

# Route-Sensitive Fuel Consumption Models for Heavy-Duty Vehicles

**Alexander Schoen,<sup>1</sup> Andy Byerly,<sup>2</sup> Euzeli Cipriano dos Santos Jr,<sup>1</sup> and Zina Ben-Miled<sup>1</sup>**

<sup>1</sup>Indiana University Purdue University at Indianapolis, USA

<sup>2</sup>Allison Transmission Inc., USA

## Abstract

This article investigates the ability of data-driven models to estimate instantaneous fuel consumption over 1 km road segments from different routes for different heavy-duty vehicles from the same fleet. Models are created using three different techniques: parametric, linear regression, and artificial neural networks. The proposed models use features derived from vehicle speed, mass, and road grade, which can be easily obtained from telematics devices, in addition to power take-off (PTO) active time, which is needed to capture the power requested by accessories in several heavy-duty vehicles. The robustness of these models with respect to the training data selection is improved by using k-fold cross-validation. Moreover, the inherent underestimation or overestimation bias of the model is calculated and used to offset the fuel consumption estimates for new routes. The study shows that the target application dictates the choice of model features. In fact, the results indicate that depending on the vocation the linear regression and neural network models, which use the same input features, are able to adequately differentiate between the fuel consumption of two vehicles from the same fleet as well as between the fuel consumption of a single vehicle over two different routes. However, the parametric model, which does not utilize PTO active time, is unable to differentiate between two vehicles from the same fleet. This latter model is more suitable for comparing fuel consumption across different fleets of vehicles. In summary, vocation-specific models should be used to optimize fuel consumption for a given fleet of vehicles, whereas general models can only provide insight into aggregated fuel consumption for entire fleets. Moreover, both the accuracy and the precision of the models as measured by their confidence interval should be taken into consideration when comparing fuel consumption estimates for two vehicles from the same fleet or the fuel consumption estimates of an individual vehicle for two different routes. This study shows that the artificial neural network models have narrow 95% confidence intervals and are therefore more precise than the equivalent linear regression models.

## History

Received: 31 Jan 2020  
 Revised: 06 Sep 2020  
 Accepted: 23 Oct 2020  
 e-Available: 10 Nov 2020

## Keywords

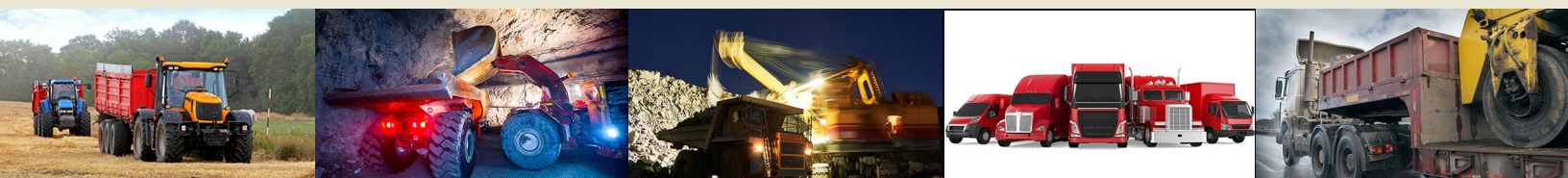
Neural networks, Linear regression, Routes, Vocations, Fuel consumption

## Citation

Schoen, A., Byerly, A., Santos dos Jr, E., and Ben-Miled, Z., "Route-Sensitive Fuel Consumption Models for Heavy-Duty Vehicles," *SAE Int. J. Commer. Veh.* 14(1):85-95, 2021, doi:10.4271/02-14-01-0006.

ISSN: 1946-391X  
 e-ISSN: 1946-3928

© 2021 Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis. Published by SAE International. This Open Access article is published under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits distribution, and reproduction in any medium, provided that the original author(s) and the source are credited.



## Introduction

The ability to analyze the performance of a physical system has various applications and challenges. Different approaches have been proposed in the literature with a varying range of complexity and utility. These approaches fall under two major categories: (1) using data collected directly from the physical system and (2) developing a digital model that replicates the behavior of the physical system. The advantage of the first approach is its reliability. Indeed, since the analysis of the system is based on real data, it typically has a high level of confidence. The disadvantage of this approach is due to the required instrumentation of the physical system over an extended period, which can be cost prohibitive especially under varying operating conditions. To overcome these limitations, digital models have emerged as a cost-effective alternative. Digital models are intended to replicate the behavior of the target physical system under different operating conditions. However, the extent to which they accomplish this goal is not well understood.

The aim of this article is to investigate the accuracy of digital models in estimating instantaneous fuel consumption for individual heavy-duty vehicles when operated on different routes. Specifically, the study evaluates the accuracy of three different models in estimating the instantaneous fuel consumption of two types of heavy-duty vehicles. The first of the three models is a parametric (PAR) model, which was previously proposed by other researchers in [1]; the second is a linear regression (LR) model, and the third is an artificial neural network (ANN) model, which was introduced by the authors in [2]. These models are applied to two different vocations: delivery trucks and refuse trucks. These vocations were selected because they have different operating profiles in terms of average speed, number of stops, vehicle mass, and acceleration.

As in the general case of physical systems, approaches available for modeling fuel consumption in heavy-duty vehicles also fall under the two categories mentioned earlier, namely, using field data from an instrumented vehicle or a digital model. Data collected from an instrumented vehicle was used in [3] to analyze fuel consumption in vehicles with different lubrication oils and fuel types. Field data was also used in [4] to compare various lubrications for engine, axle, and transmission in three 5 ton MTVs. These two studies concluded that using field data makes it difficult to attribute differences in observed performance to either the physical system configuration, the environment, or the operating conditions. For instance, in [4] data was collected from vehicles operating on a flat paved track, making it difficult to extend the results to other routes. Despite these limitations, using field-collected data supports the analysis of the behavior of the vehicle with a high level of accuracy within the boundaries of the specific context and conditions under which the data was collected. Extending the scope of this analysis and the related conclusions to a more general context may entail extensive instrumentation.

Because of the above limitations, digital models were introduced to support fuel consumption analysis in

heavy-duty-vehicles [5, 6, 7, 8, 9]. These models can be further classified into two subcategories: physics based and data driven. Physics-based models use the knowledge of the underlying dynamics of the system to express its behavior using well-defined mathematical equations. FASTSim [6] and Autonomie [10] are example frameworks that support the customization of modular physics-based models to different vehicles by modifying the equations, maps, and parameters that define each module. In the specific case of fuel consumption, a physics-based model for a conventional gasoline sedan is described in [11]. Since physics-based models rely on a deep understanding of the laws of physics that govern the behavior of the system, these models can be highly accurate [5, 7]. However they often require significant adaptation. Moreover, their ability to differentiate between the behavior of two systems that have been exposed to different operating conditions is often limited by our lack of understanding of how these conditions (e.g., aging, routine maintenance, extreme weather) may impact the response of each system.

The aim of the second subcategory of digital models (i.e., data-driven models) is to capture the abovementioned variations in the system's response, which often results from exposure to different operating conditions. Data-driven models express the target output in terms of a linear or nonlinear combination of a selected set of input features while abstracting the system's dynamics. The main advantage of these models is that they can be easily derived, adapted, and deployed for different vehicles and vocations even when knowledge of physics and engineering rules governing the system's operation is not available. However, this subcategory of models faces some challenges such as limited human interpretability [12, 13, 14], complexity, and often lower accuracy [9], compared to their physics-based counterpart.

This article addresses a third challenge related to the fidelity of data-driven models, where fidelity is defined as the ability of the model to accurately estimate the instantaneous fuel consumption of each vehicle rather than a group of vehicles. Despite being one of the main motivations behind the increasing popularity of data-driven models, this aspect has not been extensively investigated. The focus of previous studies has been primarily on demonstrating the accuracy of data-driven models for a single vehicle or a group of similar vehicles [7, 8, 15, 16]. As a result, there is limited understanding of the ability of data-driven models to, for instance, reliably estimate the difference in fuel consumption between two vehicles from the same fleet.

A survey of data-driven fuel consumption models proposed since 2000 is provided in [17]. The authors of the survey highlight the factors that can impact fuel consumption including driver behavior, vehicle, route, weather, and traffic. Moreover, these models vary based on the data being used to train the model, the sampling frequency of this data, and whether the model estimates the average fuel consumption of the vehicle over an entire trip or the instantaneous fuel consumption over a short road segment. In [18], the authors propose an instantaneous fuel consumption model that utilizes global positioning system (GPS) data. The model takes

into consideration the activity of the vehicle and distinguishes between fuel consumption when the vehicle is idle versus moving. Because of the difficulties in obtaining CAN (controller area network) bus data for all the vehicles considered in the study, the model was validated for a single vehicle and then applied to a large number of cabs within a metropolitan area. The authors emphasize the difficulty associated with identifying vehicle driving modes (e.g., acceleration, deceleration, idling) when low-precision GPS data is used. They also show that ignoring driving modes can result in a significant overestimation of fuel consumption. The model proposed in [18] improves on previous models by taking into consideration the operating modes of the vehicle. It is able to provide accurate fuel consumption estimates for a fleet of vehicles. However, its ability to differentiate between two vehicles in the same fleet was not investigated.

The accuracy of several fuel consumption models is also evaluated in [19]. The aim of most of the reviewed models was to estimate average fuel consumption for a group of vehicles (e.g., based on year, make, and model) and most of the models achieve this aim with high accuracy. The authors then introduce an instantaneous fuel consumption model using high-frequency GPS data collected from smartphones. The architecture of the model was based on a recurrent neural network. Because this model relies on GPS data, it can cost-effectively scale up to a large number of vehicles. However, as in the case of the model proposed in [18], the ability of the proposed model to accurately differentiate between the fuel consumption of two vehicles with potentially similar drive cycles was not evaluated.

In fact, regardless of the technique that is used, there is limited information about the ability of previously proposed models to distinguish between the fuel consumption of two vehicles from the same fleet. In the remainder of this section, we review a representative set of these models and their associated techniques. Several techniques can be used to derive data-driven models. We selected three popular techniques for the purpose of this study: PAR, LR, and ANN.

PAR models represent the target output by using a formula based on a set of input features that are usually selected by experts. PAR models are easy to compute. A fuel consumption PAR model based on weight and distances is proposed in [8, 15]. This model estimates average fuel consumption for a fleet of vehicles with similar operating conditions over extended distances. A second PAR model based on the probability distribution of the vehicle speed, the vehicle acceleration, and the accelerator pedal position is described in [16]. This model was developed by using two routes (control and test), six buses, and eleven drivers. The control route was used to calculate indices that correlate with fuel consumption, which are then validated against the test route. While this model is capable of distinguishing between routes, differences in fuel consumption among vehicles are not considered. Similarly, in [7], four duty cycles are used to evaluate the performance of a PAR model and three ANN models using input features derived from vehicle speed and acceleration. The models are trained using one of the duty cycles and tested against the remaining three. All of the

abovementioned models [7, 8, 16] focus on estimating the average fuel consumption over extended distances for a group of vehicles in a fleet.

An instantaneous fuel consumption PAR model was applied to 16 vehicles over 1,000 miles with varying routes in [1]. The model was developed using individual vehicles or clusters of vehicles with the same make/model. The reported errors for individual vehicles were normally distributed between  $\pm 20\%$  with a mean of 0.26%. An improved version of this model was subsequently introduced in [20]. This revised model relies on additional features such as left and right turns, which may not be readily available. For the purpose of comparison and since PAR models are typically designed by domain experts, we compare our proposed model to the PAR model proposed in [1]. This model was selected because it aligns with the aim of this study. First, it estimates instantaneous fuel consumption, and second, it compares these estimates to the actual fuel consumption of individual vehicles in a fleet.

The other two data-driven techniques considered in this study are LR and ANN. It is worth noting that the coefficients of the PAR models described above can also be optimized using linear regression. However, the PAR technique is differentiated from LR to emphasize the significance of using input features selected by subject matter experts. Examples of LR fuel consumption models for long-haul heavy-duty vehicles are described in [21]. Moreover, in a previous study [2] by the authors, an instantaneous fuel consumption model for heavy-duty vehicles using ANN was proposed. This earlier study highlighted the importance of data summarization and feature engineering in the development of an accurate machine learning model for instantaneous fuel consumption.

In summary, with few exceptions [1, 20], most previous models focus on estimating fuel consumption for an entire fleet. In fact, some of the previous studies [22] report that fuel consumption has low variance across vehicles in the same fleet over extended routes. The contribution of this article is toward understanding the ability of digital models to differentiate between two vehicles within the same fleet. Indeed, the main questions that are considered in this article are

- Can a digital model be used to accurately estimate the differences in instantaneous fuel consumption between two vehicles from the same fleet on a new route?
- How does the accuracy of the three techniques being considered differ for different heavy-duty vehicle vocations?

The second question is important because results derived using an example vocation may not translate to other vocations. In fact, it is expected that fuel consumption for a vocation that has a homogeneous operating profile exhibits fewer variations and is easier to model than a vocation that has a heterogeneous operating profile. The two vocations selected for this study are delivery trucks and refuse trucks. The remainder of the article presents the methods used to create each model, describes the datasets used for each vocation, and investigates the accuracy of the models for individual vehicles when exposed to different routes.

## Data-Driven Models

Three different approaches to deriving instantaneous fuel consumption models are compared. In this study, instantaneous refers to the ability of the model to estimate fuel consumption aggregated over a short distance of 1 km. In an earlier study [2], aggregation over time and distance were compared, and the latter was found to be more adequate for modeling fuel consumption in heavy-duty vehicles.

The first of the three models being considered in this study is the PAR model, which was introduced by other researchers in [1]. The features of this model are derived using expert knowledge. Feature selection was guided by a force balance equation that estimates the force acting on a vehicle by the engine, friction, air resistance, and gravity. The authors expressed these forces using basic units of measure (e.g., mass and speed) in relation to fuel consumption. The resulting model defines fuel consumption in gallons per mile (*gpm*) as follows:

$$gpm = q_1 m \bar{v}^2 \frac{ST + vTL}{\Delta d} + q_2 m \frac{\bar{v}^2}{2\Delta d} + q_3 m \cos(\alpha) + q_4 a \bar{v}^2 + q_5 m \sin(\alpha) \quad \text{Eq. (1)}$$

where  $m$  is the vehicle mass,  $v$  is the speed of the vehicle,  $\bar{v}$  is the average speed,  $ST$  is the number of stops signs,  $TL$  is the number of traffic lights,  $\Delta d$  is the distance traveled,  $\alpha$  is the road grade, and  $a$  is the vehicle frontal area. The coefficients  $q_1$  through  $q_5$  are learned using data collected for each vehicle. Linear regression is used to derive the values of these coefficients for each vehicle. This model was modified to facilitate the comparison to the other two models under consideration in this article, as shown below:

$$lpk = q_1 m \bar{v}^2 \frac{ST}{\Delta d} + q_2 m \frac{\bar{v}^2}{2\Delta d} + q_3 m \cos(\alpha) + q_4 \bar{v}^2 + q_5 m \sin(\alpha) \quad \text{Eq. (2)}$$

where  $ST$  is equated to the number of stops,  $TL$  is omitted because it is not available,  $a$  is omitted because it is a constant for a given vehicle, and the fuel consumption was converted from gallons per mile to liters per 100 km (*lpk*). The features of the PAR model in Equation 1 are summarized in Table 1.

**TABLE 1** Features of the PAR model.

Feature	Description
$x_1$	$m \cdot \bar{v}^2 \cdot \frac{ST}{\Delta d}$
$x_2$	$m \cdot \frac{\bar{v}^2}{2\Delta d}$
$x_3$	$m \cdot \cos(\alpha)$
$x_4$	$\bar{v}^2$
$x_5$	$m \cdot \sin(\alpha)$

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

**TABLE 2** Features for the LR and ANN models.

Feature	Description
$x_6$	Number of stops
$x_7$	Idle time
$x_8$	Average moving speed
$x_9$	Change in kinetic energy
$x_{10}$	Change in potential energy
$x_{11}$	Aerodynamic speed squared
$x_{12}$	Characteristic acceleration
$x_{13}$	PTO active time

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

The second model uses linear regression with input features that are calculated from mass, vehicle speed, road grade, and power take-off (PTO) active time. These features are listed in Table 2. The ability of these features to accurately estimate fuel consumption for heavy-duty vehicles was demonstrated in [2]. These features are able to capture the vehicle dynamics as well as the driver's behavior, as shown in [2]. Moreover, since the features are derived from three basic measurements: vehicle speed, road grade, and mass; the resulting model is applicable to a wide range of vocations.

The features of the LR model (Table 2) include the number of stops, the idle time, and the average moving speed over a step distance of 1 km. The expressions for change in kinetic energy ( $x_9$ ) and change in potential energy ( $x_{10}$ ) are included in Equations 3 and 4, respectively.

$$x_9 = \frac{1}{2} m (v_N - v_1)^2 \quad \text{Eq. (3)}$$

$$x_{10} = m g \sum_{i=1}^{N-1} \Delta h_{i,i+1} \quad \text{Eq. (4)}$$

where  $m$  is the vehicle mass,  $v_N - v_1$  represent the difference in speed over the 1 km distance,  $g$  is the gravitational constant,  $\Delta h_{i,i+1}$  is the difference in elevation between two consecutive samples, and  $N$  is the total number of samples in the 1 km space step. The difference in elevation is derived using the road grade between two consecutive samples in the 1 km space step [i.e.,  $\Delta h_{i,i+1} \approx \sin(\alpha_i)$ ].

The aerodynamic speed ( $x_{11}$ ) and characteristic acceleration ( $x_{12}$ ) are calculated according to the definitions introduced in [25]. The importance of these features in generating accurate fuel consumption estimates is analyzed in [2, 26]. Aerodynamic speed and characteristic acceleration are defined as follows:

$$x_{11} \approx \frac{1}{\Delta d} \sum_{i=1}^{N-1} \left( \frac{v_{i+1} + v_i}{2} \right)^3 \quad \text{Eq. (5)}$$

$$x_{12} = \frac{1}{\Delta d} \sum_{i=1}^{N-1} \left( \frac{v_{i+1}^2 - v_i^2}{2} + g \cdot \Delta h_{i,i+1} \right) \quad \text{Eq. (6)}$$

The last feature in [Table 2](#) is PTO ( $x_{13}$ ), which represents the additional power requested from the engine to operate accessories. Two vocations are considered in this study: delivery trucks and refuse trucks. Refuse trucks use PTO for trash compaction and bin pickup. These operations demand a high percentage of engine power. Delivery trucks do not have accessories requiring PTO. Therefore, this feature is included in the refuse truck models and omitted from the delivery truck models.

The third model under investigation in this study is a three-layer ANN model [2], which was trained using back-propagation. The input layer consists of the features listed in [Table 2](#), and the hidden layer includes five nodes. This network configuration was selected after analyzing different network architectures with varying numbers of hidden layers and nodes in each hidden layer. The results showed that no significant improvements are achieved with additional hidden layers or nodes.

## Data Collection and Processing

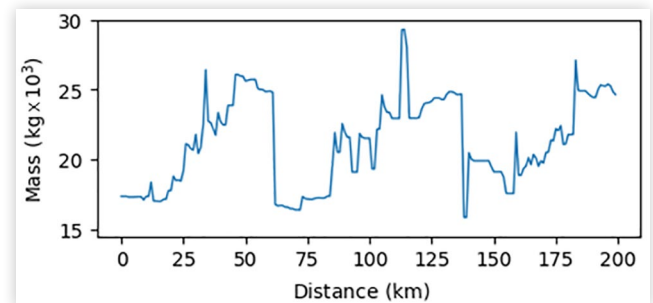
The above three models are trained to estimate instantaneous fuel consumption in liters per 100 km for each 1 km of distance traveled. The data used to train and validate the models is collected from two vocations: delivery trucks and refuse trucks. For the two vocations, data is collected at a rate of 1 Hz and then aggregated over a space step of 1 km. The characteristics of the vehicles for each of these two vocations are shown in [Table 3](#). The delivery truck data (DT) was collected from a single vehicle with twelve drivers over two routes with both city and highway segments. The truck load does not vary over the route, and therefore the mass of the vehicle is relatively constant. Drivers were instructed to exhibit either bad or good driving behavior through coasting and anticipating braking. Due to the limited sample size of the DT dataset, additional data was synthetically created. Trip data for each route was separated into segments defined by consecutive vehicle stops. These segments were then randomly sampled with replacement to generate 15 km trips as described in [2].

**TABLE 3** Heavy-duty vehicles' characteristics.

	DT	RT
Description	Delivery truck	Refuse, front-end loader
Class	5	8
Gross vehicle weight	8,700 (kg)	28,000 (kg)
Engine type	2014 Hino J05E-TP	2015 ISL-G, 320 HP
Number of axles	2	3

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

**FIGURE 1** Vehicle mass versus distance for a sample refuse truck.



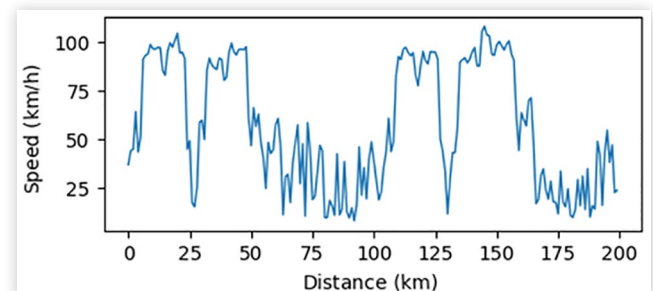
© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

The refuse truck data (RT) was collected from five vehicles over an extended distance. Therefore, synthetically augmenting the data was not necessary. Driver behavior was also not available for this vocation. The operating profile of the refuse trucks differs considerably from that of the delivery trucks. For instance, the mass of each refuse truck varies throughout the route due to trash pickup and unloading in the depot ([Figure 1](#)). However, there is no sensor on the vehicle that directly measures the mass of the vehicle. Therefore, the mass was estimated using the online estimator developed in [23, 24]. While these are widely accepted methodologies for mass estimation, these estimates may include errors as exemplified by the peaks in [Figure 1](#).

Refuse trucks also mainly operate over two ranges of speed: low speed during trash collection and high speed while traveling from/to the depot, as shown in [Figure 2](#). In contrast, the duty cycle for the delivery trucks varies intermittently between low and high speeds, as shown by the sample duty cycle in [Figure 3](#).

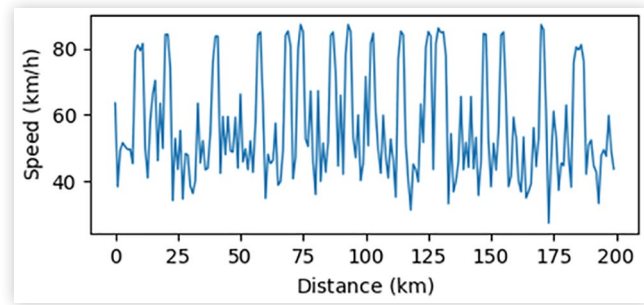
Moreover, as mentioned earlier, refuse trucks are equipped with PTO for bin pickup and trash compaction, which requires significant power. Bin pickup and trash compaction occur while the vehicles are operating in their residential service areas. As in the case of mass, there is no sensor that captures PTO active time ( $x_{13}$ ) for refuse trucks.

**FIGURE 2** A sample duty cycle for a refuse truck over a distance of 200 km.



© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

**FIGURE 3** A sample duty cycle for a delivery truck over a distance of 200 km.



© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

This parameter was estimated as the total time the fuel rate is greater than 5 liters per hour while the vehicle is at low speed. This threshold was empirically derived from the available data. The data shows that when the vehicle is idle, the baseline fuel rate is less than 1.5 liter per hour. When PTO is active, the additional load requires more engine fueling, and therefore higher fuel rates are observed. This analysis was performed for all refuse vehicles on all segments, and the threshold of 5 liters per hour at low speed was established as an indicator of PTO active time.

## Performance Evaluation

Three fuel consumption models (PAR, LR, and ANN) and two vocations (RT and DT) are studied in this article. A methodology is needed to compare these models and to assess the accuracy of each model in estimating instantaneous fuel consumption for each individual vehicle. This methodology is described next.

Let  $\phi(\cdot)$  be a function that represents the actual fuel consumption of a target vehicle, and let  $\hat{\phi}(\cdot)$  represent the fuel consumption estimate generated by a model for the same vehicle. An input data point to the model is a vector  $\bar{X}$  consisting of the normalized values of the features in [Table 1](#) for PAR and [Table 2](#) for LR and ANN. The training ( $R$ ) and testing ( $S$ ) datasets are a set of input vectors that are used to train and test the models, respectively. The  $R$  and  $S$  datasets follow a 70/30 split from a total of 4,650 km for each vehicle and route in both the DT and RT vocations, where each 1 km corresponds to a data point, that is, the  $R$  and  $S$  datasets include a total of 3,255 and 1,395 data points, respectively. These data points represent feature values that are aggregated over a 1 km space step. The aggregation follows the equations in [Table 1](#) for the PAR model, where  $v$  is the average speed over 1 km,  $\Delta d = 1$  km, and  $ST$  represents the number of stops in each 1 km segment. Similarly, the features in [Table 3](#) for the LR and ANN models are aggregated over 1 km segments. The number of stops, idle time, average moving speed, and PTO active time are calculated over each 1 km space step. The change in kinetic energy, change in potential energy,

aerodynamic speed squared, and characteristics acceleration are also calculated for each 1 km space step according to [Equations 3, 4, 5, and 6](#), respectively.

To develop a model, a subset of the data is selected for training, and the remaining portion of the data is used to test the model. This selection is performed at random. Therefore, each selection can yield a different model. Moreover, a large variance can be observed from one model to the next [\[27, 28\]](#), depending on the selected training data. One approach to overcoming this potential instability is to perturb the training data, create multiple models, and use the average of the models [\[27\]](#). This approach is commonly known as k-fold cross-validation, where  $k$  represents the number of models or consequently the number of different sampling of the training data. In this study, a fivefold cross-validation was used. Each of the fivefold models is trained with 70% randomly selected data points from the training dataset ( $R$ ). The overall fuel consumption model is an ensemble of the five models. Moreover, the fuel consumption estimate is the average of the estimates generated by each of the five models. To simplify the notation, a model is used to refer to the above-described ensemble of five models for each technique in the remainder of the article.

The fuel consumption estimated by the model is subject to deviation from the actual fuel consumption. The relationship between the actual and estimated fuel consumption is given by

$$\hat{\phi}(\bar{X}) = \phi(\bar{X}) + \epsilon(\bar{X}) \quad \text{Eq. (7)}$$

The error term ( $\epsilon$ ) in the above equation is generally used to evaluate the accuracy of the model in estimating fuel consumption for data points that were not observed during training. However, when applied to the training dataset ( $R$ ), this error also provides an indication of any inherent bias in the model to overestimate or underestimate the actual fuel consumption of the vehicle. In fact, one of the objectives of k-fold cross-validation is to reduce this inherent bias [\[28\]](#). However, previous analysis [\[29\]](#) has shown that this bias still exists even when an ensemble of models is used. The bias for each model is estimated as follows:

$$\beta = \frac{1}{r} \sum_{i=1}^r \epsilon(\bar{X}_i) \quad \text{Eq. (8)}$$

where  $\bar{X}_i \in R$  and  $r$  is the number of data points in  $R$ . The bias  $\beta$  as defined by [Equation 8](#) is an indication of whether the fuel consumption estimate generated by the model is expected to be lower or higher than the actual fuel consumption of the vehicle. The model is then exposed to the testing dataset ( $S$ ), and the fuel consumption estimates generated by the model are adjusted by the bias value in [Equation 8](#). This adjustment corrects for the model's tendency to overestimate or underestimate fuel consumption [\[29\]](#). The model average testing error ( $\mu$ ) is then computed over all the data points in  $S$  as shown below:

$$\mu = \frac{1}{s} \sum_{i=1}^s \epsilon(\bar{X}_i) - \beta \quad \text{Eq. (9)}$$

where  $\bar{X}_i \in S$  and  $s$  is the number of data points in  $S$ . Finally, the 95% confidence interval (CI) of the average testing error  $\mu$  is evaluated by using empirical bootstrapping [29, 30] over 5,000 samples, each with  $s$  data points, randomly selected with replacement from  $S$ , as shown below:

$$\left[ \mu - \delta_{.025}^*, \mu - \delta_{.975}^* \right] \quad \text{Eq. (10)}$$

where the sequence of differences  $\delta^* = \mu^* - \mu$  is computed for each bootstrap sample and  $\delta_{.025}^*$  and  $\delta_{.975}^*$  are considered to represent estimates of  $\delta_{.025}$  and  $\delta_{.975}$ , respectively. The bootstrapping technique used to evaluate the confidence interval of the model error and the k-fold technique used in the training are both statistical sampling techniques [27] that help provide more accurate statistical estimates of the distribution of the target parameter. In the case of training, k-fold reduces the variability that may result from the choice of the training data. In the case of testing, bootstrapping is performed by sampling with replacement the test data 5,000 times and evaluating the distribution of the prediction error. For each model, the 95% confidence interval has a probability of 0.95 of containing the mean error of the fuel consumption estimates generated by the model.

## Results

The first vocation under consideration in this study consists of a single delivery truck that was operated over two routes. Two models  $\phi^A$  and  $\phi^B$  are developed for this vehicle using the datasets  $R^A$  and  $R^B$  from routes  $A$  and  $B$ , respectively. Table 4

**TABLE 4** Actual and estimated fuel consumption for the test dataset ( $S$ ) from routes  $A$  and  $B$  for the delivery truck vocation. All values are in liters per 100 km.

Model	Route	Actual average fuel consumption	Estimated average fuel consumption	Error ( $\mu$ )
PAR				
$\phi^A$	A	22.41	21.86	0.55
$\phi^A$	B		22.00	-3.09
$\phi^B$	B	18.91	18.60	0.31
$\phi^B$	A		15.24	7.17
LR				
$\phi^A$	A	22.41	22.43	-0.02
$\phi^A$	B		19.22	-0.31
$\phi^B$	B	18.91	19.37	-0.46
$\phi^B$	A		22.47	-0.06
ANN				
$\phi^A$	A	22.41	22.40	0.01
$\phi^A$	B		18.41	0.50
$\phi^B$	B	18.91	18.88	0.03
$\phi^B$	A		21.54	0.87

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

shows the actual and estimated average fuel consumption of the two models for the testing datasets  $S^A$  and  $S^B$ . The fuel consumption and the errors are reported in terms of actual liters per 100 km. Since fuel consumption varies widely across the two vocations, using actual fuel consumptions and absolute errors instead of percent errors facilitates the comparison between the two vocations and allows the error to be considered with respect to the confidence interval of the model.

Several observations can be derived from Table 4. The actual average fuel consumption for route  $A$  is higher than that of route  $B$ , and this trend is estimated correctly by the LR and ANN models. The PAR model fails to capture this difference between the two routes. When the PAR model is trained using route  $A$  ( $\phi^A$ ), the fuel consumption estimates for routes  $A$  and  $B$  are nearly the same and when the PAR model is trained on route  $B$  ( $\phi^B$ ), the fuel consumption estimate for route  $A$  (15.24 liters per 100 km) is lower than the fuel consumption estimate for route  $B$  (18.60 liters per 100 km). The opposite is expected according to the actual recorded fuel consumption. The improved performance of the LR and ANN models compared to the PAR model is mainly due to the use of features that can better capture variations in fuel consumption. Both the LR and ANN models use the same features.

The confidence intervals of the two models for this vocation are shown in Table 5. For both routes, the confidence intervals of the ANN models are smaller than the confidence interval of the LR and PAR models. A smaller confidence interval indicates that the corresponding model is more precise.

The DT was collected from a single vehicle under relatively controlled conditions with respect to the routes, vehicle load, and driver behavior. Therefore the above results may not be applicable to other vocations under typical operating conditions. The RT was collected from five different vehicles in the same fleet during their routine operation. Refuse trucks consume more fuel per unit distance than delivery trucks due to their frequent stop-and-go operation and their use of the PTO for bin pickup and trash compaction. In fact, as discussed in Section 3, their operating profile varies considerably from that of delivery trucks (Figures 1-3).

The actual and estimated fuel consumption for the five refuse trucks are shown in Table 6. The actual average fuel consumption for  $\phi^A$  is higher than that of  $\phi^B$  by more than 10 liters per 100 km. This shows a significant variation in fuel consumption across the vehicles and their respective routes compared to the delivery truck. In contrast, the difference in fuel consumption across the two routes for the delivery truck is approximately 3.5 liters per 100 km (Table 4).

**TABLE 5** 95% confidence intervals for the delivery truck fuel consumption estimates in liters per 100 km.

Model	PAR	LR	ANN
$\phi^A$	[0.245, 0.890]	[-0.183, 0.124]	[-0.132, 0.115]
$\phi^B$	[0.063, 0.569]	[-0.547, -0.375]	[-0.055, 0.112]

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

**TABLE 6** Actual and estimated fuel consumption for the test dataset (S) from routes A through E for the refuse truck vocation. All values are in liters per 100 km.

Model	Route	Average actual fuel consumption	Estimated average fuel consumption	Error ( $\mu$ )
PAR				
$\phi^A$	A	88.55	89.07	-0.52
$\phi^B$	B	85.63	86.65	-1.02
$\phi^C$	C	83.65	83.96	-0.31
$\phi^D$	D	78.37	80.08	-1.71
$\phi^E$	E	73.15	73.70	-0.55
LR				
$\phi^A$	A	88.55	88.20	0.35
$\phi^B$	B	85.34	85.30	0.04
$\phi^C$	C	83.21	83.35	-0.14
$\phi^D$	D	77.77	77.76	0.01
$\phi^E$	E	73.15	73.65	-0.50
ANN				
$\phi^A$	A	88.55	88.01	0.54
$\phi^B$	B	85.34	85.49	-0.15
$\phi^C$	C	83.21	83.20	0.01
$\phi^D$	D	77.77	77.82	-0.05
$\phi^E$	E	73.15	73.75	-0.60

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

As in the case of the delivery truck, [Table 6](#) also shows that the PAR models for the refuse trucks have higher error than the LR and ANN models in most cases. However, as opposed to the delivery truck, the PAR model fuel consumption estimates for the refuse trucks follows the increasing trend of the actual fuel consumption across the vehicles. This indicates that the PAR model is able to better differentiate between the vehicles and routes for the refuse truck vocation than for the delivery truck vocation. The increasing fuel consumption estimates remain consistent for this vocation with the actual fuel consumption for the LR and ANN models.

The confidence intervals of the refuse truck models are included in [Table 7](#). As in the case of the delivery truck vocation, the PAR models for the refuse truck vocation have the widest confidence intervals, and the confidence intervals for the ANN models are either smaller or comparable to those

**TABLE 7** 95% confidence intervals for the refuse truck fuel consumption estimated in liters per 100 km.

Model	PAR	LR	ANN
$\phi^A$	[-3.079, 1.962]	[-0.360, 1.068]	[-0.061, 1.130]
$\phi^B$	[-3.658, 1.567]	[-0.566, 0.625]	[-0.698, 0.370]
$\phi^C$	[-2.892, 2.071]	[-0.783, 0.516]	[-0.500, 0.509]
$\phi^D$	[-4.167, 0.765]	[-0.593, 0.594]	[-0.539, 0.411]
$\phi^E$	[-2.656, 1.461]	[-1.145, 0.140]	[-1.217, -0.008]

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

of the LR models. This observation confirms that the ANN models tend to be most precise and the PAR models tend to be the least precise.

An important question in fleet management application is how much fuel would vehicle A consume if it is operated on the route that is practiced by vehicle B. [Tables 8, 9, and 10](#) show the estimated fuel consumption when the model for a given vehicle is tested using the test duty cycles derived from the routes driven by the other refuse trucks for the PAR, LR, and ANN models, respectively. For example, model  $\phi^A$  was created using training data from vehicle A collected on route A. This model is applied to the duty cycle of route B. The test input features for route B are derived using the mass, the road grade, the speed, and the PTO active time. However, since vehicle A is not actually operated on road B, the actual fuel consumption is not available for comparison. The fuel consumption values included in [Tables 8, 9, and 10](#) are the estimates generated by the models. The last two rows and columns of these tables include the average and the standard deviation of the fuel consumption estimates for each vehicle across all the routes and for each route across all the vehicles, respectively.

[Table 8](#) shows that the variation in the PAR fuel consumption estimates for each vehicle across the routes is insignificant especially when taking into consideration the wide confidence interval of this model ([Table 7](#)). This is not true for the estimates of the LR and ANN models ([Tables 9 and 10](#)). In fact, the difference in estimated fuel consumption by these two models for a vehicle can be as high as 13 liters per 100 km (i.e.,  $\phi^C$  on route B versus route E). [Table 8](#) also shows that the PAR model for vehicle  $\phi^E$  has significantly lower estimated average fuel consumption across all the routes than vehicle  $\phi^B$ . However, the estimated average fuel consumption across all the routes for vehicles  $\phi^B$  and  $\phi^E$  by the LR and ANN models are comparable.

With respect to a given route, the estimated fuel consumption across the vehicles varies significantly for all models ([Tables 8, 9, and 10](#)). However, the models have different trends for different routes. For example, the average estimated fuel consumption across all the vehicles by the LR and ANN models for route E is the lowest. This is not the case for the PAR model.

Several other trends can be identified from the above results. In general, the proposed methodology can be used to help assign a vehicle to a given route by a fleet manager. Nonetheless, the ability of each model to differentiate between routes and vehicles and the confidence interval of these models must be taken into consideration. For example, the difference in fuel consumption estimates for vehicle A ( $\phi^A$ ) on routes B and C generated by the PAR model is about 1 liter per 100 km. This difference is not significant given the 95% confidence interval of [-3.079, 1.962] of this model. However, the difference for the same example vehicle and routes in the case of the LR and ANN models is more than 6.5 liters per 100 km, where the 95% confidence interval for these models is [-0.360, 1.068] and [-0.061, 1.130], respectively.

**TABLE 8** PAR-estimated average fuel consumption when a given vehicle model uses test data from other routes. All values are in liters per 100 km.

		$\phi^A$	$\phi^B$	$\phi^C$	$\phi^D$	$\phi^E$	Avg.	Std.
Route	A	89.07	88.35	85.15	79.05	72.64	82.85	6.95
	B	86.63	86.65	83.78	77.55	69.71	80.87	7.26
	C	87.66	87.66	83.96	77.61	70.72	81.52	7.30
	D	89.71	87.48	86.04	80.08	73.70	83.40	6.49
	E	89.59	87.59	85.80	79.70	73.70	83.27	6.50
	Avg	88.53	87.55	84.94	78.80	72.09		
	Std	1.34	0.61	1.03	1.17	1.81		

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

**TABLE 9** LR-estimated average fuel consumption when a given vehicle model uses test data from other routes. All values are in liters per 100 km.

		$\phi^A$	$\phi^B$	$\phi^C$	$\phi^D$	$\phi^E$	Avg.	Std.
Route	A	88.20	78.71	81.86	77.52	77.65	80.79	4.50
	B	96.45	85.30	90.47	84.01	84.26	88.10	5.35
	C	89.80	79.86	83.35	78.81	78.79	82.12	4.68
	D	86.95	77.83	80.96	77.76	77.44	80.19	4.04
	E	83.60	74.55	77.22	73.80	73.65	76.56	4.19
	Avg	89.00	79.25	82.77	78.38	78.36		
	Std	4.75	3.92	4.86	3.67	3.83		

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

## Conclusions

Data-driven models for physical systems offer considerable advantages compared to physics-based models. These advantages include the ability of these models to cost-effectively generalize to multiple contexts and operating conditions. However, this advantage may come at the cost of lack of accuracy especially when analyzing similar systems, such as two vehicles, from the same fleet.

In this article, different data-driven models are applied to two heavy-duty vehicle vocations. These models estimate instantaneous fuel consumption over a space step of 1 km. Moreover, to allow the models to estimate fuel consumption

for each individual vehicle on a given route the following methodology was adopted:

- Five different models are developed for each vehicle using k-fold cross-validation. An ensemble model is then constructed by averaging the fuel consumption estimates from each of the five models. The goal of this ensemble learning approach is to limit the variabilities in the estimates of the model, resulting from the use of different training data.
- Once the model is trained, its bias with respect to any inherent tendencies of the model toward overestimating or underestimating fuel consumption is evaluated using the training data.

**TABLE 10** ANN-estimated average fuel consumption when a given vehicle model uses test data from other routes. All values are in liters per 100 km.

		$\phi^A$	$\phi^B$	$\phi^C$	$\phi^D$	$\phi^E$	Avg.	Std.
Route	A	88.01	79.07	82.27	77.33	78.38	81.01	4.32
	B	96.37	85.49	89.87	83.31	85.34	88.08	5.22
	C	89.50	80.47	83.20	78.18	79.22	82.11	4.54
	D	87.89	78.97	82.08	77.82	78.63	81.08	4.14
	E	83.34	75.05	77.42	73.29	73.75	76.57	4.11
	Avg	89.02	79.81	82.97	77.99	79.06		
	Std	4.71	3.76	4.46	3.57	4.13		

© Allison Transmission, Inc. and Indiana University-Purdue University Indianapolis.

- The ensemble model is then applied to the held-out testing data, and the estimated fuel consumption produced by the model is adjusted by the bias value established in the previous step. Besides, bootstrapping is used to determine the 95% confidence interval of each model.
- The estimated fuel consumption for a given vehicle on a new route is then derived, and the results for two vehicles on the same route or the same vehicle on two routes are analyzed while taking into consideration the confidence interval of the model.

The methodology outlined above was first used on a single delivery truck that was operated on two routes. Each of the routes was used to develop an independent model of the vehicle. Because actual fuel consumption data was available for the vehicle over the two routes, this experiment was used to validate the proposed methodology. The results show that the PAR model is not able to estimate the fuel consumption of the vehicle for a different route, whereas the LR and ANN models produce estimates with an error of less than 1 liter per 100 km.

The same methodology was then applied to five refuse trucks over five routes. Compared to the delivery trucks, this vocation shows higher variability in fuel consumption. The LR and ANN models have lower error and smaller 95% confidence interval than the PAR model. Moreover, the PAR model was unable to differentiate among the routes for a given vehicle. Therefore, this model is more suitable for comparing aggregated fuel consumption among fleets of vehicles. In contrast, the LR and ANN models were able to distinguish between two vehicles and two routes. They can be used to assign a vehicle to a given route for the purpose of optimizing the overall fuel consumption in the fleet.

The advantage of the LR and ANN models results from the use of features that better capture the operating profile of the vehicle. Most of these features are derived from vehicle speed and road grade. These parameters are widely available. The exceptions are mass and PTO. Mass was estimated using a widely accepted estimator, and an empirical rule was developed to estimate the PTO active time. In addition to outlining a methodology for estimating the instantaneous fuel consumption of a given vehicle on a specific route, this study shows the need for adapting each model to the target vocation. The methodology was applied to two vocations with different operating profiles. While most of the features are similar for both vocations, the omission of PTO from the PAR model resulted in refuse truck models with much lower precision than the delivery truck models.

There are several directions for future work. These include testing the proposed methodology for other vocations, identifying techniques for improving the confidence intervals of the models, and developing a methodology that can determine the minimum number of data points needed to adequately train the model to achieve the desired target precision.

## Contact Information

### Alexander Schoen<sup>1</sup>

Electrical and Computer Engineering Department

<sup>1</sup>Indiana University Purdue University at Indianapolis, USA

### Euzeli Cipriano dos Santos Jr<sup>1</sup>

Electrical and Computer Engineering Department

<sup>1</sup>Indiana University Purdue University at Indianapolis, USA

### Zina Ben-Miled<sup>1</sup>

Electrical and Computer Engineering Department

<sup>1</sup>Indiana University Purdue University at Indianapolis, USA

## Acknowledgments

This research was supported in part by Allison Transmission, Inc.

## Definitions, Acronyms, Abbreviations

**DT** - Delivery trucks

**ANN** - Artificial neural networks

**LR** - Linear regression

**PAR** - Parametric model

**PTO** - Power take-off

**R** - Training set

**RT** - Refuse trucks

**S** - Testing set

## References

1. Ganti, R.K., Pham, N., Ahmadi, H., Nangia, S. et al., "GreenGPS: A Participatory Sensing Fuel-Efficient Maps Application," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, San Francisco, 151-164, 2010.
2. Schoen, A., Byerly, A., Hendrix, B., Bagwe, R. et al., "A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles," *IEEE Transactions on Vehicular Technology* 68(7):6343-6351, 2019.
3. Haggis, S.F., Hansen, T.A., Hicks, K.D., Richards, R.G. et al., "In-Use Evaluation of Fuel Economy and Emissions from Coal Haul Trucks Using Modified SAE J1321 Procedures and PEMs," *SAE Int. J. Commer. Veh.* 1(1):210-221, 2008, <https://doi.org/10.4271/2008-01-1302>.
4. Brandt, A.C., Frame, E.A., and Warden, R.W., "SAE J1321 Testing Using m1083a1 FMTVs," No. TFLRF-404, Tardec Fuels and Lubricants Research Facility, Southwest Research Inst., San Antonio, TX, 2010.

5. Gao, D.W., Mi, C., and Emadi, A., "Modeling and Simulation of Electric and Hybrid Vehicles," *Proceedings of the IEEE* 95(4):729-745, 2007.
6. Gonder, J.D., Brooker, A.D., Wood, E.W., and Moniot, M., "Future Automotive Systems Technology Simulator (FASTSim) Validation Report," No. NREL/TP-5400-71168, National Renewable Energy Lab. (NREL), Golden, CO, 2018.
7. Wang, L., Duran, A., Gonder, J., and Kelly, K., "Modeling Heavy/Medium-Duty Fuel Consumption Based on Drive Cycle Properties," SAE Technical Paper 2015-01-2812, 2015, <https://doi.org/10.4271/2015-01-2812>.
8. Kopfer, H.W., Schönberger, J., and Kopfer, H., "Reducing Greenhouse Gas Emissions of a Heterogeneous Vehicle Fleet," *Flexible Services and Manufacturing Journal* 26(1-2):221-248, 2014.
9. Bhoraskar, A., "Prediction of Fuel Consumption of Long Haul Heavy Duty Trucks Using Machine Learning and Comparison of the Performance of Various Learning Techniques," Delft University of Technology, 2019.
10. Halbach, S., Sharer, P., Pagerit, S., Rousseau, A.P. et al., "Model Architecture, Methods, and Interfaces for Efficient Math-Based Design and Simulation of Automotive Control Systems," SAE Technical Paper 2010-01-0241, 2010, <https://doi.org/10.4271/2010-01-0241>.
11. Wood, E., Gonder, J., and Jehlik, F., "On-Road Validation of a Simplified Model for Estimating Real-World Fuel Economy," *SAE Int. J. Fuels Lubr.* 10(2):528-536, 2017, <https://doi.org/10.4271/2017-01-0892>.
12. Breiman, L., "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)," *Statistical Science* 16(3):199-231, 2001.
13. Freitas, A.A., "Comprehensible Classification Models: A Position Paper," *ACM SIGKDD Explorations Newsletter* 15(1):1-10, 2014.
14. Doshi-Velez, F. and Kim, B., "Towards a Rigorous Science of Interpretable Machine Learning," arXiv preprint arXiv:1702.08608, 2017.
15. Xiao, Y., Zhao, Q., Kaku, I., and Yuchun, X., "Development of a Fuel Consumption Optimization Model for the Capacitated Vehicle Routing Problem," *Computers & Operations Research* 39(7):1419-1431, 2012.
16. Huang, D., Xie, H., Sun, Q., Ma, H. et al., "A Statistical Evaluation Model for Driver-Bus-Route Combinatorial Optimization," in *2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, Hangzhou, 1423-1428, IEEE, 2015.
17. Zhou, M., Jin, H., and Wang, W., "A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing," *Transportation Research Part D: Transport and Environment* 49:203-218, 2016.
18. Kan, Z., Tang, L., Kwan, M.-P., and Zhang, X., "Estimating Vehicle Fuel Consumption and Emissions Using GPS Big Data," *International Journal of Environmental Research and Public Health* 15(4):566, 2018.
19. Kanarachos, S., Mathew, J., and Fitzpatrick, M.E., "Instantaneous Vehicle Fuel Consumption Estimation Using Smartphones and Recurrent Neural Networks," *Expert Systems with Applications* 120:436-447, 2019.
20. Saremi, F., Fatemieh, O., Ahmadi, H., Wang, H. et al., "Experiences with Greengps—Fuel-Efficient Navigation Using Participatory Sensing," *IEEE Transactions on Mobile Computing* 15(3):672-689, 2015.
21. Almér, H., "Machine Learning and Statistical Analysis in Fuel Consumption Prediction for Heavy Vehicles," KTH Royal institute of Technology, 2015.
22. Hansson, K., "Data-Driven Analysis of the Fuel Saving Potential of Road Vehicle Platooning: A Data-Driven Approach for Quantifying the Fuel Saving Effects of Platooning Based on Data Collected in Real Traffic Conditions," KTH Royal institute of Technology, 2013.
23. Fathy, H.K., Kang, D., and Stein, J.L., "Online Vehicle Mass Estimation Using Recursive Least Squares and Supervisory Data Extraction," in *2008 American Control Conference*, Seattle, WA, 1842-1848, IEEE, 2008.
24. Winstead, V. and Kolmanovsky, I.V., "Estimation of Road Grade and Vehicle Mass via Model Predictive Control," in *Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005*, Toronto, Ontario, 1588-1593, IEEE, 2005.
25. O'Keefe, M.P., Simpson, A., Kelly, K.J., and Pedersen, D.S., "Duty Cycle Characterization and Evaluation towards Heavy Hybrid Vehicle Applications," SAE Technical Paper 2007-01-0302, 2007, <https://doi.org/10.4271/2007-01-0302>.
26. Kelly, K., Bennion, K., Miller, E., and Prohaska, B., "Medium-and Heavy-Duty Vehicle Duty Cycles for Electric Powertrains," No. NREL/PR-5400-66228, National Renewable Energy Lab. (NREL), Golden, CO, 2016.
27. Breiman, L., "Heuristics of Instability and Stabilization in Model Selection," *The Annals of Statistics* 24(6):2350-2383, 1996.
28. Nadeau, C. and Bengio, Y., "Inference for the Generalization Error," in *Machine Learning* 52(3):239-281, 2003.
29. Bengio, Y. and Grandvalet, Y., "No Unbiased Estimator of the Variance of k-Fold Cross-Validation," *Journal of Machine Learning Research* 5(Sept.):1089-1105, 2004.
30. Giné, E. and Zinn, J., "Bootstrapping General Empirical Measures," *The Annals of Probability* 851-869, 1990.

