentially resulting in the sampling of both the input and output of the model in the distance domain. The mean absolute error of the predicted fuel consumption was $\approx 3\%$. A high R^2 value (> 0.8) also shows that, as opposed to the two studies discussed above, the model presented in [7] is able to make good point-wise predictions.

The models proposed in [15] highlight some of the difficulties faced by machine learning models when the input and output have different domains. In this study, the input is aggregated in the time domain over 10 minutes intervals and the output is fuel consumption over the distance traveled during the same time period. The root mean square error of the predicted fuel consumption over the entire duty cycle was ≈ 7.4 l/100km for a mean fuel consumption of 30 l/100km. This error, which is measured point-wise for each 10 minutes time interval, is relatively high compared to models that calculate the error over the entire trip [3], [4].

III. A DIGITAL MODEL USING NEURAL NETWORKS

As mentioned above Artificial Neural Networks (ANN) are often used to develop digital models for complex systems (e.g., [4], [16], [17]). The complex system is represented by a transfer function $F(\vec{p}) = o$, where $F(\cdot)$ represents the system, \vec{p} refers to the input predictors and o is the response of the system or the output. The system is then approximately defined by

$$\hat{o} = f(\vec{p}) \quad (1)$$

where $f(\cdot)$ is the predictive model and \hat{o} is an approximation of the actual output o . The model $f(\vec{p})$ in (1) is not limited to ANNs. For instance, $f(\vec{p})$ may be expressed by using a set of differential equations [18] or by a weight matrix derived using support vector machines [19].

The ANNs used in this paper are Feed Forward Neural Networks (FNN). FNNs are defined by (2) where the matrix W and the vectors \vec{w} , \vec{b} , and \vec{c} represent adjustable weights and \vec{p} is an input vector consisting of a set of predictors that

capture the state of the system and are therefore believed to be able to accurately predict o . As opposed to linear models (e.g., linear regression), FNNs are capable of representing the nonlinear characteristics of a complex system through the use of a nonlinear activation function (σ) as defined in (3).

$$\hat{o} = \sigma(\sum_k w_k \cdot \sigma(\sum_l w_{lk} \cdot p_l + b_l) + c_k) \quad (2)$$

$$\sigma(\tau) = \frac{1}{1 + e^{-\tau}}. \quad (3)$$

The weight coefficients $w_{lk} \in W$, $w_k \in \vec{w}$, $b_l \in \vec{b}$, and $c_k \in \vec{c}$ define the model where $l = \{1 \dots n\}$, $k = \{1 \dots m\}$, n is the number of input predictors and m is the number of neurons in the hidden layer. Weights are chosen such that the difference, defined in (4), between the system response (o) and the model output (\hat{o}) is minimized. The weights W and \vec{w} define the relationship between a neuron's input and output, while the biases \vec{b} and \vec{c} allow for flexibility in the model. In order to determine weights and biases, the network undergoes a process known as training. Training requires a known set of input predictors (\vec{p}) and their corresponding output (o).

$$E = \frac{1}{2} (o - \hat{o})^2. \quad (4)$$

The available input predictors and corresponding output are considered a feature set (\mathbb{F}) and typically consist of two disjointed subsets, \mathbb{F}_{tr} and \mathbb{F}_{ts} . The first set, \mathbb{F}_{tr} , is used to train the network and derive the weight and bias values in the FNN. The second set, \mathbb{F}_{ts} , is used to validate the model. This ensures that the model is capable of generalizing by predicting the system's responses to input values that were not observed during the training process.

Training is an iterative process and can be performed using multiple approaches including particle swarm optimization [20] and back propagation [21]. This paper adopts the latter. Other approaches will be considered in future work in order to evaluate their ability to improve the model's predictive accuracy.

Each iteration in the training selects a pair of (input, output) features from \mathbb{F}_{tr} at random and updates the weights in the network. This is done by calculating the error between the actual output value and the value predicted by the model as shown in (4). This error is then propagated back from the output layer to the input layer of the network. The gradient of this error with respect to each weight in W , \vec{w} , \vec{b} , and \vec{c} is calculated as shown in (5).

$$\frac{\partial E}{\partial w_{l,k}} = \frac{\partial \left(\frac{1}{2} (o - \hat{o})^2 \right)}{\partial w_{l,k}}. \quad (5)$$

For each training iteration step s , the new weights in $W^{(s)}$ are updated by adding $\Delta w_{l,k}^{(s)}$ as defined by (6) to the previous value of the weights at step $s-1$ according to (7). The values for $\vec{w}^{(s)}$, $\vec{b}^{(s)}$, and $\vec{c}^{(s)}$ are calculated in an identical manner by replacing $w_{l,k}^{(s)}$ with $w_k^{(s)}$, $b_l^{(s)}$, and $c_k^{(s)}$, respectively.

The constants ϵ and μ in (6) represent control parameters that dictate the model's learning rate. The constant μ controls the amount of resistance to change with respect to the previous

training iteration and ϵ controls the amount of change in the direction of the gradient (5) that should be added to the weights.

$$\Delta w_{l,k}^{(s)} = -\epsilon \cdot \frac{\partial E}{\partial w_{l,k}} + \mu \cdot \Delta w_{l,k}^{(s-1)} \quad (6)$$

$$w_{l,k}^{(s)} = w_{l,k}^{(s-1)} + \Delta w_{l,k}^{(s)} \quad (7)$$

The training process is followed by a validation process that is used to measure the performance of the model and in particular its ability to generalize to (input, output) features from \mathbb{F}_{ts} that were not used in training the model.

$$C_D = 1 - \frac{\sum_i (o_i - \hat{o}_i)^2}{\sum_i (o_i - \bar{o})^2} \quad (8)$$

Several metrics can be used to evaluate this performance over the test data set. The first is the Coefficient of Determination (8) where the numerator is the sum of the squared error of the actual response of the system compared to the predicted response by the model. The denominator is the difference between the actual response and the mean of the actual response of the system. The coefficient of determination (C_D) is an indication of how well the response predicted by the machine learning model tracks the actual point-wise response of the system [22] for each sample i in \mathbb{F}_{ts} . High C_D values, close to 1, indicate accurate point-wise predictions. Low values, close to 0 or negative, indicate poor point-wise predictions. The range of possible C_D values is between negative infinity and one.

$$MAE = \frac{1}{N} \sum_{i=1}^N |o_i - \hat{o}_i| \quad (9)$$

$$MAPE_{pk} = \frac{1}{N} \sum_{i=1}^N \frac{|o_i - \hat{o}_i|}{o_{pk}} \cdot 100 \quad (10)$$

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (o_i - \hat{o}_i)^2} \quad (11)$$

In addition to the coefficient of determination, the accuracy of the model is evaluated by using the mean absolute error (MAE) as shown in (9) [23] and the mean absolute peak-to-peak percent error ($MAPE_{pk}$) shown in (10) where o_{pk} is define by

$$o_{pk} = \max(\vec{o}) - \min(\vec{o}) \quad (12)$$

and N is the number of test points in \mathbb{F}_{ts} . $MAPE_{pk}$ was preferred over $MAPE$ (Mean Absolute Percent Error) in order to avoid the adverse effects of unsymmetrical results when o_i is less than \hat{o}_i and biasing when o_i is close to zero [23]. MAE is used to measure the error in the model predictions while $MAPE_{pk}$ reports this error as a percentage of the highest possible error the predictions could have had. In addition to these performance metrics, $RMSE$ (11) is reported when relevant. All of the selected metrics evaluate the performance of the model point-wise rather than based on the cumulative performance over an entire trip.

The relative importance of each predictor in the model is estimated based on the percentage of the weight (IW) of the corresponding input compared to all the input of the model as defined in (13) [22].

$$IW_l = \frac{\sum_j |w_{lj}|}{\sum_l \sum_k |w_{lk}|} \cdot 100 \quad (13)$$

The above metric was adjusted in order to account for input features with widely varying means. The adjusted IW metric used in this paper is given by

$$AIW_l = \frac{IW_l}{\sum_k IW_k} \cdot \bar{p}_l \quad (14)$$

where \bar{p}_l is the mean of the input feature p_l over the test data set \mathbb{F}_{ts} .

The AIW_l for a given input feature (p_l) indicates the proportional excitation provided by the feature to the first layer of the network compared to the other input features. This is only an approximation of the significance of the feature. It is possible that input features with low AIW values contribute significantly to the output of the model as their excitation propagates through the hidden layer of the network.

IV. DATA COLLECTION AND SUMMARIZATION

The model is developed by using duty cycles collected from a single truck, with an approximate mass of 8,700 kg exposed to a variety of transients including both urban and highway traffic in the Indianapolis area. Data was collected using the SAE J1939 standard for serial control and communications in heavy duty vehicle networks [24].

Twelve drivers were asked to exhibit good or bad behavior over two different routes. Drivers exhibiting good behavior anticipated braking and allowed the vehicle to coast when possible. Some drivers participated more than others and as a result the distribution of drivers and routes is not uniform across the data set.

This field test generated 3,302,890 data points sampled at 50 Hz from the vehicle CAN bus and a total distance of 778.89 km over 56 trips with varying distances. Most of the trips covered a distance of 10 km to 15 km. In order to increase the number of data points, synthetic duty cycles over an extended distance were obtained by assembling segments from the field duty cycles selected at random. Moreover, a set of drivers are assigned to the training segments and a different set of drivers are assigned to the testing segments, thereby ensuring that the training (\mathbb{F}_{tr}) and testing (\mathbb{F}_{ts}) data sets derived from the respective segments are completely separate.

A. Model Predictors

Several processing steps were needed in order to generate the predictors of the model. These predictors are derived from two measurements, namely, road grade and transmission output speed.

The first processing step consisted of down sampling the road grade and obtaining the vehicle speed from the transmission output speed. The road grade was measured using an on-board inclinometer and down-sampled to 1 Hz.

A review of the data also showed that there is a linear relationship between the vehicle speed and the transmission output speed given by the following equation:

$$\text{VehicleSpeed} \approx 59.3 \times \text{TransmissionOutputSpeed} \quad (15)$$

In order to reduce the noise in the variable, a moving average low pass filter was applied to the vehicle speed obtained by using (15) and the variable was down-sampled from 50 Hz to 1 Hz .

The purpose of the second processing step was to derive the synthetic duty cycles. Towards this objective, the duty cycles in the real data were split into segments defined by intervals between consecutive vehicle stops (Figure 1). A total of 455 real data segments were obtained from all the twelve drivers in the study. Out of these, 358 segments from nine drivers were used to derive the training data set (\mathbb{F}_{tr}) and the remaining 97 segments, obtained from the remaining three drivers in the study, were used to derive the testing data set \mathbb{F}_{ts} .

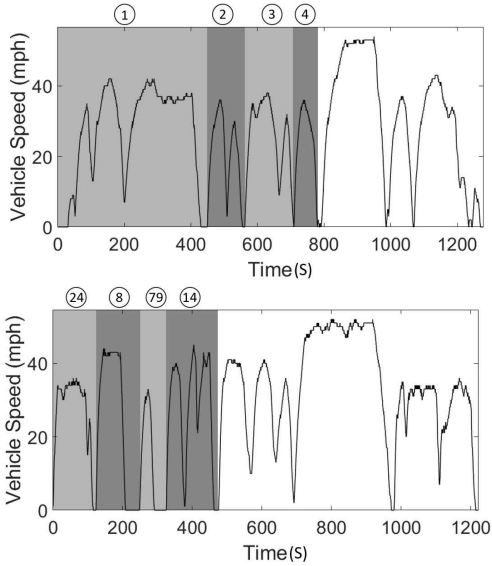


Fig. 1. The first four segments of a sample real duty cycle (top). A sample synthetic duty cycle created by concatenating segments 24, 8, 79, and 14 from the real data (bottom).

One synthetic duty cycle is generated by sampling, without replacement, from the real data segments and concatenating the selected segments until a total distance of 15 km is reached. The total distance of 15 km was selected in order to mimic the real routes used for the field data collection. It was found that an average of five segments are needed to create 15 km of data. Figure 1 shows an example synthetic duty cycle generated using this process.

Combining segments using the above approach resulted in a continuous vehicle speed. However, discontinuities were observed in road grade from one segment to the next as shown in the example of Figure 2.

These duty cycles are then aggregated over a fixed distance traveled based on the desired window (x). Table I shows the total number of data points (i.e., windows) as well as the total distance for each data set and for each window size being considered in this paper.

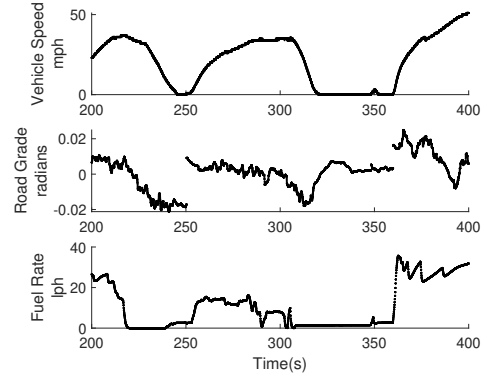


Fig. 2. Sample duty cycle with three joined training data segments. The duty cycle was created such that vehicle speed is continuous. However, road grade and fuel rate may be discontinuous.

TABLE I
NUMBER OF DATA POINTS (I.E., WINDOWS) AND TOTAL DISTANCE FOR THE TRAINING ($\mathbb{F}(x)_{tr}$) AND THE TESTING ($\mathbb{F}(x)_{ts}$) DATA SETS WITH VARYING SIZE WINDOWS (I.E., 1, 2, AND 5 km .)

Window size	$\mathbb{F}(x)_{tr}$		$\mathbb{F}(x)_{ts}$	
	Number of Points	Distance (km)	Number of Points	Distance (km)
$x = 1 \text{ km}$	20,000	20,000	32,089	32,089
$x = 2 \text{ km}$	20,000	40,000	23,106	46,212
$x = 5 \text{ km}$	20,000	100,000	6,061	30,305

The third step in the input data processing consists of generating the predictors for the proposed model. As previously mentioned, these predictors are calculated for each window and derived from vehicle speed and road grade. The selected predictors consist of:

- number of stops,
- time stopped,
- average moving speed,
- characteristic acceleration,
- aerodynamic speed squared,
- change in kinetic energy and
- change in potential energy.

The above predictors were selected because they are believed to capture the vehicle dynamics as well as the driver's behavior and the impact of the route on the target output of the model (i.e., fuel consumption). In particular, a previous study [25] states that characteristic acceleration (16) and aerodynamic speed squared (18) are highly indicative of the fuel consumption for a given duty cycle. This study argues that characteristic acceleration is directly related to the inertia work needed to accelerate the vehicle and aerodynamic speed squared captures the impact of aerodynamics on fuel consumption [25].

$$\tilde{a} = \frac{1}{x} \cdot \sum_{i=1}^{N-1} \left| \frac{(v_{i+1}^2 - v_i^2)}{2} + g \cdot \Delta h_{i,i+1} \right| \quad (16)$$

where x is the total distance of the window, g is the gravitational constant and, v_i and v_{i+1} represent the vehicle speeds at the consecutive times t_i and t_{i+1} within the window, respectively. The difference in elevation ($\Delta h_{i,i+1}$) is the product of the distance traveled ($\Delta d_{i,i+1}$) between time t_i and t_{i+1}

multiplied by the $\sin(\cdot)$ of road grade (α_i) at time t_i , as shown below.

$$\Delta h_{i,i+1} = \Delta d_{i,i+1} \cdot \sin(\alpha_i) \quad (17)$$

$$v_{aero}^2 = \frac{1}{x} \cdot \sum_{i=1}^{N-1} \frac{v_{i+1}^3 + v_{i+1}^2 v_i + v_{i+1} v_i^2 + v_i^3}{4} \cdot \Delta t_{i,i+1} \quad (18)$$

where $\Delta t_{i,i+1}$ corresponds to the fixed sample period of 1 *second* in this study. Using this fact, and in order to reduce the amount of on-board computation needed to calculate v_{aero}^2 , the initial expression (18) given in [25] was simplified to the approximation shown in (19).

$$v_{aero}^2 \approx \frac{1}{x} \cdot \sum_{i=1}^{N-1} \left(\frac{v_{i+1} + v_i}{2} \right)^3 \quad (19)$$

Inspired by the above predictors introduced in [25] to capture the transient dynamics of the vehicle, we added two predictors that can capture trends in the average dynamics of the vehicle over the distance traveled for a given window. These predictors are the change in kinetic energy (*CKE*) and the change in potential energy (*CPE*) as described by (20) and (21), respectively.

$$CKE = \frac{1}{2} \cdot m \cdot (v_N - v_1)^2 \quad (20)$$

where m is the mass of the vehicle which is constant in this study and, v_1 and v_N are the vehicle speeds at the beginning and end of the window, respectively.

$$CPE = m \cdot g \cdot \sum_{i=1}^{N-1} \Delta h_{i,i+1} \quad (21)$$

It is important to capture the change in kinetic and potential energy during the duty cycle because these changes in the energy state of the vehicle can be significant for short distances when compared to the amount of total energy consumed by fuel. Over an extended distance, the percentage of fuel energy converted to kinetic and/or potential energy is reduced.

B. Model Output

The output of the model is average fuel consumption in l/100km for each window. In order to obtain the average consumption, fuel rates are collected from the CAN bus. As in the case of road grade, and because synthetic duty cycles are derived from a random selection of real duty cycle segments, discontinuities in the fuel rate are observed from one segment to the next (Figure 2). The impact of these discontinuities is not significant because the fuel rates are averaged over the entire window in order to calculate the output of the model (i.e., average fuel consumption).

An analysis of the segments in the real data collected from the field shows a variance in average fuel consumption over all the trips. For example, a 20% difference in fuel consumption was observed between good and bad driver behavior over entire trips. Moreover, variances in average fuel consumption are also observed for different window sizes. Table II shows the

mean and standard deviation of the average fuel consumption for the 1, 2, and 5 *km* windows across all trips. While the mean fuel consumption across all windows is relatively constant, the standard deviation decreases as the window size increases.

TABLE II
MEAN AND STANDARD DEVIATION OF AVERAGE FUEL CONSUMPTION OVER 1, 2, AND 5 *km* WINDOWS IN THE REAL DATA. THE LAST ROW OF THE TABLE SHOWS THE AVERAGE FUEL CONSUMPTION FOR ALL THE TRIPS.

	<i>mean</i> l/100km	<i>std</i> l/100km
1 <i>km</i>	20.42	9.17
2 <i>km</i>	20.69	5.94
5 <i>km</i>	20.98	3.97
trips	20.88	3.96

In summary, all the input features of the proposed model are derived using the above methodology from the vehicle speed and the road grade sampled at a rate of 1 *Hz*. These variables can be obtained from a telematics device. In this study, these variables were derived from sensor values broadcasted on the CAN bus. The accuracy of the model will vary depending on the source of the data and the sampling frequency.

The accuracy of the model is also subject to the accuracy of the output feature. Fuel consumption obtained from the CAN bus can have an error as high as 5% compared to the actual fuel consumption [26]. Better accuracy can be obtained by using flowmeters. However, flowmeters are more expensive. Fuel consumption levels from the CAN bus are used in [7] and [15] as well as in this paper and high precision fuel sensors are used in [3]. Aspects related to the accuracy of the data sources will be explored in future work.

V. MODEL VALIDATION

The seven predictors listed in Section IV are used as input to the neural network model. This constitutes the first layer of the network. The first layer then feeds into a hidden layer with 5 neurons. In turn, the hidden layer feeds into an output player with a single neuron. Figure 3 shows the *RMSE* (11) during training for three models with window sizes 1, 2 and 5 *km*. In the top plot, each data point corresponds to the *RMSE* values after training the model with a group of 500 windows. This plot indicates that all models converge to a *RMSE* value less than 0.2 l/100km. However, the convergence rates for the models are different. In fact, the 5 *km* starts with a *RMSE* value of 0.16 l/100km after 500 training windows and this *RMSE* value reaches 0.08 l/100km when the model converges. The corresponding values for the 1 *km* model are 0.34 l/100km and 0.14 l/100km, respectively. When coupled with the difference in standard deviation of the average fuel consumption for the 1 *km* and the 5 *km* windows (Table II), this trend indicates that aggregating the input and output data over 5 *km* provides a stable profile for the fuel consumption of the vehicle over the routes and this profile does not necessitate extensive learning.

This finding aligns with previous studies. For example, in [14], it was found that the trip distance is an important indicator and that predicting fuel consumption over long route

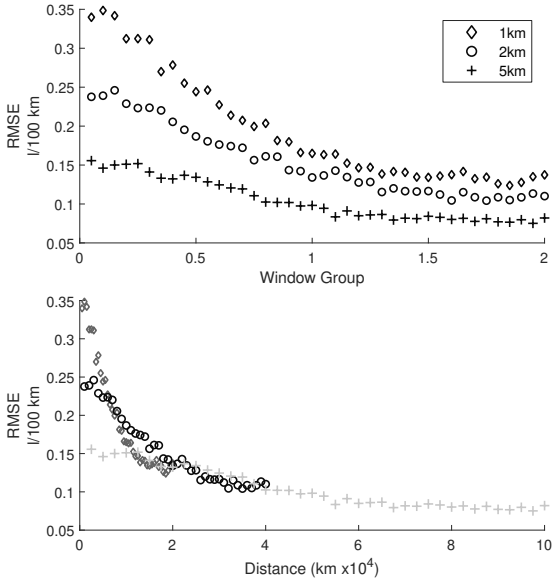


Fig. 3. *RMSE* during the training process over 20,000 km, 40,000 km, and 100,000 km for window sizes = 1, 2 and 5 km. *RMSE* plotted (top) against the number of window groups where each window group consists of 500 consecutive windows (bottom) against distance.

segments for small vehicles in urban areas has better accuracy. In this previous study, 64% of the trips covered a distance ≤ 5 km. Similarly, in [15], it was found that collecting data over 10 minutes intervals resulted in a better accuracy than 1 minute intervals. In either case, we believe that extending the data collection interval promotes a linear relationship between fuel consumption and distance traveled. While this approach yields a good average fuel consumption prediction over long distances, point-wise predicted fuel consumption may not adequately track actual values.

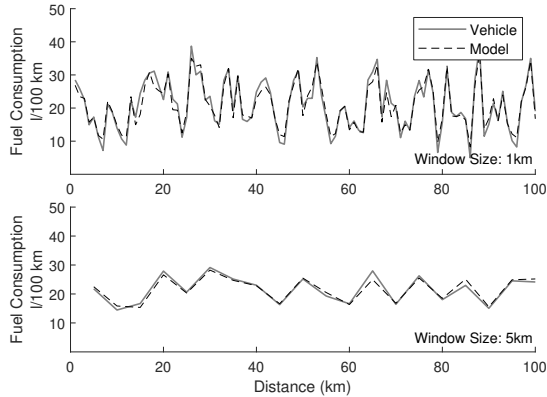


Fig. 4. Predicted versus actual fuel consumption over the first 100 km from \mathbb{F}_{ts} for a 1 km window (top) and 5 km window (bottom).

To illustrate this behavior, Figure 4 shows the predicted and actual fuel consumption over the first 100 km of the test data set (\mathbb{F}_{ts}) for window sizes 1 km and 5 km. The 1 km model is able to better track fuel consumption on a per window-basis. The same observation was made for the entire test data set as shown by the predicted versus actual fuel consumption plots in Figure 5. The results in figures 4 and 5 are for models trained

with the training data set (\mathbb{F}_{tr}) described in Table I.

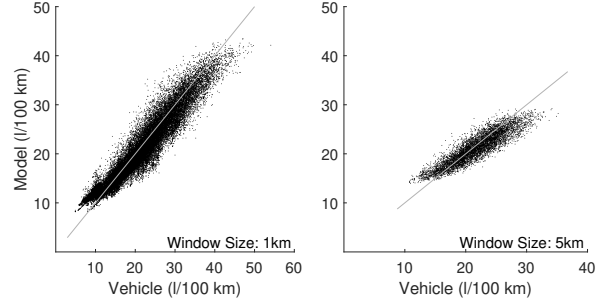


Fig. 5. Predicted versus actual fuel consumption for the entire test data set (\mathbb{F}_{ts}) for the 1 km window (left) and the 5 km window (right).

In order to compare the point-wise predictive performance for varying window sizes, models are trained five different times with the weights of the neural network initialized to random starting values. The reported performance metrics (Table III) correspond to the averages of these five training/testing runs for each window size. The standard deviation over the five runs is shown in between parenthesis below the corresponding performance metric in Table III. These standard deviations indicate that the models derived using the proposed approach are stable.

TABLE III
PREDICTIVE ACCURACY OF THE FUEL CONSUMPTION MODELS FOR 1, 2 AND 5 km AGGREGATION WINDOWS.

Window	1 km	2 km	5 km
<i>CD</i>	0.91 (0.0066)	0.87 (0.0085)	0.79 (0.0136)
<i>RMSE</i> (l/100km)	0.0132 (0.0005)	0.0142 (0.0005)	0.0234 (0.0008)
<i>MAE</i> (l/100km)	1.88 (0.0626)	1.69 (0.0515)	1.43 (0.0466)
<i>MAPE_{pk}</i>	3.74% (0.12%)	4.20% (0.13%)	5.83% (0.19%)
<i>Points</i>	32,089	23,106	6,061

Table III shows that the 1 km model has better performance than the other two window sizes across all metrics. As previously mentioned, these performance metrics evaluate the performance of the model point-wise. In particular, the coefficient of determination (*CD*) for the 1 km model is equal to 0.91 which indicates that the model is able to track the actual fuel consumption for each 1 km of distance traveled. As the window size increases, the *CD* decreases.

In terms of *MAE* and *CD*, the proposed model shows an improvement over [3], [7] despite the fact that high precision fuel sensors are used in [3]. The *RMSE* of the models is also less than 0.025 l/100km which is lower than the results obtained in [15]. That said, the test distance in this paper is higher than the one used in [15]. Longer distances favor lower *RMSE*. The *MAPE_{pk}* values for the models are also within ranges of fuel consumption accuracy for models reported in [4]. However, in this paper the error between actual and predicted is compared at the level of each window whereas in [4], the error is reported for the entire trip.

The performance metrics shown in Table III seem to indicate that the proposed models are using highly predictive input

features and that these features are adequately mapped to the output space of the model. In order to understand the contribution of each predictor, the *AIW* values of the predictors are calculated and summarized in Table IV.

TABLE IV
ADJUSTED INFLUENCE OF WEIGHTS (*AIW*) FOR THE PREDICTORS IN THE PROPOSED MODEL.

Window	1 km	2 km	5 km
No. of Stops	1.49	2.29	4.63
Stop Time	0.62	1.24	3.44
Avg. Moving Speed	13.73	10.78	8.98
\bar{a}	12.47	14.32	12.98
v_{aero}^2	11.73	11.64	10.30
<i>CKE</i>	17.04	16.13	12.26
<i>CPE</i>	13.73	11.45	9.38
Bias	29.21	32.15	38.03

The importance of the number of stops and the stop time increases as the window size increases. This is expected since fewer stops are observed in the 1 km window compared to the 2 or 5 km windows. All the remaining predictors have high *AIW* across all window sizes. In fact, eliminating any of these predictors resulted in models with lower predictive accuracy.

The increase in the *AIW*s for the number of stops and the stop time with increasing window sizes coupled with the decrease in *AIW*s for the remaining predictors indicates that as the window size increases, the model relies less on the vehicle's dynamics and more on events related to the distance traveled in order to estimate fuel consumption. Moreover, Table IV indicates that the two new predictors introduced in this paper have comparable contribution towards fuel consumption prediction to that of average moving speed [15], characteristic acceleration and aerodynamic speed [25].

VI. CONCLUSION

This paper presented a machine learning model that can be conveniently developed for each heavy vehicle in a fleet. The model relies on seven predictors: number of stops, stop time, average moving speed, characteristic acceleration, aerodynamic speed squared, change in kinetic energy and change in potential energy. The last two predictors are introduced in this paper to help capture the average dynamic behavior of the vehicle. All of the predictors of the model are derived from vehicle speed and road grade. These variables are readily available from telematics devices that are becoming an integral part of connected vehicles. Moreover, the predictors can be easily computed on-board from these two variables.

The model predictors are aggregated over a fixed distance traveled (i.e., window) instead of a fixed time interval. This mapping of the input space to the distance domain aligns with the domain of the target output, and produced a machine learning model for fuel consumption with an *RMSE* < 0.015 l/100km.

Different model configurations with 1, 2, and 5 km window sizes were evaluated. The results show that the 1 km window has the highest accuracy. This model is able to predict the actual fuel consumption on a per 1 km-basis with a *CD* of 0.91. This performance is closer to that of physics-based models and the proposed model improves upon previous

machine learning models that show comparable results only for entire long-distance trips.

Selecting an adequate window size should take into consideration the cost of the model in terms of data collection and on-board computation. Moreover, the window size is likely to be application-dependent. For fleets with short trips (e.g., construction vehicles within a site) or urban traffic routes, a 1 km window size is recommended. For long-haul fleets, a 5 km window size may be sufficient. In this study, the duty cycles consisted of both highway and city traffic and therefore, the 1 km window was more adequate than the 5 km window. Future work includes understanding these differentiating factors and the selection of the appropriate window size. Expanding the model to other vehicles with different characteristics such as varying masses and aging vehicles is being studied. Predictors for these characteristics will be added in order to allow for the same model to capture the impact on fuel consumption due to changes in vehicle mass and wear.

Future work also includes investigating the minimum distance required for training each model and analyzing how often does a model need to be synchronized with the physical system in operation by using online training in order to maintain the prediction accuracy of the model.

REFERENCES

- [1] B. Lee, L. Quinones, and J. Sanchez, "Development of greenhouse gas emissions model for 2014-2017 heavy-and medium-duty vehicle compliance," SAE Technical Paper, Tech. Rep., 2011.
- [2] G. Fontaras, R. Luz, K. Anagnostopoulos, D. Savvidis, S. Hausberger, and M. Rexeis, "Monitoring co2 emissions from hgv in europe-an experimental proof of concept of the proposed methodological approach," in *20th International Transport and Air Pollution Conference*, 2014.
- [3] S. Wickramanayake and H. D. Bandara, "Fuel consumption prediction of fleet vehicles using machine learning: A comparative study," in *Moratuwa Engineering Research Conference (MERCon)*, 2016. IEEE, 2016, pp. 90–95.
- [4] L. Wang, A. Duran, J. Gonder, and K. Kelly, "Modeling heavy/medium-duty fuel consumption based on drive cycle properties," SAE Technical Paper, Tech. Rep., 2015.
- [5] *Fuel Economy and Greenhouse gas exhaust emissions of motor vehicles Subpart B - Fuel Economy and Carbon-Related Exhaust Emission Test Procedures*, Code of Federal Regulations Std. 600.111-08, Apr 2014.
- [6] *SAE International Surface Vehicle Recommended Practice, Fuel Consumption Test Procedure - Type II*, Society of Automotive Engineers Std., 2012.
- [7] F. Perrotta, T. Parry, and L. C. Neves, "Application of machine learning for fuel consumption modelling of trucks," in *Big Data (Big Data)*, 2017 IEEE International Conference on. IEEE, 2017, pp. 3810–3815.
- [8] S. F. Haggis, T. A. Hansen, K. D. Hicks, R. G. Richards, and R. Marx, "In-use evaluation of fuel economy and emissions from coal haul trucks using modified sae j1321 procedures and pems," *SAE International Journal of Commercial Vehicles*, vol. 1, no. 2008-01-1302, pp. 210–221, 2008.
- [9] A. Ivanco, R. Johri, and Z. Filipi, "Assessing the regeneration potential for a refuse truck over a real-world duty cycle," *SAE International Journal of Commercial Vehicles*, vol. 5, no. 2012-01-1030, pp. 364–370, 2012.
- [10] A. A. Zaidi, B. Kulcsr, and H. Wymeersch, "Back-pressure traffic signal control with fixed and adaptive routing for urban vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2134–2143, Aug 2016.
- [11] J. Zhao, W. Li, J. Wang, and X. Ban, "Dynamic traffic signal timing optimization strategy incorporating various vehicle fuel consumption characteristics," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3874–3887, June 2016.
- [12] G. Ma, M. Ghasemi, and X. Song, "Integrated powertrain energy management and vehicle coordination for multiple connected hybrid electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 2893–2899, April 2018.

- [13] A. C. Brandt, E. A. Frame, and R. W. Warden, "Sae j1321 testing using m1083a1 fmvts," SOUTHWEST RESEARCH INST SAN ANTONIO TX TARDEC FUELS AND LUBRICANTS RESEARCH FACILITY, Tech. Rep., 2010.
- [14] W. Zeng, T. Miwa, and T. Morikawa, "Exploring trip fuel consumption by machine learning from gps and can bus data," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 11, pp. 906–921, 2015.
- [15] H. Almér, "Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles," 2015.
- [16] S. McBride, C. Sandu, A. Alatorre, and A. Victorino, "Estimation of vehicle tire-road contact forces: A comparison between artificial neural network and observed theory approaches," *SAE Technical Paper*, pp. 01–0562, 2018.
- [17] 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.
- [18] C. M. Atkinson, S. Petreanu, N. N. Clark, R. J. Atkinson, T. I. McDaniel, S. Nandkumar, and P. Famouri, "Numerical simulation of a two-stroke linear engine-alternator combination," SAE Technical Paper, Tech. Rep., 1999.
- [19] C.-M. Vong, P.-K. Wong, and Y.-P. Li, "Prediction of automotive engine power and torque using least squares support vector machines and bayesian inference," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 3, pp. 277–287, 2006.
- [20] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, May 1998, pp. 69–73.
- [21] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.
- [22] S. Shah, M. Hosseini, Z. B. Miled, R. Shafer, and S. Berube, "A water demand prediction model for central indiana," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [23] B. E. Flores, "A pragmatic view of accuracy measurement in forecasting," *Omega*, vol. 14, no. 2, pp. 93–98, 1986.
- [24] *SAE Serial Control and Communications Heavy Duty Vehicle Network - Top Level Document*, Society of Automotive Engineers Std., jun 2012. [Online]. Available: <https://doi.org/10.4271/J1939-201206>
- [25] M. P. O'Keefe, A. Simpson, K. J. Kelly, and D. S. Pedersen, "Duty cycle characterization and evaluation towards heavy hybrid vehicle applications," SAE Technical Paper, Tech. Rep., 2007.
- [26] S. E. Marx, J. D. Luck, R. M. Hoy, S. K. Pitla, E. E. Blankenship, and M. J. Darr, "Validation of machine can bus j1939 fuel rate accuracy using nebraska tractor test laboratory fuel rate data," *Computers and Electronics in Agriculture*, vol. 118, pp. 179–185, 2015.



Alexander Schoen IEEE student member, completed his B.S in electrical engineering at the Purdue School of Engineering and Technology in Indianapolis (IUPUI). He is currently pursuing his Masters Degree in Electrical and Computer Engineering. His research interest include machine learning and complex systems modeling.



Andy Byerly graduated with Bachelors and Masters of Science degrees in Electrical Engineering from Purdue University in 2003 and 2005. Subsequently he graduated with an MBA degree from Indiana University. Andy is currently the Director of Digital and Connected Services at Allison Transmission, Inc., leading a team to identify future controls products and services in the areas of electrification, advanced driver assistance systems, and digital system integration. Andy is a member of both IEEE and SAE.



Brent Hendrix Brent Hendrix graduated with a BS in Mechanical Engineering from Rose-Hulman Institute of Technology and subsequently earned a MS in Manufacturing Systems Engineering from Stanford University. Brent is currently the Director Medium Duty 9-Speed Transmission at Allison Transmission, Inc., prior to that he was Director Electronic Controls Software.



Rishikesh Mahesh Bagwe IEEE student member, received his B.E. (Hons.) degree in Electronics and Instrumentation Engineering from Birla Institute of Technology and Sciences, Pilani, India in 2016. He is currently pursuing an MS degree in Electrical and Computer Engineering at the Purdue School of Engineering and Technology, IUPUI. His research interests include control systems and system modeling.



Euzeli Cipriano dos Santos Jr. received his B.S., M.S., and Ph.D. degrees in electrical engineering from the Federal University of Campina Grande, Campina Grande, Brazil, in 2004, 2005, and 2007, respectively. He is currently an associate professor of Electrical and Computer Engineering in the School of Engineering and Technology at IUPUI. He co-authored more than 150 technical papers in IEEE conferences and IEEE transactions, with two conference papers winning the best paper award and the best student presenter award in the 27th IEEE-ISIE and 43th IEEE-PVSC conferences, respectively. He is the author of two books in the field of power electronics, one of them published by Wiley-IEEE Press. In 2015 he served as Guest Associate Editor in the IEEE Transactions on Power Electronics. His research interests include power electronics, renewable energy systems, control systems and electrical drives. He is a member of the IEEE Power Electronics Society, IEEE Industrial Electronics Society, and IEEE Industry Applications Society.



Zina Ben Miled received her B.S. from Oregon State University in 1988 and her M.S. and Ph.D. degree in computer engineering from Purdue University in 1990 and 1997, respectively. She is currently an associate professor of Electrical and Computer Engineering in the School of Engineering and Technology at IUPUI. Her research interests include data analytics and knowledge discovery, data models and large scale service-oriented software architectures. She has co-authored several technical papers in journals and conference proceedings. She was also the recipient of the Indiana Women in High Tech Award and the National Science Foundation Career Award. She is a senior member of IEEE and a member of IEEE CIS.