MECHANICAL ENGINEERING

UNIVERSITY *of* WASHINGTON

PACCAR Capstone Design Project
Winter, 2022 – Spring, 2022

MD and HD Vehicle Range Estimation
PACCAR

Range Finders

Anthony Rudasics[1], arudasic@uw.edu
Chia Agrawal[1], chit02@uw.edu
Peter Ma[1], rma29@uw.edu
Ruizhe Zhao[1], zhaor98@uw.edu
[1]Mechanical Engineering

With guidance by

Nick Hertlein[1], Technical Lead Powertrain Definition, nick.hertlein@paccar.com
Maarten Meijer[1], Advanced Technology Manager, maarten.meijer@paccar.com
Raeef Barsoum[1], Systems Engineer, raeef.barsoum@paccar.com
Krithika Manohar[2], Faculty Advisor, kmanohar@uw.edu
Ashis Banerjee[2&3], Faculty Advisor, ashishb@uw.edu
[1]PACCAR Employee, [2]Mechanical Engineering, [3]Industrial & Systems Engineering

BE BOUNDLESS

Executive Summary

The purpose of our project is to design an optimization algorithm for medium-duty (MD) and heavy-duty (HD) vehicles so that it can help determine optimal component sizing for extending driving range. These components include batteries, e-motors, etc. We do this by incorporating new powertrain technologies and a generic optimization algorithm into a modeling tool based on MATLAB Simulink to address range anxiety which is a critical operation a vehicle manufacturer performs as the consumer's trust is critical.

We used four regression models viz., Linear regression, Ridge regression, Lasso regression, and Gradient Boosted Decision trees. We tested our model for data for three different routes: city, highway, and pickup & delivery (P&D). Our results state that the Lasso regression shows the lowest error implying it can be used to best predict power consumption. From our results, we also found that the city and the P&D routes had the best prediction compared to the highway route, with less than half the error.

Linear regression and ridge regression use a physics-based model which includes only three variables to predict power consumption; therefore, the predictions were less precise. However, Lasso regression identifies additional drive cycle variables, such as pedal position and cruise control, which leads to higher accuracy. Also, we believe that the predictability was affected by nonlinear regenerative braking, which can explain why the highway route has the most error. Last, regression coefficients from lasso regression were validated by the physics-based model.

Table of Contents

List of Figures

List of Tables

1 Project Introduction

1.1 Need Statement

A way to design an optimization algorithm for medium-duty (MD) and heavy-duty (HD) trucks so that it can determine required component sizing such as batteries, e-motors, etc.

1.2 Project Scope and Objective

The scope of our project includes creating a range estimation model based on collected sensor data from PACCAR's electric trucks and optimizing this model to allow engineers to select the most suitable components for each truck. The model is not intended to predict range in real-time, but rather to analyze data following its collection. PACCAR already has an existing tool used to predict the range within its completed trucks. The goal for our model is for it to be capable of showing the effects of using different components within a truck to compare them.

The objective of this project is to incorporate new powertrain technologies and a generic optimization algorithm into a modeling tool based on MATLAB Simulink to determine the required component sizing for medium-duty (MD) and heavy-duty (HD) trucks. The tool should be able to evaluate and optimize powertrains based on different drive cycles and performance requirements.

1.3 Significance and Background

Range anxiety is the vehicle owner or driver's fear that their vehicle does not have sufficient resources (battery or fuel etc.) to reach its final destination before it can be restocked. Addressing range anxiety is a critical operation a vehicle manufacturer performs as the consumer's trust is critical.

We implemented a Physics model for the truck which can be represented using the following equation:

$$HighPower = \frac{1}{2} \times C \times FA \times V_{air}^{3} + RRC \times M \times 9V + \frac{M(V_k - V_{k-1})}{\Delta t} \times V$$

Our range estimation model is divided into various stages that include:

Stage 1: data acquisition and preprocessing. For this, we assumed to ignore the regenerative braking and assumed the travel to be on flat ground, and ignore turning, and pedal position.

Stage 2: Build an initial model. Build a physical model and based on that build a drivetrain model and a battery model.

Stage 3: Use drive data to fit/train the model and validate the model

Stage 4: Make use of the trained model by using different drive cycles, and changing the truck parameters and battery size.

With the current state of electric vehicle infrastructure across vast road systems, charging stations are uncommon. According to the US Department of Energy, there are 49,796 EV charging locations across the United States [1]. The majority of these locations are concentrated around urban centers with very few locations interspersed across less populated areas where trucking lanes commonly pass through [2]. This infrastructure is likely to improve, but until then it is critical to have a high degree of certainty that a charging station is within range of your electric vehicle. This requires a consistent and accurate range estimation model for electric vehicles that use collected sensor data to predict driving range. Our particular focus is on electric trucks, which must be modeled differently than electric cars due to their larger vehicle and load masses. When transporting heavy loads across long distances, weight optimization is important to understand the range and limitations of electric trucks.

1.4 Existing Solutions

A team of researchers from North Carolina State University developed a similar model to what we are pursuing [3]. The model collected vehicle data, terrain data, and meteorological data to create a range estimation model for EVs [3]. The model was able to successfully estimate the range based on historical and real-time data [3]. This model was developed generally for EVs and uses a constant mass for its calculations; however, this is not a practical representation for trucks that have frequently changing load masses [3].

1.5 Stakeholders

The stakeholders of our product include MD and HD vehicle owners, drivers, utility companies, and PACCAR or other vehicle manufacturers. The expectations, interests, and resulting strategies of stakeholders relate to and depend upon the specific configurations/conditions they require the vehicle to function in.

In general, the stakeholders of our model do not seem overly concerned about the short-term returns on investments but look for a long-term solution to their problem.

Paccar's work towards the development of multiple new powertrain architectures, such as a battery, fuel cell, and diesel hybrid, for their medium-duty (MD) and heavy-duty (HD) trucks required to optimize component sizing and select the right configuration for different real-world vehicle applications. Our solution is a model based on MATLAB to estimate the power consumption and the vehicle range based on different drive-cycles for the same model of vehicle. Using MATLAB to build optimization algorithms reduced computational costs that could have been high due to a large number of variables. If we get an accurate prediction of the vehicle range, the vehicle owner or driver can get an estimated idea of whether and where a stop needs to be planned to avoid "range anxiety".

2 Problem Statement

2.1 Needs Research

Through observation of a similar model from the North Carolina State University research group, we found aspects of their model which would still apply to larger trucks with variable mass [3]. Through conversation with our industry partners, we probed for their needs and standards of quality to make a satisfactory and useful model.

2.2 Customer Requirements

The requirements of our model include a tool that allows a user to change certain parameters including the battery size, load, air resistance coefficient, and input parameters like different drive cycles and speed.

2.3 Engineering Specifications

The specification applies to the quality of our predictive mathematical model is a root mean square error (RMSE). RMSE serves as a measure of discrepancy between real measured data and predicted data. The target value is to have 0 RMSE, but this level of RMSE is unreasonable to achieve, so we instead aimed to minimize RMSE. The model could be declared satisfactory if the magnitude of the RMSE (watts) was less than 10% of the magnitude of the average power consumption (watts).

2.4 Codes and Standards

During the process of developing this model, we were not required to adhere to any codes or standards. The model was created at the discretion of our PACCAR industry partners according to their needs and judgment of satisfactory performance.

2.5 Design trade-offs and challenges

Our PACCAR industry partners needed an algorithm for MD and HD electric vehicles to determine optimal component selection for maximizing vehicle range capabilities. The two main metrics to maximize during the scope of this project were prediction accuracy and versatility. Accuracy was quantified in terms of RMSE, while versatility was noted in the model's performance across various route scenarios. No significant tradeoffs had to be made throughout this project. The objective was to maximize these two metrics and no other metrics were significant enough to consider. A challenge that arose in our attempts to create models is that one drive scenario may have a satisfactory error whereas another scenario under the same model may not be satisfactory. Our final model has eliminated this issue and is satisfactory by our standards.

3 Conceptual Design, Evaluation, and Selection

3.1 Data Pre-processing

The data is measured by sensors, which means noise exists. By pre-processing the raw data, we can make it more usable for calculations or looks better visually.

- **Moving Average Filter:** Take the average of a window of samples. Can be used to smooth out the spikes.
- **Sensor Fusion:** In the data given by PACCAR, there are two sets of data that measure the vehicle speed: wheel speed and motor speed. By fusion of these two data sets, we can get a more accurate fusion measurement. Methods like Kalman filter and weighted fusion can be used.
- **Discrete Wavelet Transform Filter:** A wavelet-based denoising technique
- **Numeric Derivation:** The acceleration is not directly given in the dataset. We can use numeric derivation to calculate the acceleration based on speed data.

3.2 Possible Training Methods

We are given specifications of the truck and its drive cycle data. There are 2000 sets of sensor data that first thought would require some feature selection algorithm.

- **LASSO:** A penalty factor λ is used to minimize the unrelated term to avoid overfitting. Then find a sweet spot for the number of terms vs training error. Can be used for feature selections. However, it might not be able to find the non-linear relationship or combination of the variables.
- **CINDy:** Automatically use a library of combinations of variables to fit the dynamic data. Great for dynamic systems but may not work well for us.
- **Linear Regression:** The most straightforward method. Can be used if there is a linear relationship between variables and the output.
- **Ridge Regression:** Penalty term k is used to minimize the unrelated term to avoid overfitting. When k is equal to zero it is a linear regression. Use varies of k to final smallest MSE and Error.
- ***k* Fold Cross-Validation:** Randomly cut the dataset into k sections and use about 70% for training and 30% for validation.
- **CV Partition:** return randomly selected sections of data

- **Stepwise Regression:** Generate combinations of input variables and find their *P* values of them. For example, the input is speed, acceleration, and time, it will generate speed*acc, speed*time, and speed^2*acc terms, and find the *p*-value of each of them. The small *p*-value means that the variable or that combination of variables has a significant effect on the output.
- **Gradient Boosted Design Trees:** Use multiple features and multiple branches to determine one parameter.

3.3 Training

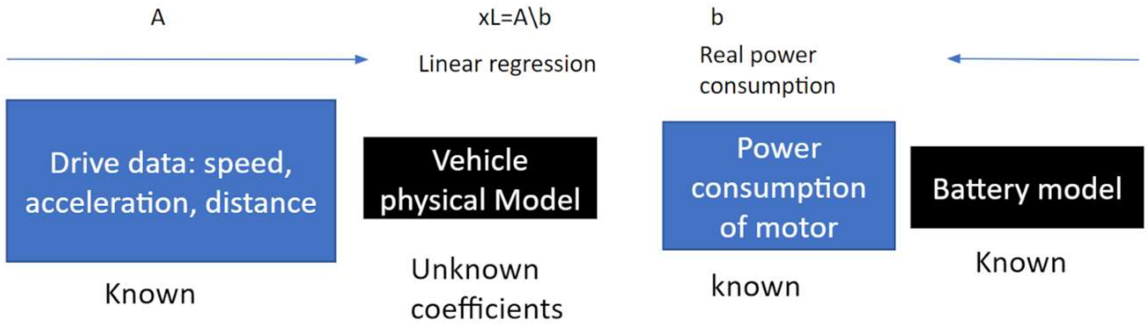


Figure 1: Linear Regression Process Flowchart

$$P_{road} = \frac{1}{2} * \rho * C_{D*}FA * v^3 + RRC_0 * M_{veh} * g * v + M_{veh} * v * \frac{dv}{dt} + M_{veh} * g * \frac{dh}{dt}$$

There are around 2000 variables in the drive dataset. Feature selection algorithms like lasso can be used but are necessary and do not guarantee success. Features were selected based on these physics' equations.

Terms	Coefficients	Estimated Coefficient	Variable
Aero resistance term	Air density (ρ) drag coefficient (Cd) Frontal Area (FA)	$0.5 * 1.225 \frac{kg}{m^3} * 0.67 * 8.052(m^2) * v\left(\frac{km}{hr}\right)^3 * \left(\frac{1}{3.6^3}\right) = 0.0706$	Velocity(v)^3
Rolling resistance term	Rolling resistance coefficient (RRC0) Mass of vehicle (M_{veh}) gravity(g)	$0.87\left(\frac{kg}{ton}\right) * \frac{1}{1000} * 11794kg * 9.8 * \frac{v}{3.6} = 27.93$	Velocity(v)
Acceleration term	Mass of vehicle (M_{veh})	909	Velocity(v)*Acceleration(dv/dt)

Table 1: Power Equation Table

A table is created to better illustrate the equation. The variables can be found in drive cycle data. All the constants can be found online or provided by PACCAR.

3.4 Validation

To validate the model, the following methods can be used:

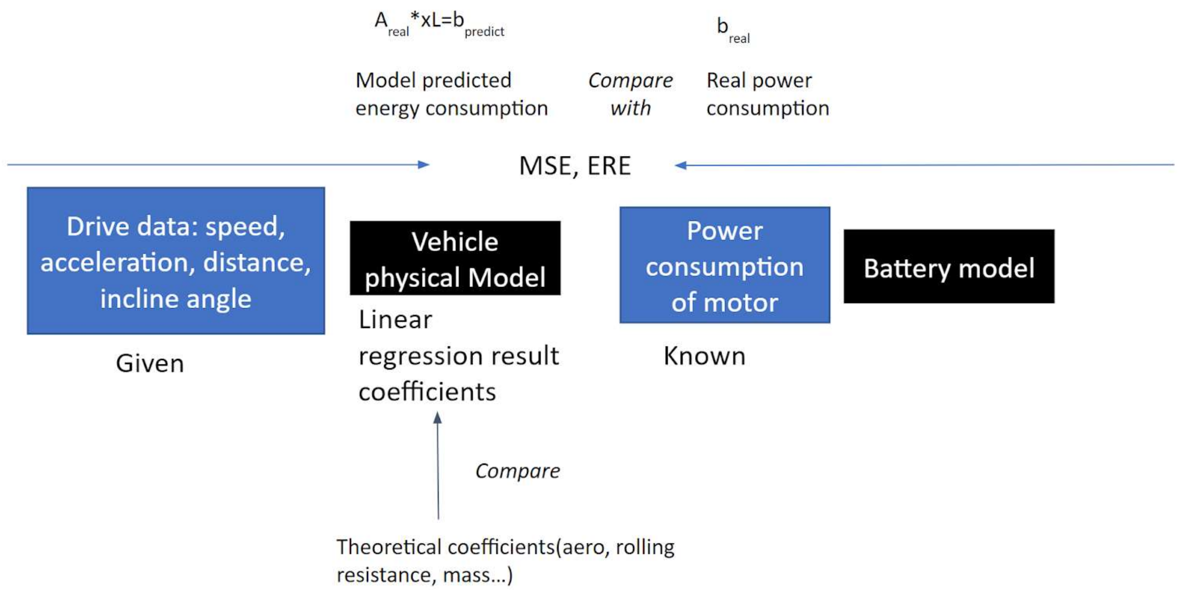
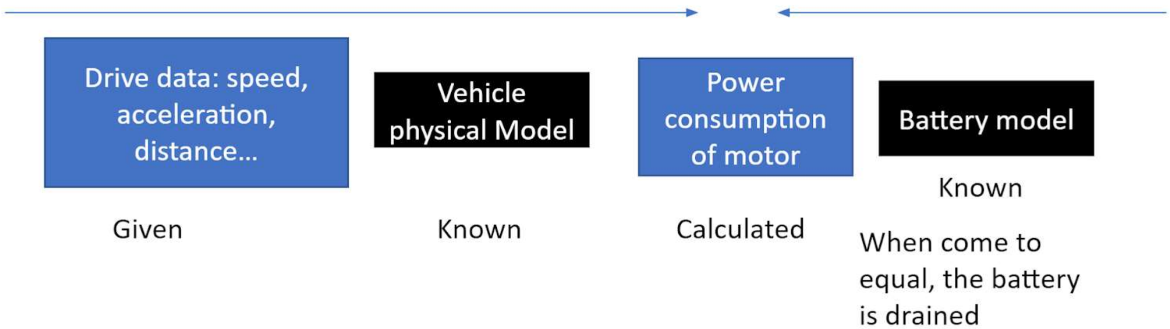


Figure 2: Model Comparison Process Flowchart

- Test for a different set of data that the model has never seen: for example, we have 3 sets of data: train1 test23; train2 test13; train3 test12
- ERE (Energy consumption Relative Error): Numeric integrates the power consumption by time to get the energy consumption and compare the predicted value to the real value.
- MSE (Mean Squared Error): It is used to measure how is fitted data deviated from real data
- Sequential 10-fold: Can be used to visualize the distribution of coefficient
- Check coefficient with theory: The linear regression model will return 3 coefficients. We can compare them with the theoretical value.
- Check overfitting: Overfitting can occur if too many variables are used. The overfitted model may look great when testing with the data that trained the model. However, it will do poorly when fitted with new data.
- P-value: Can be used to verify the significance of the featured term
- Pie chart: Visualize the contribution of each term
- Error vs time plot: Visualize the error vs time

3.5 Use The Model for Other Purposes



Possible use cases:

- Use different drive cycles: For example, The EPA Urban Dynamometer Driving Schedule to see what's the range
- Change truck physical parameters: mass, frontal area, or air drag coefficient.
- Change battery size: a smaller/bigger battery. Smaller battery capacity means lower range but it cost less and has less weight
- Find the optimized battery size: cost-effective, reasonable range for different drive cycle

3.6 Selection

Since we only obtained data from the same truck, the machine learning approach was impossible due to the limited data. Therefore, we picked a physical model with linear regression instead.

For our model, linear regression is a helpful analysis method. From the sensor data provided by PACCAR, not every data variable collected is relevant to estimating range. We have selected a small number of relevant variables which our model uses to estimate the range. Linear regression is one of the most simple but useful analysis methods because it is effective at minimizing least squares error.

The ridge regression is very similar to the linear regression and is also a very helpful analysis method for our model. The ridge regression helps to create a more generalized model when a simple linear regression is overfitting the data. The ridge regression has an additional regularization factor to assist with this generalization. We input a series of potential ridge parameters in the model and may choose the optimal parameter by plotting the ridge parameters against their corresponding mean squared error.

The lasso regression is an extension of the ridge regression that can be very useful for selecting variables. The lasso regression features a penalty term that may change the slope of the variables in a way that biases them based on their significance to the model. Unlike ridge regression, lasso regression may set the slope of a variable to zero. While this makes the lasso regression useful for selecting relevant parameters for our model, we have already hand-selected our parameters based on relevant physics equations for estimating range. Therefore, the lasso regression is not suited for an already filtered dataset.

4 Detailed Design

4.1 Data Pre-processing

We are given 3 sets of data at first; they are from the same truck and were tested in a similar route

Data set	Name	Length(samples)	Variables
1	20210923_143850_AllMessages.mat	11709	1048
2	20210924_024543_AllMessages.mat	31554	1048
3	20210924_034231_AllMessages.mat	22057	1048

Table 2: Data File Info

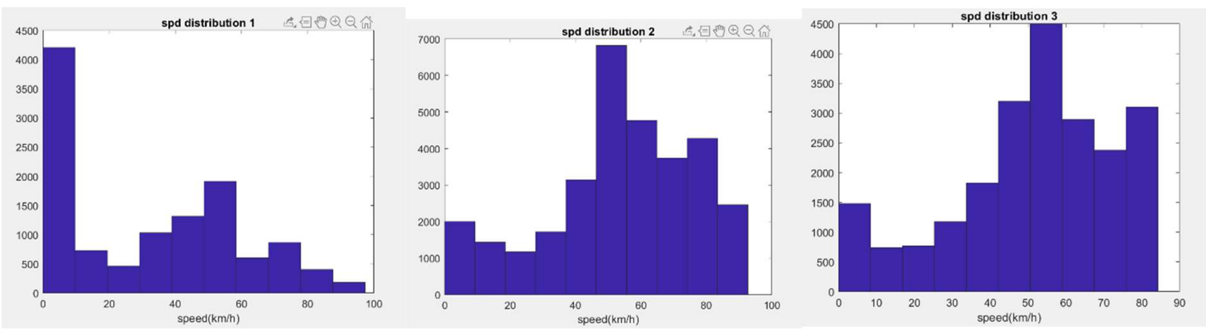


Figure 3: Histogram of Speed

We can see data 1 spending a lot of time parking. Data 2 and 3 are mostly mid and high-speed.

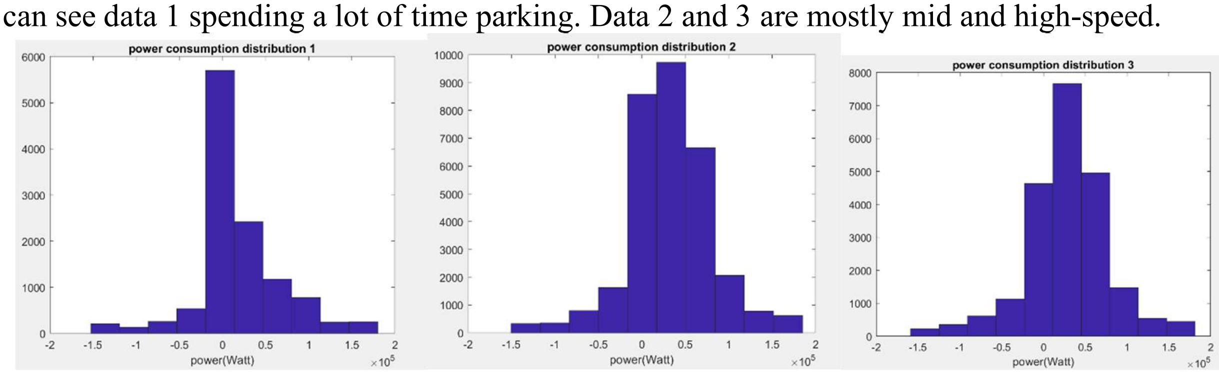


Figure 4: Histogram of Power Consumption

We can tell all of them have a mean above zero, which is to be expected. The negative power consumption means it is regenerating the power.

Using the physics equation, we have identified three featured variables: velocity, velocity^{3}, and velocity*acceleration. However, velocity^{3} and velocity*acceleration are not directly given in drive data, so we generate them before we put them into the model.

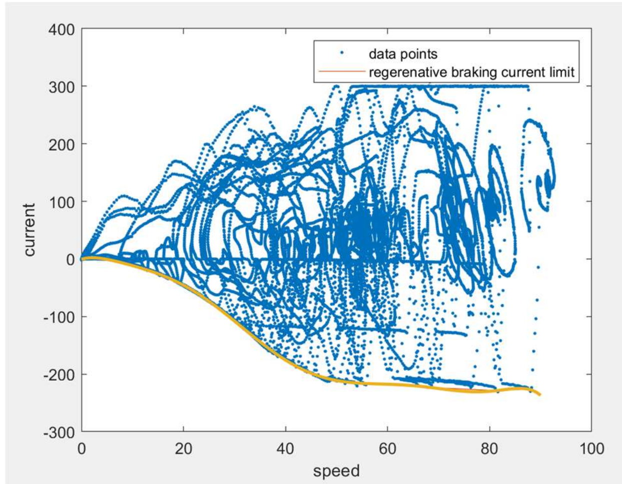


Figure 5: Current vs Speed

There is a clear boundary at the negative current region. It is the predetermined regenerative braking current limit; it is a function of speed. Poly fit is used to fit this boundary. In the later modeling section, this limit is applied to prevent the model from generating too much power. The plot without and with the limiter is shown below.

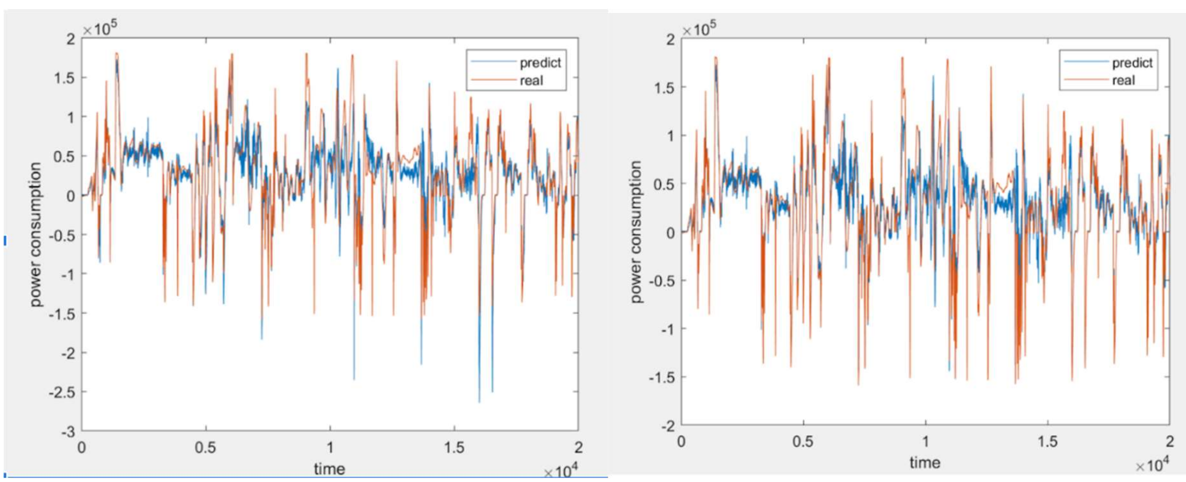


Figure 6: Linear Regression Model Testing, Before and After Limiter is Applied

There is a clear reduction of blue line spikes out of the orange line.

4.2 Modeling and Analysis

Linear regression:

The linear regression was used to fit the physical model coefficients to predict the vehicle's power consumption. We used the CV partition function from MATLAB to define training and test sets for validating our physical model using cross-validation. The CV partition will randomize our data and use 70% as a training set.

Ridge Regression:

Ridge regression was used to shrink the coefficients and produce a less overfit model.

Lasso Regression:
The Lasso regression was used to shrink the coefficients and select variables that are related to predicting power consumption.
Gradient Boosted Tree Regressor
It is a type of decision tree. It uses multiple features and multiple branches to determine one parameter when a certain threshold is met. For each tree generate, find the gradient of the loss function. Generate a new tree based on the gradient of the loss function

4.3 Prototypes and Testing (Validation and Visualization)

Linear Regression:
The following plot is one of the plots representing the actual power consumption versus the predicted power consumption using linear regression. According to the figure, we observed that the prediction follows well with the actual power consumption. We calculated the mean square error (MSE) and the error between the prediction and actual power consumption (ERE) to validate the finding, as shown in the table below. Each error was calculated using the average of 10 sets of data. According to the table, linear regression showed a good fit overall.

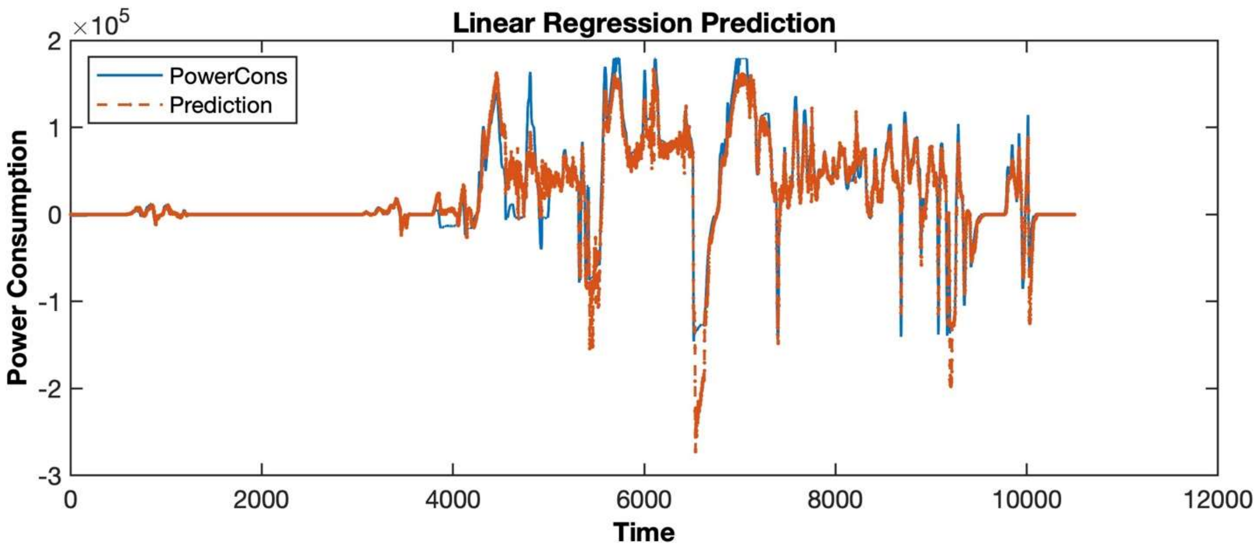


Figure 7: Linear Regression Actual Power Consumption vs. Predicted Power Consumption Plot

LINEAR MSE	City	Highway	P&D
1	4.33E+08	6.35E+08	3.16E+08
2	5.21E+08	6.40E+08	3.61E+08
3	5.73E+08	1.11E+09	2.89E+08
4	4.80E+08	1.19E+09	1.86E+08
5	5.34E+08	1.31E+09	4.33E+08
6	6.38E+08	1.30E+09	4.04E+08
7	7.63E+08	1.17E+09	1.98E+08
8	3.89E+08	1.24E+09	5.71E+08
9	4.79E+08	1.48E+09	5.52E+08
10	5.91E+08	1.43E+09	3.96E+08
Average	5.40E+08	1.15E+09	3.71E+08

LINEAR ERE	City	Highway	P&D
1	-3.12E+00 %	-3.24E+00 %	1.13E-02 %
2	-2.45E+00 %	-3.65E+00 %	-2.98E-02 %
3	-1.07E+00 %	1.40E-01 %	1.21E-01 %
4	-3.89E+00 %	-1.54E-01 %	-1.86E-01 %
5	-2.67E+00 %	-1.44E-01 %	-3.12E+00 %
6	-2.29E+00 %	-2.71E-01 %	-1.26E+00 %
7	-5.76E+00 %	-2.86E-01 %	-5.60E-01 %
8	-1.71E+00 %	4.96E-02 %	-1.65E+00 %
9	-5.29E+00 %	-5.97E-01 %	-1.85E+00 %
10	-3.06E+00 %	-1.76E-01 %	-3.28E+00 %
Average	-3.13E+00 %	-8.33E-01%	-1.18E+00

Table 3: Linear Regression Error Calculation

To further validate our model, estimated coefficients from each data set were used to test on other sets of data. For example, the estimated coefficients from one set of data were used to predict the power consumption of two other data sets. According to the table [4], we observed that the linear regression method could predict the actual power consumption with low MSE and ERE, which indicated that the linear regression model could predict the power consumption well. However, since linear regression uses a physical-based model which includes only three variables to predict power consumption, our predictions were less precise.

Training	**Data 1**		**Data 2**		**Data 3**	
Testing	**Data 2**	**Data 3**	**Data 1**	**Data 3**	**Data 1**	**Data 2**
MSE	2.3008E+4	2.4421E+4	1.2162E+4	2.4677E+4	1.2565E+4	2.1618E+4
ERE	10.32%	9.62%	2.73%	2.88%	5.93%	2.99%

Table 4: Linear Regression Model Validation

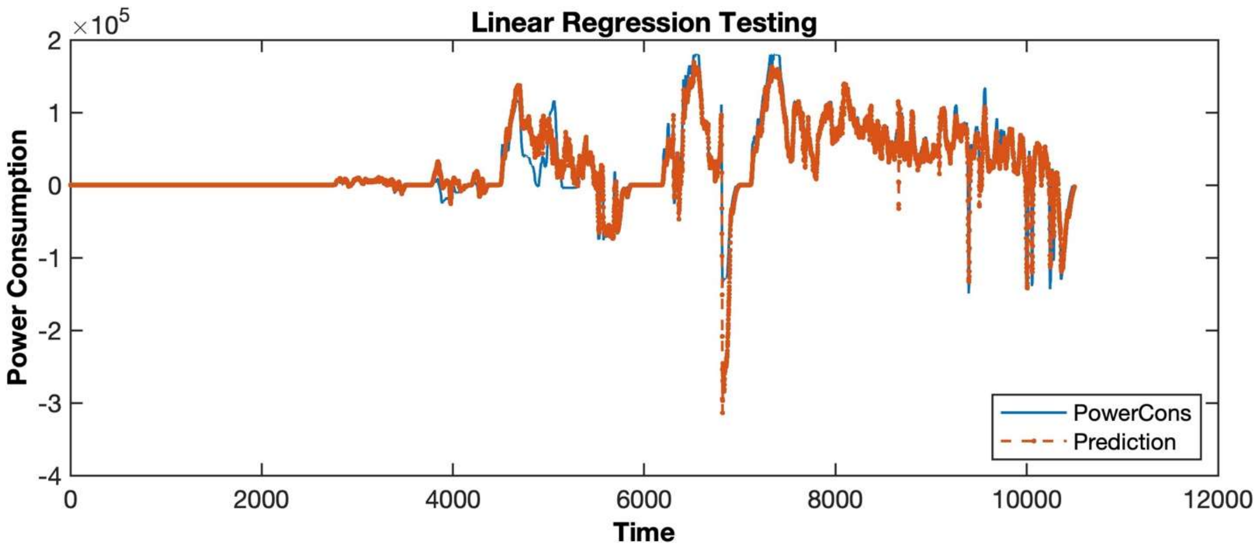


Figure 8: Estimated Coefficients Used on Different Data Sets

Lasso regression:

The following plot is the prediction plot of Lasso regression. Our prediction has a tighter fit to the actual power consumption than linear regression. It follows the trend almost exactly, even in sections that linear regression cannot predict well. We calculated the mean square error (MSE) and the error between the prediction and actual power consumption (ERE) to validate the finding, as shown in the table below.

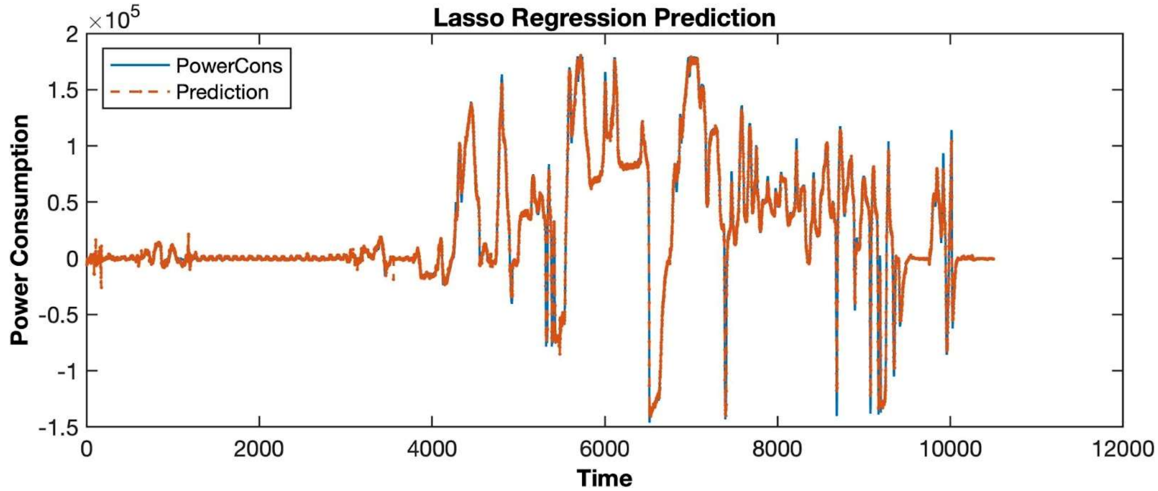


Figure 9: Lasso Regression Actual Power Consumption vs. Predicted Power Consumption Plot

LASSO MSE	City	Highway	P&D
1	1.70E+07	3.38E+08	2.50E+07
2	1.64E+07	4.08E+08	1.23E+07
3	9.10E+06	1.40E+07	1.61E+07
4	1.56E+07	1.20E+07	9.94E+06
5	1.32E+07	1.09E+07	1.70E+07
6	1.23E+07	3.67E+08	1.16E+07
7	5.53E+08	1.18E+07	1.54E+07
8	1.29E+07	1.32E+07	1.44E+07
9	1.68E+07	4.01E+08	1.48E+07
10	1.61E+07	4.70E+08	3.96E+08
Average	6.82E+07	2.04E+08	5.33E+07

LASSO ERE	City	Highway	P&D
1	-6.63E-05 %	-7.00E-04 %	1.12E-04 %
2	6.87E-05 %	-1.10E-03 %	-3.20E-05 %
3	2.33E-04 %	-1.60E-05 %	-2.72E-06 %
4	7.32E-06 %	-2.11E-05 %	-3.82E-05 %
5	3.58E-05 %	-2.17E-05 %	-6.63E-05 %
6	2.08E-05 %	-1.50E-03 %	1.40E-05 %
7	-5.48E-04 %	-5.37E-05 %	5.24E-04 %
8	3.96E-05 %	2.96E-05 %	-2.07E-05 %
9	2.80E-05 %	-7.21E-04 %	3.43E-06 %
10	2.02E-06 %	-8.11E-04 %	-2.40E-03 %
Average	-1.79E-05 %	-4.91E-04 %	-1.91E-04 %

Table 5: Lasso Regression Error Calculation

The following table shows identified important variables using Lasso regression. To validate our findings, we manually take a look at our variables names and check the meaning behind them to back sure the variables that we found were useful. The problem with Lasso regression is that it identified multiple variables of same type. For example, we were given multiple speed data, and Lasso identified all of them.

Identified Variables	Meaning
'V2BCMDVehicleSpeed'	Vehicle Speed
'Absolute_Speed'	Absolute Speed
'Cruise_Control_Enab'	Cruise Control
'RelativeSpeedFrontA'	Relative Front Axle Speed
'RelativeSpeedRearAx'	Relative Rear Axle speed
'Condenser_Fan_RPM_R'	Condenser Fan RPM
'AccelPedalPos2'	Accelerator pedal position
'v_cube'	Variables from physical model
'VtimesAcc'	Variables from physical model
40 More Additional variables	

Table 6: Partial Identified Variables

Ridge regression:
The following plot is the prediction plot of Ridge regression, which is almost identical to linear regression. Ridge regression was used initially to shrink the coefficients and produce a less overfit model. However, using CV partition, we observed that linear regression could already do a good job of predicting the actual power consumption.

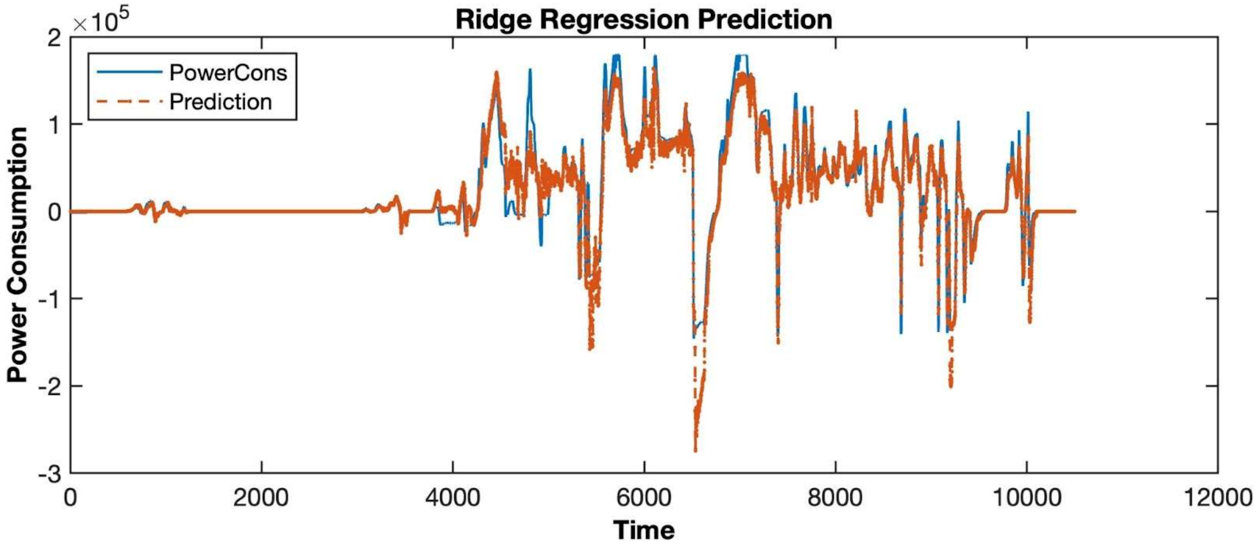


Figure 10: Actual Power Consumption vs. Predicted Power Consumption Using Ridge Regression

RIDGE			
MSE	City	Highway	P&D
1	4.38E+08	6.51E+08	3.16E+08
2	5.26E+08	6.60E+08	3.61E+08
3	5.74E+08	1.11E+09	2.89E+08
4	4.84E+08	1.19E+09	1.86E+08
5	5.39E+08	1.31E+09	4.38E+08
6	6.41E+08	1.30E+09	4.04E+08
7	7.81E+08	1.17E+09	1.99E+08
8	3.90E+08	1.24E+09	5.79E+08
9	4.84E+08	1.48E+09	5.87E+08
10	5.97E+08	1.43E+09	4.14E+08
Average	5.45E+08	1.15E+09	3.77E+08

LASSO			
ERE	City	Highway	P&D
1	-6.63E-05 %	-7.00E-04 %	1.12E-04 %
2	6.87E-05 %	-1.10E-03 %	-3.20E-05 %
3	2.33E-04 %	-1.60E-05 %	-2.72E-06 %
4	7.32E-06 %	-2.11E-05 %	-3.82E-05 %
5	3.58E-05 %	-2.17E-05 %	-6.63E-05 %
6	2.08E-05 %	-1.50E-03 %	1.40E-05 %
7	-5.48E-04 %	-5.37E-05 %	5.24E-04 %
8	3.96E-05 %	2.96E-05 %	-2.07E-05 %
9	2.80E-05 %	-7.21E-04 %	3.43E-06 %
10	2.02E-06 %	-8.11E-04 %	-2.40E-03 %
Average	-1.79E-05 %	-4.91E-04 %	-1.91E-04 %

Table 7: Ridge Regression Error Calculation

We also observed that as the *k* value increases logarithmically, the MSE and error decreased linearly, which indicates that the model is no longer overfitting. However, since we only used three variables for Ridge regression, the model cannot shrink anymore, therefore this method is not as useful as Linear regression and Lasso regression, and the *k* value is not useful in this case.

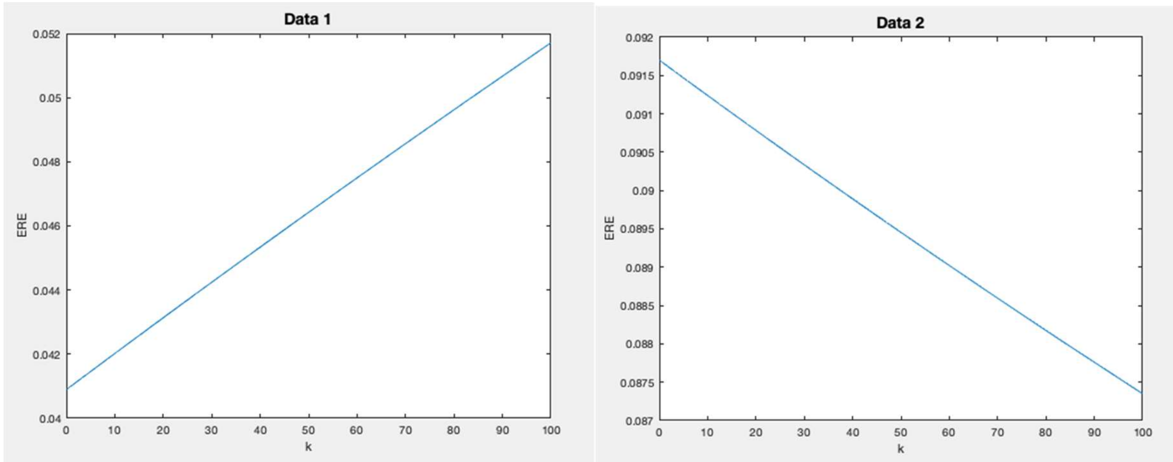


Figure 11: Error vs k plot

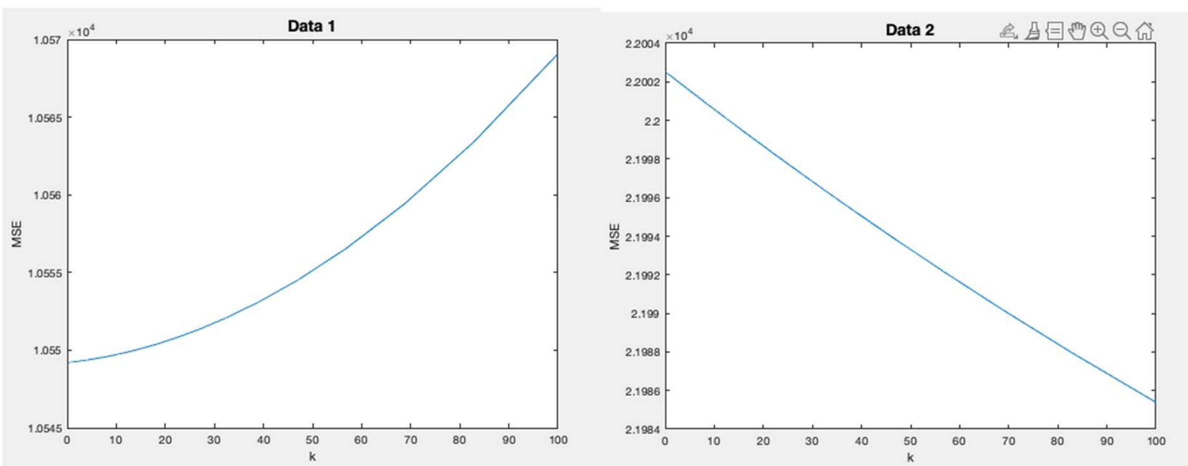


Figure 12: MSE vs k plot

Gradient Boosted Tree Regressor

This part was done in Python with the Gradient Boosting Regressor library from Sklearn. The focus for this part is to train the model with different hyper parameters that can achieve minimum MSE and ERE. This plot gives a general idea of how it fits. The error fluctuates around zero and occasionally reaches more than 100000 for a short time.

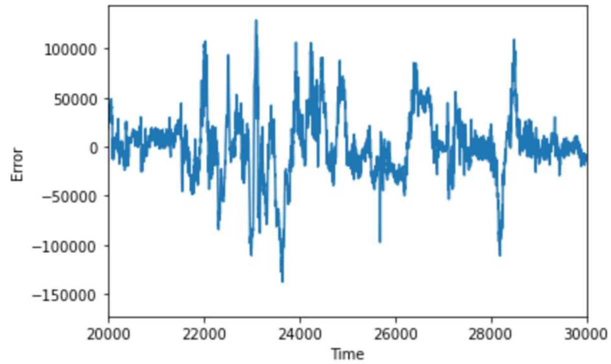


Figure 13: Error vs time

The first parameter to deal with is learning rate. In this case as long as the learning rate is above 0.1, it will perform well.

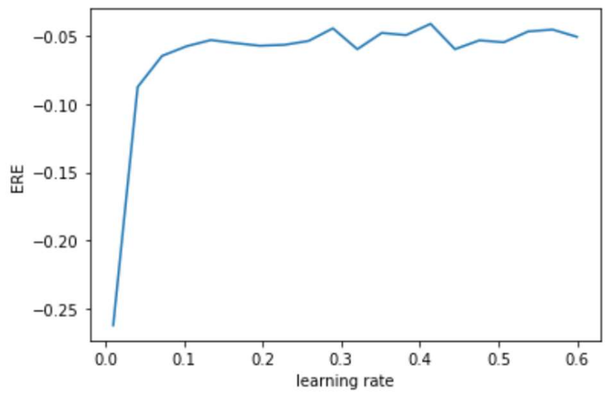


Figure 14: ERE vs learning rate

The next case is Alpha vs ERE. The loss function was changed to Huber loss function (default alpha=0.9)
Huber loss function is less sensitive to outliers than the squared error loss. An alpha value at around 0.4 seem to yield best results. Test also done on another dataset, yield similar result.

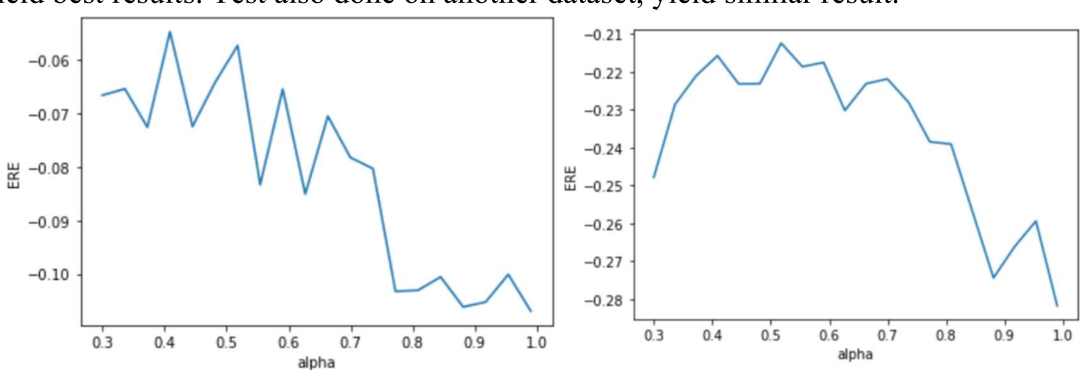


Figure 15: ERE vs alpha

However, when testing with one of the city routes, there is no clear trend of the optimal alpha value

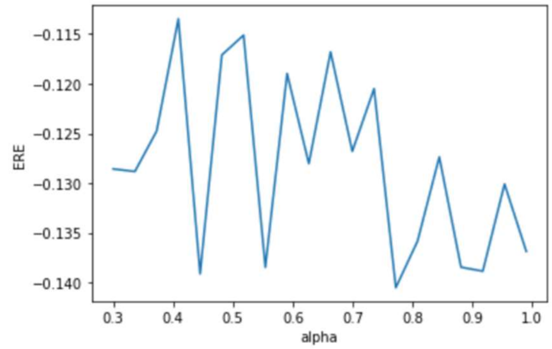


Figure 16: ERE vs alpha

The other parameter we have tested is Max depth. According to official documentation [5], the maximum depth of the individual regression estimators. The maximum depth limits the number of nodes in the tree. To tune this parameter for best performance; the best value depends on the interaction of the input variables. The default value works great for us.

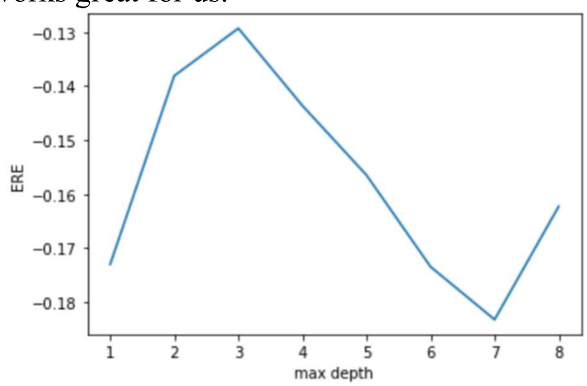


Figure 17: ERE vs max depth

According to official documents [5], Gradient boosting is fairly robust to over-fitting so a large number of boosting stages usually results in better performance. The plot below shows the effect of boosting stage on both ERE and MSE. Higher boosting stages provide a better energy consumption estimate. However, higher boosting stages make MSE slightly increases. 140 could be a reasonable boosting stages in this case.

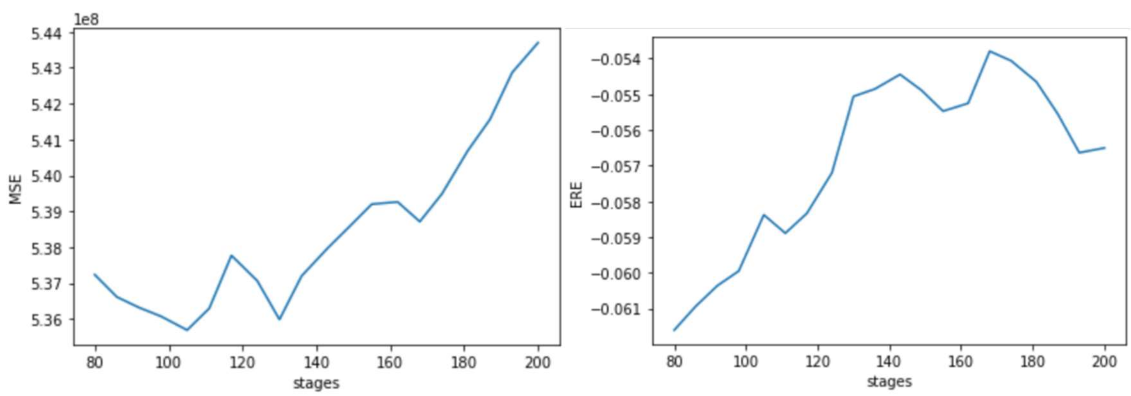


Figure 18: MSE vs stages (left); ERE vs stages (right)

4.4 Additional Validation

Check Coefficient with Theory:

Coefficient 1	**909**
Coefficient 2	**27.93**
Coefficient 3	**0.0706**

Table 8: Theoretical Coefficients

	Data 1 (20210923_143850)	**% Diff**	**Data 2**	**% Diff**	**Data 3**	**% Diff**
Coefficient 1	919	1.10	816	10.2	791	13.0
Coefficient 2	433	**1450**	312	**1017**	351	**1157**
Coefficient 3	0.0364	48.4	0.0624	11.6	0.0526	25.5

Table 9: Linear Regression Coefficients vs Theoretical

	Data 1	**% Diff**	**Data 2**	**% Diff**	**Data 3**	**% Diff**
Coefficient 1	919	1.10	816	10.2	791	13.0
Coefficient 2	406	**1354**	229	**720**	106	**280**
Coefficient 3	0.0389	44.9	0.0689	2.41	0.0756	7.08

Table 10: Ridge Regression Coefficients vs Theoretical

The aero and acceleration terms are at a reasonable level.
As highlighted in the table. We have seen a massive difference in rolling resistance terms. The theoretical rolling resistance coefficient was calculated by the wheel RRC0*weight. This only accounts for the rolling resistance of the wheel. There is more in an actual truck. For example, motors and speed controllers generate heat, gearbox, and axles have friction. The power consumption of these terms also has a linear relationship with speed. In other words, the rolling coefficient generated by linear regression includes all kinds of resistance.
We observed that the rolling resistance accounts for 91% of the total energy consumption.

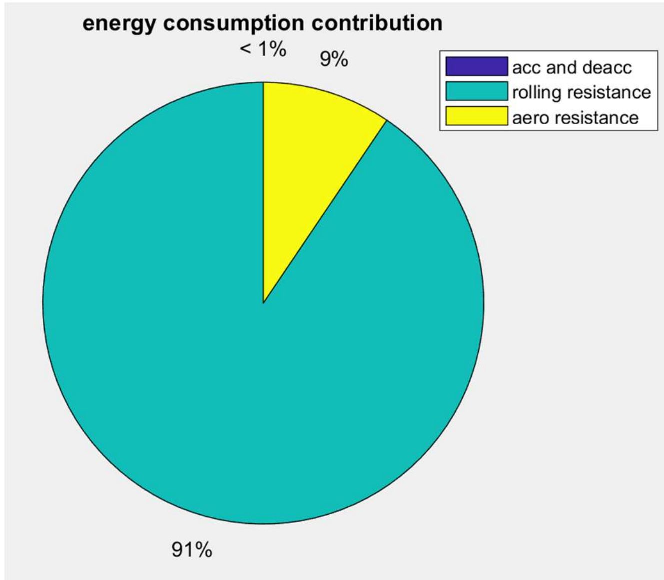


Figure 19: Energy Consumption Contribution

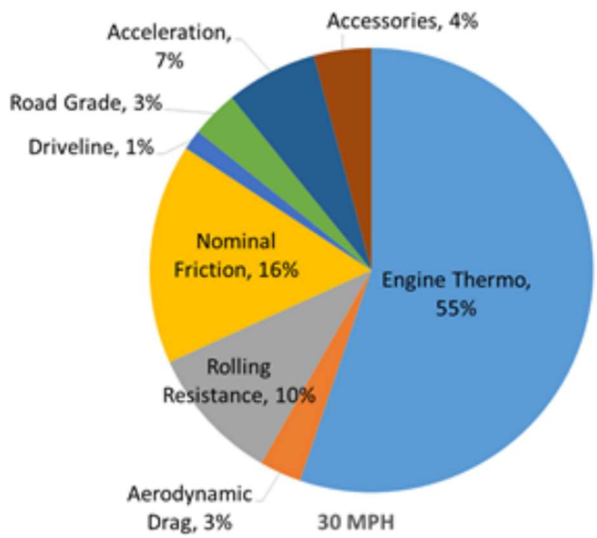


Figure 20: Energy Consumption Contribution of a Diesel Truck [4]

By comparing a diesel truck pie chart, we see a huge difference. For the aero term, the Electric Truck is much higher. We believe it is due to the Electric Truck being much more efficient than the diesel truck, so the aero is at a bigger portion of the power consumption. For rolling resistance terms, the diesel truck is at 10 percent. That's the value for the wheel rolling resistance only. The rolling resistance in our case can be expressed as: Wheel rolling resistance + Nominal Friction + Motor and controller heat.
The acceleration energy consumption contribution in our case is close to zero. This is due to our linear regression model regenerating all the braking energy, and the heat during this process was included in the rolling resistance term.

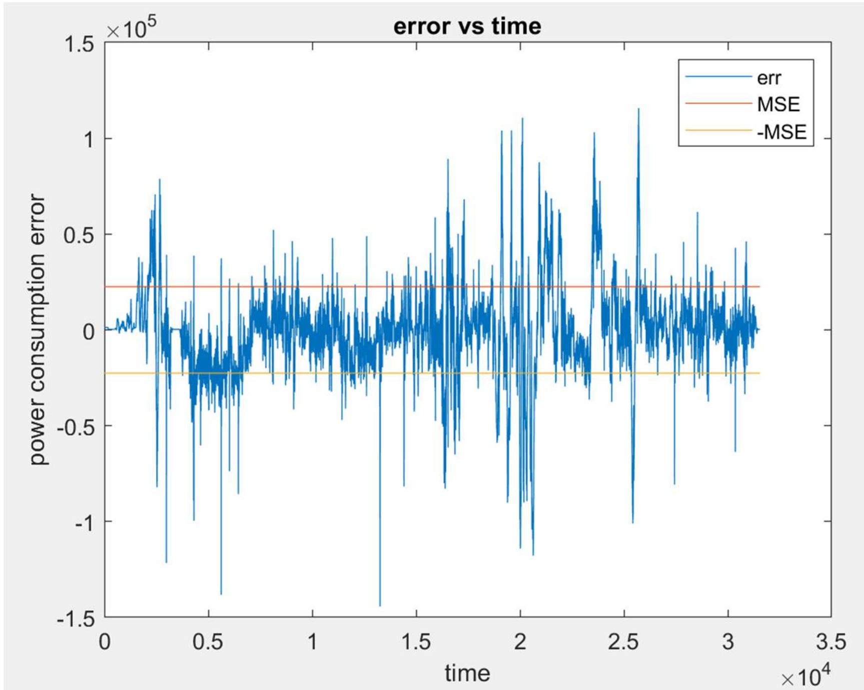


Figure 21: Power Consumption Error vs Time

In this plot, most of the errors remain within the MSE line. The error exceeding MSE only lasts for a short time. Our MSE is around 2.5E+04. This may be considered to be high. However, we need to integrate the power in the final calculations to get energy consumption. As long as the area above zero is close or equal to the negative area, the total energy consumption would be close to the real value. All our test cases above have shown energy consumption within 6%.

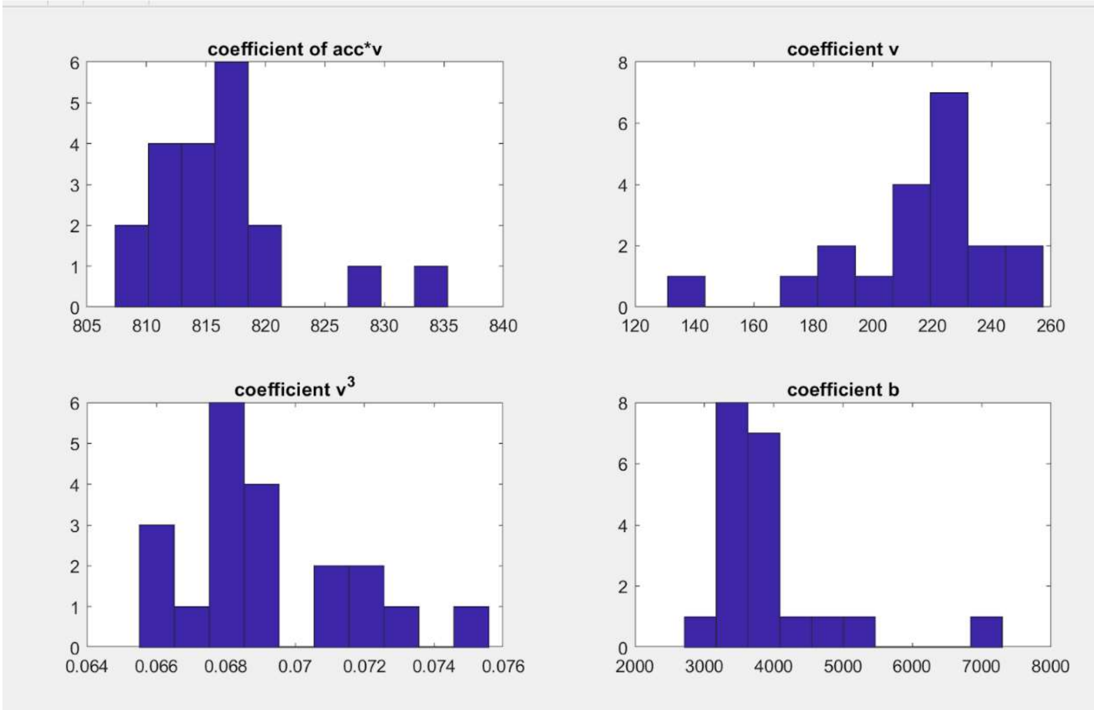


Figure 22: Distribution of Coefficient when Training with Different Sections of Data

This figure has a normal distribution trend. Which can be evidence that the linear regression method should work.
Using stepwise regression, we can obtain *p* values for each variable.

Terms	p value	Significance
Speed	0	Reject null hypothesis. It is significant
Speed^3	3e-52	Reject null hypothesis. It is significant
Acceleration*Speed	0	Reject null hypothesis. It is significant
Time	0.0122	Didn't reject null hypothesis. It is an unrelated term

Table 11: p Value of the Featured Term and an Unrelated Term

This figure has a normal distribution trend. Which can be evidence that the linear regression method should work.

ERE	PD	HW	CITY
1	0.98%	3.46%	5.73%
2	-6.56%	3.02%	6.40%
3	-11.82%	-2.12%	-4.79%
4	1.06%	-4.55%	0.38%
5	-6.60%	-4.09%	4.70%
6	-6.12%	-11.21%	-2.38%
7	-24.53%	-1.91%	6.42%
8	-2.45%	-6.19%	5.15%
9	4.86%	-8.61%	1.35%
10	2.99%	-11.84%	-5.85%
Average	-4.82%	-4.40%	1.71%
Largest Error	-24.53%	-11.84%	-5.85%
Smallest Error	0.98%	1.90%	0.38%

MSE	PD	HW	CITY
1	3.79E+08	1.36E+09	5.33E+08
2	4.28E+08	9.69E+08	5.12E+08
3	4.35E+08	1.13E+09	3.62E+08
4	3.48E+08	1.30E+09	5.70E+08
5	5.20E+08	1.34E+09	4.77E+08
6	5.46E+08	1.77E+09	6.44E+08
7	2.93E+08	1.15E+09	4.63E+08
8	4.92E+08	1.29E+09	5.33E+08
9	3.66E+08	1.50E+09	4.76E+08
10	4.59E+08	1.67E+09	5.45E+08
Average	4.27E+08	1.35E+09	5.11E+08
Min	2.93E+08	9.69E+08	3.62E+08
Max	5.46E+08	1.77E+09	6.44E+08

Table 12: ERE vs Routes(left) MSE vs Routes(right)

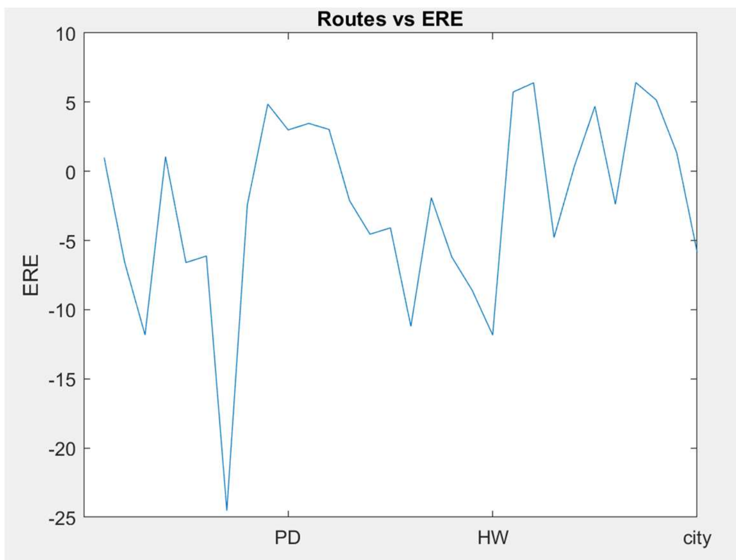


Figure 23: ERE for different routes

Additional verification was made for linear regression model after more data set was given. There is total 30 routes tested, 10 for each drive type. Total energy consumption relative difference was calculated for each route. The mean ERE is -2.5% which means the model is under estimating power consumption overall.

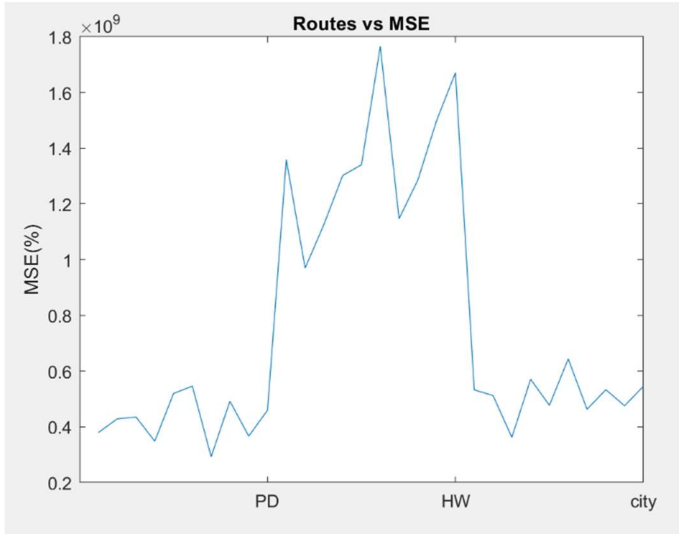


Figure 24: MSE for different routes

The MSE represent how well the predicted data is fitting to the real data. We have observed the highway route has significantly higher MSE. We suspect that is due to error in rolling resistance and aero term being magnified in high speed. The MSE in PD and city route is similar and quite consistent.

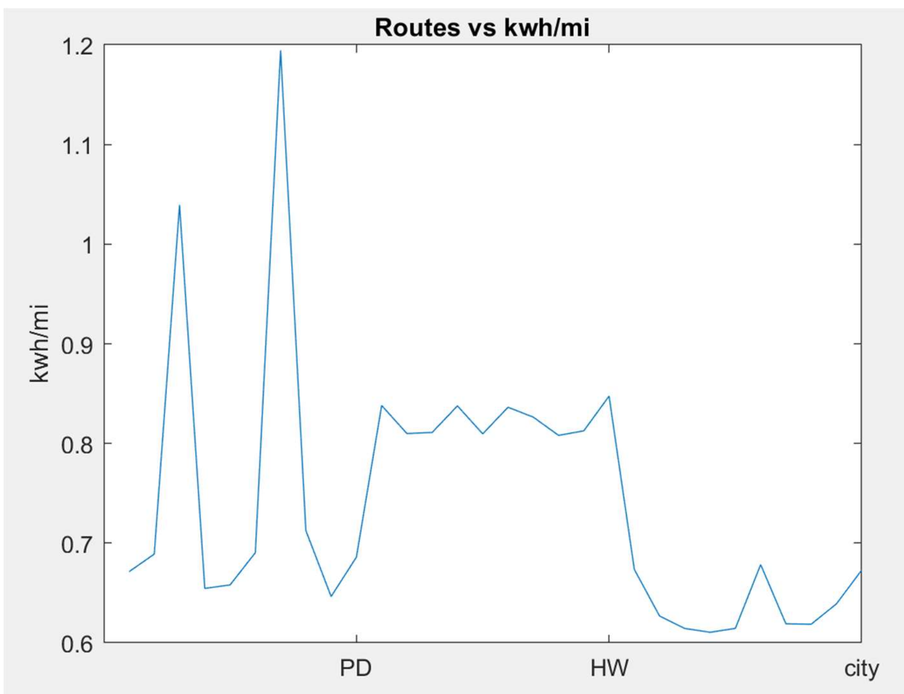


Figure 25: Energy efficiency of vehicle for different routes

2021 Tesla Model 3 Standard Range Plus RWD	0.24 kWh/mi
2021 Audi e-tron	0.43 kWh/mi
PACCAR testing truck	0.74 kWh/mi

kWh/mi represent the energy efficiency of the vehicle. Two other EV is included for reference. It makes sense that a medium duty vehicle consumes about 3 times energy per mile as Tesla Model 3. In the plot,

PD route has two big spikes in kwh/mi measurement. It was due to the total distance traveled and power consumption is low, see figure below. We need large set of data to achieve a reasonable accuracy.
The figure below represents the power consumption and distance traveled for different route. We can see a similar trend between two variables. As mentioned above, two routes in PD have particular low travel distance and causing inaccuracy. There are more discrepancy in highway routes. The car is consuming more kwh of electricity than in other road types. In high speed, more power is used to overcome rolling resistance and aero resistance.

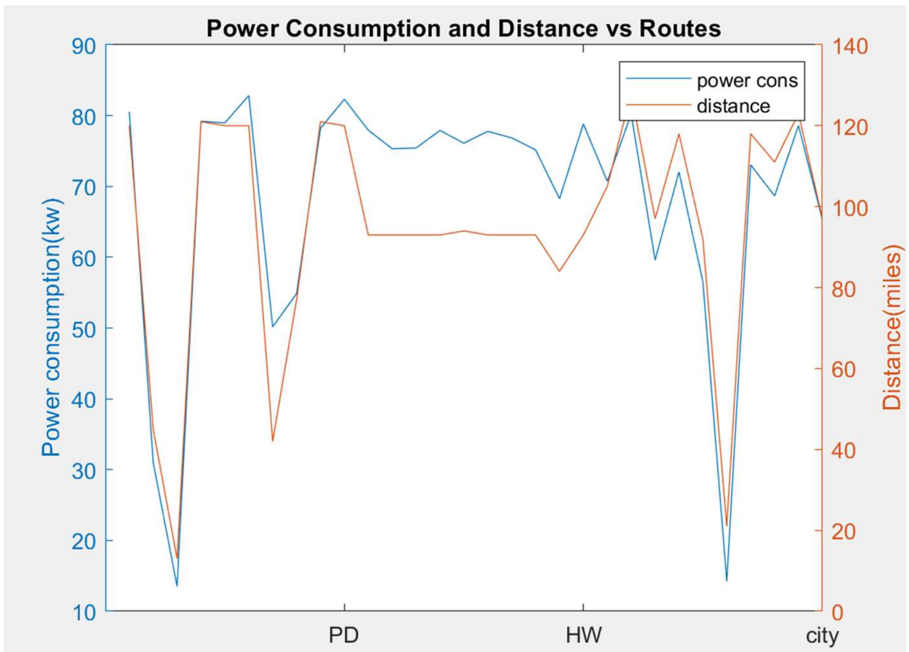


Figure 26: Power Consumption and Distance for different routes

State of charge (SOC) measures the level of charge of an electric battery relative to its capacity. The unit is in percentage.
The figure below shows the SOC vs time for one route. The predicted SOC and real SOC follows a similar trend. However, there are built in algorithms in battery management system measuring the SOC. Just by subtracting energy consumption does not yield the similar result as actual SOC. The final SOC in this chart is largely different.

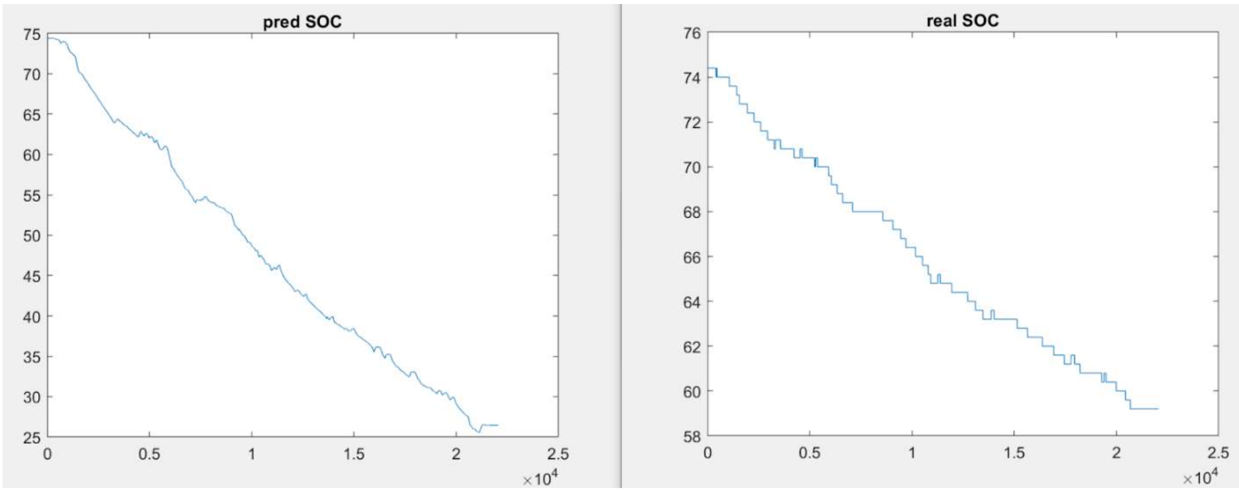


Figure 27: Predicted SOC vs time (left); Real SOC vs time (right)

5 Discussion

5.1 Environment

Running our mathematical model has no direct carbon footprint, but it does have an indirect positive climate impact. The purpose of our model is to help PACCAR engineers maximize range capabilities of electric vehicles and consequently reduce range anxiety. Reducing range anxiety will help ease industry into the adoption of electric trucks. Increasing the use of electric vehicles as a substitute to diesel trucks will reduce the carbon footprint of MD and HD trucks.

5.2 Safety and Liability

Given the performance of our model, we feel that the accuracy of its predictions constitutes safe usage in the component selection for PACCAR's vehicles. If PACCAR engineers deem that the model is inaccurate to a level which may cause design failure or safety concerns, the model may be modified to eliminate these risks or completely discarded. We trust that PACCAR has performance testing and safety precautions in place to eliminate these risks.

5.3 Ethics

Our model was developed with the intention of increasing the adoption of electric trucks. We believe this intention is ethical and benefits both the climate and society by reducing range anxiety. Our chief focus in the scope of this project was creating an accurate model. The accuracy of the model is critical to ensuring the usefulness of the model. Therefore, this the ensures the ability of PACCAR engineers to successfully engineer vehicles using accurate predictions. The accuracy of this model increases the likelihood that these vehicles are both capable and safe.

5.4 Society

Climate change and fossil fuel pollution have many negative impacts on the wellbeing of society, personal health, and environments around the world. With many governments, national, state, and local, setting targets for lower emissions and increased usage of electric vehicles, we hope that our model will prove useful in creating a path to achieving some of these goals. We hope to see industry across the globe begin to adopt electric vehicles in their operations and take actions to reduce their carbon footprints.

5.5 Intellectual Property

Our model was developed solely on PACCAR vehicle data and may be applicable to trucks made by other manufacturers. Given that this tool is largely internal for the foreseeable future we do not see this model being eligible for a patent, but could be filed as a trade secret.

6 Conclusion

6.1 Summary

The three regression models are all useful models for predicting range capabilities of EV's, with the lasso regression producing the least error. The models are all deemed satisfactory, but we observed that the models have higher error in driving scenarios with less frequent regenerative breaking such as highway driving. This higher error still does not exceed the level at which the model would be rated unsatisfactory but is still a significant difference that should be noted when using the model.

6.2 Recommendations and Next Steps

The next steps are all towards improving the accuracy of the model with regards to factors that we were unable to address in the duration of this project.

- Develop a battery model based on battery characteristics to calculate state of charge. Any input parameters regarding the battery were provided by the manufacturer. It would be advantageous to calculate the battery performance with a separate model to ensure greater accuracy.
- Integrate current models with Global Positioning System (GPS), and incline angle (IMU) will make the model more accurate. The model could include data from Google Maps to predict variations in performance over incline and decline grades over specified routes.
- Consider adding meteorological variables to the model to predict performance in varied climates. Factors such as temperature, air pressure, and wind speed could alter the performance of components within the vehicles and it would be advantageous to account for these factors.

Acknowledgements

Special thanks to faculty advisors Krithika Manohar and Ashis Banerjee, as well as our industry partners Nick Hertlein, Maarten Meijer, and Raeef Barsoum for their support and invaluable guidance.

References

[1] “Alternative fueling station locator,” *Alternative Fuels Data Center: Alternative Fueling Station Locator*. [Online]. Available: https://afdc.energy.gov/stations/#/analyze?country=US&fuel=ELEC&ev_levels=all&access=public&access=private. [Accessed: 18-Mar-2022].

[2] “Electric vehicle charging station locations,” *Alternative Fuels Data Center: Electric Vehicle Charging Station Locations*. [Online]. Available: https://afdc.energy.gov/fuels/electricity_locations.html#/find/nearest?fuel=ELEC. [Accessed: 18-Mar-2022].

[3] H. Rahimi-Eichi and M.-Y. Chow, “Big-Data Framework for Electric Vehicle Range Estimation,” *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*, 2014.

[4] “Rolling resistance measurements at the MNROAD facility,” Minnesota Department of Transportation, Mar-2012. [Online]. Available: https://www.dot.state.mn.us/mnroad/projects/HMA_Surf_Chara/pdfs/201207.pdf. [Accessed: 19-Mar-2022].

[5] “3.2.3.3.6. Sklearn.ensemble.gradientboostingregressor¶,” *scikit*. [Online]. Available: https://scikit-learn.org/0.15/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html. [Accessed: 08-Jun-2022].

Appendices

A Project Retrospection

A.1 Engineering Design Process

Because many of us had relatively little experience with machine learning algorithms, our first step was learning as much as we could. When we approached the process of developing our mathematical model, we decided to use a trial-and-error approach. We focused on a single regression type to begin with and observed what results were produced. We then tried several other regression types to produce different results and compared which of them yielded the most accurate and versatile results. This approach yielded several satisfactory models.

A.2 Teamwork and Project Management

For the duration of this project, we met regularly with our faculty and industry advisors once a week to discuss which tasks were completed and what next steps would be. We met one to two times a week to work together as team and coordinate on tasks. Outside of these sessions we worked individually as needed to complete our tasks. We communicated as a group via Slack and Discord. We generally made decisions at the advice of our advisors and reevaluated the product of those decisions after. Our advisors informed us if our work had been largely successful or if perhaps, we might alter our approach for better results.

A.3 Individual Contributions and Responsibilities

Name	Contribution
Peter Ma	**I was responsible for purposes, problem statement, section 4 of our test results and validations, and part of section 3. In addition, I have written linear regression, ridge regression, lasso regression, Discrete Wavelet Transform and matlab to csv conversion.**
Ruizhe Zhao	**3.2-3.6, 4.1, 4.4 and last paragraph of 4.3 Matlab file: mat_to_csv, linear_reg, stepwise39, tenfoldtrain. Gradient boosted decision tree. Testing of 30 routes for linear regression**
Anthony Rudasics	**I was responsible for 1.3 significance, 2 problem statement, 3.1 existing solutions, 3.7 selection, as well as formatting the report. I have written code for data prepossessing in the lasso regression. I communicated with capstone department faculty and designed the slide deck and poster visuals.**
Chia Agrawal	**I was responsible for background, stakeholders, customer requirements, executive summary, references, indexing the figures and tables, and final editing.**

Table 13: Team Contribution

B Concept evaluation iteration

Initially we intended to develop both a battery model in addition to the physics-based model which we did produce. Ultimately the battery model was cut from the scope of our project due to time constraints. We do however recommend the development of this battery model as a next step for this project.