PROCEEDINGS OF THE CANADIAN SOCIETY FOR MECHANICAL ENGINEERING INTERNATIONAL CONGRESS 2022
CSME CONGRESS 2022
JUNE 5-8, 2022, EDMONTON, AB, CANADA

ESTIMATING INSTANTANEOUS FUEL CONSUMPTION OF VEHICLES BY USING MACHINE LEARNING AND REAL-TIME ON-BOARD DIAGNOSTICS (OBD) DATA

Amir Ansari[1], Hamidreza Abediasl[1], Parth Rakeshkumar Patel[1], Vahid Hosseini[2], Charles Robert Koch[1], Mahdi Shahbakhti[1*]
[1]Department of Mechanical Engineering, University of Alberta, Edmonton, Canada
[2]School of Sustainable Energy Engineering, Simon Fraser University, Surrey, Canada
[*]mahdi@ualberta.ca

***Abstract*—Estimation of instantaneous fuel consumption of fleet vehicles to identify the causes of high fuel consumption and determine the optimum vehicle type for different applications and driving cycles is essential for the design of an intelligent fleet management system. Developing a practical and reliable method to estimate instantaneous fuel consumption of fleet vehicles is the focus of this study. The proposed method uses real-time on-board diagnostics (OBD) data from a vehicle and applies machine learning models that are trained based on actual fuel consumption measurements. Two machine learning models, including random forest and artificial neural network (ANN), are developed for fuel consumption estimation based on OBD and fuel consumption data. The data are collected during real-world urban and highway driving in a 100-km route for a Ford Escape PHEV and a Ford F-350. The OBD data used for machine learning models include engine load, engine speed, intake manifold absolute pressure, air-fuel equivalence ratio, and throttle position. The validation results show that the random forest method is more accurate than the ANN method, with a estimation accuracy of 99% for the two tested vehicles.**

***Index Terms*—Artificial Neural Networks, Random Forest, Instantaneous Fuel Consumption, Machine Learning, Fleet Management**

I. INTRODUCTION

Reducing fuel consumption of fleet vehicles is one of the primary goals for intelligent fleet management since fuel is one of the major operational costs and directly affects the emission of greenhouse gases and air pollutants from fleet vehicles.

Methods for estimating "average" fuel consumption based on the cumulative fuel consumed and the distance traveled are currently available [1]. However, cumulative or average fuel consumption is typically insufficient for fleet management systems to reduce fuel consumption. The information for identifying driving conditions with high fuel consumption, finding powertrain efficiency in different operating conditions, selecting optimum routes, or assessing the effect of fleet driver behavior on fuel consumption is lacking in the average data. Direct and accurate measurement of vehicle fuel consumption requires the installation of a flow sensor in the fuel flow line of vehicles with a spark ignition (gasoline) engine or installing two fuel flow meters in supply and return fuel lines of vehicles of a compression ignition (diesel) engine. This requires a separate data acquisition system and regular maintenance of the system which is costly. An alternative method is to develop data-driven mathematical models using measured data to estimate fuel consumption. In [2], regression models were developed for estimating fuel consumption and emission where the data was collected from dynamometer testing of a vehicle and the model inputs were vehicle speed and acceleration.

Fuel consumption is correlated to engine operating parameters. Through on-board diagnostics (OBD), many performance parameters of the vehicle engine in the electronic control unit (ECU) are available. These parameters, which are reported to control vehicle emissions, show real-time engine operation [3]. OBD-II is mandatory for all light-duty vahicles and trucks in North America since 1996 [4]. Some parameters provided by OBD-II affect fuel consumption, so these parameters can be used to estimate a vehicle fuel consumption. In [5], vehicle speed and intake mass air flow (MAF) were collected from OBD, and the fuel consumed was estimated from the MAF, assuming the air-to-fuel ratio remained constant. Estimated fuel consumption from OBD data using MAF is compared with fuel consumption determined through carbon balance measurements from dynamometer chassis testing in [6]. They obtained sufficient accuracy for estimating fuel consumption during stoichiometric operation. This method seems acceptable for applications where the combustion is at stoichiometric conditions, but is inaccurate for non-stoichiometric operations.

Machine learning estimation and modeling methods can be useful to improve the accuracy of estimation models. In a study on an articulated truck, three models of support vector machine (SVM), random forest (RF), and artificial neural network (ANN) were used to model fuel consumption in terms of parameters such as road gradient, torque, acceleration, and weight. In general, the random forest model showed the best performance to estimate fuel consumption [7]. A support vector machine (SVM) model to predict fuel consumption by vehicle throttle position and engine speed obtained from OBD is described in [8]. MAF is used to calculate fuel consumption and compare it with the estimated fuel flow. A universal OBD module is presented in [9] to predict fuel consumption using OBD parameters including air flow rate, coolant temperature, engine load, ignition timing, engine speed, vehicle speed, throttle position, and control module voltage. The deep learning models are trained for three different vehicles driving in different routes. The test results from the trained vehicles with Elman NN method show an accuracy of 96%. A Recurrent Neural Network (RNN) model was developed in [10] to monitor fuel consumption using global positioning system (GPS) data, speed, altitude, and acceleration. The results confirmed the superior performance of the proposed fuel consumption estimation model.

The University of Alberta's fleet consists of 173 vehicles that are monitored daily using cellular OBD readers and also equipped with GPS for intelligent fleet management system of the university fleet vehicles. This study aims to develop a reliable and accurate method for monitoring instantaneous fuel consumption of the university fleet vehicles using OBD data. The analysis of instantaneous fuel consumption along with GPS and other collected vehicle operational data will enable intelligent fleet management by i) identifying driving behaviors with high fuel consumption, ii) selecting the optimum routes based on fuel consumption, iii) finding powertrain efficiency in different operating conditions, iv) identifying optimum vehicle types for different vehicle applications and driving cycles at the university. Random Forest and artificial neural network methods are used in this study to model instantaneous fuel consumption using real-time OBD data.

The new contributions of this study include:

- Findings are based on direct measurement of instantaneous fuel consumption to provide accurate and sufficient data for training of the machine learning models.
- Measuring accurate fuel consumption along with real-time OBD data over a wide range of operating conditions including cold start and non-stoichiometric conditions allows for more accurate machine learning models.
- Applying an ultrasonic fuel flow meter allows to measure ultra low-volume fuel flow operation with a high sampling frequency of 5 Hz.
- Fuel consumption data from two MY2021 modern vehicles are presented.

This paper is organized as follows. The next section describes the experimental setup, data collection tools, tested vehicles, driving route, and machine learning methods. In section III, the modeling results for both vehicles and both models are presented, compared and discussed. Finally, summary and conclusions are given in section IV.

II. Methodology

A. Tested Vehicles

Two MY2021 vehicles were selected for this study; a Ford compact SUV and a Ford full-size pickup truck. The two vehicles represent two major categories of fleet vehicles owned by the University of Alberta. Among the 173 vehicles in the university fleet, there are 74 trucks and 19 sedans or SUVs. The specifications of the two selected vehicles are shown in Table I. The powertrain systems of these two vehicles are different as one is a plug-in hybrid electric vehicle (PHEV) with a 2.5-litre gasoline engine and the other is a vehicle with a conventional 6.2-litre gasoline engine.

TABLE I: Tested Vehicles

Vehicle Make/Model	**Ford Escape PHEV**	**Ford F-350**
Model Year	2021	2021
Vehicle Body Style	Compact SUV	Pickup Truck
Fuel Type	Gasoline / Battery	Gasoline
Engine Size	2.5 L	6.2 L
Engine Power	221 hp (@ 6,250 rpm)	385 hp (@ 5,750 rpm)
Battery Capacity	14.4 kWh	-

B. Test Route

A driving route, shown in Fig. 1 was selected to cover a broad range of driving conditions. The route includes residential areas (maximum speed of 40 km/h), urban areas (60 - 100 km/h), and highway areas (110 km/h). Road grades and traffic lights are parts of the route. The test starts at the University of Alberta, South Campus to Leduc, Alberta and returns to the South Campus, as shown in Fig. 1. The total distance covered is 100 km with an approximate duration of 1 hour and 33 minutes.

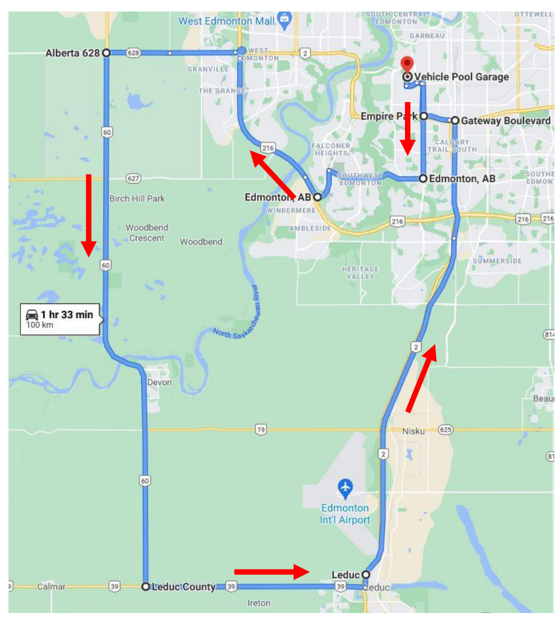
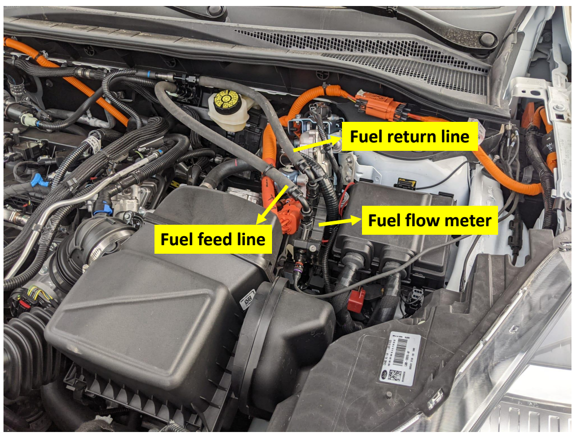


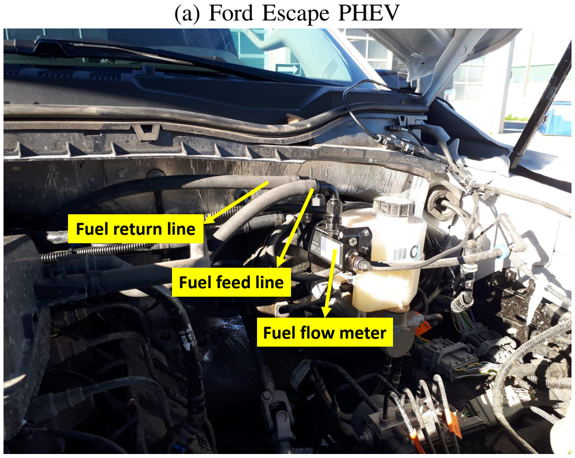
Fig. 1: 100-km driving route, image taken from Google Maps.

C. Fuel Measurement

To measure instantaneous fuel consumption, An ultrasonic fuel flow meter by Sentronics (FlowSonic LF Low-Flow Sensor) was installed on each vehicle as shown in Fig. 2. Specifications of the Sentronics fuel flow meter are listed in Table II. The fuel flow meter was selected due to its capability to measure low-volume fuel flow (e.g., idling condition of a small fleet vehicle), the ability to measure different fuels (e.g., gasoline, diesel), robustness against vibrations and pulsating flows, high measurement accuracy, small and lightweight to install easily on any engine. The flow data were measured at 5 Hz and transferred by via CAN communication to the data acquisition system.



(a) Ford Escape PHEV



(b) Ford F-350

Fig. 2: Fuel flow meter installation in the fuel path of the two vehicles.

D. CAN Data Collection

Fig. 3 shows the schematic of the data collection process, using CANedge2 data logger. Through the CAN bus, OBD data was collected and synchronized with the Sentronics sensor fuel measurements data. In this study, the CSS Electronics CANedge2 CAN bus data logger was used to collect CAN data, including OBD-II and fuel flow measurement data. The OBD data was collected by the CAN bus data logger at sampling frequency of 2 Hz and synchronized with fuel flow CAN data. Table III shows the OBD parameters that were used to calculate fuel consumption.

TABLE II: Specifications of the fuel flow meter in this study

Parameter	**Value**
Repeatability	+/- 0.15% of reading
Uncertainty	+/- 0.5% of reading
Operating flow range	8 - 4000 ml/min
Max. measurement rate	2.2 kHz
Pressure drop at maximum flow	<20 kPa (4000 ml/min gasoline @ 20°C)
Fluid temperature range	-20°C to +120°C
Ambient temperature range	-40°C to +120°C
Fluid compatibility	Gasoline, Diesel, Bio-diesel Ethanol, Methanol

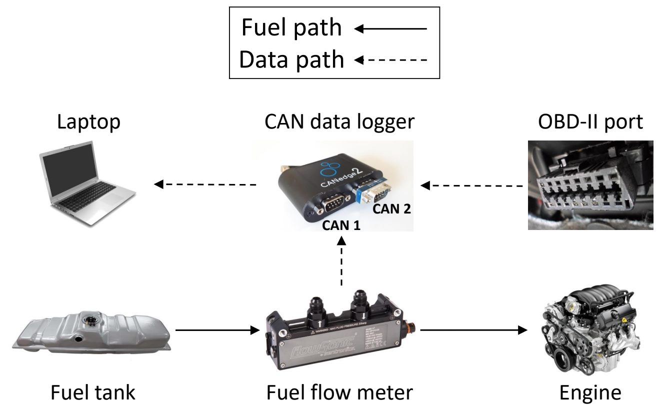


Fig. 3: Schematics of the data collection process.

TABLE III: Parameters collected by CANedge2 to estimate fuel consumption

Parameter Name	**Unit**
Engine Speed	RPM
Engine Load	Percent
Intake Manifold Absolute Pressure	kPa
Throttle Position	Percent
Air-Fuel Equivalence Ratio (λ)	-

Air-fuel equivalence ratio (λ) is one of the parameters that affects calculating the fuel consumption. Instead of using commanded air-fuel equivalence ratio from OBD data, the short and long-term fuel trims from OBD data were used to calculate air-fuel equivalence ratio accurately. The λ calculation equation is:

$$\lambda = \frac{1}{(1 + \frac{Short\text{−}term\ Fuel\ Trim}{100}) \times (1 + \frac{Long\text{−}term\ Fuel\ Trim}{100})} \tag{1}$$

E. Machine Learning Models

Initially, four machine learning models including random forest, ANN, SVM, and k-nearest neighbors (KNN) were tested for estimating fuel consumption. The initial analysis showed that random forest and ANN offer the highest accuracy. This paper focuses on presenting the results for random forest and artificial neural network (ANN) to estimate fuel consumption. Random forest training is fast and resistant to over-fitting when there are many features. It also does not require normalized data and works well with features with different ranges of values [11]. ANN is a robust method to measurement noises and works well for non-linear data sets [12]. Here, an ANN method with one hidden layer is used with the design parameters listed in Table IV. The optimal number of hidden layer neurons for each model is determined during training and validation. For the random Forest method, the design parameters are listed in Table V and the number of decision trees for each model is optimized during training and validation.

TABLE IV: Design parameters of ANN model

Parameter Name	Value
Nunmber of hidden layers	1
Activation function for the hidden layer	Relu
Solver for weight optimization	Adam
Learning rate	0.001
Numerical stability criteria	10^{-8}
Maximum number of iterations	200

TABLE V: Design parameters of random forest model

Parameter Name	Value
Split criterion function	Squared error
Min. number of samples to split	2
Min. number of samples for a leaf node	1

Unlike the random forest model, the ANN model is sensitive to differences in the range of feature values, so it is necessary to normalize the data as part of the data preparation process. The normalized data of each feature is formed as:

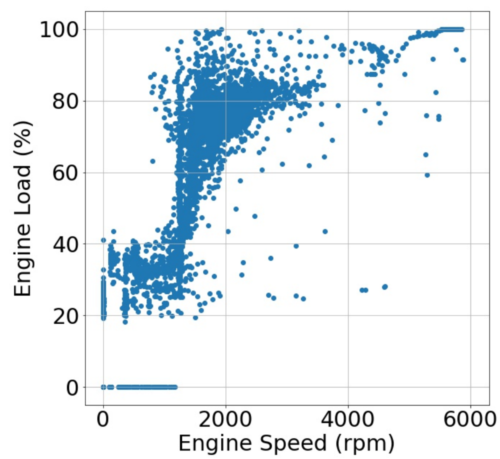
$$x_s = \frac{x - min(x)}{max(x) - min(x)} \tag{2}$$

Where x is the actual and x_s is the normalized data, between 0 and 1.

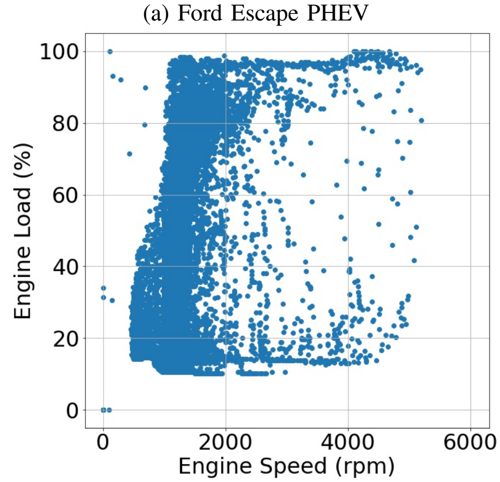
To develop the models, the data collected with CANedge2 was divided into two parts. The first part, consisting of 70% of the total data, was used in a 5-fold cross-validation method. To do this, 70% data was divided into five parts and four parts were used for the first repetition and the fifth part was used for validation of the model. The remaining 30% of the data was used to test the two trained machine learning models.

III. Results and Discussion

To get an accurate fuel consumption model, it is critical that the training data set covers all the possible operating points. Engine load and engine speed are the most important factors describing the engine operating points. Fig. 4 shows the collected data over the drive cycle in Fig. 1 for both vehicles cover a broad engine load (0 - 100%) and engine speed (idle - 5000 rpm) operating conditions. The number of points and the engine operating range in Fig. 4a is less than 4b, because Ford Escape is a plug-in hybrid electric vehicle, and in some operating conditions, instead of the gasoline engine, the electric battery powers the vehicle.



(a) Ford Escape PHEV

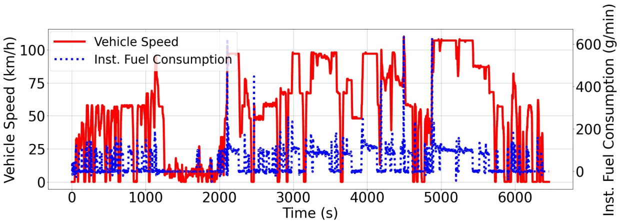


(b) Ford F-350

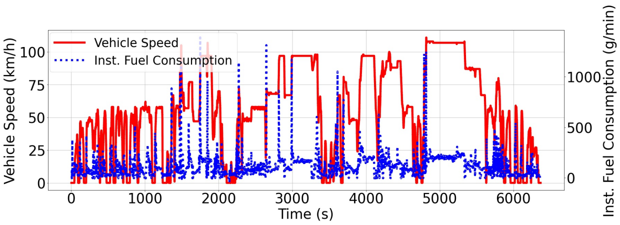
Fig. 4: Engine load and speed for the collected vehicle data.

In Fig. 5, time series of vehicle speed and instantaneous fuel consumption in the trip are shown for both vehicles. Speed profiles show that driving cycles of both vehicles were close to each other with differences due to the road traffic at the time of vehicle testings. When vehicles accelerate, instantaneous fuel consumption increases significantly and peaks, indicating that driving behavior and operating conditions are very effective in fuel consumption.

According to the validation results of the ANN model (Table VI), the validation results are acceptable for both vehicles (slightly better for Ford Escape). The hidden layer size is optimized for each model by minimization of the root mean square error (RMSE) of the estimated fuel consumption.



(a) Ford Escape PHEV



(b) Ford F-350

Fig. 5: Time series of vehicle speed and recorded fuel consumption.

TABLE VI: Specifications of ANN fuel consumption model cross validation

Parameter	Ford F-350	Ford Escape PHEV
Hidden Layer Size	160	200
RMSE (g/min)	18.88	8.65
R^2	0.96	0.98

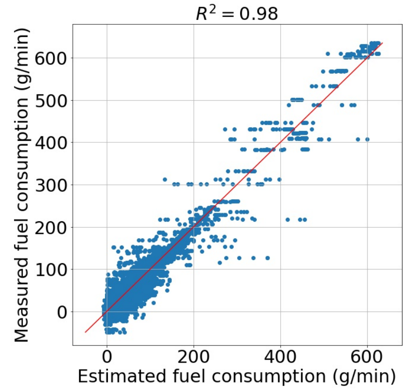
The random forest validation results (Table VII) show this model has better validation results on the fuel consumption of both vehicles compared to ANN. The higher RMSE of Ford F-350 models does not necessarily mean that those models are less accurate, but because of the larger fuel consumption in Ford F-350, which has a larger gasoline engine.

TABLE VII: Specifications of random forest fuel consumption model cross validation

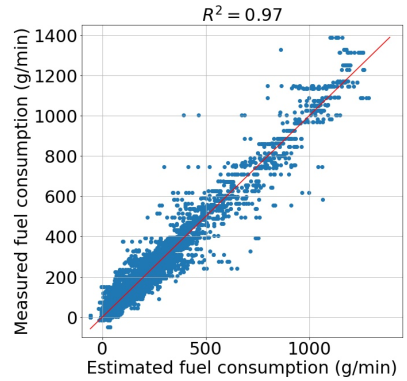
Parameter	Ford F-350	Ford Escape PHEV
No. of Decision Trees	100	90
RMSE (g/min)	4.62	2.94
R^2	1.00	1.00

The results of the performance of trained models on the test data listed in Table VIII show the random forest model provides better estimation of fuel consumption on both vehicles. Both RMSE and the coefficient of determination (R^2) show that using the random forest model leads to a more accurate estimate of fuel consumption. Fig. 6 shows that the points of the random forest model plots (Fig. 6c and Fig. 6d) are closer to the diagonal line, indicating that the estimated results are more consistent with the actual values.

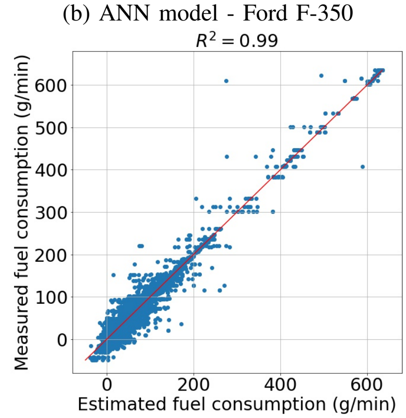
Looking at part of the drive cycle in Fig. 7 shows that the random forest model deviates less from the actual values. Further, Fig. 7 shows that machine learning estimation models are helpful tools to estimate instantaneous fuel consumption based on OBD parameters. Overall, the performance of the



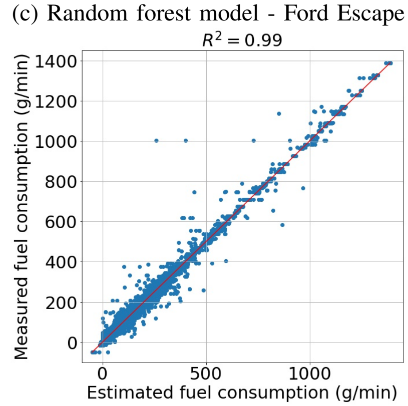
(a) ANN model - Ford Escape



(b) ANN model - Ford F-350



(c) Random forest model - Ford Escape

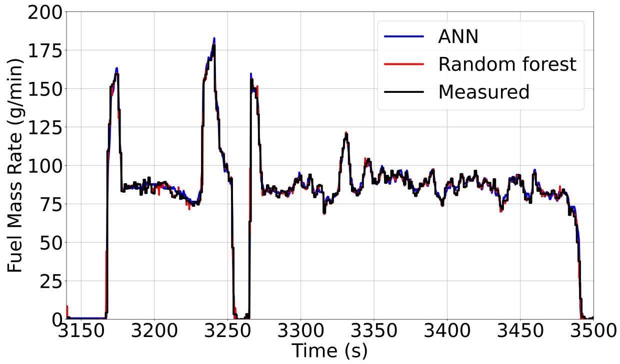


(d) Random forest model - Ford F-350

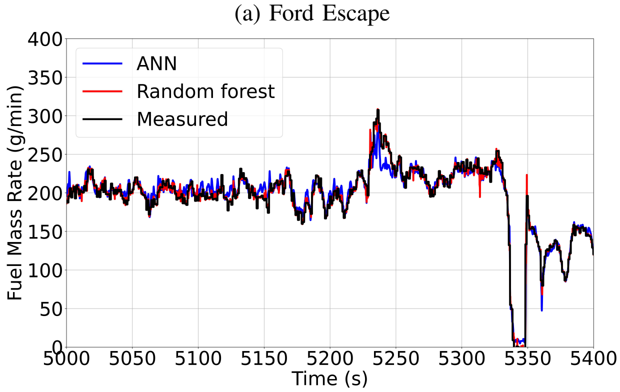
Fig. 6: Performance of the ANN and random forest models for estimating instantaneous fuel consumption for the test data.

TABLE VIII: Results of the models on the test set

Model	Vehicle	RMSE (g/min)	R^2
ANN	Ford F-350	20.40	0.97
ANN	Ford Escape Hybrid	9.07	0.98
RF	Ford F-350	11.72	0.99
RF	Ford Escape Hybrid	6.096	0.99



(a) Ford Escape



(b) Ford F-350

Fig. 7: Comparing the estimation models.

machine learning models on Ford Escape PHEV is slightly better than Ford F-350. This could be due to the higher fuel consumption of Ford F-350, which causes bigger noise in the fuel flow data.

IV. Summary and Conclusions

A method to estimate and monitor real-time instantaneous fuel consumption of fleet vehicles using easily available OBD data is described. The models were trained with OBD and measured fuel consumption data. Once the models are trained, OBD data is used and a real-time instantaneous fuel consumption is available. Two different vehicles including a Ford Escape PHEV and a Ford F-350 were equipped with a fuel flow measurement sensor. A CAN data logger was used to collect fuel and OBD data and the vehicles were driven through a selected route including a wide range of driving conditions. To develop fuel consumption estimation models, 5 features of engine load, engine speed, air-fuel equivalence ratio, intake manifold absolute pressure, and throttle position are selected as inputs for machine learning algorithms. Random forest and artificial neural network (ANN) are used as the machine learning methods and 5-fold cross validation approach is implemented on the training data. These models can be used to estimate instantaneous fuel consumption using only OBD data.

ANN model can estimate Ford F-350 fuel consumption with an accuracy of 97%. Its estimation accuracy for Ford Escape is 98%. Random forest model has a higher estimation accuracy of 99% for both vehicles. Both methods have the acceptable accuracy to be used for instantaneous fuel consumption estimation of the fleet vehicles.

Acknowledgment

The authors would like to thank Jim Laverty, the fleet manager, and the staff of Transportation Services at the University of Alberta for their support for vehicle instrumentation and testing. The authors also thank Energy Management and Sustainable Operations (EMSO) of the University of Alberta, especially Michael Versteege and Shannon Leblanc for supporting this project.

References

[1] H. Tong, W. Hung, and C. S. Cheung, "On-road motor vehicle emissions and fuel consumption in urban driving conditions," *Journal of the Air & Waste Management Association*, vol. 50, no. 4, pp. 543–554, 2000.

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

[3] R. Nandal and N. Awasthi, "OBD-II and Big Data: A Powerful Combination to Solve the Issues of Automobile Care," in *Computational Methods and Data Engineering*, pp. 177–189, Springer, 2021.

[4] I. Hascic, F. P. de Vries, N. Johnstone, and N. Medhi, "Effects of environmental policy on the type of innovation: the case of automotive emissions control technologies," *OECD Journal: Economic Studies*, vol. 2009, no. 1, pp. 49–66, 2008.

[5] R. Malekian, N. R. Moloisane, L. Nair, B. T. Maharaj, and U. A. Chude-Okonkwo, "Design and implementation of a wireless OBD II fleet management system," *IEEE Sensors Journal*, vol. 17, no. 4, pp. 1154–1164, 2016.

[6] T. H. DeFries, M. Sabisch, S. Kishan, F. Posada, J. German, and A. Bandivadekar, "In-Use Fuel Economy and CO Emissions Measurement using OBD Data on US Light-Duty Vehicles," *SAE International Journal of Engines*, vol. 7, no. 3, pp. 1382–1396, 2014.

[7] F. Perrotta, T. Parry, and L. C. Neves, "Application of machine learning for fuel consumption modelling of trucks," in *2017 IEEE International Conference on Big Data*, pp. 3810–3815, IEEE, 2017.

[8] T. Abukhalil, H. AlMahafzah, M. Alksasbeh, and B. A. Alqaralleh, "Fuel consumption using OBD-II and support vector machine model," *Journal of Robotics*, vol. 2020, 2020.

[9] M. Yen, S. Tian, Y. Lin, C. Yang, and C. Chen, "Combining a Universal OBD-II Module with Deep Learning to Develop an Eco-Driving Analysis System," *Applied Sciences*, vol. 11, no. 10, p. 4481, 2021.

[10] S. Kanarachos, J. Mathew, and M. E. Fitzpatrick, "Instantaneous vehicle fuel consumption estimation using smartphones and recurrent neural networks," *Expert Systems with Applications*, vol. 120, pp. 436–447, 2019.

[11] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, p. 272, 2012.

[12] P. Daponte and D. Grimaldi, "Artificial neural networks in measurements," *Measurement*, vol. 23, no. 2, pp. 93–115, 1998.