

Fuel Consumption for Double-Stack Intermodal Trains

Multi Linear Regression model and results

Written by

Miquel Onofre Sastre Arbós

Supervised by

Dr. John Doucette

Prepared for:

Industrial Engineering thesis

Submitted on:

May 2015



Table of Contents

Abstract.....	1
1. Introduction to railway transportation and energy consumption.....	3
1.1. Current situation of freight transportation.....	3
1.2. Intermodal railway transportation	3
1.2.1. Intermodal containers.....	3
1.2.2. Railway transportation business.....	4
1.2.3. North American railway network	5
1.2.4. Freight trains	6
1.3. Energy consumption	7
1.3.1. Importance of energy-efficiency	7
1.3.2. Energy consumption explanation.....	8
1.3.3. Resistance force factors.....	9
2. Study of fuel burn prediction for intermodal trains.....	13
2.1. Introduction.....	13
2.1.1. Goal of this project.....	13
2.1.2. Data.....	13
2.1.3. Aerodynamics study	14
2.2. Multi Linear Regression (MLR) models	17
2.2.1. Introduction MLR model.....	17
2.2.2. First model.....	18
2.2.3. Improved MLR model.....	23
2.2.4. Last model, bootstrap.....	27
3. Summary	31
4. Acknowledgments.....	31
5. List of References:.....	32

Abstract

This report paper offers the reader an outline of the development of a MLR model to predict fuel consumption for double-stack intermodal trains in Canada. Results suggested the necessity to classify the data in a non-finished project and design categories for various factors. Furthermore, the conclusion extracted is the requirement to compose small-scale models for specific conditions of track and train in order to obtain the results for an entire railroad section.

Whereas the lack of success on the original goal of this research, which was to predict fuel consumption for intermodal trains in any random configuration under any condition of railroad track, the scope of the project was shortened to a very specific piece of section of the Canadian railway network for trains that travel full power. This was as a result of complications disclosed during the development of the project and troubles discerning the vast data

Keywords: Intermodal transportation, energy consumption, freight trains, fuel burn prediction, MLR analysis



1. Introduction to railway transportation and energy consumption

1.1. Current situation of freight transportation

The freight transportation business moves billions of dollars in cargo every year; the main reason is the increasing necessity of shipping goods and products from certain locations in the globe to other locations. According to the United States Department of Transportation, the nominal amount of imports and exports in the US in January 2015 was higher than US\$35 billion [1]. Part of this trend is the increase of transshipments, which are shipments of merchandise from a country of origin to a country of ultimate destination through an intermediary country. This has become the most dynamic aspect of container shipping in recent years. Increasing transshipment and consequently growth of the transportation industry is because of the introduction of much larger vessels, a consequent progress of the international relations between countries and the constant improvement of the railway and road infrastructure [2].

1.2. Intermodal railway transportation

1.2.1. Intermodal containers

At the end of the 18th century, England started to use containers to ship coal through the canals and then ship it by railway to different locations in the country. However, it was not until the 1950s that the United States Department of Defense designed a new standardized steel intermodal container that began to revolutionize freight transportation. Two decades after, the International Organization for Standardization (ISO) issued standards for intermodal containers [3]. ISO defined the standard measurements for intermodal containers by TEU, twenty-foot equivalent unit. For example, a 20-foot container is equivalent to 1 TEU while a 40-foot container is equivalent to a 2 TEU, as shown in Table 1.1.

Table 1.1: TEU equivalences to feet and meters [4]

Length	Width	Height	Volume	TEU
20 ft (6.1 m)	8 ft (2.44 m)	8 ft 6 in (2.59 m)	1,360 cu ft (38.5 m ³)	1
40 ft (12.2 m)	8 ft (2.44 m)	8 ft 6 in (2.59 m)	2,720 cu ft (77 m ³)	2
45 ft (13.7 m)	8 ft (2.44 m)	8 ft 6 in (2.59 m)	3,060 cu ft (86.6 m ³)	2.25
48 ft (14.6 m)	8 ft (2.44 m)	8 ft 6 in (2.59 m)	3,264 cu ft (92.4 m ³)	2.4
53 ft (16.2 m)	8 ft (2.44 m)	8 ft 6 in (2.59 m)	3,604 cu ft (102.1 m ³)	2.65

Thereafter, intermodal cargos have also revolutionized the supply chain logistics industry, reducing shipping costs significantly. Statistics show that global TEU capacity forecasts a rise of almost 20% from 2013 to 2016. In 2013, the world capacity was 17 million TEU and the prospective capacity for 2016 is 20 million TEU [5].

Intermodal transportation has become the methodology used to transport ISO containers using the combination of multiple methods of transportation (ground, sea and air). Intermodal containers are very convenient when talking about economy of scale. It has resulted in a lot of savings in terms of cost and time for transportation companies. Hence, standardized containers facilitate sorting, classifying and shipping the cargo from the origin to the destination [3].

1.2.2. Railway transportation business

According to the American Association of Railroads (AAR), the strong demand for rail service demonstrates the freight rail industry is continuously growing in North America. Freight rail is playing a central role in positive economy trend by providing cost-effective transportation goods. In 2015, the US government and private companies are predicted to invest US\$29 billion to maintain and modernize the rail network and equipment in the continent [6]. Figure 1.1 argues a positive correlation between increasing net income and spending on the American railway over a period of 10 years, that profitability enables the increasing spending on infrastructure and rail operations [6].

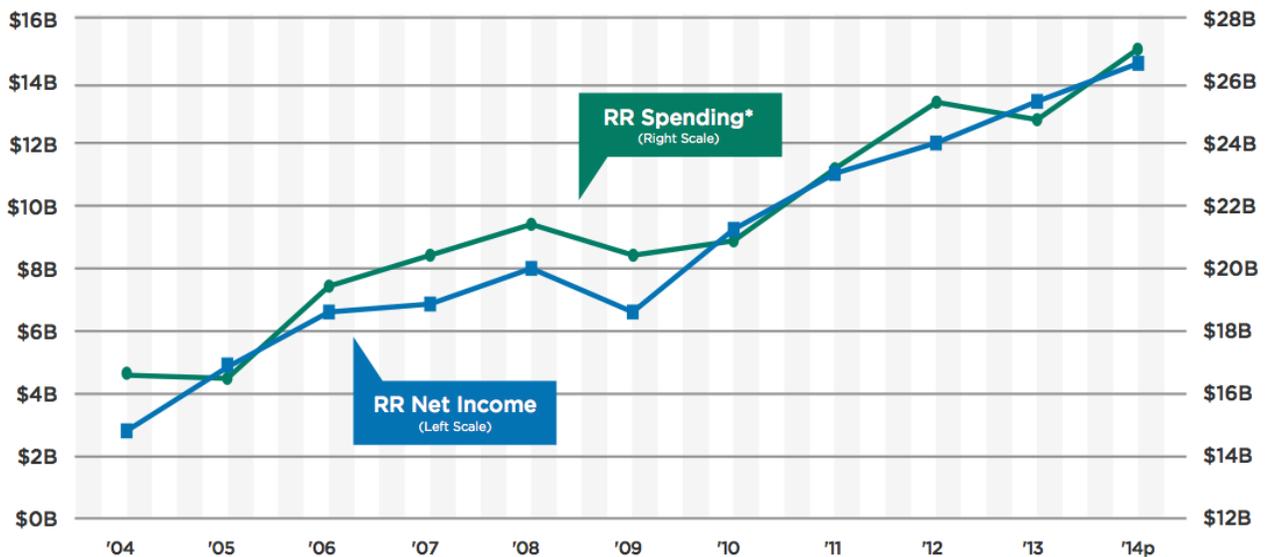


Figure 1.1: Graph of the increasing benefits of the US railway industry, from AAR [6]

Intermodal “containerization” of cargo has actively influenced the acceleration of railway transportation. Using double-stack technology, a freight train of a given length can carry approximately twice as many containers, clearly reducing costs per container. Use of standard containers with same width and height, as shown in Table 1.1, facilitates operating with them. ISO containers are easily stored, sorted and classified in the yard before being loaded on the train to be shipped. Once classified, according to the preference of the shipment, containers are placed and secured on the well cars using a crane. This simple and quick process to load or unload ISO containers on the train yields a lot of savings in costs. Transshipments by rail

The rail industry is an indicator of the health of the economy. Everything from technology products, to agricultural goods to energy products, railroads are in a continuous to movement, so freight volume rises every year. The European Rail Freight Association (ERFA) encouraged acceleration towards expanding to new markets in their 2014 annual event [7]. Worldwide business production and consumer demand have not stopped growing.

1.2.3. North American railway network

With over than 21,000 miles of track, Canada has one of the largest railway networks in the world. Transport Canada regulations, standards and programs work to make the railway system safe, secure, accessible, competitive and more environmentally responsible. According to Canadian National Railway, the North American railway network is able to cover over 75% of the US population and all the major markets in Canada [8]. This large coverage of North America demonstrates the importance of railway freight transportation for Canada.

Moreover, according to AAR, unlike other modes of transportation, North American railway operates over infrastructure built mostly by private companies. As shown in Figure 1.1, since 2004, freight railroads haven’t stopped increasing the spending (including an estimated US\$29 billion in 2015) to meet the demand requirements for rail service and network improvements. However, that huge spending is balanced because of the sustaining profitability of railroad industry [6].

Furthermore, even during the economic recession, North America’s freight railroads continued investing in the rail network and infrastructure. These private investments help to sustain many jobs and ensure the industry can afford the growing demand. Figure 1.2 plots the cumulative investment in infrastructure and equipment is US\$575 billion in private funding since 1980 in the United States [6].



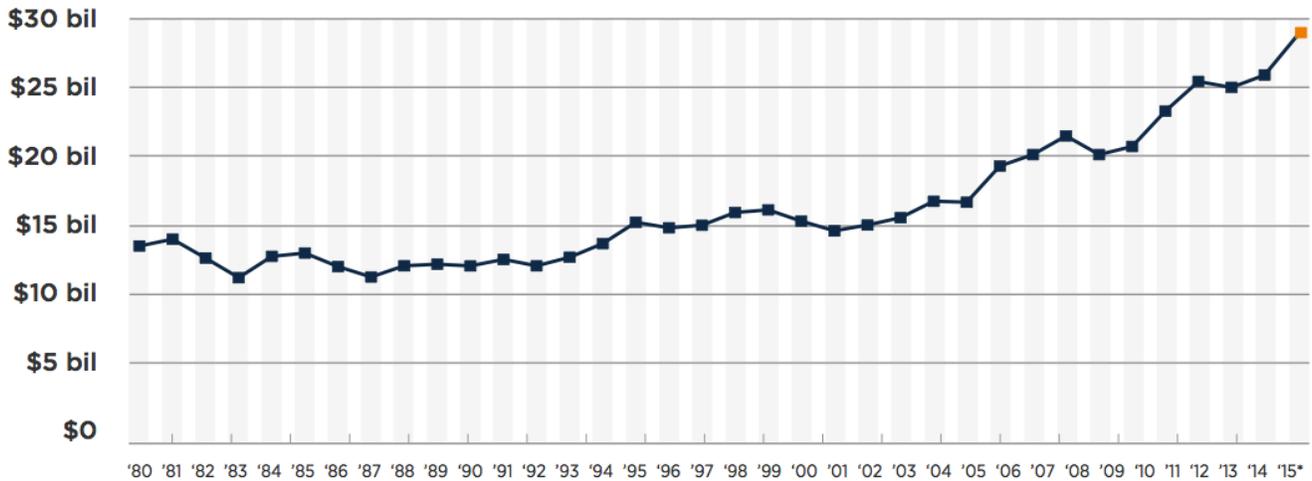


Figure 1.2: Graph of the investment in rail infrastructure in North America, from AAR [6]

1.2.4. Freight trains

In general, freight trains are very efficient for economy of scale. However, train transportation is subjected to lower flexibility than the truck transportation because of the railroad [9], [10]. Double-stack intermodal trains offer efficiency and give importance in freight transportation. Using this methodology clearly reduces costs per container because a freight train of a given length can carry approximately twice as many containers. Intermodal traffic has quadrupled over the past 25 years, with over 11 million shipments per year in North America. Moreover, the flexibility of the intermodal containers also makes them valuable; depending on the kind of cargo that needs to be shipped, ISO containers can carry a wide range of consumer goods, such as clothing, appliances, housewares, electronics and so on, and industrial and agricultural products too. Figure 1.3 shows a two 48 ft. (2.4 TEU) containers stacked on a well car from the railcar pooling company TTX [11].



Figure 1.3: Double-stack intermodal well car [11]

1.3. Energy consumption

1.3.1. Importance of energy-efficiency

Energy consumption is a big concern for railway companies to study, not only because of the huge economic cost, but also because of the exhaust pollution. The 2014 edition of the *Railway Handbook* presented by the International Energy Agency (IEA) revealed that 27.6% of the global energy use was by the transportation sector with 2.2% of this energy was consumed by rail, a 0.6% of the world’s energy [12]. Although there was a 130% increase in rail transshipments since 1975 and a 53% increase in the actual CO₂ emissions between 1990 and 2011 by the transport sector, railway specific energy consumption decreased by 50% between 1975 and 2011 (Figure 1.4) and CO₂ specific emissions dropped by 40% for the same period (Figure 1.5), thanks to the technological improvements in the trains and infrastructure [12]. The importance to railway sector in North America motivated a huge investment on research to lower costs, to improve train efficiencies and to build better infrastructure, which yielded an improvement in energy consumption and CO₂ emissions.

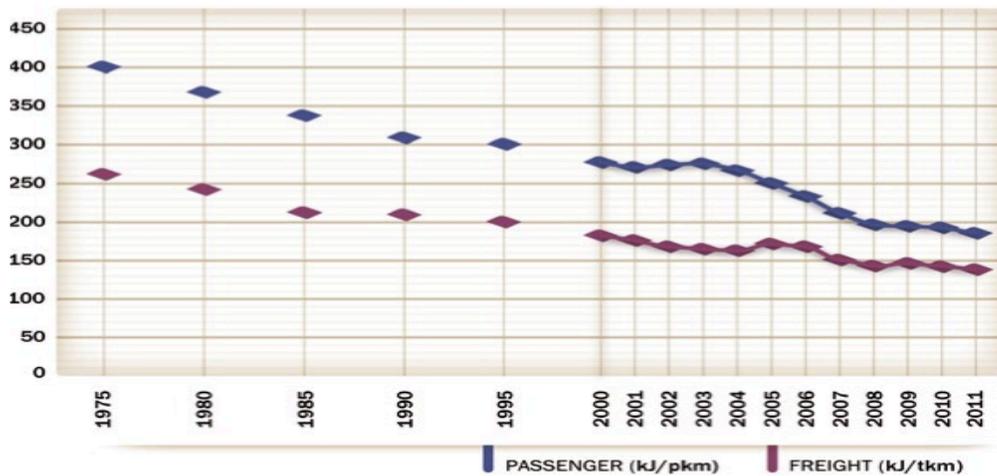


Figure 1.4: World railway specific energy consumption, 1975 – 2011 [12]

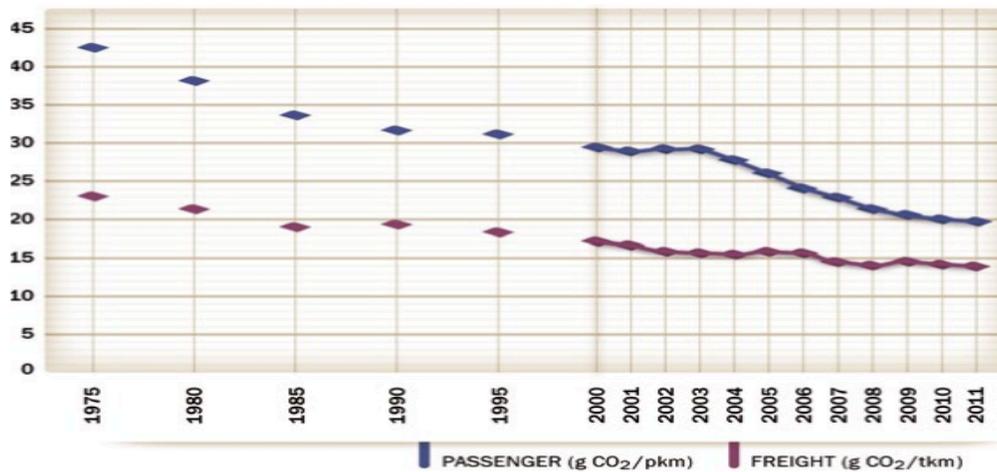


Figure 1.5: World railway specific CO₂ emissions, 1975 – 2011 [12]

1.3.2. Energy consumption explanation

Exploring the factors that affect how much fuel is burnt by the locomotives of an intermodal train, we found that energy consumption is strongly related to the resistance force of the train. The concept we try to capture to predict fuel consumption manifests from the combination of the following three equations [13], [14], [15]:

$$V_{fuel} = E * HR_{fuel} \quad (\text{Eq. 1.1})$$

$$E = \frac{W}{Efficiency} \quad (\text{Eq. 1.2})$$

$$W = \int F ds \quad (\text{Eq. 1.3})$$

Equation 1.1 explains that the volume of fuel burnt (V_{fuel}) is directly proportional to the total energy the locomotive burns (E) and the heat rate of the fuel (HR_{fuel}). In North America, the most common fuel used for freight trains is diesel. However, the environmental problems it causes and the incremental cost of the oil is a good reason to do research on different alternatives.

Equation 1.2 is the energy efficiency equation; it reveals that the energy output divided by the energy input (or energy consumed) gives the efficiency of the system. This efficiency is derived from different internal resistances of the engines, such as bearing friction, windage in motors and or efficiency of generators and cylinders [15].

Equation 1.3 is the work definition equation from thermodynamics: the integral of force (F) and distance (s) defines the work that locomotives produce (W). Then the resultant force is the traction force applied to haul the train subtracting the resistance forces that are applied to the train [13], [14], [16].

$$V_{fuel} = f(F_{traction}, F_{resistance}, F_{internal}, \dots) \quad (\text{Eq. 1.4})$$

Therefore, combining equations 1.1, 1.2 and 1.3 the concept we try to capture for our fuel prediction model is presented in equation 1.4. This equation 1.4 reveals that the volume of fuel burnt (V_{fuel}) is function of traction forces, resistance forces, internal forces and inter-alia. Hence, equation 1.4 is a key statement for the development of the prediction model because these three equations try to prove the idea that volume of fuel burnt is directly related to resistance forces and other internal factors.

1.3.3. Resistance force factors

As proven before, energy consumption is strongly related to the force applied to the trains. It was therefore relevant to include and consider resistance for factors in this study. The main factors that affect to the resistance of the train can be split in different groups. Those factors gathered are the following from [15], [17] and [18]:

- **Internal resistance of the locomotives**

Internal resistance of the locomotives is always present during train movements, it contains the efficiency of the engines, the windage in motors and generators and is influenced by the electric equipment, lighting, heating system and inter-alia. Although energy losses are inevitable, the constant development of new technologies increases the efficiency of the trains.

- **Train resistance components**

Train resistance is mainly defined by three components, Davis equation explains the influence of train aerodynamics [15], [17], [19]:

$$R = A + B * V + C * V^2 \quad (\text{Eq. 1.5})$$

Component A includes mechanical resistances. It is defined by the weight and is independent to speed. Therefore, it varies with axle load and involves bearing friction (rolling resistance, track resistance and journal resistance). Secondly, component B is defined from the weight and the length of the train. It varies directly with speed and includes primarily flange friction and effects of sway and oscillation. At low speeds, components A and B have low influence on the resistance force. Finally, even though aerodynamics is way more important for high-speed trains, it is still a very relevant factor for freight trains. Component C is defined by the cross-sectional area, weight and length of the train. As shown in Figure 1.6, component A is independent of speed but component C presents very high influence when trains travelling at higher speeds [15].

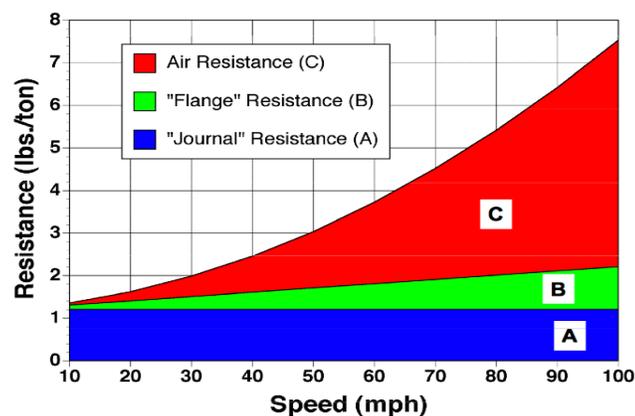


Figure 1.6: Components A, B and C in Davis equation for conventional freight trains [19]

There are many empirical models derived from Davis equation that explain the resistance components:

$$R_u = 1.3 + \frac{29}{w} + b * V + \frac{C * A * V^2}{w * n} \quad (\text{Eq. 1.6})$$

$$R_u = 0.6 + \frac{20}{w} + 0.01 * V + \frac{K_1 * V^2}{w * n} \quad (\text{Eq. 1.7})$$

$$F_{res} = K_2 * W + K_3 * W * V + (n - 1) * A * V^2 \quad (\text{Eq. 1.8})$$

Equations 1.6 and 1.7 are two different variants of the resistance. R_u is extracted from an American textbook and is expressed in [lb/ton], where w is the weight per well-car, n is the number of axles and b and K_1 are constant coefficients that depend on the kind of train is analyzed [15]. However, F_{res} from Equation 1.8 is the metric version expressed in [N], where K_2 and K_3 are coefficients that depend on the type kind of train analyzed, W is the total weight of the train, n is the number of axles and A is the cross-sectional area of the train [17].

As it can be observed, Equation 1.6 coefficients don't match with 1.4 and 1.5. Equation 1.6 uses weight in the components A and B of the Davis equation but Equations 1.4 and 1.5 use inverse of the weight. However, from these equations and others found in different research papers [13] [15], [17] we extracted which factors have influence in resistance force, some of them are weight, number of axles, cross-sectional area and the kind of train.

- **Infrastructure**

Infrastructure has an important significance on the freight trains. Minor grades, along with distance and curvature, affect to potential and kinetic energy losses on the energy consumption; and on the other hand, major grades determine the number of locomotives needed to haul the train. Moreover axles have inevitable friction losses because dynamics, with three components that contribute them: rolling resistance, journal resistance and track resistance [20].

Although the effect of minor grades may seem not important, Figure 1.7 sketches the effect of gravity in trains ascending, for example, a 500 ton well car rising on 1% grade, has around 5 ton of resistance force because of gravity. Extrapolating this to an entire train, a lot of energy is consumed when trains travel uphill.

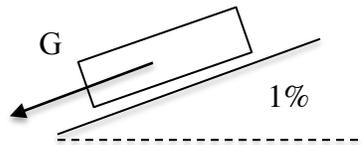


Figure 1.7: Effect of the gravity on a well car in a grade of 1%

Minor grades are frequently attributed to rise and fall. This rise and fall added to the erosion on rails and trains, contributes to increase on energy consumption of the trains. Figure 1.8 gives an outline of a train travelling on a railroad with several rise and fall, where G indicates the direction of the force applied on the train because of the gravity [20]. For example, if train is travelling east (towards the right), gravity favors movement in sections BC and ED; however, gravity opposes movement in sections AB and CD.

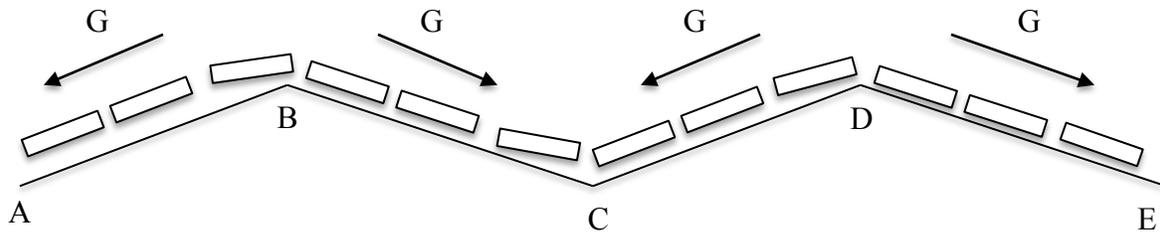


Figure 1.8: Train on a track with two rise and fall

Rise and fall gradients can be classified into three groups. Group A affects to rise and fall that produce a slight variation in speed, therefore trains don't need to brake or vary throttle. However, a long chain of group A rise and fall gradients might affect to the running time of the trains. Group B contains rise and fall on which trains don't need braking when descending the grade but trains require a modest throttle adjustment in going over the crest. This group will affect in a minor increase of the energy consumption by the engines. Group C contains grades that outcome with a marked rise in energy expenses [20].

2. Study of fuel burn prediction for intermodal trains

2.1. Introduction

2.1.1. Goal of this project

The preliminary goal for this project was to develop a model able to predict fuel consumption for double-stack intermodal trains under any configuration of the track obtaining a very small margin of error. The idea was simple, as shown in diagram of Figure 2.1, given a certain configuration of an intermodal train (known weight, length, number of axles and inter-alia), the specifications of the track where the freight train will travel through (grade, curve and inter-alia) and other external factors (such as temperature, altitude, among others) be able to predict the fuel burnt by the locomotive engines.

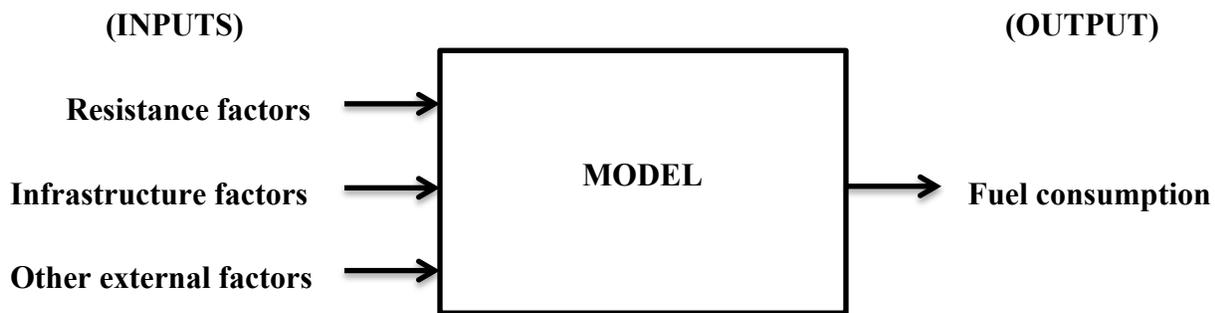


Figure 2.1: Diagram of the model

During the progress of the project we realized we wouldn't have to develop a big model with fix coefficients that covers all kinds of trains and sections. Instead of that, we may need to compose small simple models for specific conditions of road and specific configurations of intermodal trains. That meant we needed to divide the project in different parts to complete with the composition of these specific models in a large one.

2.1.2. Data

We had access to a Canadian railway company's databases to obtain all the information required about double-stack intermodal trains. It allowed us to work with real data to support the conclusions. These databases contain:

- Aerodynamics details. This database includes train configuration details that explain the aerodynamics of the freight trains as for example: containers features as length in TEU and in feet, weight or position in the train; well car specifications as position in the train, length, cargo capacity, among many others.

- Train specifications. Train database covers train specification like ID of the train, date of departure and other schedule details, direction travelling, overall weight, overall length and inter-alia. Information found in this database is useful to classify the trains and sort them.
- Fuel database. This last database contains historical data collected by a system installed in the trains called Wi-tronix [21]. This system captures actual fuel information, location in the network, distance travelled, actual fuel burnt, total power available by locomotive and many other train factors.

In order to obtain the data sets to analyze the intermodal train information, data extracted from these three databases was gathered and filtered in an excel sheet to filter duplicates, missing values and useless data.

2.1.3. Aerodynamics study

During the preliminary stage of this project, the study of Davis equation (Eq. 1.5) revealed the importance of aerodynamics in double-stack intermodal trains fuel consumption. However, including the concept of aerodynamics in our fuel burnt prediction model was a hard part to figure out. In view of intuition, we thought we could include drag force as an input in our prediction model. In fluid dynamics, as presented in equation 2.1, drag force (F_d) depends on the reference area (A), the relative speed between the object and the fluid (V), fluid properties (ρ) and the drag coefficient (C_d), this one is not constant but varies with the shape, position and size of the body and also with the state and viscosity of the fluid.

$$F_d = \frac{1}{2} * C_d * \rho * V^2 * A \quad (\text{Eq. 2.1})$$

Aerodynamic details contained in the company's database suggested we might easily calculate the actual drag force affecting the train to include it in the prediction model. In order to figure out the drag coefficient, our first idea was to measure the gap distances between two containers to calculate drag force afterwards. Although database includes well car details and container details but not gap distances between two containers, it seemed to be a simple calculation. We developed equation 2.2 to measure gap distances between two stack containers. This gap distance is calculated subtracting the length of the container (d_{cont}^c) to the overall exterior length of the well car (d_{ext}^c) from Figure 2.2. Nonetheless, the complexity of gap distance calculation increases when evaluation an entire intermodal train because not all well cars have two containers stack. Hence, we needed to figure out a different method to calculate drag force.

$$d_{gap} = \frac{1}{2} \{ (d_{ext}^{c1} - d_{int}^{c1}) + (d_{int}^{c1} - d_{cont}^{c1}) \} + \frac{1}{2} \{ (d_{ext}^{c2} - d_{int}^{c2}) + (d_{int}^{c2} - d_{cont}^{c2}) \} =$$

$$d_{gap} = \frac{1}{2} (d_{ext}^{c1} - d_{cont}^{c1}) + \frac{1}{2} (d_{ext}^{c2} - d_{cont}^{c2}) \quad (\text{Eq. 2.2})$$

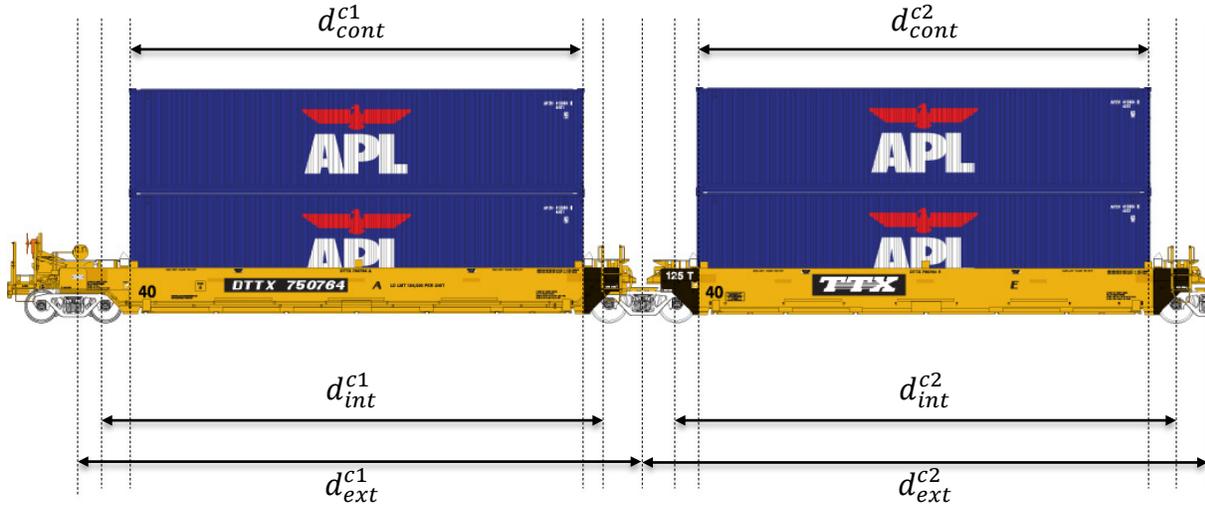


Figure 2.2: Double-stack intermodal well car lengths

In parallel to this gap length calculation study, it came up the idea of using computation fluid dynamics (CFD) software to simulate double-stack intermodal train conditions and determine drag coefficients [22]. Developing CFD simulations would result with very precise values for drag coefficients for each well car and container configuration. However, we realized that working on CFD simulations would take very long time, exceeding the longevity of this fuel burn prediction project. Moreover, there was no need to obtain that high precision for drag coefficients since the goal of this project is obtain fuel burn prediction with a certain range of tolerances.

Thereafter these preliminary aerodynamic studies, we still needed to figure out a simple approach to aerodynamics. Gap distance calculation and CFD simulation were too complex to include aerodynamics effect in our prediction model. In spite of these two options, our last idea was including a new easy and countable concept: leading edge. When two containers from different well cars are very close each other, the gap distance in between is very narrow and the drag force produced for that gap is almost negligible, as it happens in high-speed trains, where cars are very close and aerodynamic losses are lower. Thus, as shown in figure 2.3, a leading edge is produced when the gap in between two containers is significantly wide. A leading edge occurs when there is an empty car slot or two empty slots.



Figure 2.3: Two different types of leading edges

As the database revealed where each container is allocated in the intermodal train, the accounting of the number of leading edge was a very easy task. As a conclusion, we decided using “number of leading edges” as the variable that would contribute to the drag forces because we expected it would explain train aerodynamics effect in our prediction model.

2.2. Multi Linear Regression (MLR) models

2.2.1. Introduction MLR model

As the fuel database contains actual fuel burnt information, the actual volume of fuel is known. Thus, we could use these actual fuel burnt values (Y_{actual}) to build a MLR model. The methodology is very simple: predict the fuel burnt in US gallons for a set of trains and then compare those values with the actual fuel burnt to obtain the error of the prediction. The MLR consists of taking a group of random predictive variables (X_i) to find the relationship between a predicted variable ($Y_{predicted}$) using a linear equation, as shown in equation 2.3, where B_i 's are the coefficients of the linear equation. The main goal of this methodology is to find the predicted value, duty cycle in US gallons, with the minimum margin of error possible when comparing Y_{actual} and $Y_{predicted}$ [23].

$$Y_{predicted} = B_1 * X_1 + B_2 * X_2 + \dots + B_n * X_n + Error \quad (Eq. 2.3)$$

$$Y = B * X \quad (Eq. 2.4)$$

A linear equation is used to calculate the $Y_{predicted}$ from equation 2.3 with simple algebra, using matrix multiplication as shown in equation 2.4. Therefore, as B is unknown, B_i values are calculated multiplying Y_{actual} by the inverse of X (equation 2.5). Finally the $Y_{predicted}$ is calculated by multiplying B matrix by X, as shown in equation 2.6 [23].

$$B = Y_{actual} * X^{-1} \quad (Eq. 2.5)$$

$$Y_{predicted} = B * X \quad (Eq. 2.6)$$

Once found the predicted variable, the comparison of the predicted with the actual values of fuel consumption is easily calculated by subtracting $Y_{predicted}$ to Y_{actual} , as shown in equation 2.7. Then, we can calculate the percent of error to scale it.

$$Error = Y_{actual} - Y_{predicted} \quad (Eq. 2.7)$$

$$Percent. Error = \frac{Y_{actual} - Y_{predicted}}{Y_{actual}} * 100 \quad (Eq. 2.8)$$

2.2.2. First model

2.2.2.1. Data and variables

First of all we decided to focus the study on a very specific piece of section of the Canadian railway network. The reason of this simplification was to attenuate the energy losses, minimizing the influence of the grades and the curvature in the fuel consumption. Thus, we decided to study a section almost flat and completely straight. Fuel information for those trains was captured in a five-mile section west from Saskatoon, by Wi-tronix and train specifications were taken from the database. This first model consisted of a MLR analysis for a set of 221 intermodal trains travelling through that section in September 2014.

Table 2.1 shows the variables of interest selected for the first model. We subtracted the variables that our intuition told us could have influence on fuel consumption, as the number of leading edges, weight, length and inter-alia. However, there was no direct possibility to obtain that data from the database and include them in the MLR model although our intuition and research background indicated we were forgetting important variables as the speed, the speed squared and the grades of the track. Hence, we ran the first model with the factors we could easily subtract and observe the results.

Table 2.1: Table of predictor variables used for the first model

i	Variable	Symbol	Units
1	Net distance travelled	D	Miles
2	Actual power per ton	HP	HP/ton
3	Total power available	THP	HP
4	Length of the train	L	Feet
5	Number of cars loaded	#CL	-
6	Number of cars empty	#CE	-
7	Total train weight	W	Ton
8	Total moving net time	T	Hours
9	Number of leading edges	#LE	-

2.2.2.2. Results

First model presented very vague and imprecise results. As shown in figure 2.4, the MLR model presented senseless outliers; for example, there were predicted values with percent of error higher than 4,000% (equation 2.8 MLR). Moreover, figure 2.5 plots fuel burnt predicted values against actual values but range of the predicted values differs with the range of actual fuel burnt values. Horizontal axis, predicted fuel consumption, presents values around the hundreds however, the vertical axis, actual fuel burnt, are around twenty. Problems understanding the databases and the many missing values that Wi-tronix presents sometimes resulted with a mean

of percent error for the entire set of trains of 228% and a standard deviation of the error of 997%, providing a fuel burn prediction model with disastrous results.

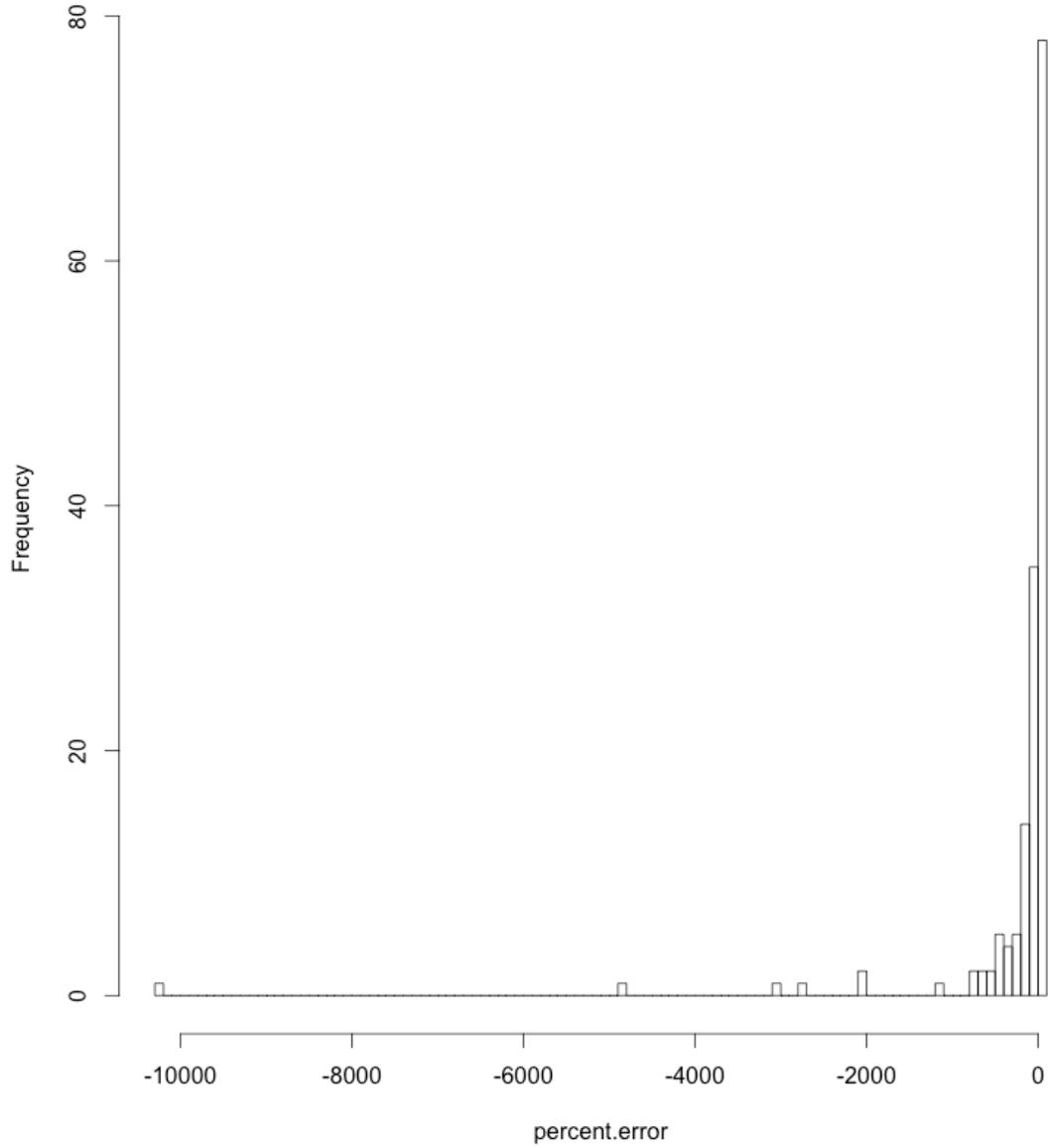


Figure 2.4: Histogram of percent of error first model

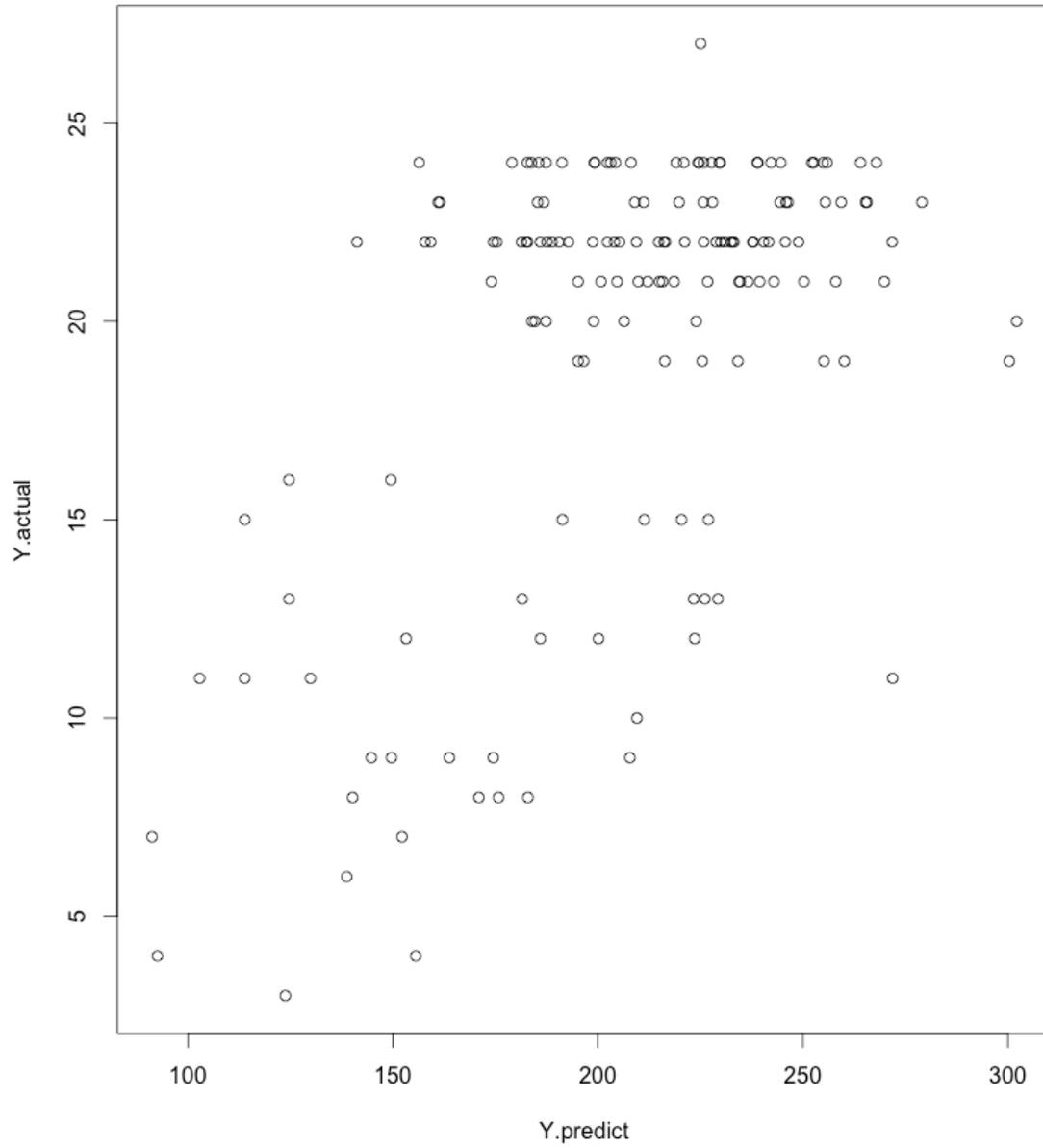


Figure 2.5: Plot of predicted duty cycle vs. actual duty cycle first model

2.2.2.3. *Conclusions for the first model*

The inaccuracy and imperfections on the execution of this very first model exposed the mistakes of the first analysis. We realized MLR doesn't contemplate the interactions between variables, we needed better and refined data and we needed to figure out the way to apply the physics background.

2.2.2.4. *Analysis after conclusions*

Before starting the second stage of the project, there was the necessity to investigate how to apply Physics on the MLR model, introduce the interactions of variables and define new filters for the data. The lack of substantial data sets and difficulties to study a piece of the road section with ideal specifications for our model brought us to a new different section between Winnipeg and Saskatoon for the subsequence stages of the project. This new section allowed collecting more accurate data without missing values and less outlier because it has longer flat and straight sections of track and Wi-tronix system presents better information.

In addition, during the transition between the first and the improved model, we realized a pattern our data was following. Figure 2.6 shows that pattern, for various datasets we analyzed there were peaks for duty cycle every 20 US gallons (20, 40, 60...). This pattern said that a third of the trains in that set burnt 20 US gallons of fuel, a third burnt 40 US gallons and a third burnt 60 US gallons for trains travelling through the same section and the same distance. After analyzing that data we realized that peaks were trains travelling full power and each locomotive in that track conditions travelling full power was burning around 20 US gallons. Therefore, a train using one locomotive would burn 20 US gallons, a train with two would burn 40 US gallons, and so on. An explanation of this pattern can be that trains travelling in a flat and straight section usually will travel with maximum power unless there is any incidence. This pattern was another reason that explained why the first model had many understandable values in the results. Hence, we decided to reconsider the study for following stages to analyze only trains travelling at full power and classify data by locomotives instead of by train.

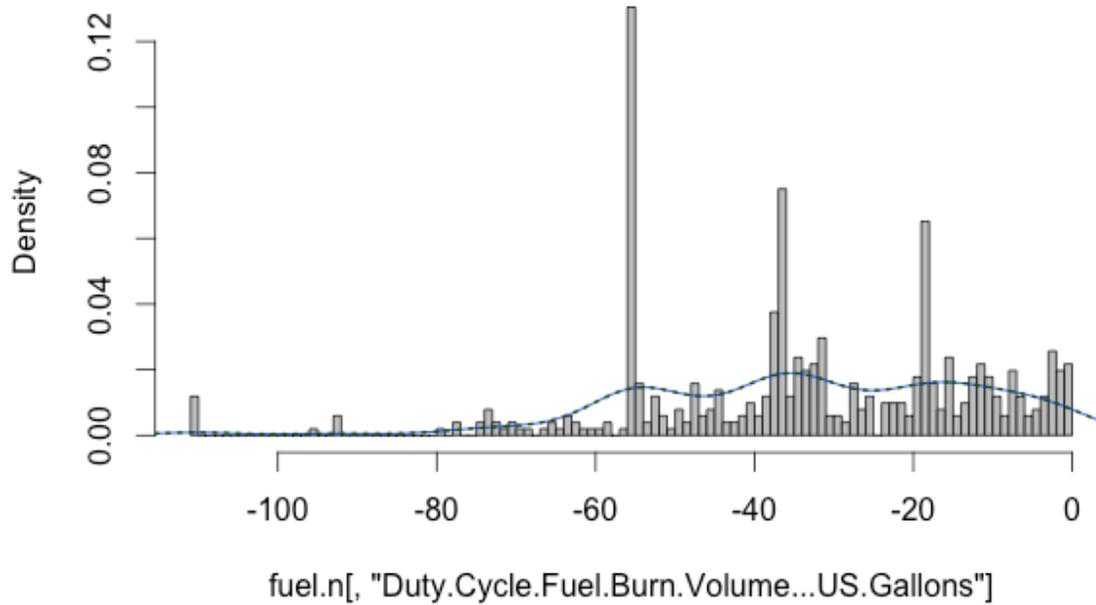


Figure 2.6: Histogram of fuel consumption for a data set we studied

Finally, first model revealed the effect of number of leading edges factor was useless. Although this lack of relevance suggested number of leading edges was not an accurate factor that contributes to aerodynamics, it didn't mean that aerodynamics has no effect in fuel consumption for double-stack intermodal trains. We thought we could remove number of leading edges factor from our MLR model but we needed to figure out an alternative. Railroad Engineering from A. Hay suggested equations used to calculate resistance forces (equations 1.6 and 1.7) tend to overstate the resultant drag force applied to the trains. Therefore, we could use certain coefficients to adjust the equations used to calculate the resistance forces. As shown in equation 2.9 and table 2.2, resistance forces coefficients (K_i) depend on the kind of freight trains, this adjustment factor can varies from 0.85 for conventional trains, where the aerodynamic efficiency is higher to 1.90 for freight trains traveling with several empty and uncovered auto racks, producing poor aerodynamic efficiency [15]

$$R_{adj} = K * R_u \quad (\text{Eq. 2.9})$$

Table 2.2: Classification of K coefficients

K_i	Classification	Value
1	Freight trains with pre-1950 equipment	1.00
2	Conventional trains with post-1950 cars	0.85
3	Freight trains with containers on flatcars	0.95
4	Freight trains with trailer-on-flatcar and hopper cars	1.05
5	Freight trains with empty, covered auto racks	1.20
6	Freight trains with loaded auto racks	1.30
7	Freight trains with empty and uncovered auto racks	1.90

Hence, we realized we wouldn't need to use a specific factor as an input to the MLR model to contribute to aerodynamics drag forces, but we could use factors such as speed, length or total number of cars instead.

2.2.3. Improved MLR model

2.2.3.1. Data and variables

This improved model was conducted with data filtered by locomotive instead of by train. The section used this time, a six-mile section in between Winnipeg and Saskatoon, has longer straight and flat portions of track meaning that Wi-tronix is able collect several times the same train, yielding a set data for 4210 trains travelling westbound. A histogram of the fuel burnt, picture 2.7, revealed that 80% of the trains travelled at max throttle. Once filtering this data set we obtained 2341 locomotives of trains to analyze.

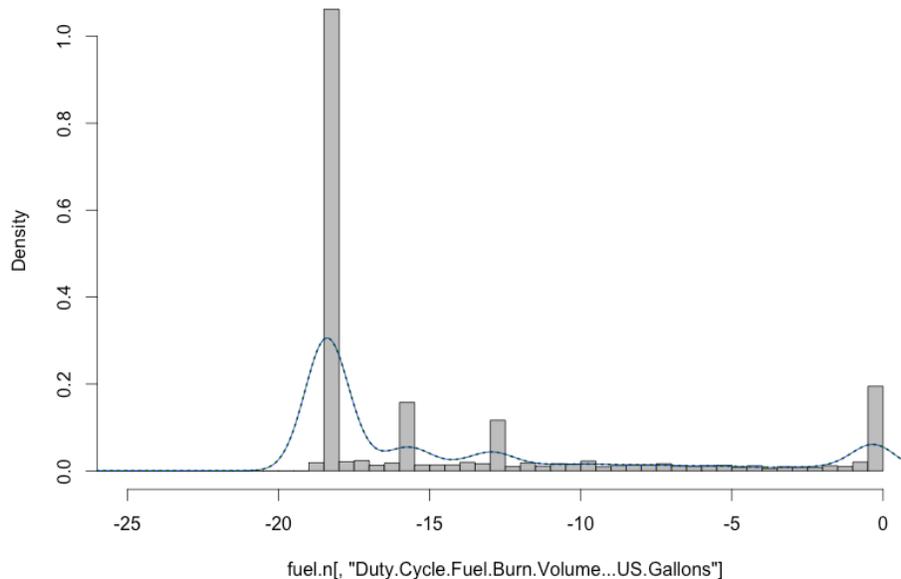


Figure 2.7: Histogram of DC for improved model before filtering by trains travelling full power

Furthermore, this new model contemplated fifteen predictor variables (equation. 2.3). The reason we included the interactions as speed divided by weight (V/W), speed multiplied by weight ($V*W$) and many other combination of interactions was because papers we read revealed people use different models to calculate force resistance [15], [17]. Accordingly, we decided to include several different combinations possible to let our model reveal the correct solution. Thus, as MLR doesn't consider interactions, we decided to build new columns from the variables in the database. From the database, we subtracted net distance the train travelled, the weight and length of the entire train, number of cars and some other variables not used in the model that allowed do derive them and create new variables. The five first columns of Table 2.3 present the new variables used to run the MLR.

Table 2.3: Table of predictor variables used for the improved MLR model

	Variable	Symbol	Units	Origin	B_i
1	Net distance travelled	D	Miles	Database	-16.10
2	Inverse of weight	1/W	(1/ton)	Derived	-6.07
3	Total weight of the train	W	Ton	Database	-1.27
4	Length of the train	L	Feet	Database	-0.01
5	Average speed squared	V^2	(mph) ²	Derived	8.49
6	Total number of cars divided by weight	#C/W	1/ton	Derived	0.17
7	Average speed divided by weight	V/W	mph/ton	Derived	2.55
8	Average speed multiplied by weight	V*W	mph*ton	Derived	5.40
9	Weight divided by grade of the track	W/G	Ton	Derived	0.003
10	Total number of cars multiplied by weight	#C*W	Ton	Derived	0.37
11	Average Speed	V	mph	Derived	15.03
12	Total number of cars	#C	-	Database	-0.45
13	Average acceleration	A	m/h ²	Derived	-7.85
14	Average speed squared divided by weight	V^2/W	(mph) ² /ton	Derived	-2.74
15	Average speed squared multiplied by weight	V^2*W	(mph) ² *ton	Derived	-5.90

2.2.3.2. Results for the improved model

After applying several filters on the data, we ran a model with interesting results. As shown in figure 2.8 this MLR model with 15 predicting variables presented improved results. There were no outliers in the histogram of the margin of error obtained (equation 2.8 MLR), the mean was almost 0% and the standard deviation of these results was 0.5%. Moreover, this dataset presented

In addition, B_i coefficients obtained, shown in last column from table 2.3, indicated there is a major influence by distance ($B_1 = -16.10$) and speed ($B_5 = 8.49$) in fuel consumption and factors as inverse of weight, acceleration and speed multiplied by weight have minor influence. However, we discovered that other factors as weight divided by grade or length of the train have null influence on the model.

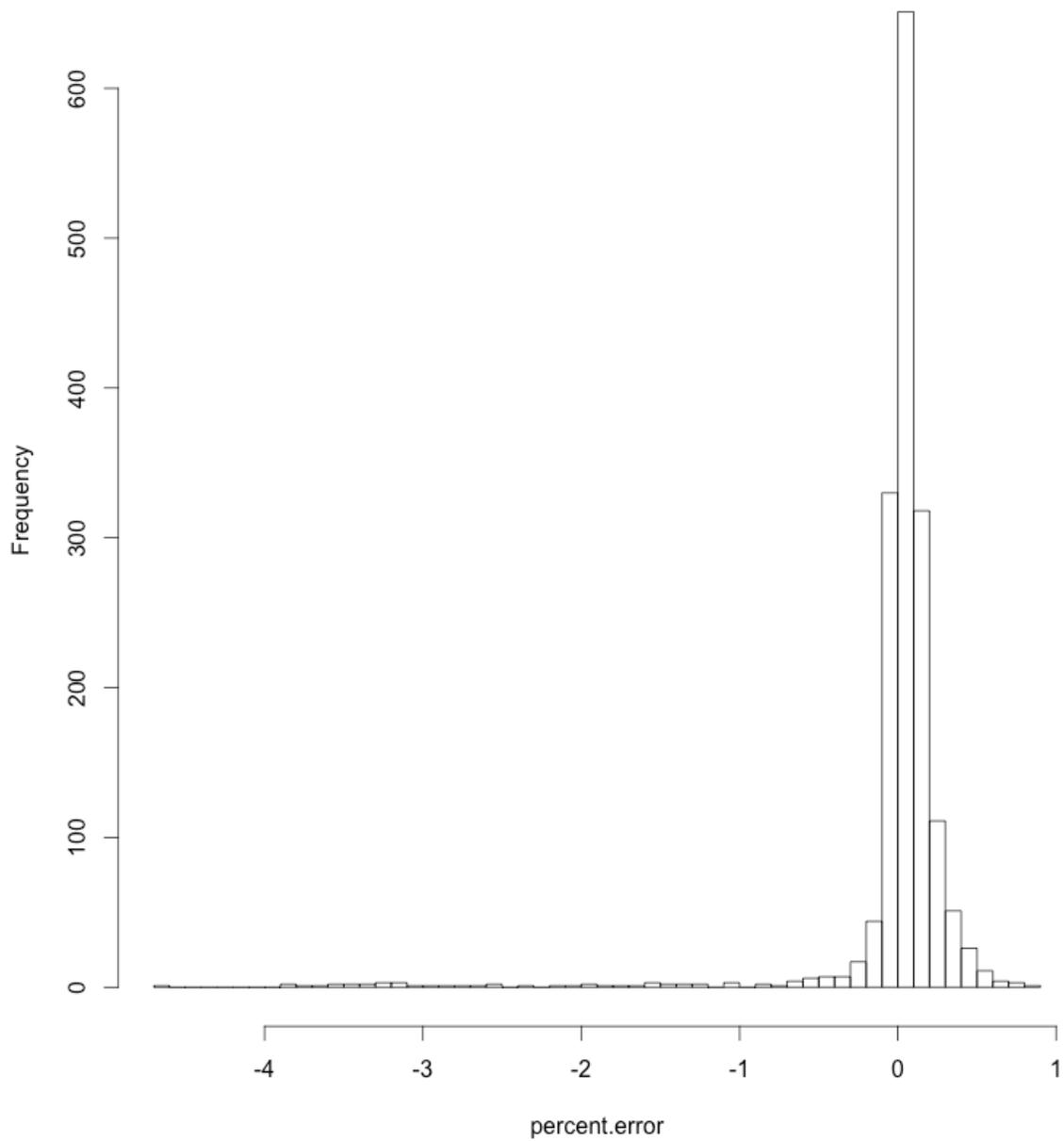


Figure 2.8: Histogram of percent of error improved model

2.2.3.3. Conclusions for the improved model

Although the application of Physics provided improved results, we thought this yielded influence of the distance should be negligible because the data was selected for a five-mile piece of track where almost all intermodal trains travelled for five miles. However, that high influence of distance and speed explained there is high correlation between fuel consumption and trains travelling at full power between 40 and 60 mph.

Additionally, from the Railroad Engineering textbook we extracted figure 2.9 to go one step further understanding this high influence of the speed [15]. Figure 2.9 plots the train resistance curves obtained experimentally by professors E. C. Schmidt and J. K. Tuthill on the Illinois Central Railroad. They plotted individual curves for car weights in increments of five tons for a range of speed from 10 mph to 70 mph. As a result it came up this exponential shape that matches with equations 1.6 and 1.7 previously described. However, the conclusion extracted here, together with the improved model, is the necessity of classifying our MLR models by weight and speed because as shown in figure 2.9, the resistance applied to a 75-ton car travelling at 10 mph is six times lower than the resistance applied to a 20-ton car travelling at 70 mph.

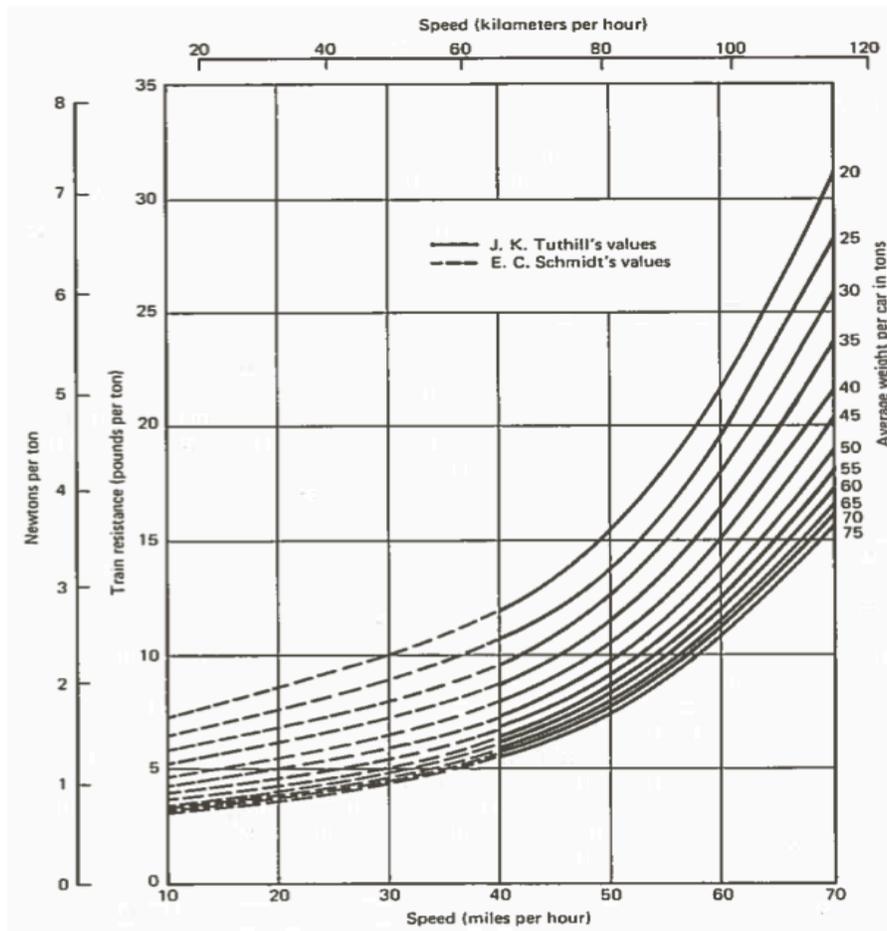


Figure 2.9: High speed freight train resistance [15]

2.2.4. Last model, bootstrap

A bootstrap was introduced in order to prove the previous results were reliable. We needed test different data sets and evaluate the possibility to add a new classify, classify data by speed. Hence, it arose the idea of introducing a bootstrap to the MLR model. The bootstrap selects randomly 70% of the set, predicts the fuel consumption for that 70% and compares the results with the actual fuel burnt (as the previous models). Then, the bootstrap repeats this process 10,000 times for random sets of 70% of the parent set. At the resolution, the bootstrap yields with an interval of confidence for the “ B_i ” coefficients (for the standard confidence 95%).

2.2.4.1. Results for the last model

Table 2.4 and Table 2.5 show the intervals of confidence for the B_i coefficients. Set 1 in Table 2.4 are the results obtained for the same set of 2341 locomotives travelling westbound used in the previous model. Most remarkable results of this test revealed the mean for the B_i coefficients is similar than the B_i coefficients in the previous model (Table 2.2). However, when filtering average speed in different ranges, some B_i coefficients suffer big variation (see mean of B_1 for set 3 and set 4 in Table 2.4, it is 2.5 times bigger).

Moreover, range of B_i for distance, average speed and average acceleration is very wide (highlighted rows in Table 2.4 and Table 2.5). The range for these B_i coefficients can vary from 21 to 89 units, which means that these predictive variables of this MLR model are very imprecise.

Table 2.4: Intervals of confidence for B_i coefficients (1)

		Filtered only by duty cycle (set 1)				Average speed ≥ 25 mph (set 2)			
i	Symbol	Low limit	High limit	Mean	Range	Low limit	High limit	Mean	Range
1	D	-43.7000	9.8800	-16.91	54	-36.2500	6.7500	-14.75	43
2	1/W	-6.3970	-5.7100	-6.05	1	-6.4410	-5.6460	-6.04	1
3	W	-1.3250	-1.2230	-1.27	0	-1.3210	-1.2100	-1.27	0
4	L	-0.0267	0.0028	-0.01	0	-0.0226	0.0082	-0.01	0
5	V^2	7.8490	9.1300	8.49	1	7.8620	9.5170	8.69	2
6	#C/W	0.0500	0.2895	0.17	0	0.0421	0.3464	0.19	0
7	V/W	2.3580	2.7240	2.54	0	2.3280	2.7780	2.55	0
8	V*W	5.0950	5.7240	5.41	1	5.0610	5.8470	5.45	1
9	W/G	-0.0003	0.0047	0.00	0	0.0005	0.0062	0.00	0
10	#C*W	0.1155	0.6115	0.36	0	0.02880	0.6425	0.34	1
11	V	-28.2500	61.0300	16.39	89	-22.8600	48.6800	12.91	72
12	#C	-0.8102	-0.1022	-0.46	1	-0.9600	-0.0134	-0.49	1
13	A	-27.0320	10.2000	-8.42	37	-21.9900	7.8000	-7.09	30
14	V^2/W	-2.9960	-2.4790	-2.74	1	-3.0960	-2.4610	-2.78	1
15	V^2*W	-6.3670	-5.4460	-5.91	1	-6.6130	-5.4310	-6.02	1

Table 2.5: Intervals of confidence for B_i coefficients (2)

i	Symbol	Average speed between 40 and 50 (set 3)				Average speed between 45 and 55 (set 4)			
		Low limit	High limit	Mean	Range	Low limit	High limit	Mean	Range
1	D	-46.8800	4.7500	-21.07	52	-23.5500	6.6600	-8.45	30
2	1/W	-8.3250	-5.1870	-6.76	3	-7.3770	-4.1470	-5.76	3
3	W	-1.4520	-0.9830	-1.22	0	-1.8830	-1.2870	-1.59	1
4	L	-0.0086	0.0150	0.00	0	-0.0076	0.0225	0.01	0
5	V ²	7.7040	10.1510	8.93	2	6.3890	7.7700	7.08	1
6	#C/W	-0.0585	0.2050	0.07	0	-0.0887	0.1664	0.04	0
7	V/W	2.1910	3.7080	2.95	2	1.6250	2.9050	2.27	1
8	V*W	4.3010	6.4610	5.38	2	4.9960	7.4790	6.24	2
9	W/G	-0.0028	0.0024	0.00	0	0.0000	0.0093	0.00	0
10	#C*W	-0.2962	0.3467	0.03	1	-0.2676	0.2764	0.00	1
11	V	-20.5500	66.1200	22.79	87	-21.8970	28.3850	3.24	50
12	#C	-0.5587	0.2339	-0.16	1	-0.5374	0.2712	-0.13	1
13	A	-29.1700	6.9800	-11.10	36	-13.3070	7.6230	-2.84	21
14	V ² /W	-4.2160	-2.3000	-3.26	2	-2.9090	-1.5750	-2.24	1
15	V ² *W	-7.2010	-4.6870	-5.94	3	-7.3810	-4.8710	-6.13	3

2.2.4.2. *Conclusions for the last model*

In conclusion, results still can't be relied. Although these four different sets maintain that some variables have influence in the fuel consumption for the intermodal trains (distance, speed, speed multiplied by weight and inter-alia) and other variables have null or almost null influence (length of the train, weight divided by grade and inter-alia), the variability of the means of B_i coefficients had wide ranges for certain variables and the lack of testing different sets of locomotives in several locations didn't allow us to take for granted that this results are conclusive.

In conclusion, the huge ranges for B_i coefficients that resulted for the variables distance, average speed and average acceleration revealed the imprecision of this new model. It suggests two ideas for future work in this project: deeper study of the bootstrap intervals of confidence and design categories to classify data. The meaning of a wide interval of confidence for a variable is still unclear at this stage of the project. As average speed and average acceleration data are derived from known data, distance and time, we still need to go one step further on this analysis. On the other hand, during the conclusions for the improved model came up the idea of classifying by speed and weight. At the end of this stage we are still wondering and analyzing the possible categories for future stages of work. One early idea we got was coming back to leading edges and using it as a factor for the design of the categories. As presented in table 2.6, models could be divided in eight different categories, where each of the three factors has two levels. Speed, weight and number of leading edges could be classified in two levels. For example, high level (+) of the factor Speed could be filtering the data set for trains travelling faster than 50 mph and the low level (-) trains travelling slower than 50 mph, (+) for factor Weight could be trains heavier than 10,000 ton and (+) for Number of leading edges could be a certain amount of leading edges

Table 2.6: Matrix design for categories

Category	Speed	Weight	# Leading Edges
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

3. Summary

By virtue of the lack of success on the original goal of this research, which was to predict fuel consumption for intermodal trains in any random configuration under any condition of railroad track, the scope of the project was shortened to a very specific piece of section of the Canadian railway network for trains that travel full power. This was as a result of complications disclosed during the development of the project and troubles discerning the vast data. We spent very long time understanding the data, there was frequent confusing information, data was crowded of outliers and there were countless missing values that made our trials biased and unreliable in terms of Physics. Moreover it was very problematic including the aerodynamics in the MLR model, there was no specific data in the database and research we did didn't revealed any solution.

The conclusion extracted during the development of the project is the requirement to combine small-scale models for specific conditions of track and train in order to obtain the results for an entire railroad section. Furthermore, results suggested the necessity to classify the data and test the model for several sets of locomotives in different sections. Nonetheless, the scope for this project leave this tasks for a future work.

4. Acknowledgments

Dr. J. Doucette
M. Ruhl
Canadian Rail Research Laboratory (CaRRL)
Dr. C. Lange
Dr. M. Secanell



5. List of References:

- [1] United States. Department of Transportation, Bureau of transportation statistics (BTS), *North American trans-border freight data*. Washington, DC: Jan. 2015. [Online]. Available: The North American trans-border freight database, <http://transborder.bts.gov/programs/international/transborder/>.
- [2] Scottish government, “Container transshipment and demand for container terminal capacity in Scotland” (chap. no. 4) Edinburg: Sept. 2004. [Online]. Available: <http://www.gov.scot/Publications/2004/09/19885/42551>.
- [3] “Intermodal freight transport”, *Wikipedia*, Feb. 2015. [Online]. Available in Wikipedia, http://en.wikipedia.org/wiki/Intermodal_freight_transport.
- [4] “Twenty-foot equivalent unit”, *Wikipedia*, Mar. 2015. [Online]. Available in Wikipedia, http://en.wikipedia.org/wiki/Twenty-foot_equivalent_unit.
- [5] Statista, Transportation & logistics, *Forecast for global TEU capacity of container ships from 2013 to 2017*, [statista.com](http://www.statista.com), Jan. 2015. [Online]. Available in the statistics portal, <http://www.statista.com/statistics/198254/forecast-for-global-number-of-containership-teus-from-2011/>.
- [6] American Association of Railroads, “2015 outlook report”, *American Association of Railroads*, p. 2-5. [Online]. Available in AAR website, <https://www.aar.org/Documents/Outlook%202015/2015OutlookReport.pdf>.
- [7] European Rail Freight Association, *News and events*, Mar. 2015. [Online]. Available in ERFA website.
- [8] Canadian National Railway, *Transportation Services*, 2015. [Online]. Available in CN website: <https://www.cn.ca/en/our-business/transportation-services>.
- [9] A.S. Fowkes, C.A. Nash and G. Tweddle, “Investigation the market for intermodal freight technologies”, *Transpn. Res*, vol. 25A, no. 4, pp 161-172, 1991.
- [10] C. Macharis and Y.M. Bontekoning, “Opportunities for OR in intermodal freight transport research”, *European Journal of Operational Research*, vol. 153, is. 2, pp. 400- 416, Mar. 2004 [Online] available in ScienceDirect.

- [11] TTX pooling company, *TTX equipment, Intermodal*, 2015. [Online]. Available: <http://www.ttx.com/AboutTTX/ttx-equipment.aspx#merchandise>.
- [12] International Energy Agency, “Railway handbook 2014: Energy consumption and CO₂ emissions”, *International Energy Agency*, pp. 18, 24, Dec. 2014. [Online] Available at [uic.org](http://www.iea.org)
- [13] Y. Bai et al. “Energy-efficient driving strategy for freight trains based on power consumption analysis”, *Journal of transportation systems engineering and information technology*, vol. 9, is. 3, pp. 1-4, June 2009. [Online]. Available in ScienceDirect database.
- [14] JX. Cheng et al. “Algorithms on optimal driving strategies for train control problem”, *Proceedings of the 3rd World Congress on intelligent control and automation*, vol. 5, pp. 3523 – 3527, Jul. 2000. [Online]. Available IEEEXplore database.
- [15] W. Hay, “Railroad engineering” in *Propulsive resistance*, chap. 6, 2nd ed. United States: Urbana, IL. Wiley-Interscience publication, pp. 69 – 89, June 1982.
- [16] W. Hay, “Railroad engineering” in *Velocity profiles*, chap. 11, 2nd ed. United States: Urbana, IL. Wiley-Interscience publication, pp. 179 – 191, June 1982.
- [17] B. P. Rochard and F. Schmid, “A review of methods to measure and calculate train resistances”, *Proceedings of the institution of mechanical engineers, Part F: Journal of rail and rapid transit*, vol. 214, pp. 185 – 199, Apr. 2000. [Online] Available in SAGE journals.
- [18] Q. Wen et. al, “Energy simulation of hybrid inter-city trains”, *Proceedings of the ICE – Energy*, vol. 160, is. 3, pp. 123 – 131, Aug. 2007. [Online]. Available in ICE virtual library.
- [19] J. Armstrong, “The Railroad, what it is, what it does”, chap. 4, 5th edition 1998. Reprinted in *Railroad engineering (CE 515): notes module 3*. Champaign, IL: University of Illinois at Urbana-Champaign, 2010. [Online]
- [20] W. Hay, “Railroad engineering” in *Problems in grades*, chap. 12, 2nd ed. United States: Urbana, IL. Wiley-Interscience publication, pp. 192 – 206, June 1982.
- [21] Wi-tronix, *Products & Services, Wi-FuelSensor*, Jul. 2015. [Online]. Available: <https://www.wi-tronix.com/WiFuelSensor.aspx>.
- [22] ANSYS, *Fluid Dynamics, CFD Software*, 2015. <http://www.ansys.com/Products/Simulation+Technology/Fluid+Dynamics>.



[23] “Multiple Linear Regression”. Reprinted in (STAT 401): MLR notes. Michigan, USA: University of Michigan. [Online]. Available:
<http://dept.stat.lsa.umich.edu/~kshedden/Courses/Stat401/Notes/401-multreg.pdf>.