Comparative Assessment of Machine Learning Techniques for Modeling Energy Consumption of Heavy-Duty Battery Electric Trucks

Emmanuel Hidalgo Gonzalez, Jacqueline Garrido, Matthew Barth, and Kanok Boriboonsomsin,
University of California, Riverside

Abstract—**Efforts to decarbonize the heavy-duty vehicle sector have generated vast interest in transitioning from conventional diesel trucks to battery electric trucks (BETs). As a result, understanding energy consumption characteristics of BETs has become important for a variety of applications, for instance, assessing the feasibility of deploying BETs in place of conventional diesel trucks, predicting the state-of-charge (SOC) of BETs after specific duty cycles, and managing BET charging needs at the home base or en-route. For these applications, mesoscopic energy consumption models offer a good balance between the amount and fidelity of the input data needed, such as average traffic speed and road grade on a link-by-link basis, and the model performance. As a common intelligent transportation system (ITS) application, this paper presents a comparative assessment of mesoscopic energy consumption models for BETs developed using three different machine learning techniques. The results show that the random forest (RF) regression outperforms the extreme gradient boosting (XGBoost), the light gradient-boosting machine (LightGBM), as well as the conventional linear regression as evidenced by the resulting model having a higher coefficient of determination (R2) value than that of its counterparts. When applied to the simulated dataset, the RF regression can capture the behaviors of BET energy consumption well where the R2 value of the resulting model is 0.94.**

Keywords—Battery Electric Truck , random forest regressor, mesoscopic model, XGBoost, LightGBM, intelligent transportation system.

I. Introduction

Transportation is essential to the everyday life of people. At the same time, vehicles are one of the major sources of air pollution [1, 2]. Reducing the pollution footprint from vehicles has been a topic of research in academia for many years. One possible solution is to increase the use of electric vehicles in place of the ones that consume fossil fuels. The reason is because not only do electric vehicles reduce pollution, but they also save money [3]. The advancement in battery technology has made it more common to see electric cars on the roads today. However, it is more difficult to electrify the heavy-duty truck sector. Heavy-duty battery electric trucks (BETs) are still few and far between, and are being researched to make them more available. Some reasons as to why BETs are harder to adopt are because they have a shorter driving range than their diesel counterparts, and at the same time it takes longer time to charge them. Given the limited driving range, it is important to understand energy consumption characteristics of BETs so that routes can be planned and make the most out of the trucks.

In research, there are a variety of electric vehicle energy consumption models [4-10]. Many of these models follow the white box approach, meaning that they are developed with known physical parameters and behaviors. A drawback of these models is their complexity, which makes them difficult to develop and apply [11]. In contrast, black box models (e.g., those based on machine learning, a common application in Intelligent Transportation Systems (ITS)) are faster to create because they are based on experimental data and some data processing [12,13]. While these models are sometimes criticized for not being able to be interpreted completely [14], the fact that they can be trained and developed faster, making them more appealing in their use in a variety of applications including vehicle routing, range estimation, etc., which are important for many ITS research topics such as eco-routing and charging planning [15-20]. To date, however, there has been very few studies that apply machine learning techniques to develop mesoscopic energy consumption models specifically for heavy-duty BETs.

Previously, we developed mesoscopic energy consumption models for BETs using a machine learning technique called random forest (RF) regression [21]. The models performed better than the ones developed with the traditional linear regression and showed that it is possible to capture nonlinear behavior. The coefficient of determination (R2) values for the RF regression models were 0.86-0.89 as compared to 0.50-0.52 for the linear regression models. However, there is room for improvements on the previous models. The objective of the research presented in this paper was, therefore, to improve the performance of mesoscopic energy consumption models for BETs in several ways. First, we made some changes to the data processing procedures to generate better input data for the models. Second, we applied other machine learning techniques to develop the models and compared the results with those from RF and linear regressors. This helped us assess how different black box models perform on the same dataset and identify the machine learning technique most suitable for modeling BET energy consumption. Lastly, since black box models are entirely data driven, we examined the interpretability of the models under various conditions.

II. Methodology

This section introduces the different datasets utilized for the different models within the paper. There are two main datasets utilized. The first dataset comes from logging vehicle activity and energy consumption data directly from real-world heavy-duty trucks and the second dataset comes from a simulation of BET energy consumption based on the real-world heavy-duty truck activity using a microscopic BET powertrain model. The main reason for creating the simulated dataset is to be able to incorporate cargo weight as a variable in the mesoscopic models as it is not available in the real-world dataset. Mesoscopic models developed in this research relate the BET energy consumption required to traverse a roadway link with variables that are generally available on a link-by-link basis such as average traffic speed and road grade. These variables have a high impact in the energy consumption of BETs [22-24]. The main difference between the different models developed is the machine learning technique used to predict BET energy consumption on any roadway links.

A. Data

The first dataset comprises real-world vehicle activity and energy consumption data obtained from heavy-duty trucks in California. Specifically, these data were collected from a group of 2020 model year class 8 trucks doing drayage operations at the port of Los Angeles. Each truck's data were collected over a period of 1-2 months using a combination of GPS and engine control unit (ECU) data loggers. The setup was designed to record GPS data—including timestamp, speed, latitude, longitude, and altitude—as well as selected ECU parameters at a rate of 1 Hz. The GPS data were used to perform map matching with a digital map to identify the roadway link and road grade associated with each data point.

The second dataset was generated by simulating the energy consumption of a BET following the second-by-second vehicle activity in the first dataset at different levels of cargo weight. The simulation utilized a microscopic BET powertrain model that needs certain information about the BET and its activity. Some of the data needed were second-by-second vehicle speed, acceleration, and road-grade. The data were fed into Equation (1) to obtain the tractive power of the BET and, subsequently, BET energy consumption was calculated. Table 1 provides the values of the different model parameters. More details about the microscopic BET powertrain model are available in [30].

$$W_{tract} = m \cdot v \cdot a + \tfrac{1}{2}\rho \cdot C_d \cdot A_f v^3 + m \cdot g \cdot C_{rr} \cdot v \cdot cos(\Theta) + m \cdot g \cdot v \cdot sin(\Theta) \quad (1)$$

where v is instantaneous speed; a is instantaneous acceleration; Θ is angle of the inclination of the road (road grade); ρ is air density; C_d is drag coefficient; A_f is BET frontal area, C_{rr} is coefficient of rolling resistance of the BET tires; g is gravity; and m is mass of the BET and cargo combined.

Unlike conventional diesel trucks, BETs can recover some energy through regenerative braking when Equation (1) outputs a negative value. The consideration of regenerative braking energy in mesoscopic BET energy consumption models makes the models more complex and highly nonlinear, which are suitable for data-driven, machine learning techniques.

Table 1. Parameter list for tractive power equation

Parameter	Symbol	Value
Front area (m^2)	A_f	10
Rolling resistance coefficient	C_{rr}	0.008
Aerodynamic drag coefficient	C_d	0.56
Air density	ρ	1.161
gravity	g	9.8

B. Data Processing

Initial data filtering was performed on the second-by-second data points. A close inspection of the dataset indicated that there were data points with power of over 350 Kilowatts, which is the rated capacity of the electric motor of the BET. Thus, these data points were removed. Other data points that did not make sense in the physical world were also removed. An example of such points is regeneration when the vehicle is accelerating while going uphill.

Given the nature of roads, a roadway link may be long enough to have significant changes in road grade. Different road grade values could potentially affect BET energy consumption. To address this issue, links were split based on different criteria. For instance, it was assumed that a BET would have similar energy consumption behavior if the road grades are within certain bins. An example would be a flat road. If the road grade values of a link were within -0.5 to 0.5 percent, then we assumed that the BET energy consumption would be similar within that part of the link. If the same link contained road grade values within other bins (e.g., 0.5 to 1.5 percent), then the map matching procedure would split that link into sub-links according to the different road grade bins. Additionally, within a sub-link it is necessary to determine the average road grade as road grade values on the sub-link may not be uniform. The average road grade, $\bar{g}$, was obtained by calculating the length of each second-by-second road grade value within the sub-link and performing a weighted average based on the distance. A similar approach was taken to obtain the average velocity, $\bar{v}$, within each sub-link. The weighted average helped stabilize the speed and road-grade profiles.

Sub-links required another layer of filtering. The length of second-by-second data indicated how long the truck spent on a sub-link. Some of the sub-links were very short, lasting just one or a few seconds. While it is possible to encounter short links, keeping many one-second links would make our models be more biased towards low energy consumption as the aggregated energy would only come from one point of data. Empirically, it was determined that sub-links below ten seconds should be removed. Additionally, as mentioned before, each sub-link contains its own average speed, road grade, weight, and energy consumption rate (in kwh/mile). This allowed us to plot BET energy consumption rate as a function of these variables. While most of the data could be grouped together, there were instances where outliers may be present, which in turn would cause the models to have lower performance. The solution to the outliers of data was to use a 3D interquartile range (IQR) on different subsets or bins of the data. An example of a 3D bin is the energy consumption for the speed between 5 and 10 mph, with the road grade between

-0.5 and 0.5 percent, and the weight of 30,000 lbs. IQR statistics for each 3D bin were calculated, which allowed us to identify and remove the outliers, resulting in less noisy data.

C. Model Development

The three machine learning algorithms used were RF regression, XGBoost, and lightGBM. These algorithms were chosen because, as ensemble algorithms, they can combine different techniques, which allows them to capture non-linear relationships [25]. XGBoost is a gradient-boosted decision tree, which is very efficient for regression problems [26]. As an ensemble algorithm, it uses different machine learning techniques to obtain a better model. Its parallel nature makes it faster than other algorithms. LightGBM is similar to XGBoost as they both use the same strategy for their predictions [27]. The main difference lies in the implementation of the construction of trees. This difference can lead to very different results in the prediction performance.

The mesoscopic models use link-level input data such as average speed and road grade to predict the corresponding BET energy consumption rate (in kwh/mile) on a particular roadway link. Then, the energy consumption rate can be multiplied by the link distance to obtain the energy consumption required for the BET to traverse the link. In addition to the three machine learning-based models, a model based on linear regression was also developed for comparison.

The dataset includes a limited number of features or predictors (average speed, road grade, and total weight). However, it is possible to create additional features, for instance, by multiplying average speed and average road grade. In Equation (1), some of the variables are higher order terms of speed or interaction terms between speed and road grade. For this reason, we created higher order features of these variables. However, it would be undesirable to train a model that has a near infinite number of features. In order to avoid this issue, it was necessary to identify key features that would provide the most important information for the model.

There are multiple feature selection algorithms. For instance, for many machine learning techniques, permutation importance is used to select prospects that contribute the most information to the model [28]. In the case of RF regression, the algorithm takes all the features at once and creates a baseline R2 value. It then proceeds to permute the column values of a single feature and test how that changes the baseline value. Finally, only the features that improve or affect the baseline value are kept and used for the final training. The process identified the 13 features in Table 2 as the most important in our prediction problem. Figure 1 shows the importance of the different features. It can be seen that feature 1 (average road grade) has the highest importance while feature 6 (weight), despite having lower importance than the other features, is still relevant in the model. A similar approach was taken for the rest of the machine learning-based models.

Since we also wanted to compare the machine learning-based models with the model developed using traditional linear regression, we performed stepwise linear regression, and the results r are given in Table 3. The model has an R2 value of 0.55. In this model, both average road grade and weight are an important feature and have a positive coefficient. This is expected because as the weight carried by the BET increases, its energy consumption per unit distance should also increase at any given average speed and average road grade.

Table 2. FEATURE NAME AND SYMBOL FOR RANDOM FOREST REGRESSOR

FEATURE NAME	SYMBOL
FEATURE 0	$\bar{\upsilon}$
FEATURE 1	$\bar{g}$
FEATURE 2	$\bar{\upsilon}^2$
FEATURE 3	$\bar{\upsilon} \cdot \bar{g}$
FEATURE 4	$\bar{\upsilon}^3$
FEATURE 5	$\bar{\upsilon}^4$
FEATURE 6	w
FEATURE 7	$\bar{g}^2$
FEATURE 8	$\bar{\upsilon}^2 \cdot \bar{g}$
FEATURE 9	$\bar{\upsilon} \cdot \bar{g}^2$
FEATURE 10	$\bar{\upsilon}^2 \cdot \bar{g}^2$
FEATURE 11	$\bar{\upsilon}^3 \cdot \bar{g}$
FEATURE 12	$\bar{\upsilon} \cdot \bar{g}^3$

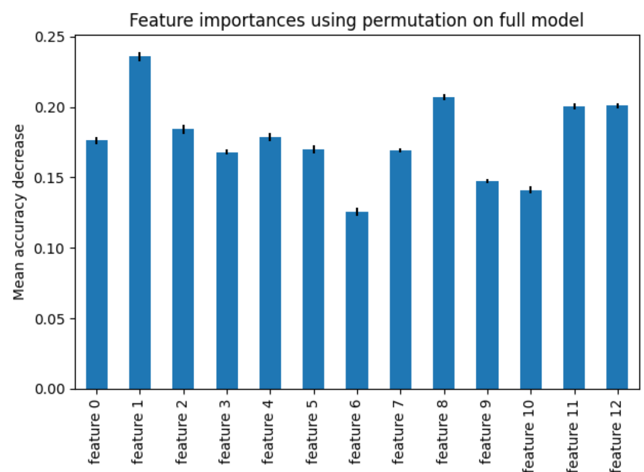

Figure 1. Importance of the different features in the simulated dataset

TABLE 3. COEFFICIENTS OF THE STEPWISE REGRESSION MODEL FOR SIMULATED DATA

Feature	b	p-val	Standard error
$\bar{g}$	0.7753	0	0.064
$\bar{\upsilon}$	-0.2355	0	0.010
$\bar{\upsilon}^2$	0.0035	0	0.000
w	4.196e-05	0	4.43e-06
$\bar{\upsilon} \cdot \bar{g}^3$	0.0028	0	0.001

III. RESULTS AND DISCUSSION

A. Model Performance

The first trained model used the XGBoost algorithm. We used an R2 value to check how well the models had been trained. The model was tested using 4-fold cross validation. The average R2 value of this model was 0.84. It is known that driving behavior can significantly affect vehicle energy consumption [29, 30]. Some possible sources of error can be attributed to XGBoost not being able to capture some characteristics of BET energy consumption. This can be visualized in Figure 2. Most of the predicted values overlap with the observed data but there are still regions not well captured, especially on the extreme ends of energy consumption rate.

The second trained model used the lightGBM algorithm. Again, 4-fold cross validation was used to assess model performance. The average R2 value of this model was 0.72. Even though lightGBM is also a gradient boosting algorithm, the performance was lower than XGBoost, as shown in Figure 3. This is due to the implementation of both algorithms. While both use the same technique to solve a problem, the implementation of that technique in the two algorithms is different in nature [27].

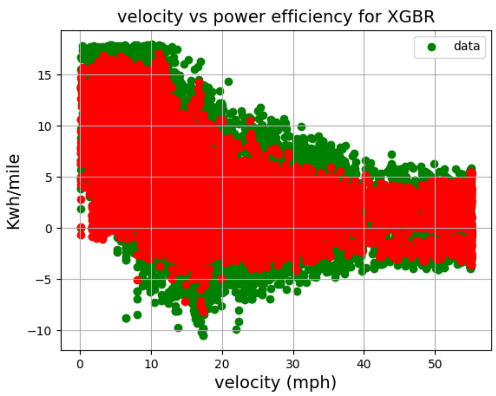

Figure 2. XGBoost model result. Green dots are the observed (simulated) BET energy consumption rates and red dots are the predicted data.

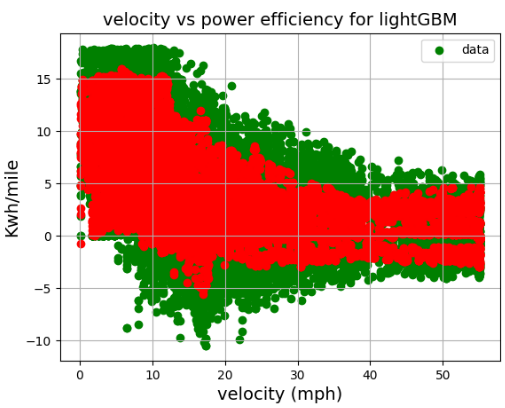

Figure 3. LightGBM model results. Green dots are the observed (simulated) BET energy consumption rates and red dots are the predicted data.

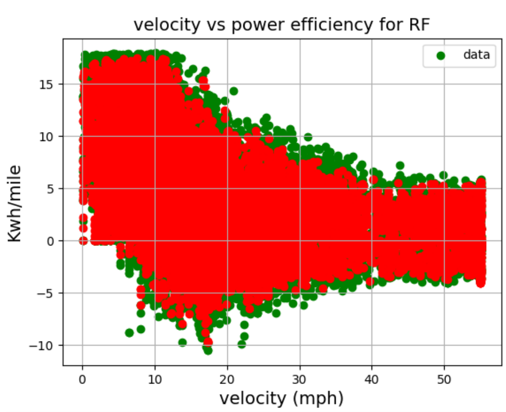

Figure 4. RF model result. Green dots are the observed (simulated) BET energy consumption rates and red dots are the predicted data

The third model was trained using the RF regression algorithm described in [32]. Similar to the previous models, a 4-fold cross validation was used to evaluate model performance. The average R2 value of this model was 0.94, which is the highest among all three machine learning-based models. This was expected as random forest is known to perform well with non-linear data. Still, it is not able to capture the behavior of the data entirely, as shown in Figure 4.

The ability of BET to recover some energy through regenerative braking makes it harder for the developed models to predict link-level BET energy consumption rates accurately. The BET energy consumption rate on a roadway link is highly dependent on the characteristics of its second-by-second speed profile on that link. Specifically, if the second-by-second speed profile includes frequent braking events, then the corresponding link-level energy consumption rate would be low or even negative. On the other hand, if the second-by-second speed profile includes mostly acceleration events, then the corresponding link-level energy consumption rate would be very high, even though its average speed is the same as for the other speed profile. According to the observed (simulated) data in Figures 2-4, the variation in energy consumption rate is highest for the average speed of 10 to 20 mph. This is because this range of average speed usually involves stop-and-go driving in congested traffic.

A summary of R2 values and mean squared error (MSE) for all the models is given in Table 4. The main difference between the model in [21] and the current model is the filtering and data processing. The extra filtering at the link-level removed noise that would not be identified otherwise. This helped the model to achieve better performance. The table indicates that the use of the machine learning techniques results in a higher model performance over traditional methods such as linear regression. This is expected as linear regression is known to perform poorly on data that is noisy, something that machine learning techniques are more capable of overcoming.

TABLE 4. SUMMARY OF THE RESULTS FOR DIFFERENT MODELS

Model	R2	MSE (kWh/mile)^2
Previous random forest model [21]	0.89	0.43
New random forest model	0.94	0.40
XGBoost model	0.84	0.78
lightGBM model	0.72	1.18
Linear regression model	0.55	1.45

B. Model Interpretability

Even if the models perform well in predicting link-level BET energy consumption rates, it is important to check if the models behave in a way that is explainable in a physical sense. Therefore, we decided to examine the RF model, our highest performing model, under different scenarios. Figure 5 shows predicted BET energy consumption rate as a function of average speed for an average road grade of 0 percent at different weights. Intuitively, the heavier weight carried by a BET, the more energy is needed to operate it. The graphs generally agree with this intuition except around 15 mph where the energy consumption rates seem to be dominated by energy regeneration from braking.

In addition, the graphs show that BETs would have relatively lower energy consumption rates between 30 and 50 mph, which is similar to the trend observed in [31]. On the other hand, unlike in [31], the trend of the graph between 0 and 15 mph is not as smooth. This is because the second-by-second speed profiles in this range of average speed have more fluctuation, as evidenced by the distributions of acceleration in Figure 6. Such noisy speed profiles cause the link-level energy consumption rates to be more varied.

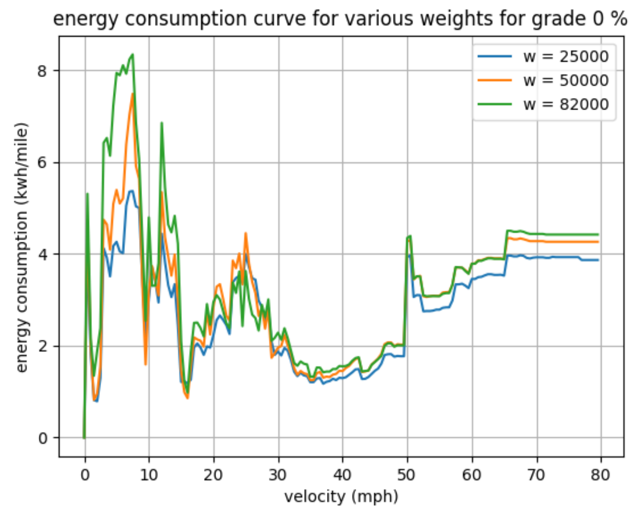

Figure 5. BET energy consumption rate vs. average speed

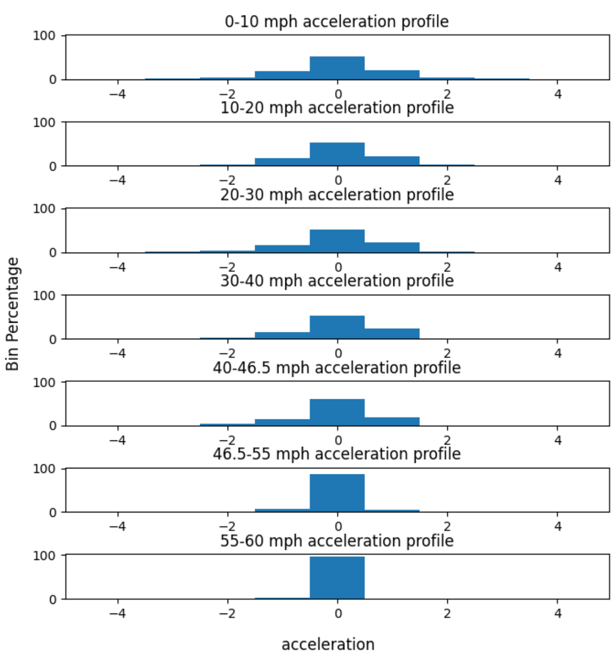

Figure 6. Acceleration distribution of different speed bins

The next variable to be examined is the average road grade. Figure 7 depicts predicted BET energy consumption rate as a function of average road grade at different velocities for a fixed weight of 50,000 lbs. Negative grade is as expected because going faster will regen more when braking. Furthermore, when the truck is climbing grade, there is less regen (slow down mainly to gravity and less braking), and that makes the energy consumption change back to what conventional diesel trucks behave (i.e., around 40 mph is the lowest energy consumption).

The last variable to be examine is weight. Figure 8 shows predicted BET energy consumption rate as a function of the combined weight of BET and cargo for an average speed of 60 mph at different levels of average road grade. As expected, when a BET is on downhill, it would mostly recover energy from regenerative braking, and the heavier weight it carries, the more energy it will generate. The trend is opposite when a BET is going uphill. However, the trend when a BET is on a flat road is counterintuitive where the graph indicates that it would consume less energy if carrying a heavier weight. This unexpected trend warrants further investigation in the future.

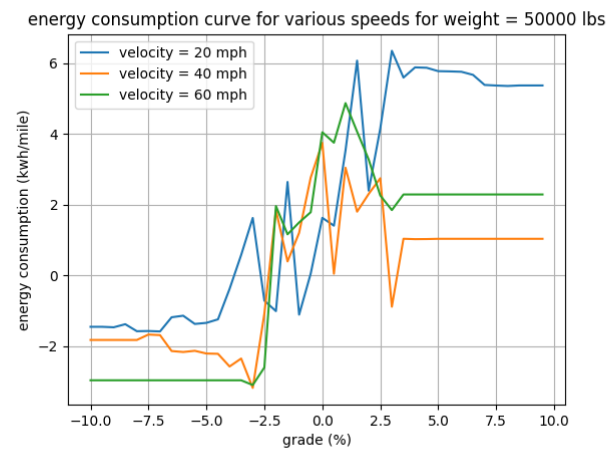

Figure 7. BET energy consumption rate vs. average road grade

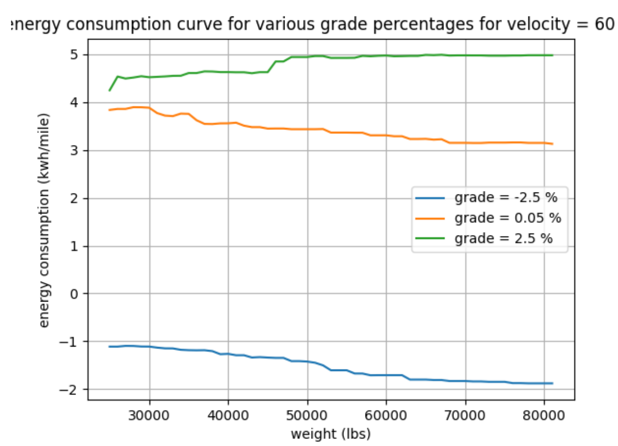

Figure 8. BET energy consumption rate vs. weight

IV. Conclusions and Future Work

This paper presents four different mesoscopic models for predicting the energy consumption of battery electric trucks, an important pre-cursor for ITS strategies. The main contribution is the development of these models using three different black-box machine learning techniques and comparing them against each other as well as against the model developed with the traditional linear regression. Indeed, the machine learning-based models outperform the linear regression model due to their nature of being able to learn from nonlinear data and at the same time be more robust to noise. It was found that random forest regression produced the most accurate results out of all the different models with an average R2 value of 0.94. The model also showed that it largely agreed with the physical phenomena that would be expected of battery electric trucks. This model can be used in a variety of BET applications, for example, predicting the remaining range, finding the most energy-efficient route, and estimating the charging need when arriving back at the depot.

In the future, we will explore incorporating additional variables affecting BET energy consumption that can be easily obtained, such as ambient temperature, into the model. We will also account for powertrain component efficiencies development of the model. Additionally, we will apply the developed model to some of the applications mentioned above to demonstrate the practicality and utility of the model.

REFERENCES

[1] California air resources board (carb). 2022 inventory of greenhouse gas emissions. https://ww2.arb.ca.gov/ghg-inventory-data. Date accessed: December 2022.

[2] Barth M, Boriboonsomsin K. “Real-World Carbon Dioxide Impacts of Traffic Congestion.” Transportation Research Record. 2008;2058(1):163-171

[3] Ardalan Vahidi and Antonio Sciarretta. 2018. Energy saving potentials of connected and automated vehicles. Transportation Research Part C: Emerging Technologies 95 (2018), 822–843

[4] Xuewei Qi, Guoyuan Wu, Kanok Boriboonsomsin, Matthew J. Barth, “Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions,” Transportation Research Part D: Transport and Environment, Volume 64, 2018, Pages 36-52,ISSN 1361-9209

[5] Chen, Y., Wu, G., Sun, R., Dubey, A. et al., "A Review and Outlook on Energy Consumption Estimation Models for Electric Vehicles," SAE J. STEEP 2(1):79-96

[6] Holden, J., Reinicke, N. and Cappellucci, J., “RouteE: A Vehicle Energy Consumption Prediction Engine,”SAE Technical Paper 2020-01-0939, 2020

[7] Xinmei Yuan, Chuanpu Zhang, Guokai Hong, Xueqi Huang, Lili Li, “Method for evaluating the real-world driving energy consumptions of electric vehicles,” Energy, Volume 141, 2017, Pages 1955-1968

[8] Y. Zhang, W. Wang, Y. Kobayashi and K. Shirai, "Remaining driving range estimation of electric vehicle," 2012 IEEE International Electric Vehicle Conference, 2012, pp. 1-7

[9] Chris Bingham, Chris Walsh, and Steve Carroll. 2012. Impact of driving characteristics on electric vehicle energy consumption and range. IET Intelligent Transport Systems 6, 1 (2012), 29–35

[10] Chiara Fiori, Kyoungho Ahn, and Hesham A Rakha. 2016. Power-based electric vehicle energy consumption model: Model development and validation. Applied Energy 168 (2016), 257–268

[11] Lino Guzzella and Antonio Sciarretta. 2007. Vehicle propulsion systems. Vol. 1. Springer

[12] Benedetto Piccoli, Ke Han, Terry L Friesz, and Tao Yao. 2013. Estimating fuel consumption and emissions via traffic data from mobile sensors. In 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 472–477

[13] Weiliang Zeng, Tomio Miwa, and Takayuki Morikawa. 2015. Exploring trip fuel consumption by machine learning from GPS and CAN bus data. Journal of the Eastern Asia Society for Transportation Studies 11 (2015), 906–921

[14] O. Loyola-González, "Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View," in *IEEE Access*, vol. 7, pp. 154096-154113, 2019

[15] K. Boriboonsomsin, M. J. Barth, W. Zhu and A. Vu, "Eco-Routing Navigation System Based on Multisource Historical and Real-Time Traffic Information," in IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 4, pp. 1694-1704, Dec. 2012

[16] J. Garrido, E. Hidalgo, M. Barth and K. Boriboonsomsin, "Home-Base Charging Load Profiles of Battery Electric Trucks Considering Tour Completion and Time-of-Use Rates," *2023 IEEE Transportation Electrification Conference & Expo (ITEC)*, Detroit, MI, USA, 2023, pp. 1-5

[17] Bart Saerens, Hesham A Rakha, Kyoungho Ahn, and Eric Van Den Bulck. 2013. Assessment of alternative polynomial fuel consumption models for use in intelligent transportation systems applications. Journal of Intelligent Transportation Systems 17, 4 (2013), 294–303

[18] Yuche Chen, Guoyuan Wu, Ruixiao Sun, Abhishek Dubey, Aron Laszka, and Philip Pugliese. 2020. A review and outlook of energy consumption estimation models for electric vehicles. arXiv preprint arXiv:2003.12873 (2020)

[19] Barth, M., Younglove, T. and Scora, G., 2005. “Development of a heavy-duty diesel modal emissions and fuel consumption model.” Transportation Research Record 1587 (1997): 56 – 62

[20] Ray Galvin, “Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers,” Transportation Research Part D: Transport and Environment, Volume 53, 2017, Pages 234-248

[21] E. H. Gonzalez, J. Garrido, M. Barth and K. Boriboonsomsin, "Machine Learning-based Energy Consumption models for Battery Electric Trucks," *2023 IEEE Transportation Electrification Conference & Expo (ITEC)*, Detroit, MI, USA, 2023, pp. 1-6

[22] Sangjun Park and Hesham A Rakha. 2006. Energy and environmental impacts of roadway grades. Transportation research record 1987, 1 (2006), 148–160.

[23] Erik Hellström, Maria Ivarsson, Jan Åslund, and Lars Nielsen. 2009. Look-ahead control for heavy trucks to minimize trip time and fuel consumption. Control Engineering Practice 17

[24] Boriboonsomsin K, Barth M. “Impacts of Road Grade on Fuel Consumption and Carbon Dioxide Emissions Evidenced by Use of Advanced Navigation Systems.” Transportation Research Record. 2009;2139(1):21-30

[25] Y. Ren, L. Zhang and P. N. Suganthan, (2016) "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]," in *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 41-53

[26] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794

[27] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

[28] André Altmann, Laura Toloşi, Oliver Sander, Thomas Lengauer, “Permutation importance: a corrected feature importance measure,” Bioinformatics, Volume 26, Issue 10, 15 May 2010, Pages 1340–1347

[29] Olavi H Koskinen. 2008. Improving vehicle fuel economy and reducing emissions by driving technique. In 15th World Congress on Intelligent Transport Systems and ITS America’s 2008 Annual MeetingITS AmericaERTICOITS JapanTransCore

[30] Orkun Karabasoglu and Jeremy Michalek. 2013. Influence of driving patterns on life cycle cost and emissions of hybrid and plug-in electric vehicle powertrains. Energy policy 60 (2013), 445–461

[31] Wang, C., Hao, P., Boriboonsomsin, K., and Barth, M. (2022) “Developing a mesoscopic energy consumption model for battery electric trucks using real-world diesel truck driving data.” Proceedings of IEEE Vehicle Power and Propulsion Conference, Merced, CA, November 1-4

[32] Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.