Using Real-Time Truck Transportation Information to Predict Customer Rejections and Refrigeration-System Fuel Efficiency in Packaged Salad Distribution

by

Carlos Seminario
Industrial Engineer, Universidad Católica Andrés Bello, 1998
MBA, Business School Sao Paulo, 2004

And

Emmanuel Marks
B.A. Visual Arts, Columbia University, 2002

Submitted to the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Logistics

at the

Massachusetts Institute of Technology

June 2011

Signatures of Authors...
Master of Engineering in Logistics Program, Engineering Systems Division
May 6, 2011

Certified by..
Dr. Edgar Blanco
Executive Director, MIT SCALE Network in Latin America
Thesis Supervisor

Accepted by..
Prof. Yossi Sheffi
Professor, Engineering Systems Division
Professor, Civil and Environmental Engineering Department
Director, Center for Transportation and Logistics
Director, Engineering Systems Division

Using Real-Time Truck Transportation Information to Predict Customer Rejections and Refrigeration-System Fuel Efficiency in Packaged Salad Distribution

By
Carlos Seminario
And
Emmanuel Marks

Submitted to the Engineering Systems Division
on May 6, 2011 in Partial Fulfillment of the
Requirements for the Degree of Master of Engineering in Logistics

ABSTRACT

Companies that operate cold supply chains can greatly benefit from information availability and data generation. The abundance of information now available to cold chain operators and harvested from every echelon of the supply chain, ranging from the procurement process to the sales and customer service processes, provides an opportunity for logistics organizations to monitor and improve their operations. It is increasingly imperative to transform data into meaningful information that creates a competitive advantage for early adopters.

This thesis attempts to determine how to make best use of and effectively interpret the information generated by trailer mounted temperature sensors and geospatial data collection devices during refrigerated transportation of packaged salads. The study covers only the transportation segment from the manufacturer's distribution center to the customer's (grocery retailer) distribution center.

This thesis uses regression analysis in an effort to create a model that effectively uses real-time transportation information to identify the elements that can create a competitive advantage for cold chain operators. The main performance measurements subject to analysis in this thesis are reefer-unit fuel consumption and rejections of salad products at the customer's drop location.

Regression yields a formula that can predict more than 70% reefer fuel consumption. However, with the independent variables available in the data at our disposal, it is not possible to build a model the effectively predicts product rejections.

The findings of this thesis can help operators of transportation cold chains better manage fuel consumption by isolating and improving the independent variables we identified.

Thesis Supervisor: Dr. Edgar Blanco
Title: Research Director, MIT Center for Transportation and Logistics

AKNOWLEDGEMENTS

Firstly, we would like to thank two individuals without whom this endeavor would not have been possible, our wives Kristin and Ximena.

We would like to thank the following individuals for their support, guidance, patience and unfaltering belief in our abilities: Neal Lerner, Josh Merrill, Ted Holcombe, Larry Verdelli, Rick Chandonnet and Cheryl Schondek. Last but not least, we would like to thank our advisor Dr. Edgar Blanco for his valuable suggestions, which greatly improved the quality of this thesis.

For Macarena, Micaela, Milo & Joaquin.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1. INTRODUCTION

Technology for monitoring the supply chain environment, particularly container temperature and reefer fuel efficiency, is more affordable and becoming more commonplace. Real-time cold chain monitoring is readily available and is currently deployed in a variety of industries making or using perishable goods. Monitoring technology is adding a degree of transparency to a once cloudy corner of materials handling. Accurate and timely information sharing has the potential to inform and shape stakeholders' actions and operating processes. However, the standards for interpreting and using this data are relatively immature. In large part this is due to the lack of properly aggregated data and a nascent understanding of what actionable information can be plucked from the ocean of data collected. Once the challenges of data aggregation have been met, a detailed framework for analysis can be undertaken.

Specifically, this thesis seeks to investigate how cold chain information generated in real-time during the transportation process can be interpreted and analyzed in order to reduce load rejections and improve transportation fuel efficiency.

Numerous other tactical areas are candidates for investigation; however, we have chosen to emphasize rejections and fuel efficiency, as we feel that the greatest impact can be realized in these areas. Additional effort is spent on developing operational metrics from information harvested from major data sources. This line of investigation serves to elucidate the additional potential benefits of the information generated from monitoring cold chain truck transportation.

We will focus on the cold chain that serves the commercial salad industry. Within this larger supply chain we focus our research on the specific link in the chain between raw materials processing facilities and customer distribution centers. Three processing facilities are included in the study and these facilities serve upward of one hundred and twenty customers in total. Once packaged, boxed and palletized, salad products are delivered to the diversity of food services and retail grocery customers served by company XYZ.

During the transportation process from salad pack processing centers to customer distribution centers, cold chain data is collected via a refrigerated trailer tracking, monitoring and control system manufactured by Par Logistics. Additional information is drawn from the study sponsor's transportation management system and from an Excel-based product rejection tacking system managed by the company. A diversity of data is harvested by these systems and is discussed in detail. These data sources provide an opportunity to perform regression analyses in an effort to define the key independent variables that affect load rejections and reefer unit fuel consumption.

This thesis seeks to define a model (or rationale) to identify when a rejection is likely. Such a model would allow the industry to more consistently and proactively manage supply chain performance. Furthermore, the model would allow chain managers to take timely corrective action to prevent product damage, financial losses, and respond to specific events in the most cost effective manner.

Additionally, this thesis seeks to identify the key independent variables that affect reefer unit fuel efficiency. Finally, this thesis seeks to define a set of key performance indicators from the data generated by the refrigerated trailer tracking and monitoring system to rate how effectively and efficiently the transportation operation is being managed from the perspective of the variables available in the data.

The results of these three major lines of investigation provide cold chain transportation managers with insights into how to reduce the rate of rejections of products per load, improve the rate of fuel consumption by refrigerated assets, and increase the value that can be extracted from the data generated during the refrigerated transportation process.

2. LITERATURE REVIEW

This literature review seeks to achieve two primary aims: to provide the reader with a strong foundation from which to consider the thesis we present, and to define the logical next steps called for by prior researchers that have lead us to undertake the research at hand.

A general review of current literature covering five key topics will serve as a foundation of knowledge for the readers of this thesis. These five key topic areas are as follows:

a. Purpose of real-time cold chain Information gathering
b. Current state real-time information gathering in refrigerated trucking
c. Industry measures of quality and impacts on rejection
d. Refrigerated transport fuel efficiency
e. Data analysis & modeling
f. Gaps in the literature

2.1 PURPOSES OF REAL-TIME COLD CHAIN INFORMATION GATHERING

Real-time information gathering in fresh fruit and vegetable cold chains is relatively new, and the diversity and abundance of information collected by real-time sensing devices can rapidly escalate to a point of information overload. Hundreds of thousands of lines of information can be captured from a single carrier over a relatively short period of time, sometimes no longer than several days. This abundance of information is both a blessing and a curse. Owners of this new wellspring of transportation information must invest the necessary time and resources to transform raw data into models and subsequently use those models to make meaningful decisions that justify the cost of investing in such technology in the first place.

Further compounding the importance of real-time information during the transportation of fresh fruits and vegetables is the fact that food now travels farther than ever before from producing regions to the end point of consumption (22 Christian Schenk 2010). Additionally, food is different than other cargo in as much as perishability of product and potential health risks to the end user can be affected by out-of-standard transport conditions. A leading thinker in food sciences, Schenk states that "food must be properly refrigerated or food becomes susceptible to bacteria. Just one tainted shipment could cost a company its reputation, its future orders and millions of dollars in lost sales and liability" (22 Christian Schenk 2010). Accordingly, these concerns have made real-time information collection a necessity and one that will increasingly be at the forefront of food tractability, quality control and logistics improvements.

Before delving into the details of real-time monitoring of the cold chain, let us first outline the current definition of cold chain. Sensitech states that "'Cold chain' is the term applied to a supply chain established for products that must be handled under controlled temperature conditions" (9 Sensitech, The Cold Chain Company).

Schenk outlines a categorization of the possible benefits of real-time information in cold chain. He states that Intelligent Reefer Management solutions help carriers in four critical areas. Specifically he isolates Cost Reduction, Customer Satisfaction, Safety and Compliance and Driver Performance as the key areas where potential benefits are most tangible(22 Christian Schenk 2010). We will focus on two of these areas for our discussion of the potential benefit of cold chain monitoring during transportation of fresh fruits and vegetables (Hereafter referred to as FFV). Specifically we will focus on Cost Reduction through reefer fuel efficiency, and in Customer Satisfaction through rejection reduction.

Schenk states that fleets have a significant opportunity to reduce cost by monitoring fuel consumption, temperature and other operational metric via reefer management systems (22 Christian Schenk 2010). Customer service can be improved by enabling transportation managers with the ability to provide clients with real-time information regarding the status of their goods (22 Christian Schenk 2010). Access to such information also improves safety and compliance (22 Christian Schenk 2010). In essence, the correct monitoring of FFV in transit can reduce spoilage, boost fuel efficiency and improve compliance.

2.1.1 COST REDUCTION

With regard to cost reductions, adjusting operating practices based on analyzing and modeling real-time cold chain data can yield significant benefits. In *Making the case for Cold Chain,* David Sterling states that the companies lose billions of dollars a year as a result of product waste (11 David M. Sterling 2010). In his opinion, this waste stems from two areas: loss of product shelf life and the difficulties associated with inventory planning for a product with variable shelf life (11 David M. Sterling 2010).

Sterling goes on to state that the two major drivers for cold chain improvements are economic benefits and improving a firm's ability to meet regulatory requirements (11 David M. Sterling 2010). With regard to cost benefits, Sterling states that "When maximum shelf life is attained, inventory is optimized, product waste is lessened, product return costs (if applicable) are impacted positively, and the cost of customer complaint resolution is decreased"(11 David M. Sterling 2010).

Much of the literature on the subject of cost savings in the cold chain deal with the pharmaceutical industry, as the value of pharmaceutical products transported via cold chain are orders of magnitude greater than the value of corresponding volumes of FFV. In terms of pharmaceutical applications of real-time monitoring, Schlenker and Chapman state that as a response to ever increasing fuel costs, proactive organizations are looking for new ways to reduce cold chain operating costs without negatively affecting the quality of pharmaceutical cold chains (7 Karl Schlenker, Melissa Chapman 2010). Such attention is increasingly important, as Kevan notes:

> The stakes in food, pharmaceutical, and chemical cold chains are high. The loss of a trailer of food due to improper handling or transport is measured in hundreds of thousands of dollars; a pharmaceutical shipment, in millions. Because of the financial pressure and increasing regulatory demands for better recordkeeping resulting from the Bioterrorism Act, suppliers and logistics service providers are turning to systems that combine RFID and temperature and humidity sensors(10 Kevan, Tom 2005).

Real-time information also allows for corrective action to be taken. Again speaking of the pharmaceutical industry, Kevan states that if manufacturers find that a shipment of products has been exposed to a small amount of out-of range temperature, but not to the extent that the product has been damaged, having the ability to react before the product must be considered a total loss is important (10 Kevan, Tom 2005).

Product that is slightly affected but still usable can be rerouted for use before it incurs any additional damage.

Another factor to be considered with regard to cost is incremental energy costs associated with fluctuating transport temperatures. Sterling states that cold chains often operate ineffectively, which results in fluctuations in temperature: "At vulnerable points of the chain such as at the handoffs between cold chain partners, the temperature may increase beyond acceptable levels" (11 David M. Sterling 2010). Sterling goes on to consider the energy costs associated with these fluctuations and states that "The incremental energy costs to bring the product to proper temperature are significant compared to the cost of maintaining the product at proper temperature with minimal variability" (11 David M. Sterling 2010). He also states that multi-stop routes further exacerbate these fluctuations as a result of the temperature variability inherent in frequent door openings.

With regard to these fluctuations, Sterling states that "in multiple stop distribution made to stores, the product being delivered in the final stops may be especially vulnerable to higher temperatures because of several factors such as repeated opening of doors, mixing chilled product with dry product loading in ways which impede airflow, etc. As this product is introduced to the store coolers, the cooler has to compensate for the introduction of higher temperature product into the cooler environment. Additional energy for each store cooler is necessary to bring all of the products back to the correct temperature again. If the proper temperature were maintained during delivery, the additional energy costs would be mitigated" (11 David M.

Sterling 2010). These routing patterns and the corresponding temperature fluctuations tend to be the norm in the FFV industry.

While cost reductions can be generated through the effective use of real-time information detailing the location and condition of FFV transported in refrigerated trailers, customer satisfaction can also be affected by the use of this information.

2.1.2 CUSTOMER SATISFACTION

Let us first define what "Customer" means throughout this literature review and thesis. The Customer is defined by food services providers and grocers who purchase FFV products for resale to end consumers. Food services vendors may reprocess product or use pre-processed product for resale to multiple end consumer locations. Grocers may operate a single retail location where they take receipt of the product they order, or they may operate multiple retail locations and take delivery of the product they order at central or local distribution center. These distribution centers, in turn, deliver product to store locations; however, we are not concerned with this leg of transportation, as manufacturer´s real-time monitoring does not occur during this segment of travel.

Customer satisfaction is a major factor shaping the future of the FFV industry. Planet Retail states that

> A primary driving force in the global food market is the consumer. Income growth, lifestyle changes brought about by urbanization, and changing family structures have resulted in diet changes among consumers worldwide. Because of either increases in purchasing power or the increased opportunity cost of time required

by preparing food, the demand for higher value and processed food products has expanded globally. Meeting the demands of today's critical consumer is becoming increasingly challenging (12 Planet Retail LTD 2006).

Customer satisfaction is largely determined by the quality of the product purchased by the end consumer. In the FFV industry several factors are at play in determining customer satisfaction. Primarily, specifications play a large role in ensuring and maintaining customer satisfaction (23 Brown, Martyn 2008). If customers demand a clears set of attributes for a given FFV product, and these attributes have been incorporated into that products specifications, customers are much more likely to be satisfied with in-spec products.

As with all agricultural products, seasonality, growth region, and growing conditions all play a large role in determining product quality as measured against accepted specifications. Brown states that although the existing nature of the product at hand cannot be controlled by the cold chain, the quality and effectiveness of cold chain technology on hand does impose a maximum possible quality and shelf life limitation after initial quality has been determined during agricultural production (23 Brown, Martyn 2008). Accordingly, cold chain integrity plays a vital role in preserving existing product quality and in doing so, plays a vital role in ensuring customer satisfaction as defined by product specifications.

Customer satisfaction is extremely important as consumers' brand loyalty is directly affected by the quality of the FFV product they purchase. Sterling states that "An additional major consideration is that of brand equity impact. Consider the situation of consumers who purchase products in which quality has been compromised. They may decide to no longer purchase the product or even the brand because of their experience with the compromised quality of their purchase" (11 David M. Sterling 2010). In essence consumer exposure to poor-quality products can punish a brand twice; not only will a consumer not purchase a product with poor quality, but that consumer may avoid the brand in the future. As a result, companies are increasingly turning to real-time solutions to monitor cold chain performance in an effort to control quality.

2.2 REAL-TIME INFORMATION GATHERING IN REFRIGERATED TRUCKING

The current state of real-time cold chain monitoring is rapidly changing. The recent explosion in sensing technologies and information transfer applications has been slowed only by the recent downturn in the business cycle, but forward looking companies are continuing to investigate and implement progressive monitoring strategies and for good reasons. In his article, *Hot Market – Cool Freight*, William Huffman states that, "Temperature sensitive shipping [ ] is growing at a rate of 15 percent a year" (18 Huffman 2006). Furthermore, Huffman cites the work of the Cool Chain Association stating that, "the group estimates that 30 percent of all perishables – food, pharmaceuticals, artwork and microchips -- are lost in transit" (18 Huffman 2006). Furthermore, in his *Economist* special report, John Parker states that "Rich countries waste about the same amount of food as poor ones, up to half of what is produced, but in quite different ways. Studies in America and Britain find that a quarter of food from shops goes straight into the rubbish bin or is thrown away by shops and restaurants. Top of the list come salads, about half of which are chucked away" (44 Parker 2011).

The projected growth of the perishable goods industry coupled with the real need to improve cold chain transportation warrants additional investment in monitoring. The major trends in the industry monitoring are characterized by two technology shifts. First, temperature monitoring is moving from chemical temperature indicators to electronic temperature indicators. Second, and most important, data from these electronic temperature indicators is being fed into web-based databases in real-time or near real-time for immediate retrieval and analysis.

Chemical temperature indicators are temperature sensitive strips about the size of a stick of gum that change color in response to temperature exposure above of below an established threshold. These strips have been a cost effective tool for monitoring out of range incidents in cold chain transportation; however, they do not record when or for how long an out of range event occurs. Chemical tags only measure that such an event has happened. A recent article in *World Pharmaceutical Frontiers* describes the shift from chemically based temperature indicators to electronic temperature indicators and states that chemical indicators are slowly being replaced by digital monitors as price and acceptance grow in response to the increased need to monitor last mile logistics (24 WorldPharma 2010). The article goes on to highlight the importance of this shift based on several key benefits, specifically, accuracy, validation, alarms and deployments.

The second major improvement in cold chain monitoring in the FFV industry is the advent of real-time cold chain monitoring via web-based interface. In her article, *Fresh and Fresher, How Web Monitoring Keeps Perishables From Perishing Too Soon*, Mary Wagner discusses recent developments in web-based monitoring of FFV cold chains. She states that

> With perishable products arriving at grocery distribution centers and stores from increasingly far-flung points of origin--even overseas--it's more challenging than ever to transport and handle them so as to minimize product deterioration. But now, the Internet is helping grocers and their perishables supply chain to manage the process, making big strides in reducing inventory shrink. Technology that continuously monitors container temperatures travels with shipments from growers and producers to retailers. Feeding the data into a web-based database for sharing and analysis can pinpoint exactly where in transport the container temperatures went above or below agreed-on standards by even as little as a degree (25 Wagner 2005).

These advances also allow for improvements in long-term performance and transportation planning optimization.

With regard to potential visibility into the supply chain with an eye to improvement, Wagner states that, "The web-based database and analytics can reveal, for example, if temperatures for produce delivered to a particular region show greater variance in transport than is the case in other regions. That would allow the grocer to review the trucking vendor used in that region, leading to the discovery, for example, that the vendor's refrigerator fleet is a few years older than average, or that the fleet's condition or maintenance is poor" (25 Wagner 2005). Access to tactical performance information of this nature aids cold chain partners in making strategic planning decisions moving forward.

Additionally, this new wellspring of information facilitates prioritization of necessary channel improvements according to greatest need. In quoting Steve Dirubio from Sensitech, Wagner continues, "The analytics give a picture of the relative performance of participants in cold chain management, region-to-region, supplier to supplier, and compares different times of year. . . . It allows grocers to prioritize improvements. . . . Especially if you are a large grocer spread across many states, attacking something as complicated as your perishables logistics system is an expensive proposition. So it's important to know solving which problem first will give you the biggest bang for your buck"" (25 Wagner 2005). Accordingly, forward thinking companies are using these data as a useful information source during capital expenditures planning. But linking the collected information to an accessible web-based platform is not a simple task.

Numerous wireless sensor network technologies are available to facilitate the transfer of information from the cold chain asset in questions to an online database interface (27 Shan 2004). The most commonly used platforms for the transfer of this information is ISM, which stands for Industrial, Science and Medical, and denotes a section of radio bandwidth available free of charge (27 Shan 2004). Note that the use of this bandwidth only applies to radio frequency identifier devices. The sensor technology used to collect data for the purposes of this thesis utilized hard wired electronic temperature sensors that transmitted data over cellular frequencies and did so at regular intervals or at the time of alert.

Interest in advancing cold chain performance extends beyond the involvement of industry players. In 2010, Georgia Tech created the Integrated Food Chain Center

(IFCC) which links academia, industry players and the federal government with the objective of improving cold chain management of perishable food products. The substantial investment being made in the IFCC highlights the need to develop new tools and methodologies that can be used to optimize cold chain. According to the IFC two areas stand out, performance analytics and predictive modeling. The center website has the following to say about these two areas:

Performance Analytics:

- Develop models that predict the status [quality, arrival time, etc.] of a product at a future time and allow for intervention when appropriate.
- Optimize trade-offs between waste and transportation.
- Optimize trade-offs between costs and value of additional data in improving shelf life estimates.
- Decrease transportation and handling costs with better chain design.

Predictive modeling (remaining shelf life)

- Challenge testing and simulation.
- Predict the effect on shelf life of the product caused by the mishandling of temperature and humidity along the SC.
- Develop models to forecast demand/supply and determine optimal pricing for maximizing revenue.
- Determine the value of prediction versus reactive decision making (28 Georgia Tech, Integrated Food Chain Center 2010).

2.3 INDUSTRY MEASURES OF QUALITY AND IMPACTS ON REJECTIONS

Fresh fruit and vegetable quality is hard to define. Martyne Brown's book, *Chilled Foods – A Comprehensive Guide,* states that quality may have several meanings depending on the context. He states that quality may refer to, "the degree or standard of excellence for specified characteristics, the suitability for purpose (e.g. processing, shelf-life) or the consistency of attainment of the specified properties" (23 Brown 2008). Furthermore, the diversity determinants of quality concerning specific FFV products and the variance of definitions of quality have caused government organizations to step in to regulate this area (23 Brown 2008).

Hence, although regulation has helped define product quality, large retailers are becoming more influential in determining the criterion against which product quality is measured (23 Brown 2008). In order to further investigate the definition of quality, isolating the pre and post-harvest stages of a produce item's lifecycle is beneficial.

The quality of fresh fruits and vegetables is best investigated from two perspectives relating to the two distinct stages of the product lifecycle. First, we consider the literature that discusses the general pre-harvest factors affecting the quality of fresh fruits and vegetables. Second, we consider the post-harvest care of fresh fruits and vegetables.

Agronomic characteristics of pre-harvest fresh fruit and vegetables define product quality prior to picking and processing. We will not delve into this area in great detail as our thesis deals more specifically with post-harvest care. Nonetheless a cursory tour of these determinants is warranted. Brown defines the key agronomic characteristics as, "Variety/Cultivar, Shape and Size, Color and Appearance, Flavor and

Texture"(23 Brown 2008). Essentially, the standards established to measure quality begin in the field. Brown states that the quality of a perishable ingredient may vary greatly due to varietal, production location, and the growing season (23 Brown 2008). Growing conditions determined by the season and weather, quality of arable land, and myriad other factors, can contribute to the quality of the FFV product or ingredient. Quality as established by growing conditions and other factors in the field are essentially the starting point for product quality on its journey from farm to table.

Post-harvest quality is determined by a number of factors but good post-harvest quality begins with good pre-harvest quality. Brown writes,

> Once a crop is harvested, its quality cannot be improved. At this stage, the objective must be to maintain the produce in good condition through any short- or long-term storage until it is delivered to the customer. Thus, the ultimate quality and shelf-life of a final product depends not only on growing conditions but also on harvesting and on post-harvest handling and storage (23 Brown 2008).

With regard to product handling, Brown is particularly concerned with mechanical damage stating that rough handling during transportation leads to bruising and subsequent spoilage(23 Brown 2008).

Mechanical damage aside, Brown is most concerned with ensuring that FFV products are stored and transported at proper temperatures and he goes into great detail to elucidate this point. He states with respect to proper temperature:

> Maintenance of suitable post-harvest temperature is extremely important to maximize shelf-life, both for produce for immediate use and for that to be

> stored. Crops continue to respire after harvesting, using up reserves and shortening shelf- life through wilting and yellowing. Respiration rate is temperature related and is roughly halved for every 10 °C that the temperature is reduced. The general rule is to remove field heat as quickly as possible after harvest and then to maintain the produce at chill temperature (23 Brown 2008).

In essence, Brown makes that case the improper temperatures during transport and storage exacerbate the signs of poor pre-harvest quality and mechanical damage the product has been subjected to. Proper temperatures reduce the respiration rates of FFV thus forestalling signs of poor pre-harvest quality and mechanical damage.

Brown notes that discontinuities in the cold chain occur at several key stages in the materials handling process, most notably during reception and handling and during storage (23 Brown 2008). Inefficient product handling at the time of receipt, loading or unloading can lead to reductions in product quality.

In addition to temperature control, two other techniques can be employed to preserve product quality and maximize product shelf life. They are Modified Atmosphere Packaging (MAP) and Moisture Control. Brown notes that the application of MAP essentially replaces the atmosphere in a product container with one that is more beneficial to the preservation of the product therein. Brown writes, "MAP involves the removal of air from the package and its replacement with a single gas or mixture of gases that is different from the normal composition..." (23 Brown, Martyne 2008). In lettuce manufacturing the most common MAP additive is nitrogen (23 Brown 2008). Essentially, nitrogen retards product spoilage caused by oxidation. As a result, Brown states that this leads to better product sales due to, “the attractive color and presentation of food products”(23 Brown 2008). MAP used in packaged lettuce greatly slows decay and forestalls the appearance of the negative signs of incurred mechanical damage.

The second major advance in packaging has been the advent of strategies to combat moisture buildup in the product pack. One such strategy, “pack fogging,” is described by Brown as “the condensation of air moisture on cold plastic, causing the formation of tiny droplets on the surface which scatter light and obscure the contents of the pack” (23 Brown 2008). Technological advances and economies of scale have driven the use of anti-fogging films and allowed for the deployment of MAP in a cost effective manner. Coupled with proper temperature management these strategies play a crucial role in preserving the quality of FFV in transit and storage.

The expanding trade in perishable goods requires regulation of the definition of quality to protect final consumers from standards erosion (3 Verbic 2006). European Union regulations have led to the introduction of cold traceability, which requires tools and equipment to trace disparate groups of perishable goods being produced and distributed under different cooling requirements (3 Verbic 2006). We can expect to see this trend continue to become the norm.

The new access to information will drive the move towards the monitoring of quality of FFV in real-time as stated by Planet Retail below.

> Access to reliable, meaningful data provides the visibility of cold chain performance that is required to ensure consumers are safe and their families get the high quality food they deserve. Quality and safety of

chilled foods during storage are largely determined by time-temperature. These concepts refer to the relationship between the storage temperature and the storage life. Different types of food and different mechanisms govern the rate of quality degradation of foodstuff. Time-temperature tolerance relationships can also help in predicting the effects of changing or fluctuating temperatures on shelf-life (12 Planet Retail LTD 2006).

Our research seeks to look into the relationship between quality and temperature and to understand how these variables are correlated to the level of rejections and fuel efficiency in the refrigerated transport of fresh fruits and vegetables.

2.4 REFRIGERATED TRANSPORT FUEL EFFICIENCY

Refrigerated transportation fuel efficiency deals specifically with the fuel efficiency of refrigerator gen-sets used to cool the trailer compartment that holds goods during transit. This thesis does not consider the fuel efficiency of the tractor that pulls a refrigerated trailer, but rather considered the fuel efficiency of the generator that provides power to a refrigerator unit. Most refrigerator units are comprised of a generator that produces electric power by burning diesel fuel and a refrigerator unit that uses this electric power to cool the air in the container compartment. These two components together are known as a gen-set. A gen-set unit is located at the nose of the trailer, and is mounted between the trailer and the tractor unit. A separate fuel tank for the gen-set is typically mounted under the nose (front) of the reefer unit and is filled with non-taxable red diesel fuel. Red diesel fuel gets its name from the solvent red 26 dye added to the fuel to indicate its low-tax or non-taxed status.

As fuel prices continue to rise, carriers are increasingly interested in finding ways to offset fuel expenditures. Accordingly, carriers are more aware of Reefer fuel efficiency and the market is responding to this awareness. Advances in technology as well as changes in operator behavior can greatly improve reefer generator fuel efficiency. Developments in temperature-controlled technology can improve efficiency and reduce fuel consumption (33 Frank 2005).

Fleet managers face three major issues that negatively affect reefer fuel efficiency:

- Slow pre-cooling.
- Inefficient or aging reefer fleets
- Poor Driver Training (33 Frank 2005)

Pre-cooling is the process of cooling down the inside of a refrigerated trailer before any perishable product is loaded into the asset. Slow pre-cooling indicates that a large amount of fuel is being consumed to cool the empty container. Longer-than-anticipated cooling cycles may be brought on by high ambient air temperatures or may indicate that a gen-set is not operating effectively or that a trailer is not sealed properly. Older assets tend to cool more slowly due to poor container seals and older, less efficient gen-set equipment.

The use of increasingly tailored refrigeration software allows fleet directors to program temperature settings that maximize fuel efficiency and load protection for a specific type of produce or material being transported (36 Skydel 2009). Ronn Kemm of

Extra Lease states that "[b]y implementing and following custom or pre-programmed settings, significant fuel savings can be seen to improve fuel efficiency by 10% to 30%"(36 Skydel 2009). Many produce companies have started to employ smart setting systems driven by sophisticated and produce specific cooling profiles that better protect produce in transport. In addition to use better programming to drive fuel efficiency, gen-sets that can run on diesel fuel or alternatively be plugged into electric power sources to enable cooling are also drastically improving fuel efficiency.

Manufacturers are attuned to the potential benefits being able to plug units into existing electric power. Kemm states, "The use of electric standby systems can reduce stationary fuel consumption to essentially zero. . . With the power being consumed coming from building's electrical supply, the return on investment is easily realized by comparing the stationary operating hours to the energy and maintenance savings between diesel and electric operation, including infrastructure related costs" (36 Skydel 2009).

Skydel goes on to state "depending on the price of fuel, a fleet can save 40% to 70% in energy costs by plunging into the grid versus running the reefer strictly on diesel fuel" (36 Skydel 2009). This trend in reefer gen-set design will surely continue.

In order to better understand the needs of customers, one large reefer manufacturer is teaming up with one major client in an effort to spur innovation. Wal-Mart and Thermo-king are working together in an effort to develop more energy-efficient transportation technologies. For example Thermo-Kings worked with Wal-Mart to develop Spectrum SB, a three zone, multi-temperature reefer unit for trailers with a 30-foot frozen compartment in the center of the trailer. The unit is available with electric standby, which allows it to use electricity during loading and stationary operation. All models feature bio-diesel compatible engines (34 Food Logistics 2011). This new system provides significant benefits for Wal-Mart, and the firm has retrofitted seven thousand reefer units with the new tri-pack equipment. Tim Yatsko, VP of transportation at Wal-Mart states, "this installation could help us save $25 million in fuel costs" (34 Food Logistics 2011). Other changes to trailer interiors are also creating breakthroughs in reefer fuel efficiency.

Ceiling mounted insulated bulkhead can now be installed in refrigerated trailers. These new bulkheads are adjustable and slide from the nose to the rear door of the trailer with relative ease. After each stop, the volume of refrigerated space can be reduced by partitioning the trailer using the sliding ceiling-mounted track bulkhead. Without this practice the equipment ends up cooling a lot of "dead air" as the truck delivers each subsequent load creating more and more empty space in the trailer that needs to be cooled (35 Klei, Leonard 2005).

Bulkheads also allow companies to use multi-temperature zones in a single trailer unit. Multi temp zones allow for more than one type of cargo to be carried. With multi-temps, companies are able to optimize requirements for different food types (35 Klei 2005).

Although this practice can bring saving in terms of fuel-efficiency, the pitfall is driver's safety at the moment of lifting and lowering the bulkheads. The floors inside a trailer are usually pretty slippery and drivers could fall.

All the above advances point to two major trends. First, carriers are interested in reefer fuel efficiency as a function of fuel price increases over time. Second, manufacturers of refrigerated trucking equipment are eager to address the needs of their clients by developing fuel saving technologies.

Our thesis highlights how real-time information can be used to effectively monitor reefer fuel efficiency in the refrigerated transportations of fresh fruits and vegetables.

2.5 DATA ANALYSIS AND MODELING

New technologies including the Internet, database software, and sensors for real-time data collection give companies access to a huge amount of information from every echelon of the supply chain. We are entering an era where companies are facing a new and different kind of problem: they can generate or buy an enormous amount of data, but they don't know how to unlock the value in that data (26 Albright, 2011).

The question is not how to get information but how to interpret and use information to a company's advantage. The answer depends primarily on employee's quantitative and statistical skills, and on their ability to work with available tools (26 Albright, 2011).

This thesis is focused on the refrigerated transport of processed salad products. In investigating this area, the following questions frame our data collection and analysis:

- What variables do we want to improve?
- Which analytical methods should be selected to figure out how to improve them?

The first question is already answered in the problem statement of the thesis. The second question is answered by examining the existing literature concerning methods for data analysis. Two approaches have been selected: multiple linear regression and logistic regression.

With regard on linear regression, Dimitris Bertsimas and Robert M Freund on their book, *Data, Models, and Decisions, The Fundamentals of Management Science* state: "The goal of a linear regression model is the development of a specific formula that relates the movements of one variable to those of a small set of other relevant variables..." (45 Bertsimas,Dimitris 2004).

This approach matches the objective of this thesis as we seek to understand how fuel efficiency and rejections are related to the independent variables that play a relevant role during the refrigerated transportation process. In the case of rejections, a special technique of regression methods called logistics regression will be required. The details of this technique are presented in the Methodology section.

2.6 GAPS IN THE LITERATURE AND CALLS FOR FUTURE STUDY

Although considerable effort has already been put into creating methodologies for collecting information on refrigerated assets in transit, little work has been done to produce meaningful frameworks for the analysis of the wealth of data being harvested. Data is a raw material that needs to be shaped into a useful tool for decision-making. This process occurs through careful analysis. The fresh fruit and vegetable industry would greatly benefit from such analysis. In large part, the Integrated Food Chain Center at Georgia tech was established in an effort to advance this nascent area of cold chain analysis.

According to the IFCC, performance analytics and predictive modeling must be explored. The center highlights several possible areas for research as performance analytics and predictive modeling, and calls for analysis that can achieve the following:

- Develop models that predict the status [quality, arrival time, etc. of a product at a future time and allow for intervention when appropriate.
- Reduce energy consumption and carbon footprint.
- Predict the effect on shelf life of the product caused by the mishandling of temperature and humidity along the supply chain (28 Georgia Tech, Integrated Food Chain Center 2010).

This thesis seeks to contribute in addressing the needs above by developing a methodology that helps explain reefer fuel consumption and packaged salad rejections. By doing so, the study also aims to fill several gaps existing in the current literature.

3. INDUSTRY BACKGROUND

3.1 INTRODUCTION TO REFRIDGERATED TRUCKING OF FRESH FRUITS AND VEGETABLES

Fresh fruits and vegetables (FFV) are transported from the region they are grown in to the region they are processed or consumed in by refrigerated truck. Refrigerated trucks consist of a tractor and a refrigerated trailer and are commonly called Reefer Units in the specialty trucking industry. The tractor provides locomotion for the trailer unit. The trailer unit consists of a cargo hold, generator set and refrigeration unit. The refrigeration unit is composed of a power generation unit that provides electricity, a condenser unit that cools trailer air and a fuel tank that hold diesel fuel burned by the power generator. For the purposes of this thesis, wherever generator set or gen-set is used, the entire cooling system is meant, unless otherwise stated.

Larger fresh fruit and vegetable processing and distribution companies have traditionally utilized a mix of public and private fleets to transport their produce. The fleet that is monitored in this research is a leased, private fleet in good operating condition. Additionally, important to this thesis is a general overview of the transportation of FFV via refrigerated truck.

The refrigerated transport of FFV originates at the point of production or importation of perishable goods. In the United States, most refrigerated truckloads originate in California, the Pacific Northwest, Arizona and Florida and move towards major population centers often time by way of an intermediary processing facility. The following list highlights the highest transit volume (tons) perishable fresh fruit and vegetable commodities moving in the United States:

Table 1 Commodity Volumes (Tons) by Quarter

Commodity	3rd Quarter 2010	Previous Quarter	Same Quarter Last Year	Current Quarter as % change from: Previous Qtr	Same Qtr Last Year
Potatoes	723	789	1,175	-8%	-38%
Lettuce	510	574	592	-11%	-14%
Cantaloupe	403	303	405	33%	-1%
Tomatoes	379	608	571	-38%	-34%
Watermelon	377	942	780	-60%	-52%
Grapes	345	187	431	84%	-20%
Onions	304	463	577	-34%	-47%
Strawberries	246	360	219	-32%	12%
Apples	225	350	364	-36%	-38%
Peppers	209	237	250	-12%	-16%

Source: Agricultural Marketing Service, Fruit and Vegetable Programs, Market News Branch

Table Taken From (42 Taylor, April 2010).

This thesis is concerned with the transportation of prepared salad products and deals specifically with bagged salad products traveling from processing facilities to customer distribution centers or store locations for purchase by end consumers. By weight, lettuce is the second most shipped perishable commodity in the United States, second only to potatoes. California accounts for the vast majority of lettuce produced and shipped throughout the United States. In Q3, 2010 California produced 496 tons of the 510 tons shipped, equivalent to more than 97% of lettuce produced in the stated quarter (42 Taylor, April 2010).

Lettuce requires special care in transport in order to ensure that the highest quality product possible so that the most days of remaining shelf life can be consistently delivered to customers. Accordingly, refrigerated fleets that handle salad products need to be maintained with a high degree of temperature precision in order to ensure sanitary conditions, fuel efficiency and general operating conditions conducive to the effective transport of this highly perishable product.

Of great importance in the transportation of any perishable material and even more so in the transportation of salad products specifically, is the pre-cooling of refrigerated trailers prior to loading. Pre-cooling is the process of cooling down an empty refrigerated container to a specified temperature prior to loading the container with perishable materials. Perishable goods loaded into a warm container will inevitably be damaged. Sometimes this damage is immediately perceivable, and other times, as with salad products, the damage is not manifest until several hours or days later. Accordingly, proper pre-cooling procedures are vital to the success of a perishable goods cold chain operation.

Good pre-cooling begins with a clean and dry trailer being checked for open doors and proper doors seals. Then the gen-set is powered on to run on diesel fuel or alternatively can be run off an auxiliary electric power source available at the cooling or loading location. Pre-cooling is initiated to bring the air and interior wall temperature to the desired set point of the reefer unit. The reefer set point is the target temperature the reefer gen-set is tasked to achieve and then maintain.

The pre-cooling process can range in time from an hour to several hours based on the ambient air temperature, the age and condition of the equipment at hand, and the temperature of the trailer walls at the time when pre-cooling is initiated, among other factors. Once a reefer unit is properly cooled, produce can be loaded into the container compartment with a much lower risk of damaging the product to be transported.

Loading is also a critical success factor in the transportation of FFV. Improper loading can result in undoing the positive effects of pre-cooling. Perishable produce held at appropriate storage temperatures should be loaded into a cooled reefer container via a well-designed air baffle. This condition, of course is ideal and not always possible. Loading through such an airlock allows for reefer interior temperatures to be held consistent.

Also important during this process is a clear understanding of how air is circulated around the interior of a trailer. The model of refrigerated trailer involved in this study consists of an exterior mounted refrigeration unit located at the nose (front) of the trailer. This unit recycles the cool air inside the trailer, in-taking air from the front floor of the container, cycling this air through its condenser for cooling, and then returning this air to the container. Cooled air is returned to the container from the top of the nose of the trailer and is funneled through perforated sock acting like a cold air conduit. This cloth air sock mounted to the center of the interior ceiling of the reefer is perforated at regular intervals to allow air to be both channeled to the rear of the container, but also distributed evenly to the right and left of the sock. As colder air is denser than warm air, evenly distributed cold air falls to the floor of the interior of the container. This air is slowly cycled to the front of the reefer container where the air intake for the refrigeration unit is located. This air is then reprocessed through the unit to begin the journey all over again. Figure 1 below shown an illustration of the air sock in green.

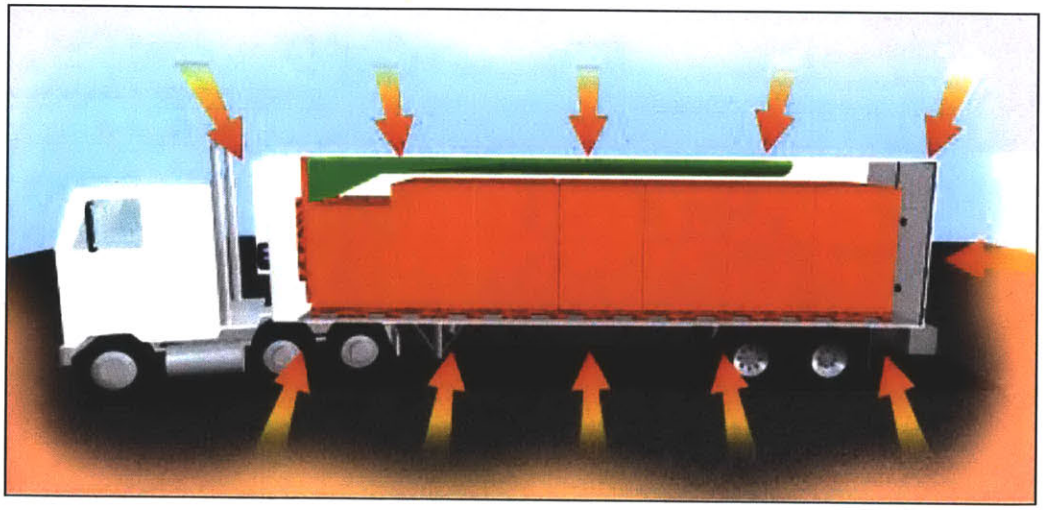


Figure 1 Trailer Side View

Understanding this process is essential to proper loading for a major reason. When loading occurs with the refrigeration unit running, instead of continuing to cool the reefer container, the unit essentially sucks warm air into the container through the open back door as the air return vacuum at the nose of the trailer draws in air. Essentially, the reefer unit only effectively cool and maintain internal container temperature when the container's doors are sealed. Otherwise, unless located in a refrigerated unloading dock, the container will be brought to ambient temperature, rapidly undoing the benefits of pre-cooling.

An important determining factor of proper refrigerated transportation is the temperature of cargo at the time of loading. Refrigerated containers are not effective at cooling cargo, but are designed to maintain the temperature of this cargo during transportation. Accordingly, effective cold chain transportation depends on the temperature of the cargo at the time of loading.

Major trends in refrigerated trucking point to the use of information systems to help address the myriad complications faced by the industry. Most notably, as fuel prices continue to rise, carriers are becoming increasingly interested in squeezing as much value from each gallon of fuel as possible. Advances in information systems have aided in this endeavor.

3.2 INTRODUCTION TO REAL-TIME INFORMATION AND AVAILABLE INPUTS

Trucking companies are increasingly interested in collecting information that can help them improve the services they offer to customers and maximize the efficiency of their operations. The major trend in specialty trucking in the past decade has been a move to more highly integrated information systems. This thesis takes advantage of this trend by using available information being collected and stored by a real-time information system installed on refrigerated trailers.

Real-time data refers to data that is collected and delivered immediately. Real-time information is not stored, filtered or assessed prior to sending the data to the user of need. There is no delay in the delivery of the data collected (37 Wade, Sommer 2006).

Real-time data is often used synonymously with dynamic data. Both terms refer to the same general idea that collected data is transmitted without delay. However, real-time data is often used in describing IT systems specifically.

Real-time data is used in a diversity of applications ranging from navigation to measurements of equipment run times on an assembly line. This thesis is concerned with real-time data collected regarding the current status of variables that describe the status of trailer assets used to transport bagged salad products from processing facilities to customer drop locations.

This thesis uses a diversity of real-time data generated from a variety of sensors. The company selected by our sponsor company (XYZ) to wire and collect data from trailer assets is Par Logistics. The wiring and initial use of this system and the data it generates is not new and has been ongoing for several years. This thesis makes use of this existing information with the aim of deriving further utility from the available rich data set. Although other data collection systems exist, in the passages that follow, the Par system is described in detail as this system plays a role in the thesis material that follows.

Par states that its system "extends information needed to monitor and control reefer temperatures, react and respond to alarms, and optimize logistics and management of cold chain assets and resources" (39 Par LMS 2009). In essence this equipment allows cold chain carriers to maximize their profit by providing better asset management, improved efficiency, and better customer service through the use of real-time information technology. The Par tools utilized in this thesis provided the following information:

GPS Antenna and Module

The GPS antenna and module collect real-time location information using commonly available GPS hardware and programing. This GPS information is linked to the rest of the data collected so that each data worthy incident can be associated with the location of occurrence as well as the time of occurrence. Exact information on the location of a fleet asset is becoming increasingly important for carriers, as their customers, shippers, are holding less inventory and very often function in a just in time or near just in time environment.

Cellular Satellite antenna

The Par unit transmits all collected information over cellular network to a central Par server that stores this information for future use or forwards alerts to the necessary party. Typically, a transportation company's fleet manager or on duty manager will receive all important alerts as a text message sent to a cellular phone and/or as an e-mail accessible from a variety of platforms.

Motion Sensors

Motion sensors measure the movement of a trailer asset. These sensors allow a person combing through available information to be able to see if a trailer asset is moving or is stationary.

Power Sensors

The Par-unit power sensor denotes the power sources being used to run the refrigerated trailers generator set. The gen-sets equipped with this technology can be run off of diesel fuel or can be run off of an auxiliary electric power source. There is a diversity of auxiliary power sources and any can be used so long as the reefer gen-set can be plugged into this power source with ease. Auxiliary power sources are not built into the trailer unit but rather are typically part of the existing infrastructure at a location where a trailer asset may spend a considerable amount of time in a status that would require it to have the gen-set running. These locations include any place a reefer unit is pre-cooling, being loaded, being unloaded or waiting while holding perishable cargo.

Temperature Sensors

A diversity of temperature sensors are wired to the Par unit. They are as follows:

Ambient Air Temperature

Ambient air temperature is a measure of the outside air temperature at the specific GPS location indicated in the data set. Outside air temperature may influence the functionality of the reefer gen-set and a diversity of other areas.

Return Air Temperature

Return air temperature measures the temperature of air being sucked back into the reefer's gen-set through the air return located at the floor level of the nose of the trailer. This air has cycled through the entirety of the trailer and is now being returned to the gen-set to be cooled and begin the circulation cycle again.

Return Temperature and Probe 1 & 2

Two additional temperature probes are installed along the interior of the trailer, one at the center and one near the rear door. These probes are attached to a twenty foot cord that carries the temperature information taken at the tip of the probe and transmitted back to the Par communication module and cellular antenna. Temperatures inside refrigerated trailers tend to vary. Accordingly, the greater the number of temperature probes the better the understanding of temperature within a given trailer asset.

Customizable Web-based Reporting

An important component of any metering device is its ability to display data in a meaningful way. The Par product comes with a very rudimentary display for data. The web-based interface does however allow for easy exporting of reefer data that can then be imported into Excel and analyzed. This web-based interface is accessible from any location with an Internet connection, ensuring that managers and analysts far removed from the actual asset in question can monitor the performance of this asset with relative ease.

Real-time Reefer Alert Message (e-mail and text messaging)

As alluded to in a previous section, all Par system alerts are transmitted to system users via e-mail and or text message. Transportation is a round-the-clock pursuit and will increasingly be so as transportation lanes become further congested. Managers are not always at their desks and Par service providers are not always at their terminals. Sending automated alerts via text and e-mail ensures that the necessary party is more likely to receive a real-time message than if it were delivered by a human.

Reefer Temperature Sense and Control Option

This option allows managers to take corrective action in real-time and control a reefer's run settings via a web-based interface. Imagine a situation where a trailer asset is getting too warm and must be rapidly cooled to save the perishable cargo on board. A manager can use a web-based interface to control the trailers refrigeration unit remotely and can lower the temperature of a reefer by adjusting the run settings.

Communications electronics

The communication electronics allow for all outgoing and incoming data and action requests to be processed appropriately. Essentially this system knits together all sensor information into a neat package for periodic upload to the web-based interface and also manages incoming information or setting changes.

Fuel-Level Monitoring and Reporting

Reefer gen-sets typically run on diesel fuel. This fuel is housed in a fifty-gallon tank mounted to the underside of the trailer near the nose. The par system measures the fuel remaining in this tank in terms of gallons and also makes this information available as a percentage of total tank capacity.

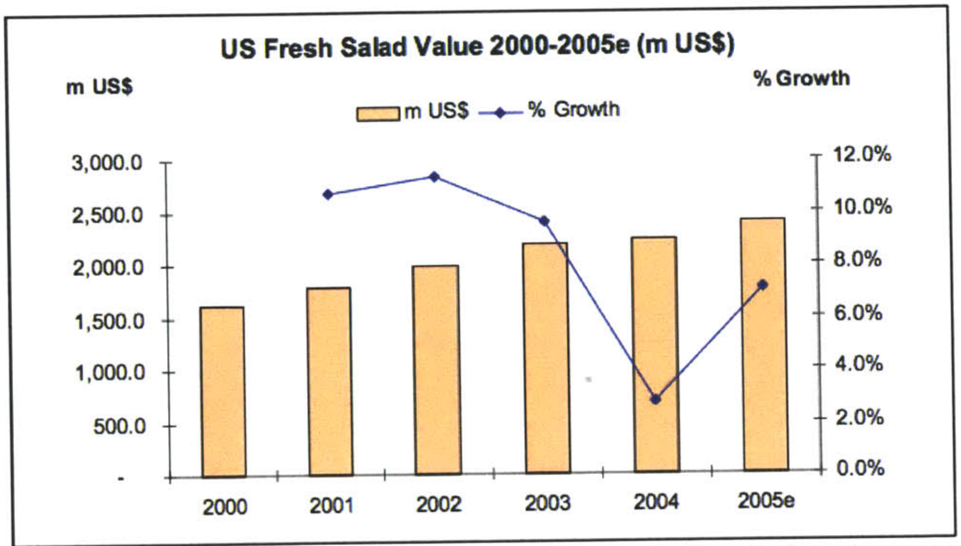
Door Sensor

Trailer tail (rear) doors are wired with proximity sensors that provide door-open and door-closed status.

This great diversity of information is collected in a centralized database and is available in both real-time and as historic data for use in future research. We have chosen to use that data as a source of historic information in an effort to reach a deeper understanding of the variables that affect fuel efficiency of reefer gen-sets and the probability of rejection on a per customer load basis. If a meaningful model of rejections or fuel efficiency can be generated, a system can be put into place to have such a working model run off of real-time data. The purpose of this thesis is to determine if such a set of models exists.

3.3 THE PACKAGED SALAD MARKET

The packaged salad market in the United States is stable but poised for growth. Total annual sales for the fifty-two week period ending on January 24^{th} 2010, was \$2.76B, down 1.8% from the previous year (29 Grocery Headquarters Editorial Staff 2010). This drop in sales dollars may be attributed to business cycle fluctuation. During better economic conditions at the beginning of the decade, the category had experienced significant growth. This growth can be attributed to the change in food preparation habits of most Americans (30 Snapshot Data International Group 2005). The American diet has changed considerably over the past two decades from one that embraced predominantly packaged foods to one that has been characterized by increasing demand for fresh products that promote a healthier lifestyle. In recent year, this has very much been the trend. "The compound annual growth rate for the period 2000-2004 was 8.5%. The strongest growth was in 2002, with a rate of 11.3%" (30 Snapshot Data International Group 2005).



Source: ACNielsen, Information Resources, Inc, Refrigerated and Frozen Foods, Snapdata Research

Figure 2 US Fresh Salad Value 2000-2005

This Figure Taken From (30 Snapshot Data International Group 2005).

Information Resources, Inc. of Chicago, Il, lists refrigerated salads as 27th in sales amongst the top 50 super market categories. Additionally, packaged salad is the second-largest category in the produce department, accounting for 7% of sales in the fifty-two week period ending on April 24th of 2010 (29 Grocery Headquarters Editorial Staff 2010).

The packaged salad category is part of the larger produce category. Some industry analysts also place packaged salads into the cut-produce category, which includes washed and prepared fruits and vegetables.

The popularity of bagged salad is no doubt a result of the convenience the product affords consumers. Bagged salads come pre-washed and cut to size. A variety of salad greens of balanced flavor and appealing color can be offered as a salad mix product and sold in a single bag as a ready to eat product.

The packaged salad category is comprised of the following sub-categories:

- Blends
- Premium Garden Salads
- Complete/Kits
- Regular Salad

These four categories account for 80% of the category dollar share with 37%, 17%, 14% and 12% contributions, respectively. The remaining 20% of the prepared salad category is composed of the following categories:

- Organic
- Coleslaw
- Spinach Salads

These categories comprise 11.7%, 3.7% and 3.7% of the market dollar share, respectively (30 Snapshot Data International Group 2005).

A mix of several salad green varieties characterizes blends. Premium garden salads are a mix of cultivars that are offered at a premium due to materials quality or rarity. Salad blends that include more exotic cultivars, as ingredients tend to have a higher price point and be sold under modified labels.

Complete/Kit salads are bagged salad products that include within the product package all the necessary ingredients to build a specific salad. These kits, or Completes as they are sometimes called, will typically include a specific salad green or mix of greens, croutons and a marketing appropriate salad dressing. Typically, the Complete salads are marketed based on the dressing they include or based on the type of well-known named salad they allow the consumer to construct.

Several national brands are now producing salad kits that include a diversity of ingredients above and beyond salad greens alone.

Regular Salad makes up the 4th largest segment of the category with just over 12% of the market and is characterized by single varietals sold in isolation of other ingredients. Lettuce varietals dominate this segment of the category with Iceberge and Romain lettuce being the predominant product offerings (29 Grocery Headquarters Editorial Staff 2010).

Organic salads account for 11.7% of the packaged salad market, but are a growing category. Organic salads can be sold as blends or individual cultivars and are differentiated by certification labeling and clear marked packaging that identifies the products organic status (29 Grocery Headquarters Editorial Staff 2010).

Coleslaws account for 3.7% of the packaged salad market (29 Grocery Headquarters Editorial Staff 2010). Rounding out the product category, broad leaf spinach salads account for 3.7% of the prepared salad market in the US (29 Grocery Headquarters Editorial Staff 2010).

Salad consumption is not evenly geographically distributed across the United States. The most common way to measure the sale of prepared salad products is in weekly sales per store. Doing so yields the following results for the 52 week period ending on April 25, 2009 (30 Snapshot Data International Group 2005).

Table 2 Average Weekly Salad Sales By Region

Region	Average Weekly $ Sales per Store
Central	$3,165.00
East	$3,840.00
South	$2,226.00
West	$2,800.00
Total US	$2,683.00

As we can see from the above table, per store sales are significantly higher in the eastern United States than in any other region of the country.

Salad demand also displays the effects of seasonality. Peak salad sales typically occur during the first or second week of January potentially as a result of New Year's resolutions for improved eating habits. The public's focus on healthy eating habits spurs the sale of prepared salads during this period of the year.

Four main players dominate the packaged salads industry in the United States. Market segmentation as of 2005 is shown in the table below (30 Snapshot Data International Group 2005).

Table 3 Salad Industry Participant Market Share

Company	Market Share
Fresh Express	37.7%
Dole	36.9%
Private Label	11.6%
Ready Pac	7.1%
Others	7.1%

As shown in the table above, two main players dominate this market, Fresh Express, a subsidiary of Chiquita Brands Int. and Dole (30 Snapshot Data International Group 2005).

In conclusion, the size and scope of the packaged salad industry coupled with the high degree of cargo perishability and the extent of available data make this industry a perfect candidate for study. This thesis is built on a foundation of deep understanding of the salad industry. By reading the literature review and industry overview, a reader of this thesis will gain a rich working knowledge of the salad industry and will be well prepared to understand the analyses performed to identify the variables that affect reefer fuel efficiency and the probability of load rejections in the cold chain transportation of bagged salad products.

3.4 QUALITY DETERMINANTS OF PACKAGED SALAD PRODUCTS

With regard to the degree or standards of excellence for specified characteristics, strict agronomic characteristics are sought. Specific cultivars of leaf salad are selected for production based on time of year, location, and a diversity of other factors. The shape and size of these cultivars may vary for genetic reasons or growth condition reasons, and growers closely monitor these anomalies.

With regard to suitability of purpose, color and appearance are the most important factors that affect the measure of product quality at the store level. Barney et al. (2007) state that consumers of produce purchase products “by eye” and as a result, an attractive product appearance is very important. This is very much the case in the bagged salad industry.

The major metrics used by bagged salad providers measure the following areas that relate directly to visual attractiveness:

1. Pinking of leaf ribs and cuts
2. Excessive bruising
3. Decay
4. Moisture in the bag
5. Discoloration Other Than Pinking
6. Drying

The department of Post-Harvest technology at UC Davis defines Pinking and Pink Rib as “a disorder associated with heads that are over mature”, and goes on to state that “(h)igher than recommended storage temperatures can also lead to an increased incidence of pink rib. In this disorder, the midribs take on a generalized pinkish coloration. Ethylene exposure does not appear to affect pink rib and low O_2 atmospheres do not control it” (40 UC Davis).

Pinking is a major problem for retailers, as consumers do not purchase discolored product. The fact that this problem is exacerbated by higher-than-optimal transport temperatures highlights the critical importance of optimal cold chain performance from farm to table.

Excessive bruising is the result of mechanical damage occurring at any point in the products life cycle. Numerous factors can cause bruising during the transportation and production processes, but one factor stands out. Barney, et. al. (2007) state that the number of times a product is handled by humans directly affects the amount of bruising that products sustain. Reducing the number of touches during distribution is therefore of importance.

Decay is caused by numerous factors. Spinach is of particular concern in the salad industry as it is highly susceptible to leaf decay. A 1994 UC Davis study, stated that "an average of 17, 28, and 45% of leaves of 16 varieties had decay after 2, 3, and 4 weeks at 5°C, respectively. After the same periods at 5°C, 18, 25, and 45% of the leaves showed some yellowing. Commercial verities such as Imperial Spring, Shasta, Polka, Spectrum and Sporter had notably longer shelf- life than did varieties Bossanova, Spark and Space" (41 UC Davis, 1994).

In addition to high temperatures, temperatures below freezing also negatively affect tender greens. Temperatures below freezing cause water in cells of leafy greens to freeze which results in expansion of water molecules, damage to plant cell walls and eventual cellular destruction. These areas of damage are manifest first as translucent areas on leaf faces and stems and once warmed to room temperature become darker green spots that are highly susceptible to decay.

In addition to unappealing visual signs of decay, the process also results in the production of offensive odor. Although this odor is not evident when bag salad products are sealed, as soon as they are opened, these odors are an indicator of the presence of decay.

The presence of decay is the most common reason for the rejection of bagged salad products and is very closely linked to temperature during transit. The presence of three or more leaves that display signs of decay in a single product bag will typically result in the rejection of the entire product order.

Another common quality indicator is the amount of moisture in the bag holding salad products. Moist environments are conducive to bacterial growth and promote decay and discoloration. Fluctuations in the storage temperature of salad products can cause leaves to give of moisture via transpiration. This moisture can then condense as storage temperatures are lowered resulting in condensation or visible moisture in the product bag. The consistency of product temperature as it moves through the cold chain can, to a great degree, control the amount of moisture buildup in product bags. Regardless of the appearance of decay or no decay, produce buyers can and do reject product on the basis of excess moisture in the product bag.

Beyond pinking, other types of leaf discoloration can result in reductions in product quality. Freezing mentioned above is one such form of discoloration. The presence of ethylene gas will also cause spotting in many green salad varieties. According to UC Davis (41 UC Davis, 1994), Romain and Iceberg lettuce are both very sensitive to Ethylene spotting. Spinach is also very sensitive to the gas. Accordingly, proper air circulation or the use of a modified atmosphere must be employed in order to forestall the appearance of ethylene spotting.

Drying is also a major source or product rejection. Dehydrated greens are not appealing to customers. A careful balance needs to be maintained so as to provide the right amount of humidity for a product to retain its freshness while not introducing so much moisture as to cause excessive condensation to occur and bring on the eventual rot that accompanies such condensation. Drying also has major implications with regard to the suitability of purpose of the product. Wilted leaves do not have good mouth feel nor do they have the requisite visual appeal demanded by discerning customers.

Also, during the course of our research we were able to meet with a Produce Quality Inspection Manager for a major national supermarket conglomerate. This individual provided our team with a wealth of information regarding quality determinants. Specifically, the Inspection Manager stated that the four major variables considered when deciding to accept or reject a product load are, in order of importance:

Product Temperature During Transportation

Evidence of Epidermal Peel

Evidence of Tip Browning

Evidence of Pinking

These accept/reject determinants are very much in line with the common understanding that end customers, store shoppers, select or reject product predominantly on the basis of appearance. Additionally, the statements by the Produce Quality Inspector also highlight the importance of temperature during transit. To a great extent this also point to increasingly sophisticated tools being deployed in order to effectively gauge quality.

This distributor requires all carriers to track temperature over time using a digital temperature recorder. On arrival, the disposable digital recorder is handed over to quality inspectors so that proper temperature during transportation can be verified. Often, within range temperature is evaluated prior to unloading the trailer and before any other quality determinants are assessed. This is particularly exciting in light of the fact that real-time remote sensing could provide an opportunity for suppliers and customers to turn back or reject loads that experienced temperatures that will cause a quality rejection before the load arrives the customer's facility. The ability to reject loads on the basis of adherence to temperature ranges during transit would greatly improve the transportation process for customers, enabling them to receive replacement product sooner, and for suppliers, allowing them to reroute potentially damaged product sooner and avoid burning fuel unnecessarily. These themes are discussed in detail later.

4. METHODOLOGY

As cited before, today companies are living the technology revolution in many ways, especially in terms of information availability and data generation. Companies have access to huge amounts of information from all echelons of the supply chain. The question today is not so much how to get information but how to interpret and use information as a competitive advantage for companies.

This thesis attempts to define how to use and interpret information generated by systems that include temperature sensors and GPS monitoring during refrigerated transportation of packaged salads. The study covers the transportation segment from the company XYZ's processing and distribution center to the customer's (retailer) distribution center.

As explained in detail in the prior section, the real-time data analyzed in this thesis is sent via cellular phone technology to a central server, owned by the logistics information provider. Data transmissions occur at least every fifteen minutes but often are logged more frequently, every time an alarm event happens. Currently this information is used by company XYZ's transportation department for two primary purposes:

1. To receive alarm messages, mainly for above-range temperatures in an effort to prompt corrective actions.
2. To identify the geographic position of the trailer.

The objective of this project is to expand the way in which real-time logistics information can be used and analyzed to generate value for the supply chain. Value can be created for both the manufacturer and the retailer.

This chapter is divided in two groups of sections; the first one include sections 4.1, 4.2, 4.3 and 4.4 and offers a brief overview of the data that was gathered for this project, and a short explanation of the way it was organized to allow for the regressions. The second group includes sections 4.5 and 4.6, which first summarizes the main statistics of the data that was finally grouped and used to run the regressions, and then explains the way the regressions were performed.

4.1 OVERVIEW ON THE ROLE OF TEMPERATURE

4.1.1 EFFECT OF TEMPERATURE ON SHELF LIFE

The first step to organize the data was to define the kind of relationship that could be expected between temperature and shelf life. By establishing this potential relationship it was possible to know how to better organize the data. Regarding Fuel Consumption, the relationship with temperature and other variables was more intuitive and easier to organize.

According to the Literature Review of this thesis, there is a realistic opportunity to model the effects of out or range temperature on the shelf life of salad products. Products with evidence of low shelf life are more likely to be rejected at the time of delivery. Therefore, it may be possible to model the effects of out of range temperature on the rejection of salad products on a per load basis. In order to establish a link between out of range temperature and shelf life, the results of two experiments were undertaken. Both experiments were performed by company XYZ's quality assurance department and provided for this thesis' usage.

The first test assessed how exposure of lettuce to temperatures above desired maximum temperatures affects produce shelf life. In this experiment salad products were exposed to four periods of higher than optimal temperature and then taken back to optimal temperature ranges. Seven different series of four out or range temperature combinations were performed and are shown in the Figure 3 below as T1 through T7. Note that the corresponding temperature for each T-test has been included in the legend in Fahrenheit degrees. Product decay was evaluated daily beginning day seven after exposure had begun. The experiment proved that exposure to high temperature causes irreversible damage to salad products with regard to shelf life. Specifically, regardless of the fact that products are brought back down to acceptable temperatures, days of shelf life have already been lost. A further complication is that shelf-life loss is not immediately evident, and is only manifest at end of a product's shelf life. This may occur a week or more after the high temperature exposure event and means that the effects of temperature damage are often not apparent until products are on store shelves or in the possession of consumers.

In the experiment, shelf life was measured in terms of decay. As defined for the purposes of the study, product rejection criterions were established as four decayed leaves per bag. Any product containing more than four decayed leaves per bag was considered to have exhausted its available shelf life. The results are shown in Figure 3 below:

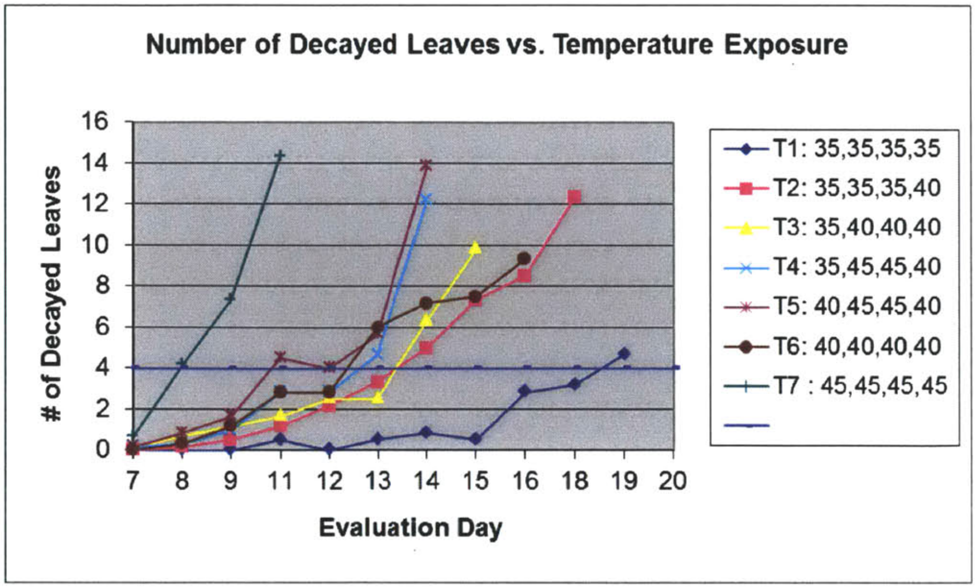


Figure 3 # of Decayed Leaves

Figure 3 above provided by company XYZ

Note that the higher the initial exposure temperature, the faster exposed product lost all its shelf life. In the case of T7 for example, the most temperature abused sample, product shelf life was exhausted in eight days. T1, the least abused sample, retained its shelf life for over 18 days. Accordingly, this study clearly displays the relationship between temperature and shelf life loss in bagged salad products.

The second experiment tested lettuce shelf-life loss behavior at four different temperatures (45°F, 50°F, 55°F, and 60°) over three different exposure durations (5h, 9h, and 12h). For each observation, the product of temperature increase and time was called "Abused Temperature Area". The average decay percentage was calculated for each Abused Temperature Area and a regression analysis was built to establish the correlation between both variables. Results are shown in Figure 4 below:

Abused Temp. Area	% Decay per bag
1,791	7.6%
2,299	10.4%
4,374	10.5%
1,998	11.0%
4,388	12.0%
4,946	12.9%
4,122	13.2%
9,496	13.7%
12,177	14.8%
2,725	10.7%
8,155	11.6%
15,711	17.9%

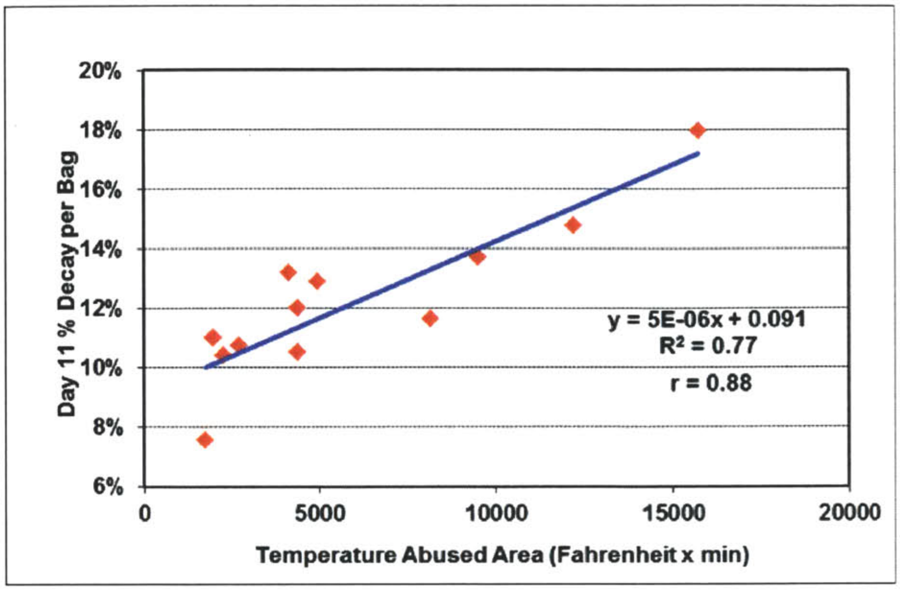


Figure 4 Temperature Abuse Test Results

Figure 3 above provided by company XYZ

Using the results of both experiments as a starting point for analysis, an initial theoretical framework was constructed with the intention of predicting rejections as a function of out or range temperature. However, it is important to highlight that, as explained before, these experiments exposed the product for temperature abuse over long periods of times (above 5 hours), which was later found to be very unlikely to happen during the transportation process.

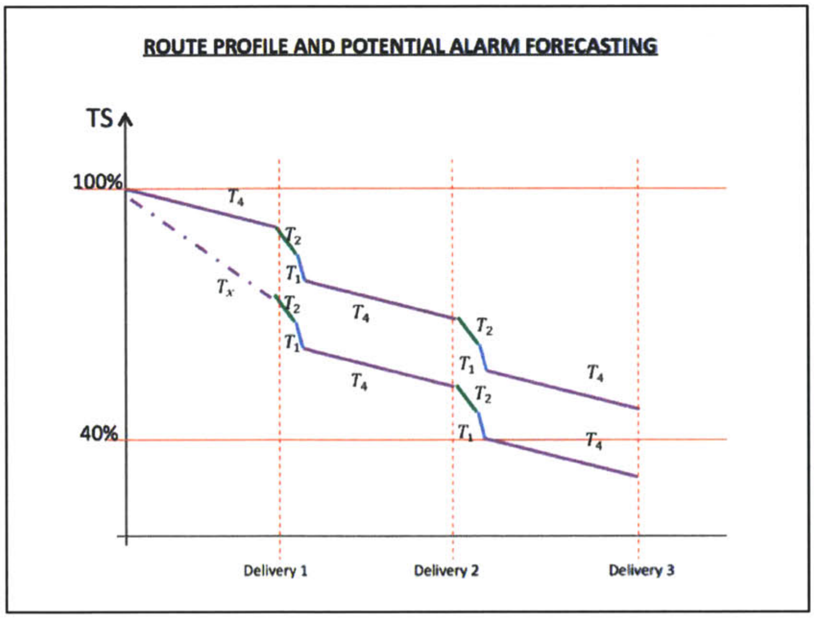


Figure 5 Potential Alarm Forecasting

With the insights obtained from the experiments described earlier, a hypothesis of the relationship between produce temperature-exposure (during transportation) and rejections probability was built and is illustrated by Figure 5 were:

- The y-axis represents the produce's shelf life starting at 100%, and being rejected by customers at 40%.
- The slope of the curves represents the temperature to which the product is being exposed during transportation (the more negative the slope, the higher the temperature).
- The vertical dashed lines represent a customer stop.

Note that the Figure 5 is for explanatory purposes only and does not illustrate an actual transportation event. Note that there is a decrease in shelf life towards delivery three. The lower curve represents a load with sub optimal temperature management. Note the change in slope of the line segment labeled Tx compared with optimal T4 located above it. This change in slope is meant to signify an out of range temperature occurrence that lasts only as long as line segment Tx. All other line segments have the same temperature signature as the optimal line located above. As this load progresses towards delivery three, the damage done by poor temperature management initially has caused a decrease in product shelf life as compared with the optimal load. Accordingly, the lower curve displays that this load's product arrives to delivery three with less shelf life than required by specification. One of the main purposes of this thesis is to find out whether a predictive model based on this simple rationale is possible to build, including other variables that might also influence customer rejection.

4.1.2 TEMPERATURE BEHAVIOR ON TRANSPORTATION

The vast amount of data available for analysis demands a simple graphical analysis approach. Accordingly, a detailed methodology was developed that allows for loads to be graphically displayed one by one for further investigation. The graphs presented below display a diversity of information most important of which is temperature on the y-axis and time on the x-axis. All data points are composed of temperature readings taken at sequential intervals by the PAR system. In addition to temperature time information, the graphical representations of loads also display location information segregated into three buckets:

- Trailer at company XYZ denoted in blue
- Trailer traveling denoted in purple
- Trailer at customer location denoted in orange.

The graphs that follow and the detailed descriptions of the loads depicted provide insights into how to interpret the data on hand and its potential uses and limitations.

4.1.2.1 LOADS WITHIN DESIRED STANDARDS

Trip Profile 1 below details a trip taken to one customer. The green line displays the trailer's average temperature throughout the duration of this load. Point 1 indicates the beginning of the pre-cooling cycle. Note that at Point 2 pre-cooling is complete and the temperature has stabilized within the acceptable range. Note that at Point 3 the load has left company XYZ and is moving to a customer location. Temperatures stay within acceptable limits while product is unloaded at the customer location. Lastly, the reefer unit is turned off when the trailer is completely empty and temperatures begin to rise within the trailer (Point 4).

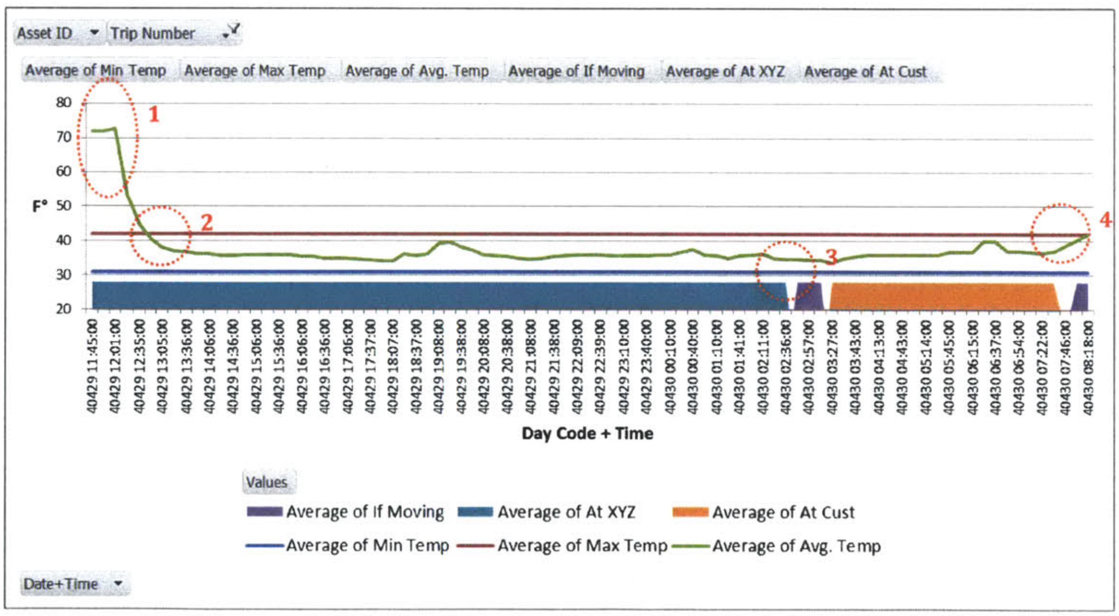


Figure 6 Trip Profile 1

The load displayed in Trip Profile 2 below shows a two stop trip with good performance characteristics. Point 1 denotes the beginning of the precooling cycle. Point 2 indicates an inflection point in temperature at the first customer location and subsequently temperatures rise, signaling that doors are open during unloading. At

Point 3 temperature is almost out of range and then begins to drop back down to the reefer unit set point. This is probably an indication that the doors have been closed prior to departure. This cycle of increasing container temperatures during unloading is characteristic of reefer units not being shut down when trailer doors are open. This is discussed in detail in another section of this thesis; however, graphical depictions of this phenomenon are displayed in the load illustrated in Trip Profile 2 below. Point 4 marks reefer unit shut down after complete unloading and subsequent departure to company XYZ's pick location.

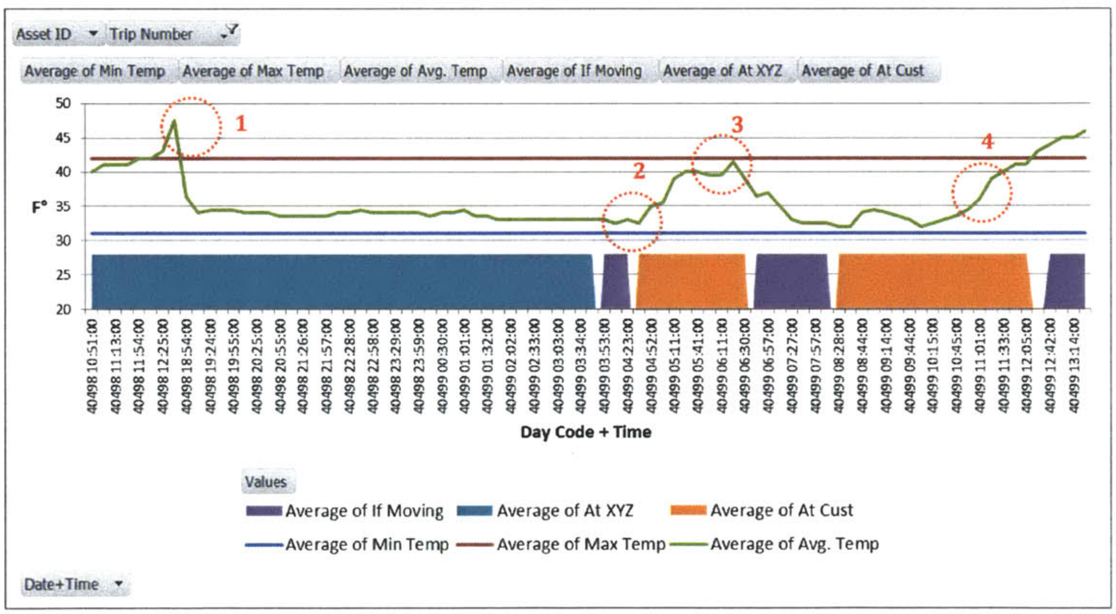


Figure 7 Trip Profile 2

Trip Profile 3 below details information gathered on another well controlled load. Points 1 and 2 denote small bumps in temperature most likely corresponding to door opening events at each of the three customers serviced by this load. Point 3 denotes the point at which product is most likely fully unloaded at customer 3 and accordingly, the reefer unit is shut off and temperatures rise.

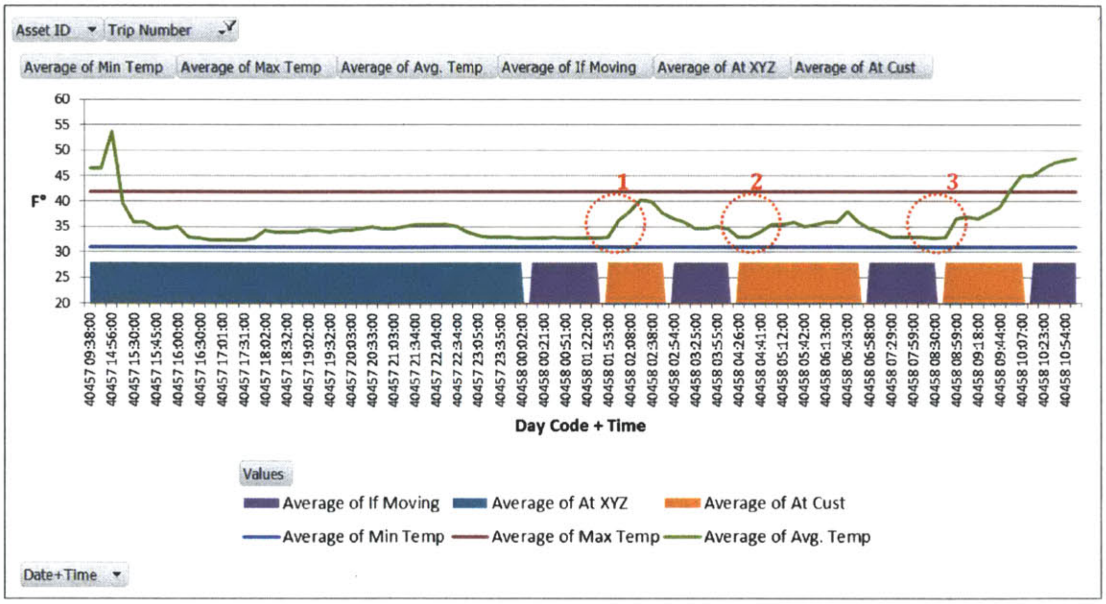


Figure 8 Trip Profile 3

4.1.2.2 LOADS OUT OF DESIRED STANDARDS:

Instances of less than ideal loads also occur. For instance, Trip Profile 4 below shows signs of excessive fuel consumption. Note that this asset stays at company XYZ from 1PM on the first day until 8AM the following day, all the while burning diesel fuel. Although no out of range temperatures are recorded during this load, fuel is apparently consumed unnecessarily.

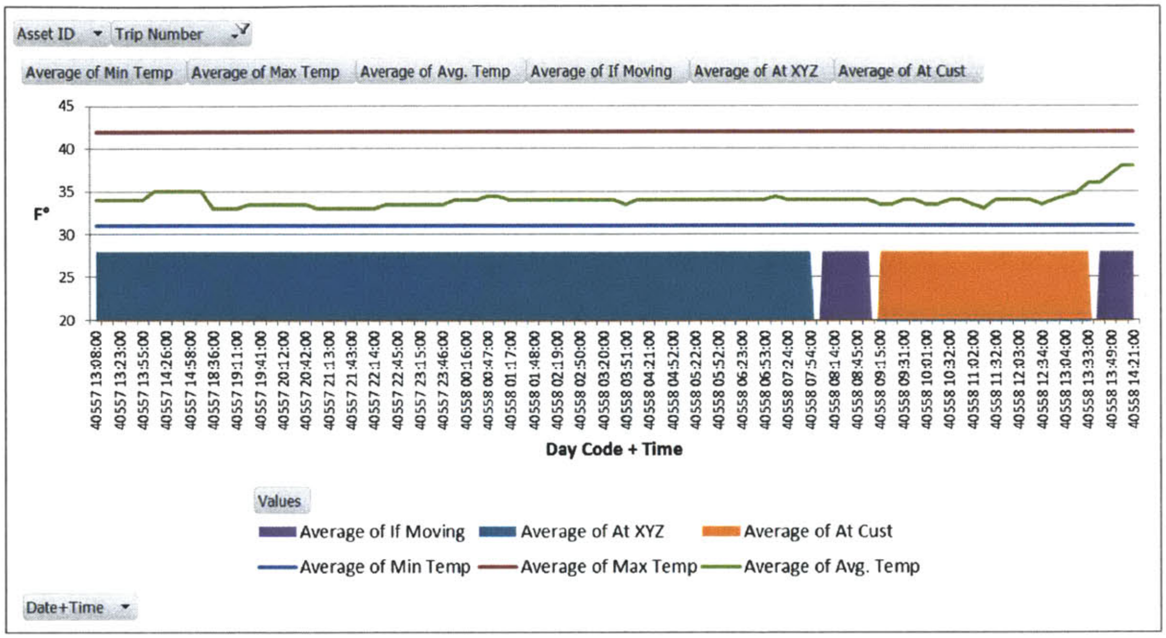


Figure 9 Trip Profile 4

Trip Profile 5 below details a load that does display and out or range temperature events. Point 1, shows a significant out of range temperature that could potentially result in product damage. This out of range temperature event last over two hours, begins at the first customer visited and is not addressed until after the second customer is visited.

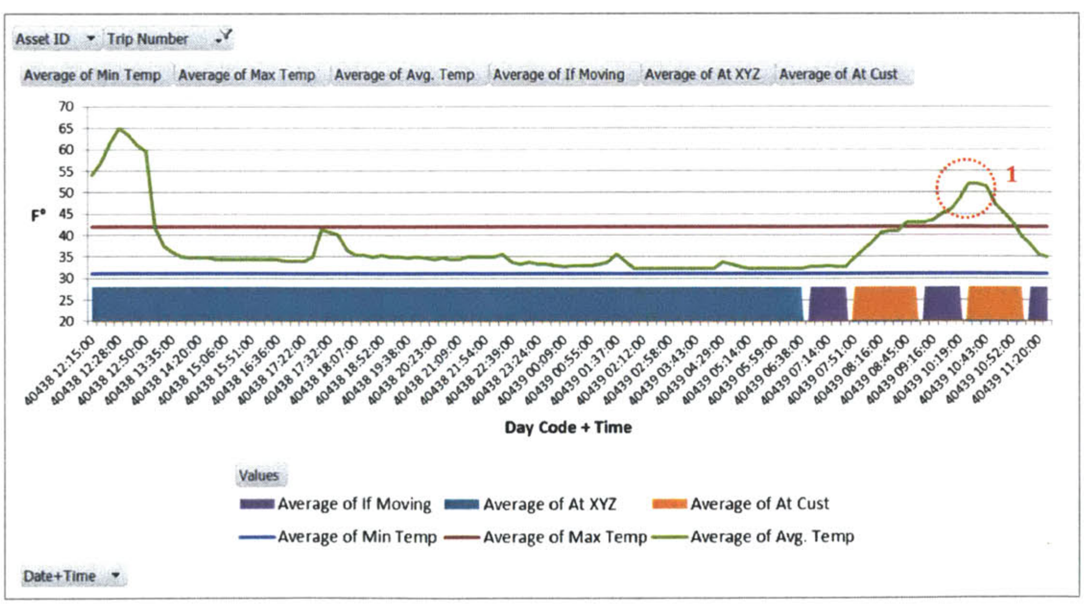


Figure 10 Trip Profile 5

Trip Profile 6 below details another out of range temperature occurrence. Point 1 denotes an out of range temperature event at customer 2. Point 2 denotes an additional short out of range event at customer 3 and then subsequent cooling down of the trailer. Product delivered to customer 3 will have experience two out of temperature events during transportation and is could be a risk for having a shorter shelf life or showing signs of package fogging or moisture in the bag.

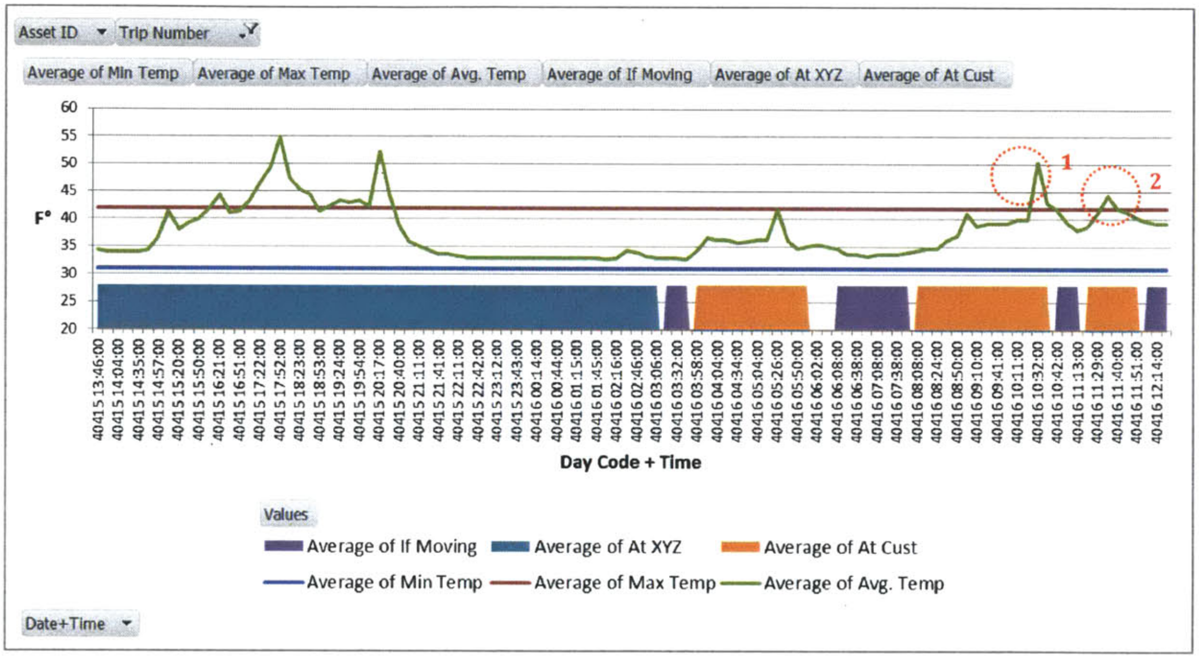


Figure 11 Trip Profile 6

As noted in the literature review, salad products are equally susceptible to low temperature exposures as they are to high temperature exposures. Trip Profile 7 below details one such low temperature exposure. Point 1 indicates a significant low out of range event while at company XYZ but there is no evidence that the product is already in the trailer. Point 2 denotes a significant low out of temperature range event while at the first customer location. Point 3 denotes another low out of temperature range event while at customer 2. These low temperature events may be the result of reefer doors being opened during significantly cold days. Again, reefer units are not shut down resulting in fridge ambient air being drawn into the trailer compartment.

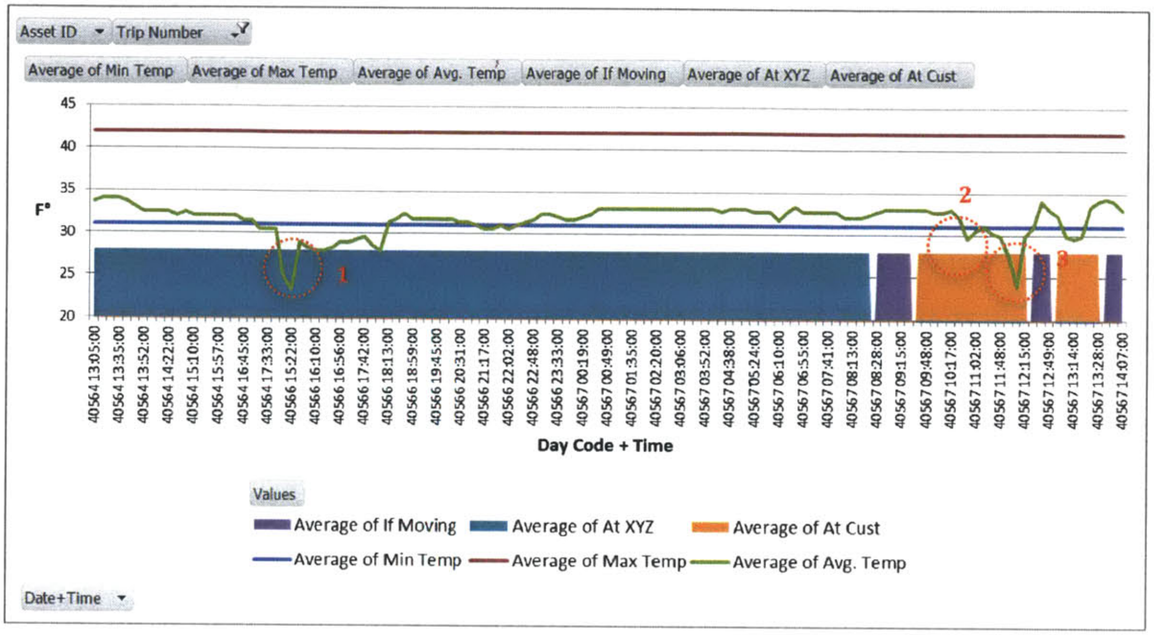


Figure 12 Trip Profile 7

4.2 DATA STRUCTURE

Three different sources of data were identified as relevant for the thesis; the primary data table is called REEFER data and refers to the real-time-transportation data that is provided by Par logistics to the FFV manufacturer. The REEFER table comprises the core data backbone that the other two data tables have been linked into. The other two data tables also contain information regarding the refrigerated transportation process. The first is called Transportation Management System (TMS) data and the second is called REJECTIONS data. Although all three data tables refer to the transportation process from company XYZ's distribution center to the customer's distribution center, they are not linked within company's information systems. The methodology used to link the three data tables is described later in this document.

The information analyzed covers the period beginning in August of 2010 and runs through January of 2011. This range contains a wide spread in ambient temperature, which allows for seasonality to be taken into account as the data set covers both warm and cold seasons of the year. Ideally it would be better to have twelve months of data in order to more effectively track the effects of seasonality in the data. As this was not possible, not much emphasis is placed on the effects of seasonality in the investigation.

Three company XYZ's distribution centers are covered by this study during the period mentioned above: Location 1 in Pennsylvania, Location 2 in Georgia, and Location 3 in Texas. Each distribution center serves a defined region and set of customers.

Each of the data tables mentioned above has a number of fields, some of which were not relevant for the study. A more detailed description of each of the data tables and of the rationale used to link them follows.

4.2.1 REEFER DATA

As mentioned before, this data table is the most important table of the study. The company currently uses temperature control and monitoring services provided by Par Logistics. Temperature is measured for all shipments that travel on Par sensor equipped trailers that move from salad processing facility to Customer's Distribution Center. Each truck measures temperature with three temperature sensors that constantly record temperature and send data every fifteen minutes to a central server through cellular phone technology. Each wired trailer is equipped with three temperature sensors. One temperature sensor is located in the nose of the trailer, a second sensor is located in the middle of the trailer, and the third sensor is located in the tail of the trailer (several feet in from the door). The information sent to the central server is of multiple types including temperature, geographical position, time, reefer unit status and other data types described later. Much of this information is required to verify the trailer's temperature behavior during product transportation from factory to customers' DCs. By analyzing the data it is possible to identify periods of time when temperature was outside of recommended temperature ranges required to preserve product shelf life and ensure that product remains within established specifications. If it is possible to build mathematical models to predict relevant logistics KPI's from this information, corrective actions could be taken on specific loads before trailers arrive to customer's facilities. The information gathered by each Par system is sent to a server on an almost real-time basis. Around 300,000 REEFER data records were originally gathered for the time period and loads included in this study. The most relevant fields contained in this real-time information database are described in Table 4 below.

Table 4 PAR System Variables

VARIABLE	DESCRIPTION
Asset ID	Refers to the trailer's unique identification number. Thirty-six trailer ID numbers were included in the analysis.
Event Time (EDT)	Contains information about the date and time a certain information record was collected.
Event Type	Contains a short description of the event that occurred prompting a record to be generated. If there is no special event happening after fifteen minutes, a record is automatically generated and transmitted using STATUS REPORT as the default description
City	Refers to the city the trailer was in at the moment the record was sent to the central server.
Latitude and Longitude	Contains the geographical coordinates the trailer was on when the record was sent to the central server.
Speed	Speed of the truck at the moment the record was sent to the central server.
Fuel Level (Gal)	Reefer unit fuel level, in gallons, at the moment the record was sent to the central server.
Ambient Temp (F)	External temperature at the moment the record was sent to the central server.
Door Status	Records whether the trailer's door is opened or closed at the moment the record is sent to the central server.
Operating Status	Refers to the status of the reefer unit. For example, the unit may be ON, OFF, or IDLE at moment the record is sent to the central server.
Zone1 Return Temp (F)	This field refers to the return temperature that is recorded by the reefer unit at the moment the record was sent to the central server. The return temperature is the temperature of the air flow (within the trailer) that moves back to the reefer unit. This return air is collected at floor level at the nose of the trailer through an air return slot.
Zone1 Set Point (F)	The Set Point refers to the temperature level the reefer unit is programmed to maintain inside the trailer.
Zone1 Supply Temp (F)	This field contains the value of the air temperature that is actually supplied by the reefer unit to the trailer.
Probe Temp1 (F)	Refers to the temperature measured by the probe located in the middle of the trailer at the moment the record was sent to the central server.
Probe Temp2 (F)	Refers to the temperature measured by the probe located at the back of the trailer at the moment the record was sent to the central server.

4.2.2 TMS DATA

The TMS data is obtained from the Transportation Management System and is inputted into the system before the trailer leaves the company's distribution centers and after the product is delivered to the customer. Each line in the TMS data corresponds to a Shipment number, which is equivalent to a customer order. A Load refers to a delivery trip done in a single trailer. One Load number may contain one or many Shipment Numbers, and a single Customer can have more than one shipment (or order) being delivered in the same Load. Usually a Load will last less than one day, will serve two customers, and accordingly will make two distinct stops. The relevant fields included in the TMS data table are operational and detailed in Table 5 that follows.

Table 5 Transportation Management System Variables

VARIABLE	DESCRIPTION
Pick Plan Calendar Week	Refers to the week number of the year in which the load occurred.
Pick Plan Calendar Date	Day in which the Load was picked from company XYZ's processing facility.
Load ID	Unique load identification number assigned by the TMS system.
Carrier Name	Name of the carrier who was responsible for a specific load. The load can be carried via private fleet or external fleet. Nevertheless, almost all loads carried on intelligent trailers are carried by company XYZ's private fleet.
Shipment	Details the unique shipment identification number assigned by the TMS system.
Pick Location	Name of company XYZ's processing facility and distribution center at which the trailer is loaded.
Load Original Pick City	City location of the company's processing facility at which the trailer is loaded.
Drop Location	Name of the customer location where a load is delivered.
Drop ID	Unique identification code of a customer drop location receiving a shipment.
Drop City	City name of a customer drop point.
Drop Appointment To Times	Refers to the date and time the shipment delivery is committed to the customer.
Drop Arrival Times	Refers to the actual date and time the shipment was delivered to the customer. The driver inputs this information manually.
Mileage Amt. (Load Level)	Mileage traveled by the trailer per load.
Shipped Qty. Amt.	The amount, in number of cases, shipped per shipment number.
Total Pallets	Refers to the number of pallets per shipment number.
Shipped Weight Lbs. Amt.	Weight, in pounds, per shipment number.

With the TMS data it is possible to find the most important routes in terms of number of loads and total amount of product shipped to each customer within the analyzed period of time. However, the major constrain in this data table is that no direct field to link TMS data with the REEFER data exists. Accordingly, several linking rules, described later in the document, were applied in order to create usable data to work from.

4.2.3 REJECTIONS DATA

The rejections data table contains the number of cases rejected by customer per item ordered. Rejections should mostly happen due to the identification of out of specification products at the moment of deliver. A rejection can be partial or total, and occurs at the moment the customer inspects a load on the dock of its distribution center. If a customer accepts a shipment, downloads the product, and decides to return it after the truck left the facility, this event is considered a return and not a rejection. Returns are out of the scope of this thesis.

Rejection information is manually inputted into a central Excel database by company XYZ's customer service team. The most relevant fields included in the data table are listed in Table 6 below.

Table 6 Rejection Data Variables

VARIABLE	DESCRIPTION
Delivery DATE	Date the rejected shipment physically returned to company XYZ's distribution center.
Customer	Name of the customer that rejected the product.
Customer #	Unique identification code of the customer to whom a shipment is delivered.
Item	Name of the rejected item.
SKU / Code	Unique code of the rejected item.
Prod Day	Production Date of the rejected item.
Cases Rejected	Cases rejected per item.
Total cases on order	Total cases per item in the original order. With this field and Cases Rejected, it is possible to calculate the percentage rejected.
PRIMARY - Customer Rejection / Quality Complaint Detail	Short description of the root cause for the rejection event. The customer service team regarded this information as not being too precise.

As in the case of TMS data, the REJECTION data table cannot be linked directly with the REEFER data. Nevertheless, using the date and the Customer Location, it is possible to link each rejection to its corresponding Load Number. Also, the data table contains information about returns but as stated before, returns are excluded from this analysis. The reason is that returns can happen several days after the product has been received and stored by the customer and from the available data, it is impossible to determine which Load Number each return corresponds to. Additionally, detailed returns information would be essentially useless for the purpose of modeling unless data that described product exposure conditions during storage at the customer DC level was collected.

4.3 IDENTIFICATION OF RELEVANT VARIABLES

As established in the thesis problem statement, this study seeks to determine whether it is possible to build a mathematical model to predict product rejections and reefer unit fuel consumption as a function of the variables that exist as a part of the real-time transportation REEFER data and the data that can be linked to it. The starting point is to identify the variables that might play a relevant role during the transportation process and could affect Rejections and Fuel Consumption regardless of whether they exist in the available databases described before. This is accomplished by using four sources of interviews:

- Personnel from company XYZ's transportation operation.
- Personnel from company XYZ's customer service department.
- Customers.
- Personnel from the reefer manufacturing company.

Twenty two variables were identified based on responses to questions posed during the interview process. It was also determined that these variables could have different degrees of separation with regard to the effect each variable has on determining rejections and fuel consumption. To be able to classify the variables into different levels depending on how directly they affect rejections and fuel consumption, some basic concepts of System Dynamics were used (46 Sterman 2000).

The approach taken made it also possible to identify the impact polarities each variable has on any other variable as well as the impact each variable has on the two target variables, Rejections and Reefer Fuel Consumption. Also, using this approach it was possible to classify the variables into a hierarchy depending on whether a variable has a direct or indirect effect on rejections and fuel efficiency. The levels assigned within this hierarchy to each variable are defined as follows:

- **Level 1:** Assigned to the two target variables being analyzed: Rejections and Reefer Fuel Consumption.
- **Level 2:** Assigned to variables that directly affect Rejections and Fuel Consumption.
- **Level 3:** Assigned to variables that directly affect Level 2 variables.

No additional lower levels were added to the analysis in order to limit the scope of the research to a manageable range. Using the System Dynamic approach, a diagraming exercise was done in which links and polarities for each variable were graphed (46 Sterman 2000). This diagram was used to define the hypothesis of how the system should work. In other words, this methodology was of help to understand how all the identified variables are interrelated. Only five variables were linked without having polarities since they represent categorical and not numerical variables, these were:

- Customer
- Asset Number
- Time of the Year
- Driver Name
- Pick Location

No reinforcing or balancing loops were identified in the system. Below is the generated system diagram, Figure 13.

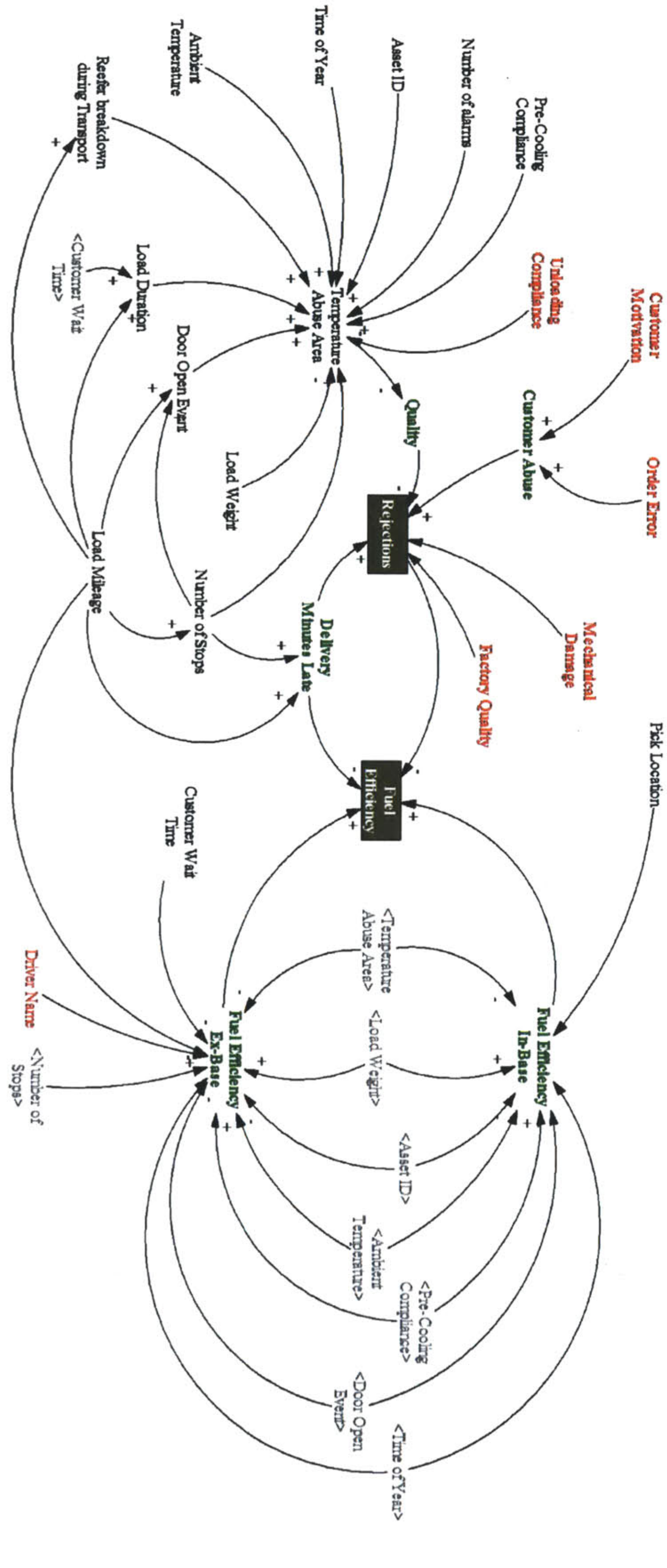


Figure 13 System Dynamics Model

In Figure 14 the resulting scheme of all variables and their relation to the problem statement can be found. Values for variables represented in red boxes were not possible to obtain from the available data tables.

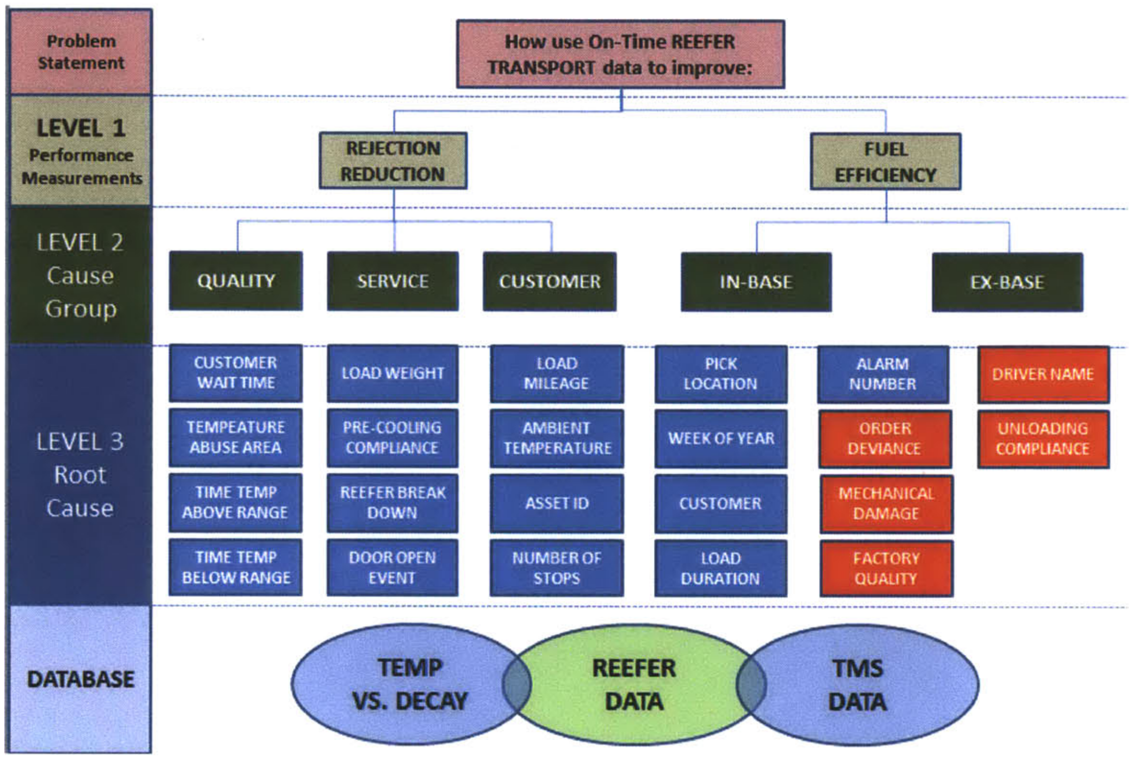


Figure 14 Scheme of Variables

Other types of classifications were assigned to the variables to obtain a better understanding on their utility in the model. Below is a summary of the three types of classifications that were used:

1. Responsible: Indicates the entity that has the highest responsibility stake for adjusting or controlling the variable. The value "External" means that the variable is uncontrollable by any of the entities considered in this research, for example Ambient Temperature.
2. Included in the Dataset: The value will be YES if the variable can be found or calculated form the available databases.

3. <u>Direct field or Calculated Field:</u> This classification indicates whether the value of the variable can be obtained directly from the available tables as a field value, or has to be calculated form available information found within the three main databases.

A summary of the variables that were identified can be found in Table 7 below.

Table 7 Identified Variables Summary

ID	Variable	Level	Responsible	Included in Dataset?	Direct or Calculated Field
1	Customer Wait Time	3	Customer	Yes	Calculated
2	Temperature Abuse Area	3	Carrier	Yes	Calculated
3	Time with temperature above range	3	Carrier	Yes	Calculated
4	Time with temperature below range	3	Carrier	Yes	Calculated
5	Load Weight	3	Company	Yes	Direct
6	Pre-cooling Compliance	3	Company	Yes	Calculated
7	Reefer Breakdown	3	Carrier	Yes	Direct
8	Door Open Event	3	Carrier	Yes	Direct
9	Load Mileage	3	Company	Yes	Direct
10	Ambient Temperature	3	External	Yes	Direct
11	Asset ID	3	Company	Yes	Direct
12	Number of Stops	3	Company	Yes	Calculated
13	Pick Location	3	Company	Yes	Direct
14	Week of the Year	3	External	Yes	Direct
15	Customer	3	Customer	Yes	Direct
16	Load Duration	3	Carrier	Yes	Calculated
17	Alarm Number	3	Carrier	Yes	Direct
18	Order Error or Deviance	3	Customer	No	N/A
19	Mechanical Damage	3	Carrier	No	N/A
20	Factory Quality	3	Company	No	N/A
21	Driver Name	3	Carrier	No	N/A
22	Unloading Compliance	3	Customer	No	N/A

As can be seen by comparing the table above with the list of the variables contained in the data sets, not all the variables that were identified as having a potential relevance for the analysis exist in the real-time transportation data or in the data tables that can be linked to it. These variables are detailed in Table 8 that follows.

Table 8 Irrelevant Variables

VARIABLE	DESCRIPTION
Order Error	This variable indicates the relative amount by which a specific order is different from the average order a customer places. Errors that are much greater or smaller than the average order required by a customer might be an error, thus a higher probability of rejection could be expected.
Mechanical Damage	This variable would indicate the degree to which the product suffered from mechanical damage during the loading or transportation process, thus resulting in a higher probability of visible damage and accordingly rejection.
Factory Quality	Although company XYZ has a quality process implemented to identify non-compliant ingredients entering the manufacturing process and non-compliant products exiting the manufacturing process, detailed statistics regarding this process are not available. It would be useful to incorporate such information in the model with the specific intent of quantifying how far from the specification limits are the outputted products contained in each load.
Driver Name	According to the surveys, driver's behavior could be an explanatory variable for both fuel consumption and rejection patterns. However, it was not possible to obtain this information from the available data.
Unloading Compliance	Refers to temperature unloading compliance at the customer's unloading dock. This information is not available from the existing data. From the existing information it is possible to know when the trailer is in the customer's facility but it is not possible to know the exact moment at which the product is being unloaded.

Although not all variables identified as having the potential of affecting the two target variables (fuel consumption and rejections) can be encountered in the available data tables, it should be possible to explain a percentage of the behavior of these two target variables from the other available variables. This thesis focuses on running different regression techniques with the variables that are present in the existing data tables. By doing so, it is possible to determine whether the existing data is sufficient to build a useful quantitative model to predict and improve rejections and fuel efficiency. In other words, one of the expected results of this thesis will be to confirm whether the fields contained in the typical real-time information data can be used to add operational value other than providing basic asset location and temperature information.

4.4 DATA COMPILATION FOR REGRESSION:

To obtain a useful and reliable database for purposes of running the desired regressions, it was necessary to first clean and link the original three data tables. Then the variables that could not be directly obtained from the available data had to be calculated. The detail of the process followed to achieve this objective can be found in **Appendix 9.1**. The list below summarizes the steps that were followed:

- Align time periods available among data tables.
- Filter pick locations not included in the three data tables.
- Exclude returns from the rejections data table.
- Eliminate duplicates from each data table.
- Link the three data tables
- Calculate variables feasible with available data

4.5 DATA SUMMARY:

With regard to the transportation information available, each trailer is considered a company asset and is referred to by the general term "asset". Additionally, each asset is deployed from one of three dedicated processing plants. Each processing plant produces bagged salad products and accumulates stores of finished goods in a climate-controlled co-located warehouse. Each processing facility is referred to as a pick location. Figure 15 details the number of assets currently working, based at each location.

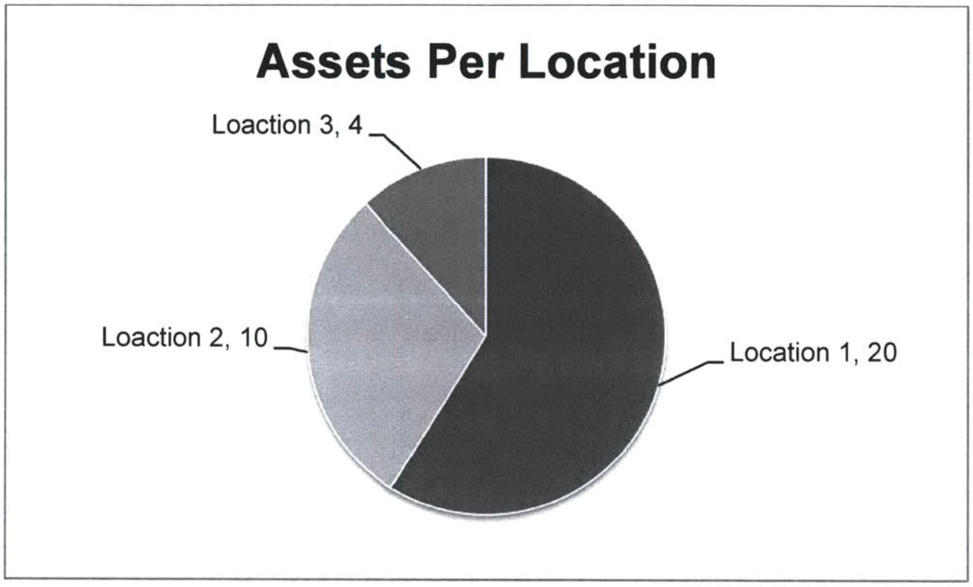


Figure 15 Assets Per Location

Figure 16 details the number of loads distributed from each pick location found in the available data.

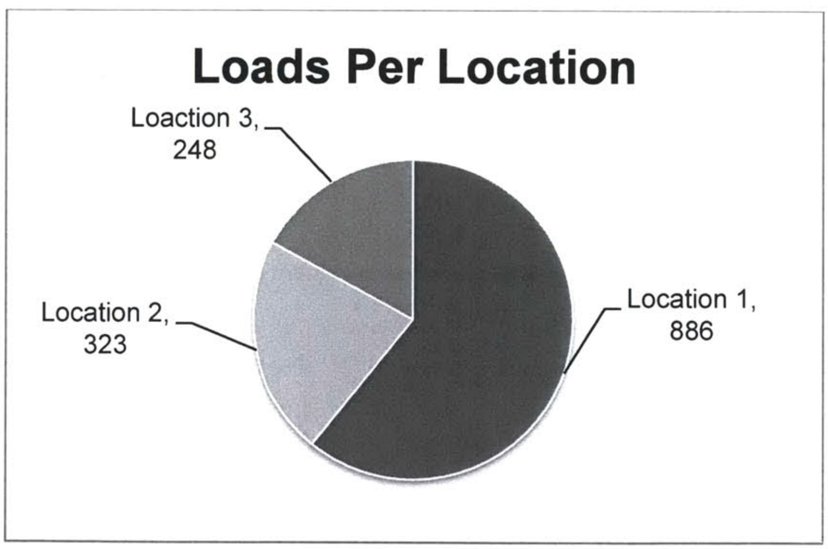


Figure 16 Loads Per Location

Most data analysis is segregated by pick location. Each location has its own set of characteristics which would be lost if data was viewed in aggregate. Summaries of the three pick locations' key distinguishing characteristics are displayed in Table 9 below, where Avg. accounts for Average, Std. Dev refers to the standard deviation of each measure, and COV accounts for Coefficient of Variance.

Table 9 Key Location Characteristics

Pick Location	Load Mileage			Load Duration (Hrs.)			Weight (Tons)		
	Avg	Std Dev	COV	Avg	Std Dev	COV	Avg	Std Dev	COV
Location 1	261	146	0.56	15.4	8.5	0.55	5.1	1.8	0.35
Location 2	127	63	0.50	9.7	5.8	0.60	5.3	3.5	0.66
Location 3	156	132	0.85	9.0	6.3	0.69	8.4	3.4	0.41
All Locations	214	143	0.67	13.1	8.2	0.63	5.7	2.9	0.50

Pick Location	Fuel Consumed per load (Gls)			Number Stops			Cust. Dwell Time (Hrs.)		
	Avg	Std Dev	COV	Avg	Std Dev	COV	Avg	Std Dev	COV
Location 1	14.5	8.6	0.60	1.7	0.7	0.42	6.8	3.6	0.53
Location 2	5.9	3.9	0.66	1.8	0.7	0.41	5.7	4.0	0.70
Location 3	6.7	4.7	0.71	1.3	0.5	0.37	3.3	1.9	0.59
All Locations	11.3	8.3	0.74	1.7	0.7	0.41	5.9	3.7	0.62

To obtain a better understanding of the variability of the typical load weight per pick location, the corresponding histograms are presented as Figures 17, 18, and 19 for Location 1, Location 2 and Location 3 respectively.

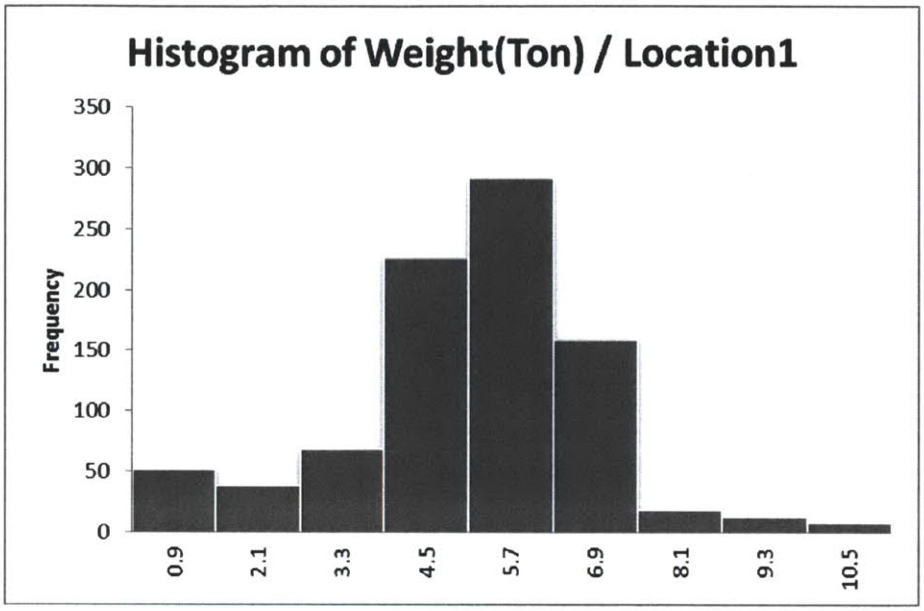


Figure 17 Histogram of Weight for Location 1

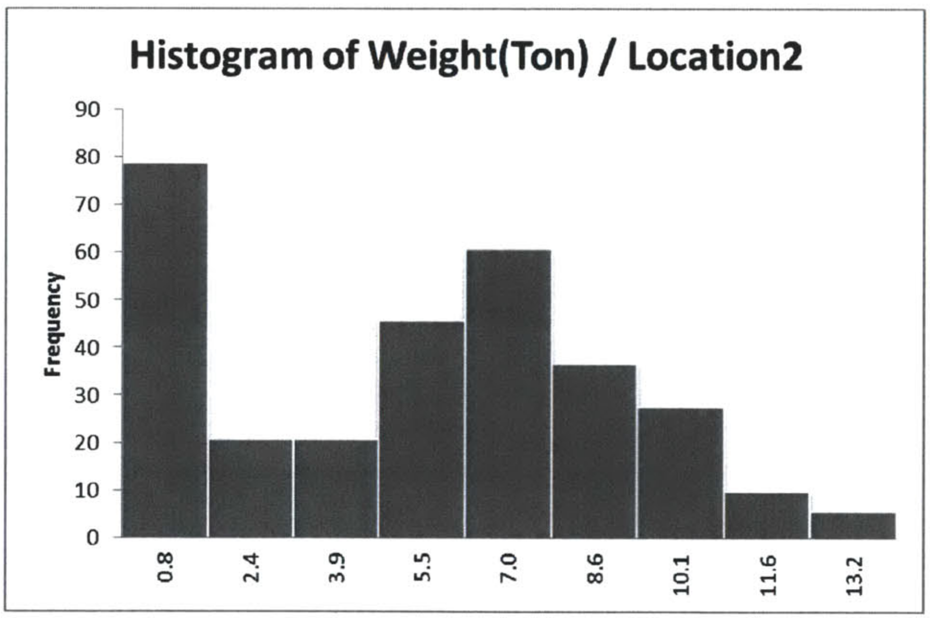


Figure 18 Histogram of Weight for Location 2

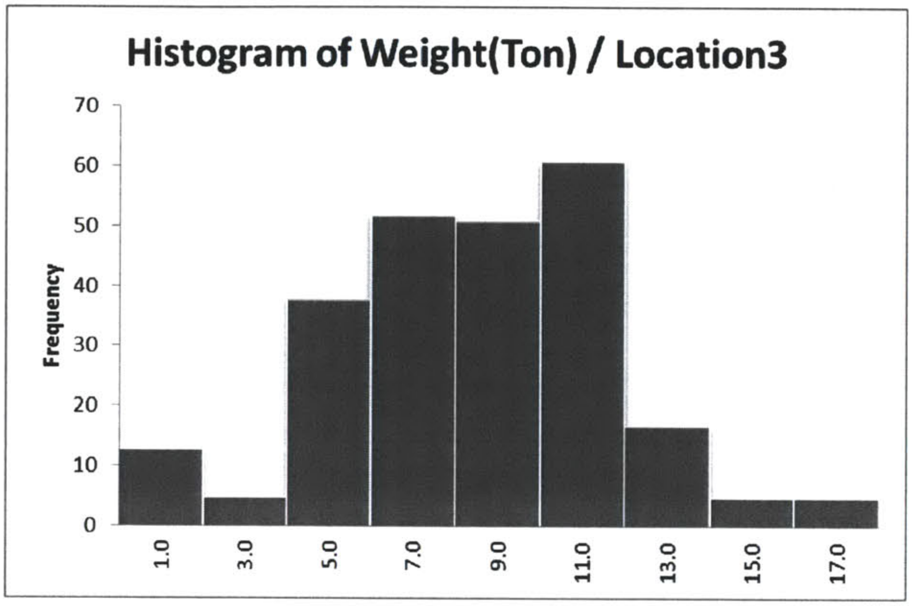


Figure 19 Histogram of Weight for Location 3

As shown by the histograms, Location 1 and Location 3 tend to have a distribution of weight per load with the mean close to the center, similar to a normal distribution. Location 2, however, has a distribution of weight per load that is skewed to the left, with a high frequency of lower weight loads being delivered.

4.5.1 REEFER FUEL CONSUMPTION

Fuel consumption by asset has also been a focus of analysis for this thesis. Basic fuel consumption information can be found in the linked TMS and PAR reefer data available. Figure 20 details the average number of gallons consumed per load per pick location. Note that the average number of gallons consumed by Location 1 is significantly high than at other locations. This is most likely due to the nature of the clients served from Location 1. Location 1 serves mostly grocery customers and accordingly, load sizes are smaller, and load durations are longer.

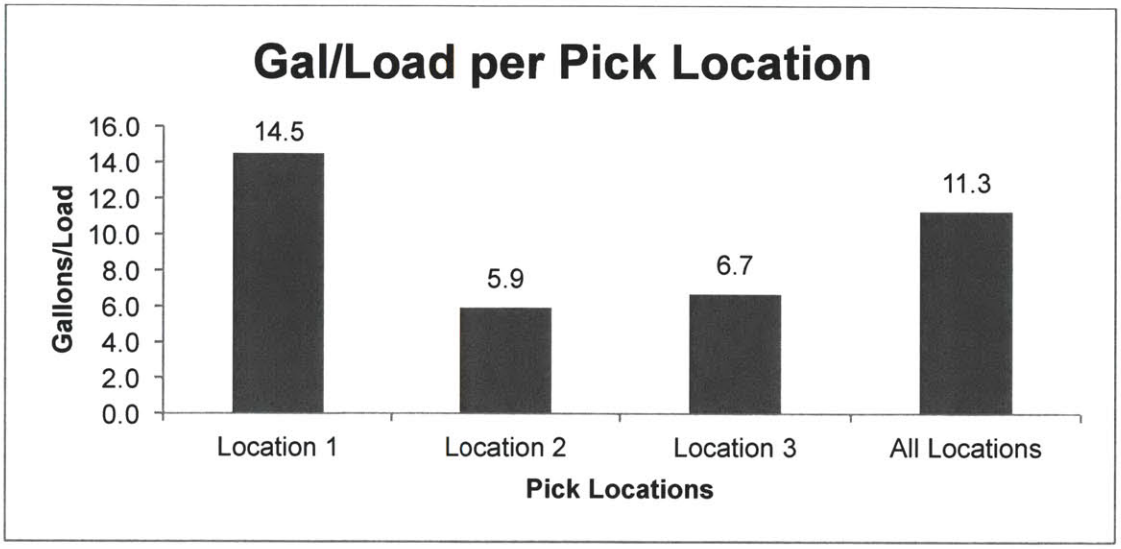


Figure 20 Gallons Per Load

4.5.2 LOAD REJECTIONS

Loads with rejections are not distributed evenly across all pick locations. Rather significant variances in rejections of total loads are seen from location to location. Figure 21 details percentage of total loads with rejections by pick location.

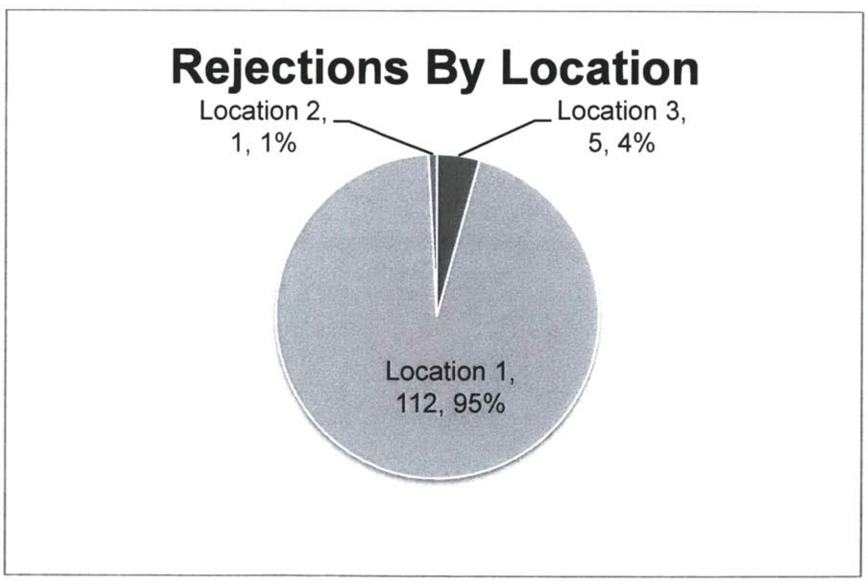


Figure 21 Rejections By Location

Note that only one rejection occurs at Location 2. The single rejection that occurs that originates from this location is not statistically significant to be included in analysis of rejections included in this study.

Figure 22, that follows below, show that the largest percentage of rejections per pick location, as found in the data available for this study, occurred for finished product originating from Location 1.

The high incidence of rejections from loads based at Location 1 may be determined by the type of customers served by this location, the rejection reporting practices and this location's proximity to salad greens growing locations.

Whereas locations 2 and 3 service a blend of food services and grocery clients, Location 1 serves mostly grocery clients and quality conscious customers. The nature of customers serviced from Location 1 and the quality consciousness of these clients may be a factor.

Concerning rejection recording practices, the Rejections Database provides no assurances regarding the quality of the information therein. Variances in the way rejections are recorded are possible. Procedures for recording rejections and their causes are not standardized. All these factors place doubt on the quality of available rejections data.

With regard to proximity to growing regions, Location 1 is significantly father away from these growing regions than the other two locations investigated. As a result, raw materials travel farther to reach Location 1 and accordingly, will have less shelf life on arrival than at the other two locations studied. As noted previously, the transportation leg from farm to processing facility has not been investigated in detail in this thesis, however, it is important to note that factors associated with this transportation segment may have significant effects on load rejections as noted here.

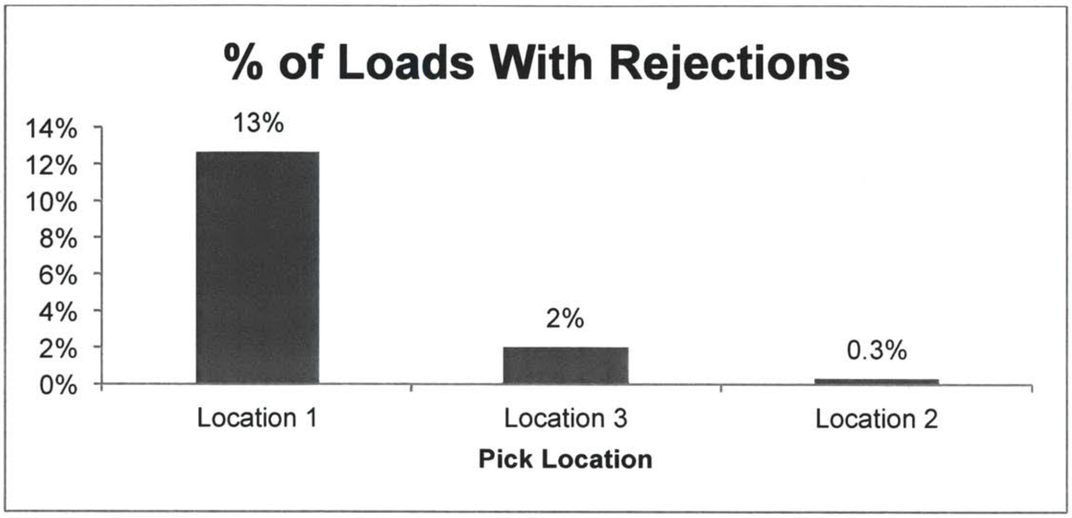


Figure 22 Percent of Loads With Rejections

4.6 REGRESSIONS:

As described in the problem statement, this thesis seeks to investigate how real-time transportation cold chain data in the FFV industry can be used to improve cold chain performance, specifically with regard to Rejections and Reefer Fuel Consumption. To achieve this objective, this thesis uses regression analysis methods.

4.6.1 REGRESSION METHOD OVERIVIEW AND SELECTION PROCESS

As stated in Albright, Winston, Zappe, *Data Analysis and Decision Making*, regression analysis can be applied to many situations as a tool to understand the relationships between variables (26 Albright 2011). Said in other words, regression analysis helps explain how the behavior of a variable is related to the behavior of other variables, and can also be used to predict the variable in question depending on the expected values of other variables.

Every regression analysis has a single variable that is being explained or predicted by the model, and a set of variables that are used to explain the first variable. The variable that the model is trying to explain is called the dependent or target variable, and the variables used to explain the target variable are called the explanatory variables (26 Albright 2011).

This thesis seeks to better understand and predict two different variables:

- Rejections
- Reefer Fuel Consumption

From now on, these two variables are referred to as the target variables. The objective of this study is to understand if it is possible to explain and predict both target variables with the available set of explanatory variables in the provided data. If this objective is possible, the focus of the resulting model should be to minimize the per load expected value of both target variables by controlling the explanatory variables that play the most relevant role during the transportation process.

It is important to highlight that the identification of a useful model to understand both target variables would not necessarily imply causation. In other words, it is possible to find a strong relationship between a target variable and one of the explanatory variables, but this relationship might be is caused by a third variable not included in this model.

Before moving forward it useful to define some additional basic terms used in regression methods:

Observed Value: is the name given to an actual value of the target variable at a specific set of values of the explanatory variables. The observed value is represented by the letter Y (26 Albright 2011).

Fitted Value: is the model's predicted value of the target variable. The fitted value is represented by the letter $\hat{Y}$.(26 Albright 2011)

Residual: is the difference between the actual and fitted value of the target variable. Residuals are represented by the letter e.(26 Albright 2011)

The relationship among the three terms defined above is illustrated by the Equation 1

Equation 1 Observed Value

Observed Value = Fitted Value + Residual

Or Equation 2:

Equation 2 Residuals

$$e = Y - \hat{Y}$$

Least Squares: a method used to find the equation that minimizes the sum of squared residuals. When the model has a single explanatory variable, it is possible to refer to the least squares line, the line that minimizes the sum of squared residuals (26 Albright 2011).

R Square (R^2): percentage of variation of the target variable explained by the regression. R^2 is calculated using the following formula:

Equation 3 R-Squared

$$R^2 = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

where $\bar{Y}$ is the average value of the target variable.

Since there are different variations of regression analysis methods, before selecting the adequate method for each one of the target variables, it is important to define each one:

Rejections: Within the available sample of 1,457 loads, only 118 were identified as having experienced a rejection event, making the sample relatively small. Also, the objective of the study is to build a model that can help understand and predict whether a specific load will have a rejection event or not. The objective of the model is not to predict the expected percentage of rejections versus the total number of cases in a load. Thus, the target variable is binary, reject or don't reject.

Reefer Fuel Consumption: All the 1,457 loads to be included in the analysis have the total fuel consumption per load as a continuous variable. Reefer fuel consumption is calculated as the consumption of fuel, in gallons, from the moment the trailer leaves the company's pick location until it comes back after delivering the produce to its customers. The reefer fuel level can be found directly for each record as a field in the REEFER data table.

As can be seen from the definition above, the main difference between one target variable and the other is that the REJECTION target variable is binary, and the REEFER FUEL CONSUMPTION target variable is continuous. Based on these inherent differences it is possible to choose the appropriate regression methods to be used in this analysis:

Multiple-Regression: This method is suitable for models containing a continuous target variable and many continuous or nominal explanatory variables. The Multiple

Regression method is used to model the REEFER FUEL CONSUMPTION target variable. Below, the general multiple regression equation is presented:

Equation 4 Predicted Y

$$Predicted\,Y = \hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

(26 Albright 2011)

Logistics Regression: This method is useful when trying to decide which group a specific record belongs to (26 Albright 2011). This method estimates the probability that an individual record is in a particular group. The logistics regression is essentially a regression with a dummy (0-1) target variable. The variable indicates whether the record belongs to group 0 or group 1 (26 Albright 2011). For this thesis, group 1 represents the group of loads with a rejection event, and group 0 represents the group of loads with no rejection events.

A special measure of error in Logistics Regression is called Deviance. Deviance is a measure of the deviation of the actual from the predicted values for a model with no explanatory variables and a constant (26 Albright 2011).

The logistics regression model uses a function to estimate the probability p that any observation will be in group 1. If the value of p is greater than 0.5, the record is usually classified as belonging to group 1. The general equation for the probability p of a record being in group 1 is given by Equation 5 below:

Equation 5 Probability p

$$p = \frac{1}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k)}}$$

(26 Albright 2011)

4.6.2 CONSTRUCTION OF THE REGRESSION TABLE

After defining the appropriate regression method to be applied to model each of the two target variables, a regression table was built to run both regressions. The regression table initially contained the 1,457 Loads to be used as the sample size. A list of the steps followed to prepare the regression table is described next.

4.6.2.1 Identify and Assign Values to Dummy Variables

Some of the variables to be used in the regression where categorical or, in other words, could not be measured on a quantitative continuous scale. These variables where converted into a number of dummy variables equal to the number of categories included in each variable. Dummy variables can have values equal to 1 or 0 depending on whether they happened on a specific record or not. For example, the Pick-Location variable had to be transformed into three different dummy variables equal to its three possible values:

- Location 1
- Location 2
- Location 3

Only one of these three dummy variables can occur on a single record, and will have a number 1 assigned. For example, for a Load that is delivered from the Location 1 facility, only the Location 1 dummy variable will have a number 1 assigned, and the other two variables will have a 0 assigned.

It is important to highlight that at the moment of running a regression, the number of dummy variables that are included is one less than the actual number of categories in the categorical variable. This means that the regression will assume the missing dummy variable as the baseline of the model.

4.6.2.2 Generate Scatter Plots to Check for Linear Relationships or Need for Transformations.

As a means to look for patterns of linear relationships between each of the target variables and the explanatory variables, and also to find outliers, individual scatter plots were generated. Some scatterplots showed non-linear regressions, these cases had to be analyzed further. Figure 23, Figure 24, and Figure 25 show some examples of the scatter plots that were generated for the Fuel Consumption target variable.

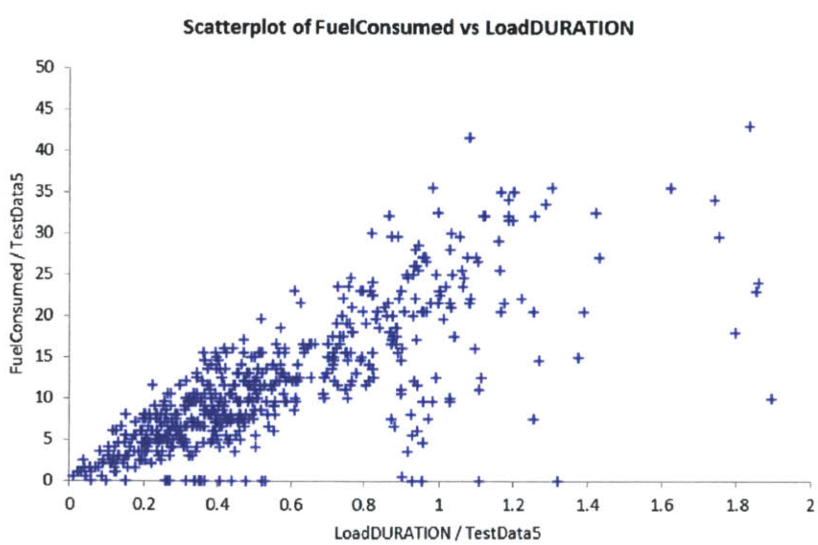


Figure 23 Scatterplot of Fuel Consumption vs. Load Duration

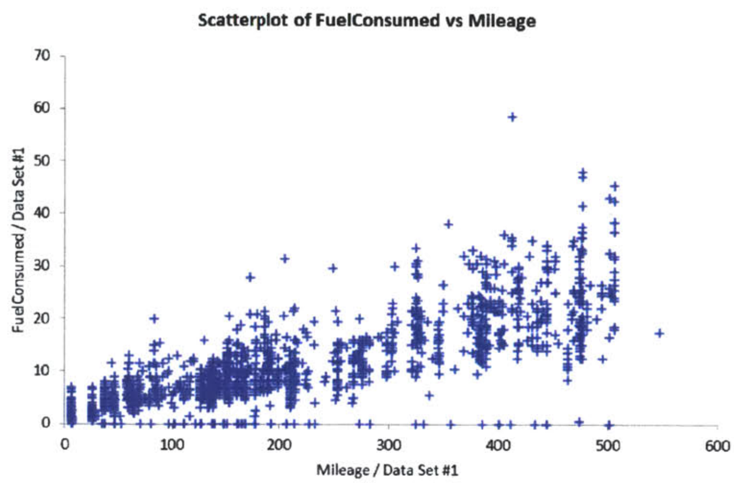


Figure 24 Scatterplot of Fuel vs. Mileage

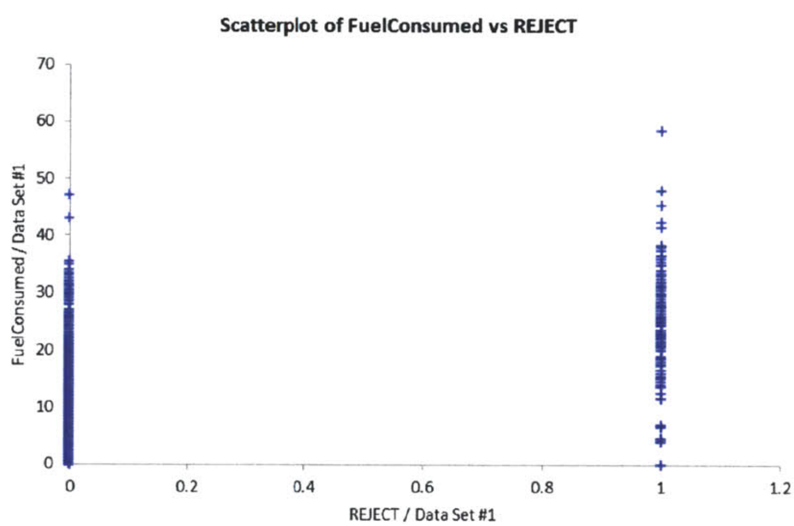


Figure 25 Scatterplot of Fuel Consumer vs. Rejections

4.6.2.3 Establish Best Possible Linear Transformations.

In order to satisfy the assumptions implied in regression methods, to ensure that errors do not increase or have a pattern as a function of the fitted values; several linear transformations were tested. The only transformation that yielded useful results was the square root linear transformation performed on Load Duration. By doing so, the correlation between this variable and the target variable improved from 0.72 to 0.75. Also, the transformation reduced the increasing pattern of variability as the explanatory had greater values.

4.6.2.4 Outliers.

An outlier is an observation that falls outside the general pattern of the rest of the observations (26 Albright 2011). An outlier can be caused by the existence of an error in the database information is pulled from or may be due to exceptional events (26 Albright 2011). Before running the regressions, 18 outliers were identified by generating scatter plots of each target variable versus each of the explanatory variable. The outliers identified had a high probability of being data errors and accordingly were excluded from the loads to be considered in the final regression table.

4.6.2.5 Summary Description of the Final Regression Table.

- The final regression table has 1,439 records.
- All records contain information about Reefer Fuel Consumption.
- Regarding Rejections, only 118 records contain a rejection event.
- Each record of the regression table corresponds to a single Load ID.
- The regression table contains information from August 2010 to January 2011 inclusive.
- The type of the target, or response, variables included in the table are:
    - Fuel Consumed: Continuous
    - Rejection event: Dummy

- The explanatory, variables included in the table are:
  - Asset ID: Dummy
  - Load Duration: Numerical
  - Average Trailer Temperature: Numerical
  - Alarm Number: Numerical
  - Customer Wait Time: Numerical
  - Number of Door Openings: Dummy
  - Number of Stops: Dummy
  - Pre-Cooling Compliance: Dummy
  - Time Above Temperature: Numerical
  - Time Below Temperature: Numerical
  - Time Out of Range: Numerical
  - Temperature Abuse Area: Numerical
  - Pick-Location: Dummy
  - Week: Dummy
  - Month: Dummy
  - Season: Dummy
  - Average Ambient temperature: Numerical

4.6.3 RUNNING THE REGRESSION MODELS

4.6.3.1 Reefer Fuel Consumption Regression.

As mentioned before, the Multiple Regression method was selected to build the model for this variable. To perform the regression, the regression table was divided into

Training Data and Testing Data. The Training Data was used to run the multiple coefficient regression and obtain the model and the Testing data was used to validate the reliability of the model against fresh, unbiased data. The following procedure was used to find the best mix of variables to explain the behavior of Fuel Consumption per Load:

1. To obtain an initial idea about which variables to include in the model, the stepwise feature of statistical software (JMP) was used. With the stepwise approach, the software successively adds and deletes variables from the model until it finds the most efficient combination of variables according to the internal procedure it uses.
2. After obtaining the list of variables from the stepwise process, these variables were combined in different permutations to run several multiple regressions in Stat Tools (a statistical Excel add-in). The objective was to identify the combination that accomplished the following with the least possible number of variables:
    - Maximizes the R-square.
    - Minimizes the Standard Error of the estimate.
    - Minimizes the p-value in the ANOVA table, indicating that under a certain degree of significance, at least one of the variables is helping explain the target variable.
3. After all of the above were provided, the next step was to verify that the coefficients of each of the explanatory variables to be included in the model had a p-value smaller than 0.05. Doing so signifies that there is at least 95%

confidence that the actual coefficient of the explanatory variable is different from zero, and thus, has a contribution in explaining the target variable.

After running several regressions using the Training Data, the best model obtained had an R-square of 0.7 approximately, which means that the model accounts for 70% of the variation of the target variable Fuel Consumption (45 Bertsimas 2004). Additionally, all the coefficients of the explanatory variables selected for the model had p-values smaller than 0.002. In other words, there is more than 99.8% confidence that these coefficients are not equal to zero and, there is more than 99.8% confidence that the selected explanatory variables belong to the regression equation and that the target variable depends in a certain degree to the value of each of those variables (45 Bertsimas 2004). In Table 10 below, these and other performance indicators of the selected regression model are shown, including the Standard Error, and F-Ratio:

Table 10 Fuel Consumption Regression Results

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate
	0.839	0.704	0.700	4.643

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value
Explained	10	43,617	4362	202.33	< 0.0001
Unexplained	852	18,367	22		

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Upper
Constant	-4.356	0.777	-5.61	< 0.0001	-5.880	-2.831
AssetGroup	1.583	0.508	3.12	0.0019	0.587	2.579
StopGroup	2.016	0.552	3.65	0.0003	0.932	3.100
REJECT	5.496	0.639	8.60	< 0.0001	4.241	6.751
TIME OUT OF RANGE	-7.977	1.382	-5.77	< 0.0001	-10.689	-5.264
Mileage	0.016	0.002	6.47	< 0.0001	0.011	0.021
Sqrt_LoadDuration	15.396	1.671	9.21	< 0.0001	12.116	18.677
Season = Autum	2.030	0.350	5.80	< 0.0001	1.343	2.717
Season = Summer	2.738	0.550	4.98	< 0.0001	1.658	3.818
Interaction(Mileage,PickLocation = Location 3)	-0.011	0.002	-4.77	< 0.0001	-0.015	-0.006
Interaction(Mileage,PickLocation = Location 2)	-0.016	0.003	-5.42	< 0.0001	-0.022	-0.011

To validate the selected model furthermore, the equation that resulted from the regression was tested using the testing data. As explained before, the testing data was not used to obtain the original model, and was randomly extracted from the original database to ensure unbiased segregation of training and testing data.

Equation 6 Predicted Fuel Consumption

$$\begin{aligned} &Predicted\ Fuel\ Consumption \\ &= -4.36 + 1.58AssetGroup + 2.02StopGroup + 5.50Reject - 7.98TimeOutofRange \\ &+ 0.016Mileage + 15.40LN_{LoadDuration} + 2.03Autum + 2.74Summer - 0.011Miles \\ &\cdot Location\ 3 - 0.016Miles \cdot Location\ 2 \end{aligned}$$

By testing this equation in the testing data, it was possible to confirm very similar results to those obtained when the same model is run against the training data. The R-square obtained was 0.744 approximately, and the Standard Error dropped from 4.64 to 4.11. Furthermore, the ANOVA F-Ratio and the p-value resulted in 164.2 and 0.0000 respectively, suggesting even more strongly the good level of significance of the model. Table 11 summarizes these results.

Table 11 Regression Significance Results

Parameter	Value
Avgerage Observer Value (Avg. Y)	11.3
Total Variation (SST)	37,319.1
Sum of Squared Residuals (SSE)	9,554.2
Sum of Squares due to Regression (SSR)	27,764.8
Number of Explanatory Variables (k)	10
Sample Size	576
Degrees of Freedom of SSE (n-k-1)	565
Standard Error of Estimate (Se)	4.11
R-square (R2)	0.744
Means Square SSR (MSR)	2776.5
Mean Square SSE (MSE)	16.9
F-Ratio	164.2
p-value	0.00000

Finally, to test and stress the validity of the model even more, two groups of scatter plots were generated. The first was a single plot (Figure 26) to graph the residuals against the fitted value to verify that residuals are randomly located around zero (26 Albright 2011).

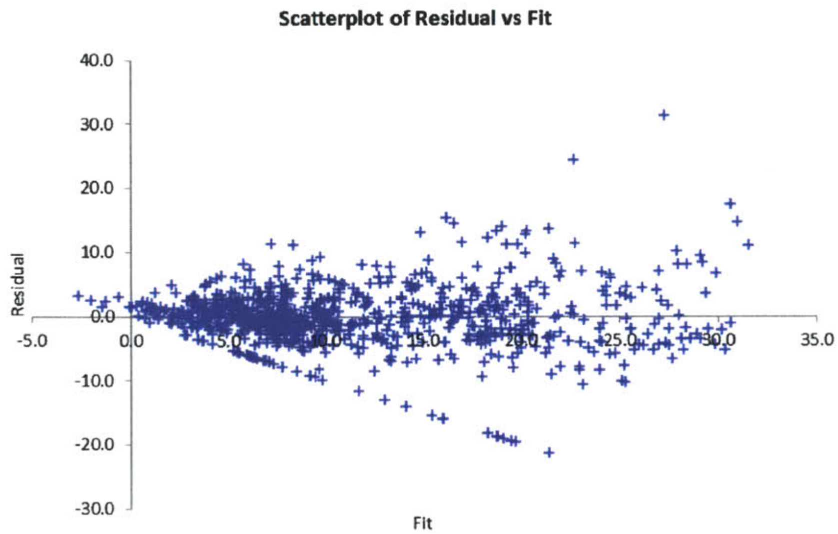


Figure 26 Scatterplot of Residual vs. Fit

The second group of scatter plots was generated to test for homoscedasticity, which establishes that the variability of the target variable should be the same for all the values of each explanatory variable (26 Albright 2011). Figures that follow below show the scatter plots for each explanatory variable.

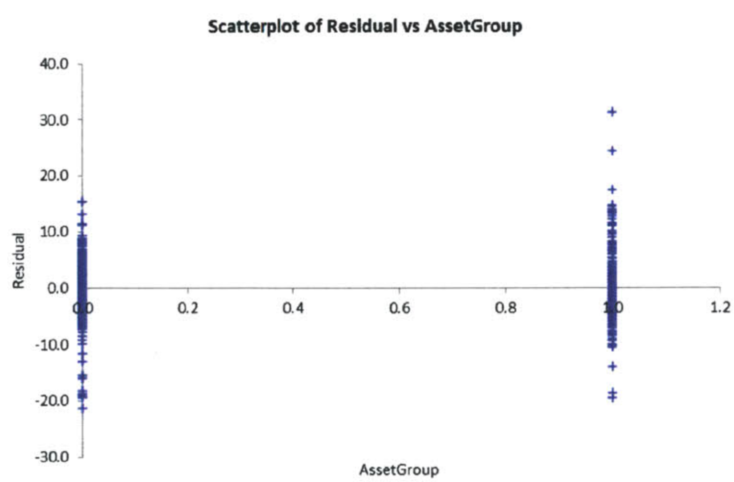


Figure 27 Scatterplot of Residual vs. Asset Group

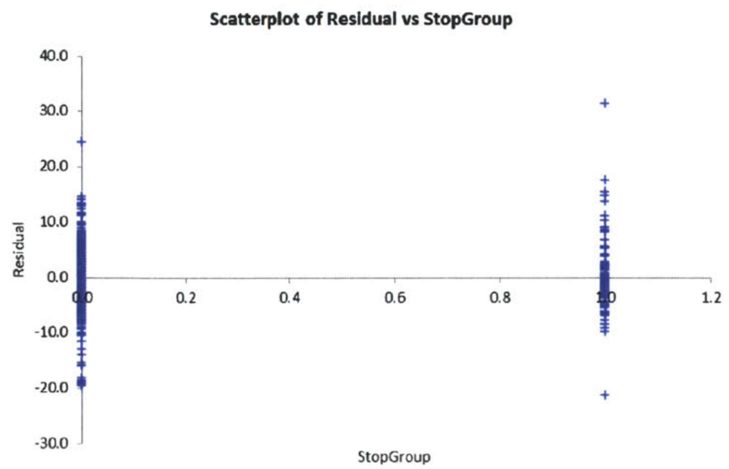


Figure 28 Scatterplot of Residual vs. Stop Group

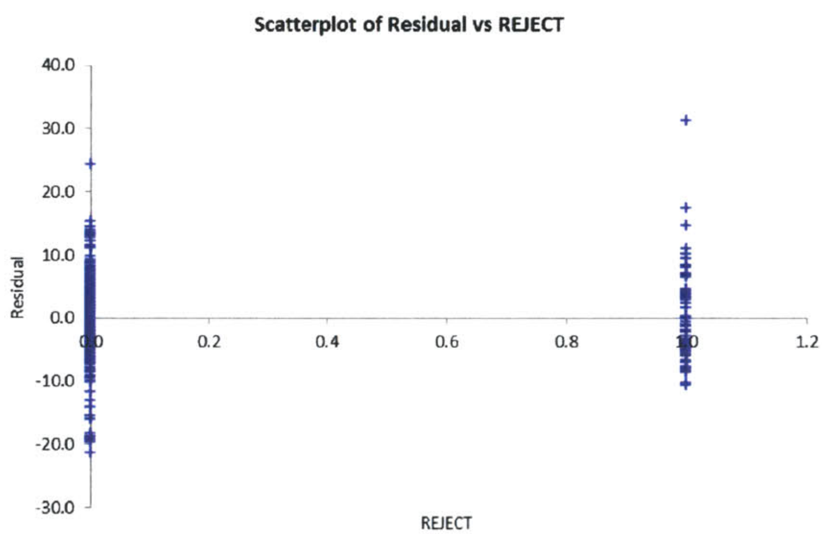


Figure 29 Scatterplot of Residual vs. Rejections

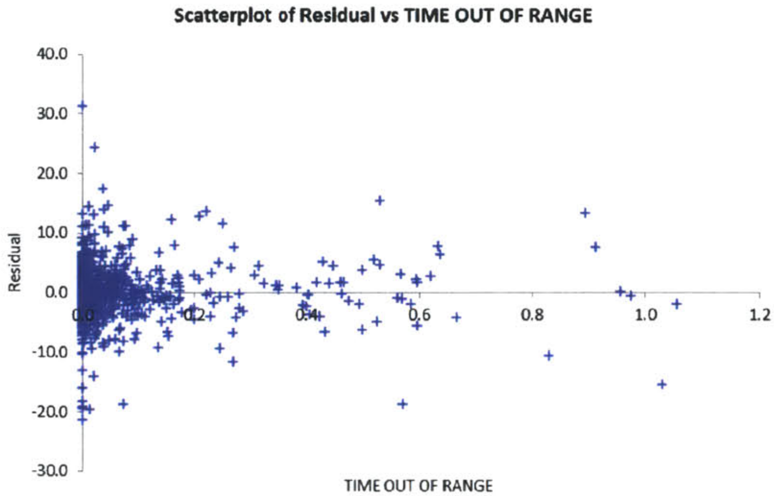


Figure 30 Scatterplot of Residual vs. Time Out of Range

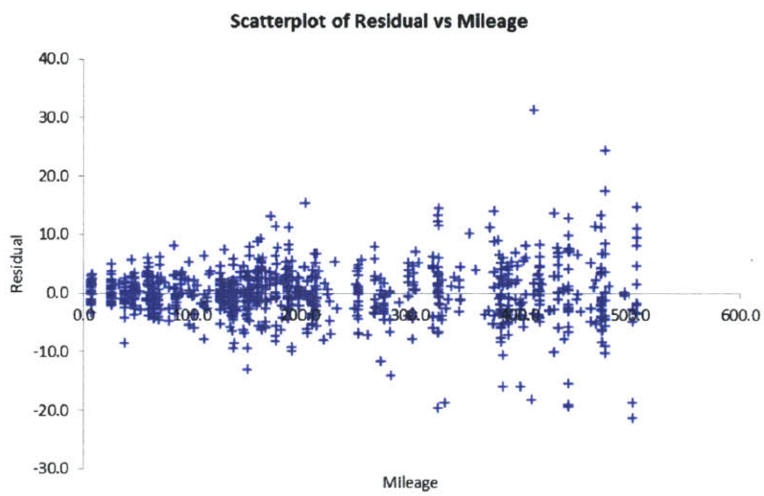


Figure 31 Scatterplot of Residual vs. Mileage

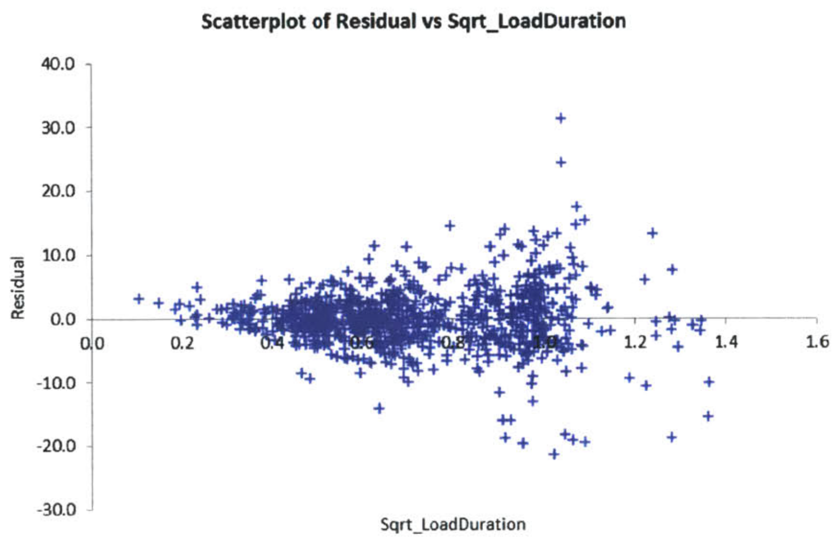


Figure 32 Scatterplot of Residual vs. Square Root of Load Duration

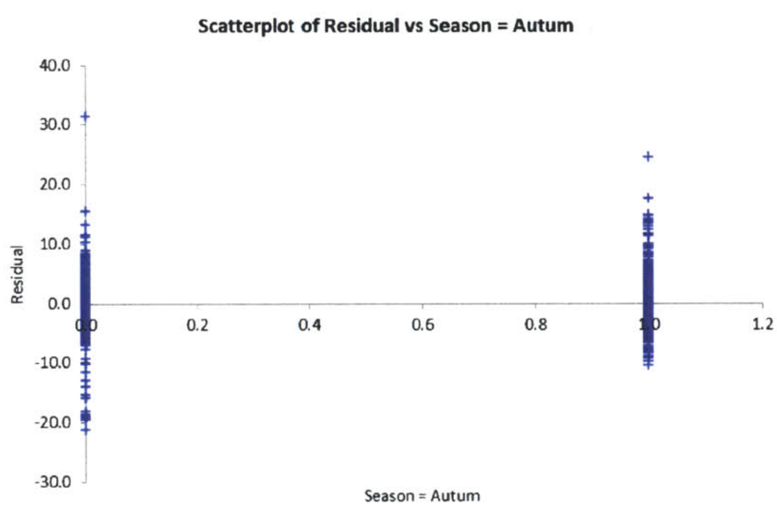


Figure 33 Scatterplot of Residual vs. Season = Autumn

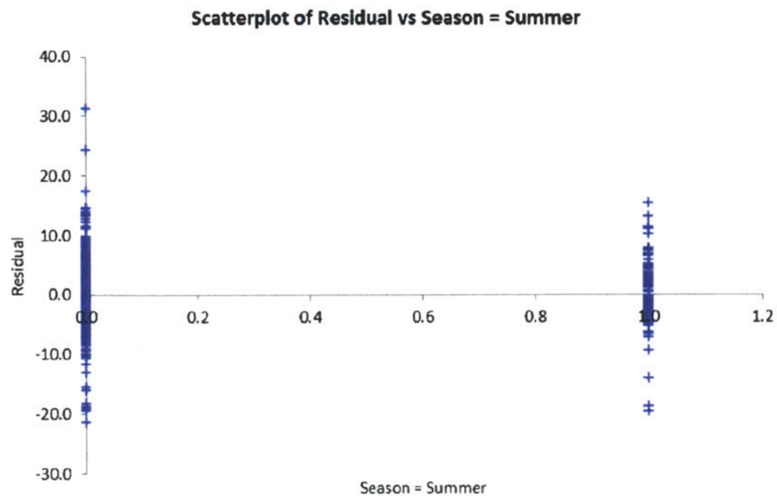


Figure 34 Scatterplot of Residual vs. Season = Summer

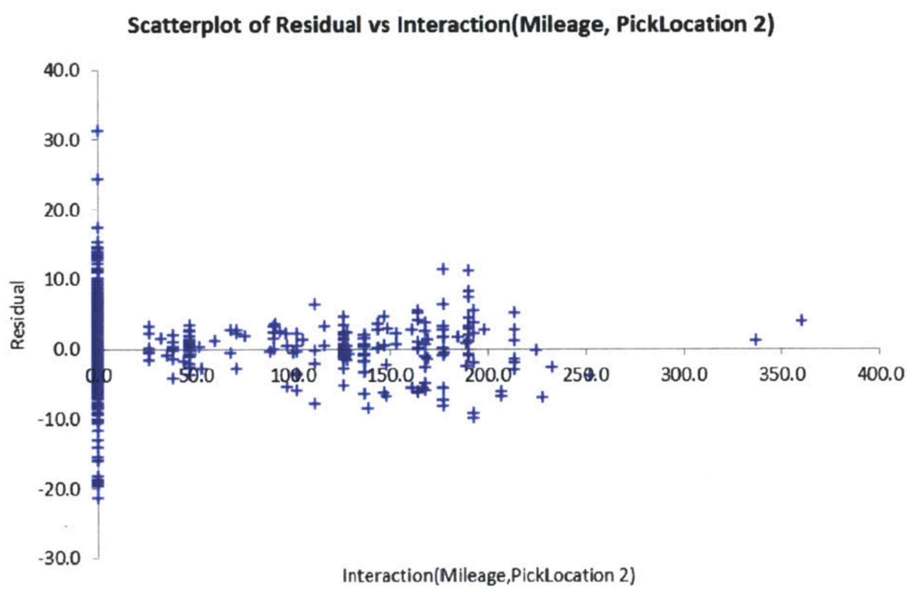


Figure 35 Scatterplot of Residual vs. Interaction Location 2

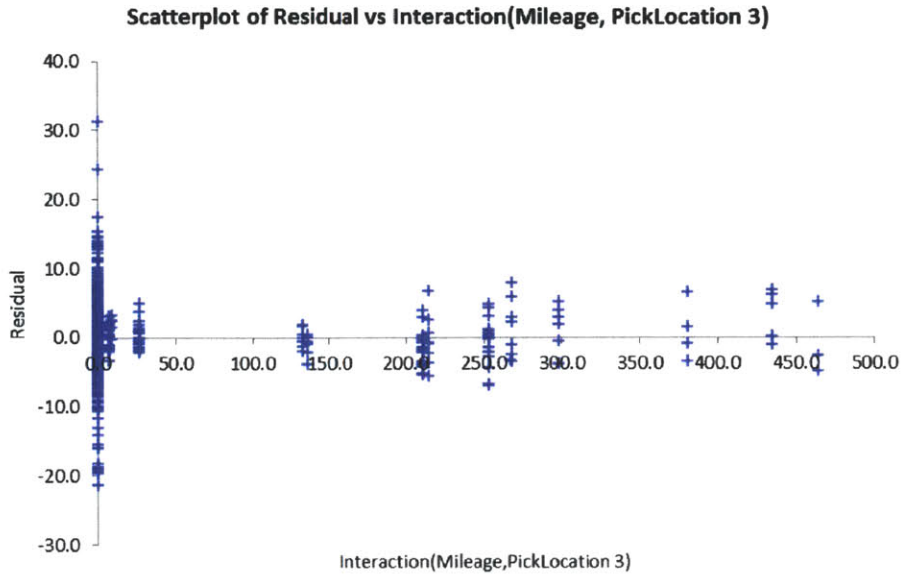


Figure 36 Scatterplot of Residual vs. Interaction Location 3

From the scatter plots it is possible to see that, in the first place, residuals are not clearly randomly located around zero when graphed versus fitted values. Also it is possible to see the lack of perfect homoscedasticity (called heteroscedasticity) for some explanatory variables, especially at extreme values, to the right in the case of Mileage and the Square Root of Load Duration, and to the left in the case of the interaction variables of Pick Location x Mileage. However, none of these non-randomness patterns and signs of heteroscedasticity were strong enough to invalidate the model. This argument is reinforced by the fact that the model worked very well with the testing data. However, the level of heteroscedasticity found for some of the explanatory variables should be used as a message of caution when interpreting the predicted values at extreme points of these explanatory variables.

4.6.3.2 Rejection Regression.

As mentioned before, the Rejections target variable was defined as a dummy variable that may have a value of 1 for a record where a rejection event happened, and 0 for a record were no rejection event happened. Given the characteristics of this target variable, the Logistics Regression method was selected to build the model. To perform the regression, again the regression table was divided into Training Data and Testing Data. The Training Data was used to run the regression and obtain the model. The Testing data was used to test the reliability of the model against fresh data. A particular point for defining the training data in this case was that the number of records with rejections was much smaller that the number of records with no rejections (118 versus 1,321 respectively). Thus, in order to obtain a model that captures the key characteristics present in records with rejections, the training data was arranged to ensure a blend a mixture composed of 50% records with rejections and 50% records with no rejections. The total number of records used in the training data was 118.

The Logistics Regression was run using different combinations of variables. The objective imposed on model construction consisted of maximizing the overall percentage of correct classifications while striving to include the fewest possible number of variables. After running several combinations, the best result was obtained by including the explanatory variables Number of Stops, and Mileage. Table 12 below summarizes the results obtained with the training data.

Table 12 Rejections Logistics Regression Results

Summary Measures	
Null Deviance	163.58
Model Deviance	82.41
Improvement	81.17
p-Value	< 0.0001

Regression Coefficients	Coefficient	Standard Error	Wald Value	p-Value	Lower Limit	Upper Limit	Exp(Coef)
Constant	-5.9403	1.213	-4.8958	< 0.0001	-8.318	-3.562	0.003
NumStops	1.0536	0.437	2.4095	0.0160	0.197	1.911	2.868
Mileage	0.0144	0.003	5.5044	< 0.0001	0.009	0.020	1.014

Classification Matrix	1	0	Percent Correct
1	53	6	89.83%
0	7	52	88.14%

Summary Classification	Percent
Correct	88.98%
Base	50.00%
Improvement	77.97%

As can be interpreted from the parameters in the table, with a decrease in Deviance of almost 50%, and a p-value close to zero, the resulting classification model adds value by increasing the level of classification accuracy compared to a random classification. This can also be interpreted from the Summary Classification table, were the classification accuracy compared to the base of 50% improves 77.97%, to 88.98%. Equation 7 obtained from this table is shown below and was used in the Testing Data to verify its accuracy on classifying records as belonging to the rejection group. It is important to highlight that in the testing data, the original ratio of records with rejections versus the total number of records (9%) had to be used to verify the reliability of the model under realistic conditions.

Equation 7 Logistics Regression Results

$$p = \frac{1}{1 + e^{(-5.94 + 1.05 NumbStops + 0.014 Mileage)}}$$

Equation 7 represents the probability that a record belongs to the rejections group. When the calculated probability is higher than or equal to 0.5, the record is classified as belonging to the rejections group, and inverse happens when the probability is less than 0.5. The summary of the error obtained from the classification done over the testing data is summarized in Table 13 below.

Table 13 Logistics Regression Error

	Analysis 1	Analysis 0	Total	% Correct
Actual 1	52	7	59	88.1%
Actual 0	178	482	660	73.0%
TOTAL	230	489	719	74.3%

Summary Classification	Percent
Correct	74.3%
Base	91.06%
Improvement	-18.44%

As shown in the table above, although the model worked well enough in the training data, the result using the Test Data does not provide an improvement over the base case of 91% (in this data set, the base case represents a scenario where every load would be classified as not having rejections). However, by looking closer at the accuracy obtained on the classification made by the model to loads with no rejections, the level is increased to from 91% to 99% (482/489). Further interpretations of these results will be undertaken in the following section.

5. RESULTS AND FINDINGS

This section is divided into two parts. The first part describes the results and findings obtained from the regressions that were performed, and the second part briefly discusses possible operational Key Performance Indicators for refrigerated transportation.

5.1 REGRESSIONS

After performing both regressions for the target variables analyzed in this thesis, it is possible to observe two completely different degrees of usefulness for each model. The Fuel Consumption regression resulted in a model that explains the target variable's behavior with a relevant degree of accuracy; however, the Rejections regression did not result in a reliable model when applying it to the data available. The results for each regression are explained separately.

5.1.1 FUEL CONSUMPTION RESULTS

The model obtained for Fuel Consumption shows that with ten variables it is possible to explain approximately 70% of the Fuel Consumption per Load. Four of these explanatory variables are dummy variables (Asset Group, Stop Group, Reject, Season), and two are Interaction Variables (the combination of Mileage and Pick Location), the rest are numerical variables.

When testing the different combinations of variables, the Mileage explanatory variable alone accounts for approximately 56% of the R-square. This result is intuitive. However other variables, when modeled together with the Mileage variable, improve the R-square by approximately 14%. Also interesting is that each variable on its own plays a significant role in determining the level of fuel consumption of each load. This is evidenced by the low p-values obtained for all variables included in the final model. The following list describes an interpretation of the effect of each variable on Fuel Consumption.

Asset Group: Assets were divided in two different groups according to guidelines defined by the JMP software. The software suggested an even more detailed degree of subdivisions that could potentially increase the level of the R-square to nearly 0.75. This suggests that there is a relevant impact associated with the reefer unit used for each load. This impact may depend on the age or level of maintenance given to the equipment, or on the behavior of the driver that uses the equipment. However, given the available data it is not possible to define the common characteristics inherent in the sub-groups as suggested by the software. Accordingly, only the first level of clustering was used in the analysis. This first level proved the potential relevance of equipment condition with regard to fuel consumption. The coefficient of this explanatory variable (1.58) can be interpreted to mean that every time a Load is carried in an asset that is part of group 1, the consumption of 1.58 additional gallons with standard error 0.5 gallons should be expected. Again, from the data available for this thesis, it is not possible to know what the common physical characteristics of these units are.

Stop Group: As could be guessed by common sense, the number of stops that occurs in each load also impacts the level of fuel consumption. This variable grouped the number of stops in two parts: Loads with 1 or 2 stops, and Loads with 3 or more stops. The coefficient of this explanatory variable (2.02) suggests that every time a Load has more than 2 stops, the consumption of 2.02 additional gallons should be expected. The standard error of this coefficient is 0.55 gallons.

Reject: According to the model, Loads with rejections result in additional gallons of fuel consumed. Every time a load experiences a rejection, the consumption of 5.5 additional gallons should be expected. The standard error of this coefficient is 0.64 gallons. This process could be related to the fact that after a rejection occurs, the trailer has to get back to the original pick location loaded with produce, and thus, with the need of maintaining the reefer unit working.

Time out of Range: According to the model, loads that experience temperature out of range will use less fuel than had been required under normal conditions. Although the coefficient looks negatively big (-7.98), it is important to take into account that the unit of measure of this variable is in days. Usually the time out or range will be around 1.5 hours, so really the impact of an hour out of range should be around 0.33 gallons, a relatively a small amount. Also, it is important to clarify that this is a variable than when improved will generate a higher consumption of fuel, since the idea is not to have time with temperature out of range.

Mileage: The model suggests that loads will consume 0.016 gallons for each mile. The average mileage per load is around 214 miles. It is important to reinforce that reefer-unit consumption is independent of truck fuel consumption. Thus, the consumption per mile can significantly change depending on the amount of time the trailer spends at each customer all the while maintaining produce within temperature specifications. The standard error of this coefficient is 0.002 gallons.

Load Duration: The total load duration is mainly a function of the mileage and amount of time the trailer spends waiting and unloading produce at the customer's facility. During this time, produce is maintained within temperature specifications. A transformation was done to this variable to improve its linear relationship with the target variable, and to reduce the way in which residuals tend to increase as the Load Duration increases. The following example is used to help understand the real meaning of the coefficient obtained for this explanatory variable using the model. Let's say Fuel Consumption needs to be calculated for a load that had duration of six hours. To calculate the expected fuel consumption driven by this variable, first the six hours are transformed into the fraction of a day, then the square root of this number is calculated, and the resulting number is multiplied by the coefficient to obtain the expected fuel consumption fixing all other variables (see Equation 8). To give some perspective of the proportions this variable has in the data available, the average Load Duration in the regression table is of 0.55 days (around 13 hours).

Equation 8 Fuel Consumption Results

$$fuel\ consumption = 15.4 \cdot \sqrt{6/24} = 7.7\ gallons$$

Season: Regarding the season of the year, the model included only six month of data from a single year. With these data, the model suggests that loads occurring during autumn and summer will have an additional consumption of 2.03 and 2.74 gallons respectively. The standard errors of these coefficients are 0.35 for autumn and 0.55 for summer. These results make sense since the reference season is winter. The model suggests that more fuel will be consumed to maintain the proper container temperature in warmer seasons.

Interaction Variables: Mileage vs. Pick Location: The 'mileage vs. pick location' interaction variable suggests there is a variation in the fuel consumption per load pattern depending on where the trailer is loaded. This difference could be caused by the load practices at each pick location or by the type of reefer unit used at each pick location. According to the model 0.011 less gallons will be consumed per mile if the load is picked in Location 3, and 0.016 less gallons will be consumed per mile if the load is picked in Location 2 (Location 1 is the base case).

5.1.2 REJECTIONS RESULTS

No significant results were obtained from the model by running the Logistics Regression used to identify the variables that govern rejection. There are many possible reasons for this outcome:

- Rejections are not always the result of the variation of the conditions measured during the transportation process. Causes that are subjective or difficult to record could occur. For example the customer could be full of inventory and not want to receive the load that was ordered. Another possible cause for a rejection could be that the order was erroneously introduced in the system, but was delivered anyway.
- The proportion of loads with rejections compared to loads with no rejections is too small to be significantly meaningful. Also, statistical modeling of transportation data might not be the best way to predict and understand rejections. Other approaches to this problem exist such as implementing root

cause analysis for rejections, where each rejection is assigned a standardized cause in a data base at the moment the rejection is accepted by the company.

- The quality and temperature control standards used by company XYZ are managed to such a fine degree that it becomes difficult to perceive the effect of other events on rejections as temperature does not seem to fluctuate significantly as a result of these events. The sample used showed that although around 10% of the time the trailer was at a customer or moving towards a customer the temperature was above the acceptable limit, the equation that resulted from the model did not even included a single variable related to temperature as being relevant for rejections.

Although the model was not useful to predict rejections, the level of accuracy to predict loads that will not experience a rejection event improved compared to the base case (from 91.1% to 98.6%). The model suggests that the more the mileage and the number of stops, the higher the probability that a rejection will happen. Although the values of the coefficients are difficult to interpret because they apply to training data where 50% of the records experienced a rejection, the insight obtained from the model could be used to offer top customers loads with characteristics that minimize the odds of having a rejection.

Of the two explanatory variables in the model, Mileage cannot be changed. It is not possible to change this variable unless the pick location or the customer's facility physically moves. Thus, the only variable company XYZ can work with to offer top customers routes with a reduced rejections probability is 'Number of Stops'. Top customers could be offered loads with a reduced number of stops, or with the priority to be the first stop in the route. Of course cost benefit analysis would be required.

To give some dimension to the potential of the proposal, see in the Table 14 below the customer distribution obtained from the cleaned TMS data. In this table, customers are classified by the weight of produce they ordered during the period of the study.

Table 14 Customer Distribution

Customer Class	Customer Count	% Customers	% Weight Bought
A	11	11%	50%
B	30	31%	45%
C	55	57%	5%
TOTAL	**96**	**100%**	**100%**

From the table it is possible to see that around 11% of customers account for around 50% of the weight of packaged salads sold. If measures that minimize rejections like the one obtained from this thesis can be confirmed and tested with these customers, potential benefits could be obtained by controlling the loads of a very small number of customers.

5.2 POTENTIAL ECONOMIC BENEFITS OF FUEL CONSUMPTION MODELING

Potential economic benefits associated with reducing fuel consumption can be estimated roughly. To do so, the two major sources of potential savings must be isolated for inspection. These are as fuel consumed while at company XYZ locations, and fuel consumed during delivery.

Firstly, it is important to note that savings associated with reducing fuel consumption can only be calculated for the assets included in the transportation management system. For the purposes of regression analysis, a selection of clean loads was used. However, many more loads exist in the six month period investigates. Total available loads for study can yield an estimate of the total gallons consumed per year, based on a calculated average of nineteen gallons per load (including time spent at pick location). An average fuel price of $4 per gallon is used to calculate total annual fuel spend for PAR equipped assets. Table 15 below details this calculation.

Table 15 Estimated Annual Fuel Spend

Source Size	Loads 6 months	Gal per Load + Pick Loc.	Annual Consumption (gal)	$ per gal	Annual Spend
Regression Load Selection	1,457	19.21	55,979	$4	$223,916
Total Available Loads (TMS)	2,820	19.21	108,346	$4	$433,386

To do a better estimation of the potential fuel savings, a calculation of the fuel consumed at each stage (from the moment the trailer is loading at company XYZ location, travels to deliver product to customers, and returns back) must be performed. This is necessary because reefer data indicates that fuel consumption is not evenly distributed among the phases of the transportation process. Table 16 below details the percentage of fuel consumed during each of these phases and highlights the importance of partitioning.

Table 16 Fuel Consumption By Transportation Phase

Phase	Gallons	% Consumed
At XYZ	11,662	42%
At Customer	5,611	20%
Traveling	10,717	38%
TOTAL	**27,990**	**100%**

Potential reductions in fuel consumption at company XYZ locations can yield significant benefits. Currently, trailers spend long hours at pick locations burning fuel. More efficiently scheduling of loading time and trip departure time could result in significant percentage reductions in fuel consumed. Table 17 explores the dollar saving potential of such improvements in increments of 20%.

Table 17 Calculated Fuel Efficiency Saving At Company XYZ Locations

Scenario	US$
20% Scenario	$36,115
40% Scenario	$72,229
60% Scenario	$108,344

With regard to potential savings through improvements in delivery based on the regression model presented, four variables have been isolated as potential target areas for improvement. An analysis of the linked data allows the percentage of possible improvement to be extrapolated and applied to all the available data. Carrying this calculation forward it is possible to develop a 95% confidence interval for the potential savings if 100% of the change is realized. Then, an expected value of savings for all variable improvements can be calculated. A summary of these findings follows in Table 18 below.

Table 18 Potential Fuel Consumption Saving Based On Regression

VARIABLE	Potential Gal Saved per Load	Standard Error	Change Potential	Lower Limit	Expected	Upper Limit
Asset Group	1.6	0.50	32%	$4,476	$11,551	$18,625
Stop Group	2.0	0.55	11%	$2,338	$5,013	$7,688
Rejection Event	5.5	0.64	8%	$7,663	$9,926	$12,190
√Load Duration	0.6	0.34	100%	$0	$14,438	$29,521
Total	**NA**	**NA**	**NA**	**NA**	**$40,928**	**NA**

The table above shows that careful selection of asset can impact fuel costs by more than $11k per year. Additionally, focusing on sending one or two stop loads versus loads with more than two stops can also generate reductions in fuel consumption per load; however, taking this action could cause the need of additional loads. Additionally, the table shows that in addition to lost business, rejection events levy an additional penalty on loads in the form of additional fuel consumption associated with backhauling product that needs to be refrigerated. Reducing rejections can lower fuel costs by around $9K. Lastly, reducing load duration can have yield savings around $14k.

As it is rather unlikely that all of these potential regression identified savings are possible, Table 19 below details the potential saving that will occur given certain levels of change adoption applied to the expected total savings.

Table 19 Potential Fuel Consumption Savings During Delivery By Scenario

Scenario	US$
20% Scenario	$8,186
40% Scenario	$16,371
60% Scenario	$24,557

Analysis performed to determine potential saving at Company XYZ locations and potential savings during delivery can be aggregated for further analysis. Table 20 that follows details the total potential savings based on three different change adoption scenarios as well as the corresponding calculations of savings as a percentage of annual fuel spending.

Table 20 Total Potential Fuel Consumption Saving By Scenario

Scenario	US$	% Annual Spend
20% Scenario	$44,300	10%
40% Scenario	$88,601	20%
60% Scenario	$132,901	31%

When viewed as dollar figures alone, these potential savings seem rather lack luster. However, when viewed as a percentage of annual reefer fuel spending, significant saving can result from investigating fuel consumption at company XYZ locations and through a careful application of the actions suggested by the generated regression model. If this methodology were applied to cold chain trucking in general, even more significant savings could be possible.

5.3 POTENTIAL OPERATIONAL KEY PERFORMANCE INDICATORS

From the available data, a detailed description of the firm's transportation operations can be created regarding load location, fuel consumption, load duration, amount of product transported per load, temperature behavior, etc. Although this data does not include financial information and only covers loads made by company XYZ's private fleet, many operational key performance indicators (KPIs) can be extracted from it. These operational KPIs could further be enhanced by adding the above mentioned missing components.

The operational KPIs that can be obtained from the available data are generated by combining five variables:

- Pick Location
- Customer
- Mileage
- Time
- Fuel Consumption

If the methodology used in this thesis to link and clean the data is implemented and automated, company XYZ can generate these KPIs on a regular basis to:

- Analyze company XYZ's overall operational efficiency.
- Benchmark the current efficiency profile of a pick location or customer.

A short example of the most useful KPIs is shown below.

5.3.1 ASSET UTILIZATION

The linked thesis data makes it possible to measure what percentage of the time the assets are located at each of the different stages in the transportation process. Figure 37 below shows that on average, assets spend almost 70% of their time at company XYZ's facilities.

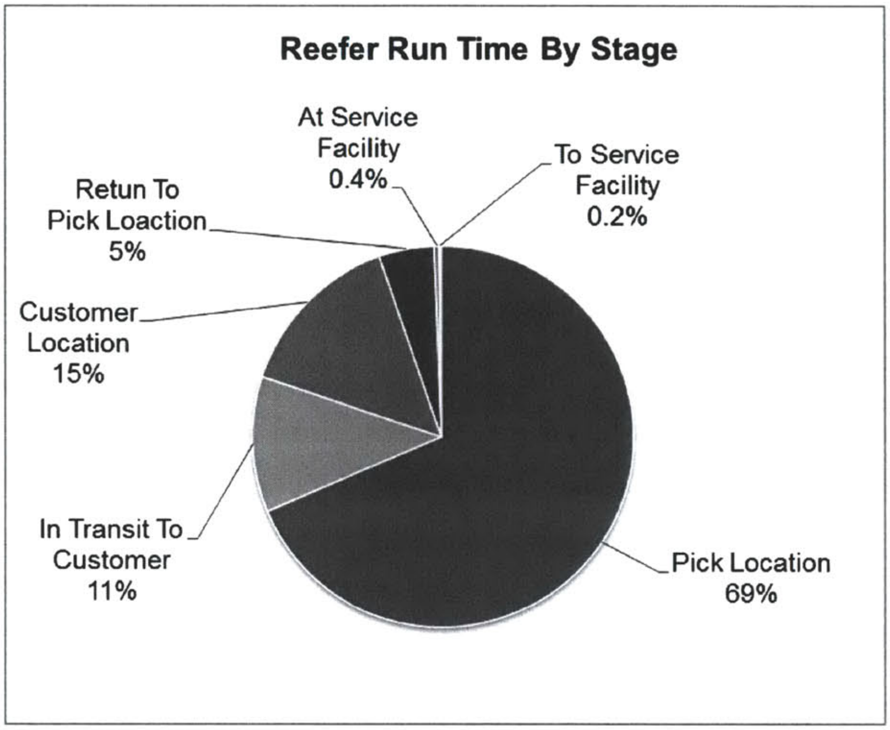


Figure 37 Reefer Run Time By Stage

5.3.2 PICK LOCATION BENCHMARKING

It is also possible to build reports to compare operational statistics and efficiency between pick locations. This information can be used to guide resource allocation and levels of control to be assigned to each location. Figure 38 illustrate an example of the type of comparisons that can be done on operational statistics.

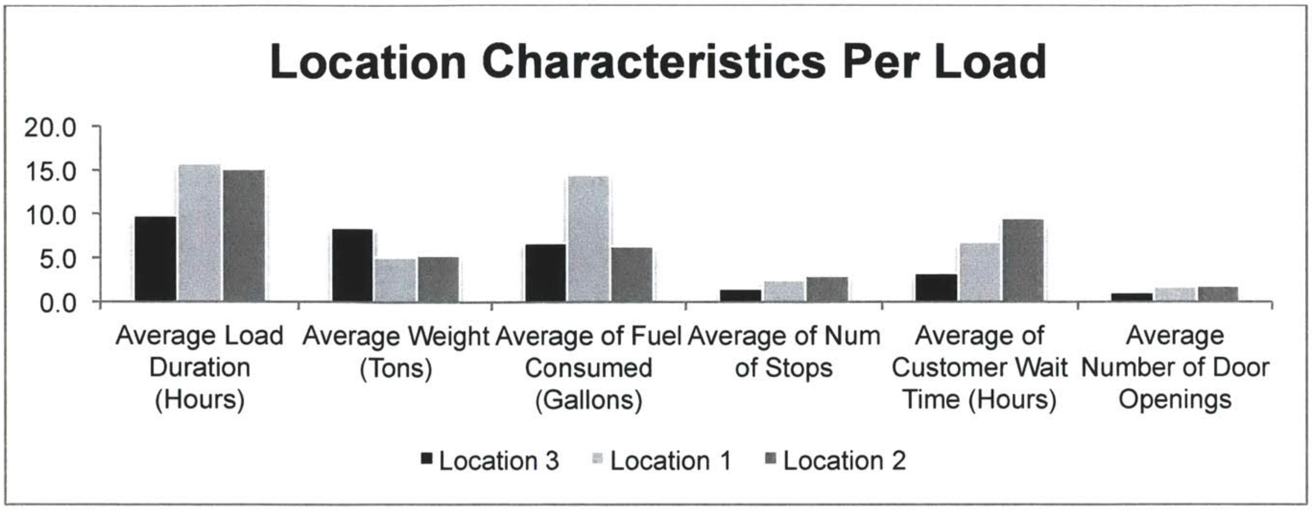


Figure 38 Pick Location Characteristics

Regarding operational efficiency, with the data tables that were built for this thesis, it is possible to compare the three pick locations' operational efficiency. Among the most meaningful indicators included in this comparison are average gallons per mile, and gallons per load weight (ton). Using these KPIs it could be possible for company XYZ's logistics management to move all locations to the highest possible standards subject to their logistical constrains. Examples of these indicators obtained from the available data are presented in the figures below.

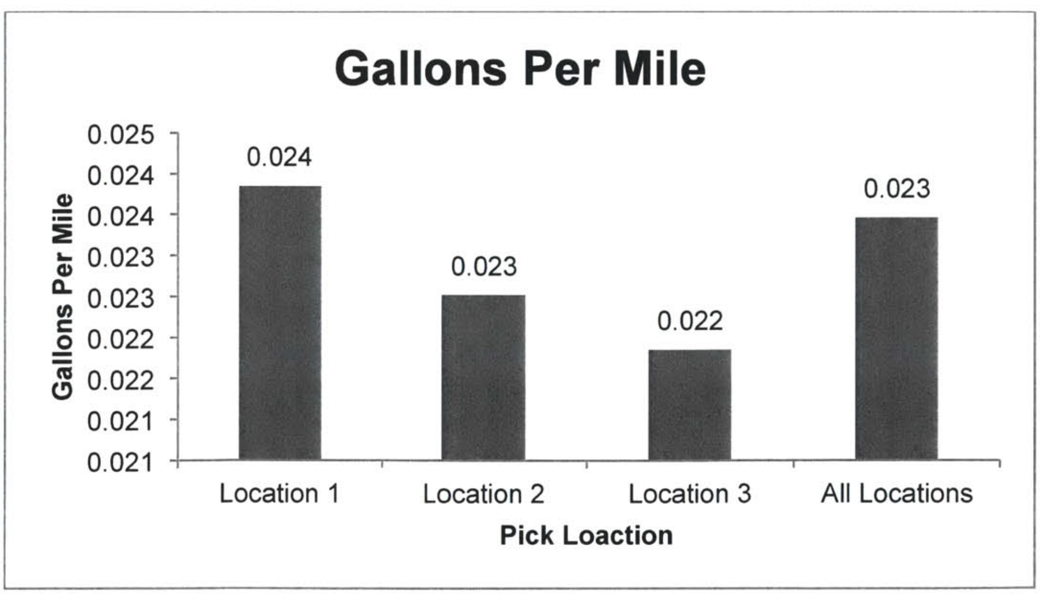


Figure 39 Gallons Per Mile

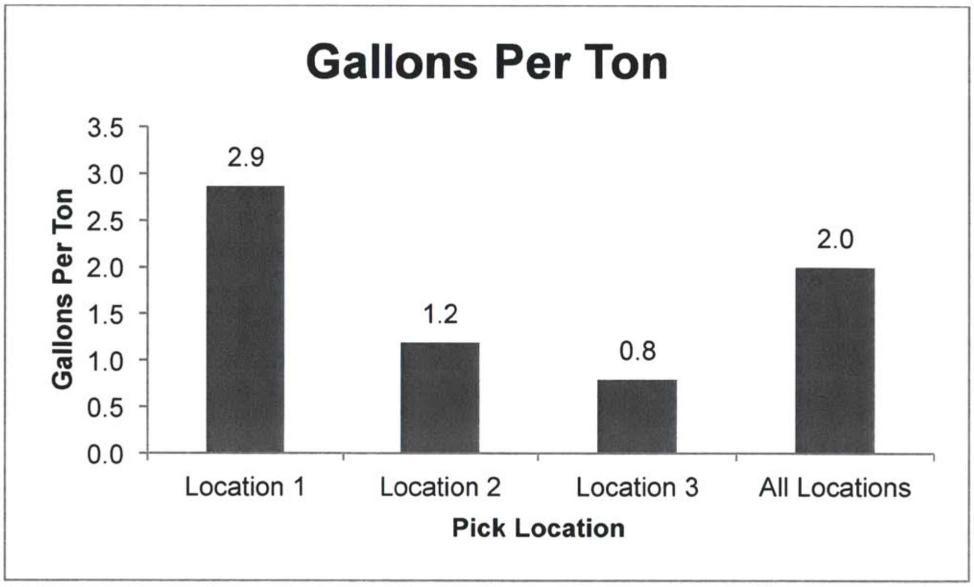


Figure 40 Gallons Per Ton

5.3.3 CUSTOMER BENCHMARKING

Finally, but as important or even more important, the following KPIs provide an example of the type of benchmarking that could be done between customers. By using this information, it could be possible to set limits defining at what point it is no longer profitable to serve a customer. Also this information could be used to set desired customer efficiency standard that could be used to improve company XYZ's operational efficiency by closely working with the most problematic customers. Figure 41 illustrates an example of the type of benchmarking that can be done. In this case, customers are ranked according to the average weight per order.

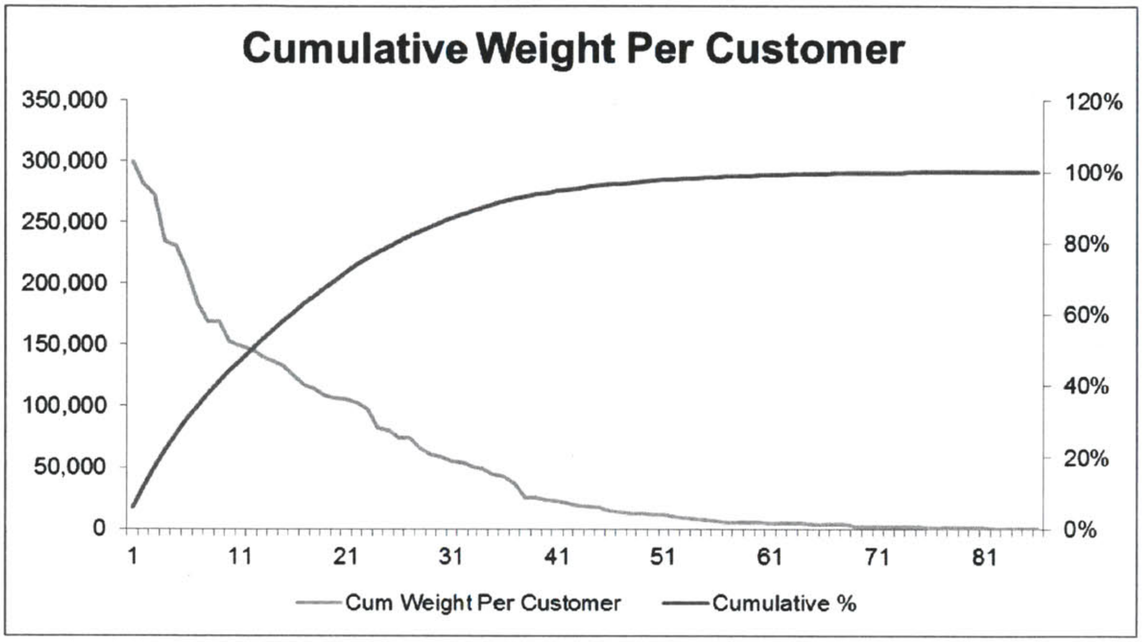


Figure 41 Cumulative Weight Per Customer

Another potential benchmarking opportunity exists with regard to Customer Dwell Time. With this indicator, it is possible to measure the average amount of time trailers spend at each customer. Also this information can be calculated as a ratio of time vs. load size. Actions can be taken to improve the time a trailer spends at customers where the time vs. load size is high. See Figure 42 as an example extracted from the available information.

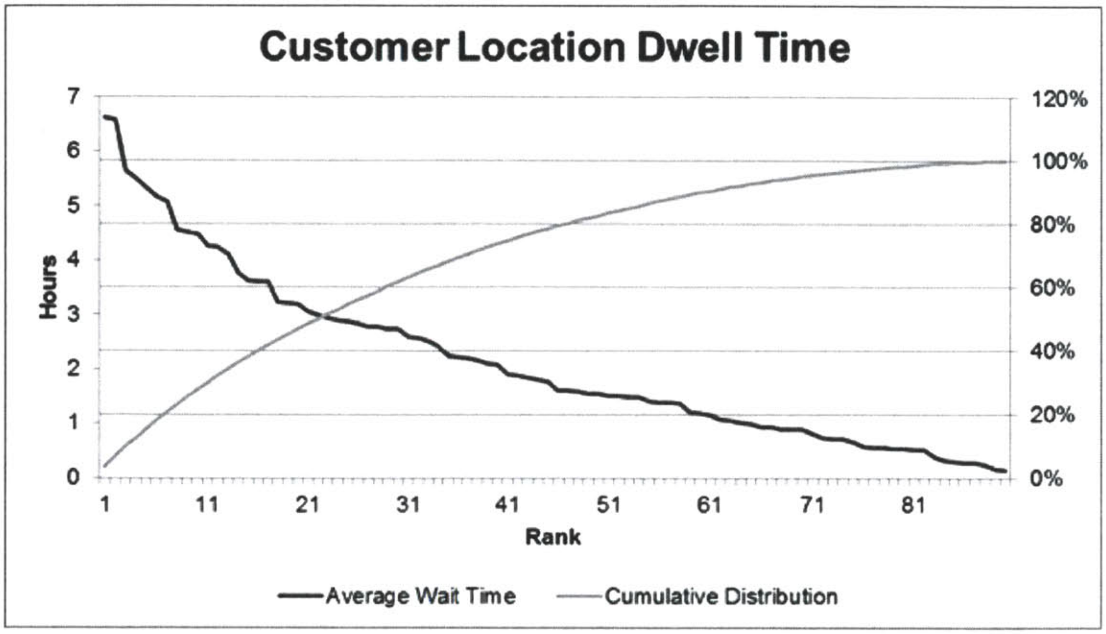


Figure 42 Average Customer Location Dwell Time

6. RECOMMENDATIONS

The results that were obtained from the methodology used in this thesis prove it is possible to obtain additional value from real-time information on refrigerated transportation other than the typical location and temperature alarms currently generated by company XYZ's monitoring system. A good model can add value by making it possible to predict the expected outcome of important performance measurements, such fuel consumption, given a set of conditions. Other relevant performance measurements, such as On-Time Delivery, could also be tested using this methodology.

If it is possible to predict the behavior of other relevant performance measurements while a load is being transported, it could be feasible to establish a set of actions to be taken to correct or mitigate an expected negative outcome before it happens.

6.1 RECOMMENDED DATA COLLECTION IMPROVEMENTS

Several key data collection improvements, if implemented, have the potential of significantly improving the quality and accessibility of the data collected and greatly aid future analysis. These possible improvements are listed in order of importance.

6.1.1 RECOMMENDED TEMPERATURE STUDY IMPROVEMENTS

The model obtained from this methodology has several opportunities for improvement that we strongly recommend. Among the most important is the need for a more granular understanding of the effects of out of range temperature on product quality. A study should be performed that results in a better understanding of the impact of temperature on shelf-life for shorter periods of time. Currently, the minimum out or range temperature exposure in existing studies is six hours. The data used for this thesis rarely showed evidence of out of range temperature durations lasting over one hour. A considerable improvement of the rejections model would result from a finer grained and quantifiable understanding of how small temperature aberrances affect product specifications and shelf life. A study should be performed that considers out of range temperature exposes in less than one hour intervals.

6.1.2 DATA TABLE LINKING

The three data tables involved in the study, REEFER data, TMS data and REJECTIONS data, are currently not unified, as detailed in the thesis. It is of great importance that these three tables be linked together in a seamless way so that a unique primary key exists for each load. Doing so will greatly improve the accessibility of the information held within these three data tables. Failing to do so would ensure that the potential revelations that live within these data tables would remain undiscovered. The current uses of these databases, as mentioned in the thesis, out of temperature warnings and historic data storage, provide only a small portion of the potential strategic advantage a linked data table could create. Collecting all load information in a single database will greatly speed up future research.

6.1.3 REJECTIONS DATA IMPROVEMENT

Currently rejections data is consolidated on an *ad hoc* basis and stored in an Excel spreadsheet. The current robustness of this record keeping system could be improved in several ways. First, the data would be better organized as a web-based database that could be easily searched and exported into various data mining tools for further analysis.

Second, the level of detail in existing records is lacking. A great deal of information can be learned from detailed accounts of product rejections. It is within the realm of possibility to imagine that more detail exists, but that the structure of the record keeping system in place limits a user's ability to appropriately document and save this information for future use. A more robust data-collection process and a more user-friendly interface would yield richer data.

Specifically, rejections data should be automated and linked to load number from the time of origin. To do that, the Transportation team should link the return receipt to the Load in the TMS system at the moment the truck arrives back at company XYZ's distribution center.

6.1.4 DRIVER DATA

Operator behavior may be a crucial variable that would allow for a more detailed understanding of reefer fuel consumption and load rejections. Currently driver data is available only in hard-copy manifests. Linking driver data to a unified data table would allow for an additional layer of analysis. Such an analysis has the potential to result in not only a better model for fuel consumption and load rejections, but also would enable the creation of a host of new key performance indicators. For instance, average load duration per driver, average same lane duration per driver and average rejections per driver would all yield information useful in managing in a proactive manner that embraces opportunities for value creation. Accordingly, adding driver information to a unified database is highly recommended.

6.1.5 ASSET PERFORMANCE DATABASE

Currently, no information on asset specific performance indicators was included in this analysis. There would be a tangible benefit to being able to track asset run hours, age, average downtime, repair history and fuel consumption rates in an effort to spot potential problems before they occur. Collecting asset performance information and incorporating this information into a regression model could prevent inefficient fuel consumption patterns and could decrease the probability of exposing loads to out of range temperatures.

6.1.6 DOOR SENSORS

As mentioned, reefer door sensors have a significant failure rate of 22%. Although this information is not crucial, it may not be necessary to pay for a monitoring feature that does not work effectively. A significant performance improvement would be necessary in order to recommend carrying this component of the sensing program forward in future asset upgrades. If not made more accurate, door sensor information adds little value to the overall sensing package.

6.2 ASSET DWELL TIME IMPROVEMENTS

Significant improvements may be possible by isolating areas of high-asset dwell time for future inspection. Asset dwell times play a significant roll in reefer fuel consumption and potentially play a role in load rejections. It is safe to say that reductions in asset dwell times will certainly not adversely affect these metrics and could potentially improve both.

6.2.1 DWELL TIMES AT COMPANY XYZ LOCATIONS

Dwell times at company XYZ locations should be monitored in an effort to reduce the percentage of fuel consumed. Operational constraints may dictate that current dwell time durations are necessary; however, a careful analysis of dwell time may possibly highlight areas for improvement on fuel consumption practices while the trailer is not moving or at a customer facility.

6.2.2 DWELL TIMES AT CUSTOMER LOCATIONS

Dwell times at customer locations vary greatly and should be further investigated and monitored. Currently the top twenty customers account for almost half of all customer location dwell time. This distribution of dwell time should be investigated in greater detail to assess the economic ramifications of such a phenomenon. Significant fuel consumption and asset utilization reductions could be realized and could potentially result in the reduction of fleet sizes and, accordingly, complexity.

6.3 TOTAL SUPPLY CHAIN STUDY

The analysis undertaken by this thesis investigates cold chain performance along only one link in the farm-to-table salad supply chain, namely, the link between company processing facilities and customer DCs. Unknown and potentially dramatic gains could be realized by performing an end-to-end supply chain evaluation using similar methodology and technology as has been employed in this thesis.

The link in the chain that this thesis investigates is very well managed, however still shows room for improvement. It is highly likely that the portion of the supply chain from customer DCs to store shelves or the link from farm to initial vacuum cooling could display significantly different behaviors that could greatly influence the ability of a model to effectively predict and explain rejections and reefer unit fuel efficiency. The rejections model provided as a deliverable of this thesis could be greatly improved by including data from all links of the salad supply chain. A detailed analysis of the end-to-end supply chain would enable this information to be collected.

7. CONCLUSIONS

This thesis has shown the extent to which real time cold chain logistics information can be used to create models that can predict and explain reefer-unit fuel consumption and load rejections in the transportation of packaged salad products.

From the data collected an effective model of fuel consumption has been created. This model has significant applicative merit and should continue to be further refined from data collected on an ongoing basis. The constructed logistics regression model of rejections does not yield results as significant as the fuel efficiency model given the variables existent in the data available for study. It is imperative to continue to refine this rejections model as new variables and data become available.

Furthermore, this thesis has shown that cold chain transportation data can be effectively aggregated and molded into useful information in the form of key performance indicators (KPIs). These KPIs have the potential to provide operations personnel with key insight into company XYZ location and customer performance.

However, the benefits this thesis identifies are short lived and transient in nature unless these tools can be further developed and integrated into the operating fabric of company XYZ. Before significant strategic advantages can be drawn from the methodology outlined in this document, several key endeavors must be undertaken.

First, all currently available data must be linked into one unified database with a single primary key. This database must be user friendly and must have a robust graphic user interface (GUI) that allows for the rapid and error free generation of exception reports and KPIs. Ideally, the available information and suggested analyses would be bundled into a transportation performance dashboard visible at a high level.

Second, the data collection must be continued and expanded in order to further refine both existing prediction models in an effort to improve the degree of accuracy and explanatory power provided by each. More data and data sources have the potential of improving the rejections model to the point of operational salience. Currently the rejections model is accurate in effectively predicting when rejections will not occur, a scenario useful when a critical customer must be served accurately on a one off basis. Nonetheless, the model should be improved in order to effectively be able to predict when a rejection will occur. Doing so would allow company XYZ to benefit from a potent operational monitor.

Finally, more research is necessary to identify the hidden drivers of rejections. Currently, 10-11% of the time, reefer temperatures are out of the acceptable range; however, according to the logistics regression model, temperature does not play a relevant role towards rejection events. Three lessons flow from this realization. (1) It is possible that variables not identified in this study play a significant role in defining the probability that a load will be rejected. (2) Salad products may be tolerant to a wider range of temperature than is prescribed by current understanding and, accordingly, reefer temperature requirements could be relaxed. (3) More damage is occurring to bagged salad products during transportation than is evident at the time of delivery and, accordingly, rejections are understated and damage does not become apparent until product is on the store shelf, or in the consumer's possession.

As a means of providing for additional data collection in relation to refining the rejections and fuel efficiency model, the type of study undertaken in this thesis should be extended to the entire salad products supply chain. Such an extension has the potential to uncover and quantify currently missing variables and performance information, which can subsequently be added to each model in an effort to improve their predictive and explanatory powers.

The intension of this thesis is to shed light on the powers of proper data aggregation, analysis, and modeling in an effort to provide meaningful strategic benefits to producers of bagged salad products. This line of investigation should be expanded to other product lines in light of the potential operational benefits for all parties involved.

Significant potential exists for application of a rejection model in the transport of perishable goods other than bagged salad products. Possible extensions of a logistics regression based approach for modeling product quality can be implemented in the transportation of meat, dairy and other fruit and vegetable products. Pharmaceutical applications may also be possible and timely, given the growth of the biopharmaceutical industry. More detailed information is required in order to better understand the variables that drive rejection. Regression modeling will yield meaningful predictive results if this research is undertaken.

8. REFERENCES

1. Albright, S. C., Winston, W. L., Zappe, C. J., & Broadie, M. N. (2011). *Data Analysis and Decision Making* (4th ed.). Mason, Ohio: South-Western/Cengage Learning.
2. April Terreri. (2010, March, 2010). Cold Chain Champions. *Food Logistics,* 14.
3. April Terreri. (2010, June 1, 2010). Cool it! *Food Logistics,* 20-23.
4. Bertsimas, D., & Freund, R. M. (2004). *Data, Models, and Decisions: The Fundamentals of Management Science*. Belmont, MA: Dynamic Ideas.
5. Brown, M. (2008). *Chilled Foods - A Comprehensive Guide* (3rd Edition ed.) Woodhead Publishing.
6. Campuzano, J. (2009, August, 2009). Packaged Salad. *Grocery Headquarters*, 47. Retrieved from www.groceryheadquarters.com
7. Carullo, A., Corbellini, S., Parvis, M., & Vallan, A. (2009). A Wireless Sensor Network for Cold Chain Monitoring. *Instrumentation and Measurement, IEEE Transactions on, 58*(5), 1405-1411.
8. Christian Schenk. (2010). Transform a Cold Chain Into an Advantage. *Food Logistics,*
9. David M. Sterling. (2010, June 1, 2010). Making the Case for Cold Chain. *Food Logistics,* Retrieved from http://www.foodlogistics.com/print/Food-Logistics/Cool-Insights/1$3710
10. Food Logistics. (2011, January 12, 2011). Wal-Mart and Thermo King Team Up on Energy-Efficient Technologies. *Food Logistics,*
11. Frank, D., & Fine, C. (2005, August, 2005). Keep Your Cool - Smart Reefers Save Money. *Xtra Lease, The Source - Newsletter for Website Members,*
12. Georgia Tech, Integrated Food Chain Center. (2010). *Home page.* Retrieved March, 2011http://ifc.scl.gatech.edu/
13. Grobler J., & Yadavalli V.S.S. (2008). Managing The Cold Chain: A Case Study At a South African Ice Cream Company. *South Afr.J.Ind.Eng.South African Journal of Industrial Engineering, 19*(2), 17-39.
14. Grocery Headquarters Editorial Staff. (2010). *2010 state of the industry almanac*
15. Hallie Frocinio. (2007, August, 2007). Protecting the Cold Chain. *Pharmaceutical Technology, Packaging Forum*, 40.
16. Hanson, K. (2010, August, 2010). Packaged salads. *Grocery Headquarters,* , 62. Retrieved from www.groceryheadquarters.com
17. Karl Schlenker, M. C. (2010, June 1, 2010). Saving Money on Cold Chain Transport. *World Trade WT 100, the Benchmark for Global Supply Chain Business, June, 2010* Retrieved from http://www.worldtrademag.com/Articles/Column/BNP_GUID_9-5-2006_A_10000000000000834875

18. Kevan, T. (2005). Control of The Cold Chain. *Business Source Complete, EBSCO. Web, Frontline Solutions 6.2 (2005)*, 30.

19. Klei, L. (2005, July, August 2005). Cool Only What's Necessary. *Food Logistics,*

20. L. Ruiz-Garcia, P. Barreiro, J. Rodriguez-Bermejo, J.I. Robla. (2007). Monitoring the Intermodal, Refrigerated Transport of Fruit Using Sensor Networks. *Spanish Journal of Agricultural Research, 5*(2), 142.

21. Lv, G., Guo, Z., Xie, S., & Pan, W. (2009). Web-based Real-time Monitoring System on Cold Chain of Blood. *I2mtc: 2009 Ieee Instrumentation & Measurement Technology Conference, Vols 1-3,* , 1268-1273.

22. Lynn Petrak. (2007, November, 2007). Cold Truths. *Farm to Plate Magazine,* , 14.

23. Par LMS. (2009). *Refrigerated Trailer Tracking, Monitoring and Control (RT-200)*

24. Parker, J. (2011). *The Economist, A special Report on Feeding the World*

25. Planet Retail LTD. (2006, May 18, 2006). COLD CHAIN INTELLIGENCE IS HOT – THE DRIVE TOWARDS GLOBALISATION. *Planet Retail,* , 13-18.

26. Richard Beilock. (1981). Towards a Balanced Delivery System for Fresh Fruit and Vegetables. *American Society of Transportation and Logistics, Transportation Journal,*

27. Rick Eyerdam. (2008). Cold Chain Special Report. *Florida Shipper, Refrigerated Cargo*

28. Rijgersberg, H., Tromp, S., Jacxsens, L., & Uyttendaele, M. (2010). Modeling Logistic Performance in Quantitative Microbial Risk Assessment. *Risk Analysis : An Official Publication of the Society for Risk Analysis, 30*(1), 20-31.

29. SENSITECH INC. WHAT YOU DON'T KNOW CAN HURT YOU: New, End-to-End RF-enabled Cold Chain Visibility Solution Helps Reduce Perishable Shrink.

30. Sensitech, The Cold Chain Company. Information for Managing the Cold Chain: Necessary v. sufficient.

31. Shan, Q. (2004). Wireless Intelligent Sensor Network for Refrigerated Vehicle, Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication (IEEE cat. no.04EX710). *Emerging Technologies: Frontiers of Mobile and Wireless Communication, 2004. Proceedings of the IEEE 6th Circuits and Systems Symposium on,* , 2 0_1.

32. Skydel, S. (2009, August 13, 2009). Fuel Efficiency: Reefer Engines. *Fleet Equipment,*

33. Smith, A. F. (2004). *Oxford Encyclopedia of Food and Drink in America.* New York: Oxford University Press.

34. Snapshot Data International Group. (2005). *US Fresh Salad 2005* No. us141095) doi:http://dx.doi.org/10.1337/us141095

35. Sterman. (2000). *Business Dynamics : Systems Thinking and Modeling for a Complex World*. Boston: Irwin/McGraw-Hill.

36. Taylor, A. (2010). Relationship Between Diesel Fuel Prices and Truck Rates. *Agricultural Refrigerated Truck Quarterly, Third Quater 2010, July-September*, 11.

37. Terry Myers. (2007). RFID Shelf Monitoring Helps Resolve Disputes. *RFID Journal,*

38. UC Davis. *Produce Facts: Lettuce: Romaine, Recommendations for Maintaining Post-Harvest Quality.*http://postharvest.ucdavis.edu/Produce/ProduceFacts/Veg/romaine.shtml

39. UC Davis. *Produce facts: Spinach*

Recommendations for Maintaining Postharvest Quality

http://postharvest.ucdavis.edu/Produce/ProduceFacts/Veg/Spinach.shtml

40. United States Internal Revenue Service. (2009). *History of Trucking Industry*

41. Verbic, M. (2006). Discussing the Parameters of Preservation of Perishable Goods in a Cold Logistic Chain Model. *Applied Economics, 38*(2), 137-147.

42. Wade, S. (2006). In Tasha Wade and Shelly Sommer (Ed.), *A to Z GIS : An Illustrated Dictionary of Geographic Information Systems / edited by Tasha Wade and Shelly Sommer.* (2nd ed.) Redlands, Calif. : ESRI Press : Independent Publishers Group [distributor], 2006.

43. Wagner, M. (2005). Fresh and fresher - How Web Monitoring Keeps Perishables From Perishing Too Soon. *Internet Retailer,* Retrieved from http://www.sensitech.com/PDFS/food/InternetRetailer_Fresher.pdf

44. William Huffman. (2006, April 3, 2006). Hot market - Cool Freight. *World Traffic Magazine, Logistics and Freight Section*, 15.

45. WorldPharma. (2010). Moving to Electronic Temperature Monitoring. *World Pharmaceutical Frontiers*

9. APPENDIX

9.1 DATA COMPILATION FOR REGRESSION

9.1.1 DATASETS CLEANING AND LINKING

Steps used to clean and link the data are described below:

TIME PERIOD ALIGNMENT

The tree data tables covered different time periods. It was necessary to identify an overlapping period among the three data tables, which turned out to cover the range running from August-2010 to January-2011, and resulted in a reasonable seasonal blend of warm-weather months and cold-weather months.

TMS PICK LOCATION FILTERING

Since Location 1 through 3 are the only distribution centers that use trailers equipped with PAR monitoring system, the other pick-locations were excluded from the database.

RETURNS FILTERING

The original REJECTIONs dataset included information regarding returns, however, for the reasons explained before, returns are out of the scope of this thesis and thus were eliminated from the database.

DUPLICATE ELIMINATION

A number of duplicates were identified and eliminated from the TMS and REEFER data tables. In the case of the TMS data the duplicates were identified by using the Load ID. In the case of the REEFER data, duplicates were identified using a combination of date, time, and Event Type markers. To avoid losing relevant information, before excluding any duplicate records from the data sets, it was verified that all of the fields in these records we duplicated thus ensuring that accurate results.

The next step was to link the three datasets given the lack of a direct field to relate them each database to the next. The rationale used to link each pair of datasets is described below:

LINKING REEFER DATA AND TMS DATA

The Reefer data table does not include a field containing either Load or Shipment number. Nevertheless it was possible to confirm that there is a very low probability of having the same customer being served from the same pick location on the same date more than once. This was confirmed via expert interview. Although in the REEFER data it was possible to directly obtain the date and time, and the pick-locations of each trip, using the Geo-fence field it was only possible find a few number of customers. The way this limitation was overpassed was by using the fields that included the Latitude and Longitude of the trailer's position during the entire load trip. To achieve this, a table including the coordinates of all the customers in the TMS data was built using Google Maps®, and then a search function in excel was used to identify, in the REEFER table, the records that contained those coordinates. Thus it was possible to assign a customer number to the appropriate records in the REEFER data. This technique had the risk of including customer as stops even when the trailer passed close to that specific coordinate but did not stop at it. To reduce this risk of assigning incorrect customer drops to the REEFER data records, a second rule was added that required the trailer to be static in that location for at least two records for a true stop to be registered.

The next step was to assign the correct Load ID to each one of the trips made from the company XYZ's Distribution Centers to their customers' locations. This was achieved by concatenating customer, date, and time on both datasets; and then using an Excel search functions to match the records.

A few cases of customers being served on a single date more than once by the same pick location were found, indicating that those trips had more than one possible Load ID to be linked with. From the available data there was no way to assign the correct Load ID to the corresponding trip, these records had to be excluded from the data table so as to ensure that correct shipment load pairs were created.

The product of the data link described above was a REEFER database with three additional fields:

- Customer ID
- Customer name
- Load ID

LINKING TMS DATA AND REJECTIONS DATA

The REJECTIONS data also did not contain the Load ID or Shipment number information; nevertheless it was possible to find a corresponding Load ID by concatenating the Customer's code number and the event date on both the REJECTIONS and the TMS Data. Only rows containing rejections were utilized. As explained before, rows containing returns were excluded from the analysis. One of the difficulties which arose by using this procedure was that the date a specific rejection was recorded in the REJECTION dataset was sometimes different from the date the rejection event actually happened. The reason for this is that the date recorded on the REJECTION data set corresponded to the date on which the trailer returned to company XYZ's distribution center and reported the rejection event and submitted the appropriate documentation required by customer service. This could happen one or two days after the actual rejection happened in the customer's distribution center, two days having elapsed due to the truck's return trip timing. To overcome this difficulty, the concatenating process was done in three rounds, subtracting one day from the rejection table date on each round. The product of the data link described above was to add the LOAD ID field to the REJECTIONs data. Doing so allowed specific loads to be linked to specific product rejections.

9.1.2 CALCULATED FIELDS

Some of the variables required to run the regressions could be obtained directly from the fields existing in the available data tables. However, other variables had to be calculated from the existing information, or could not be obtained at all. This section briefly describes the rationale that was used to obtain the calculated variables listed below.

- Customer Wait Time
- Temperature Abuse Area
- Time with Temperature Above Range
- Time with Temperature Below Range

- Pre-cooling Compliance
- Number of Stops
- Load Duration

All the variables that needed to be calculated were part of the REEFER data. Accordingly, the following calculations were possible to perform in the REEFER database. As mentioned in the Data Structure section, each record in the REEFER data contains information concerning the precise time and event occurred. Since the information is ordered in chronological order, it is possible to calculate the time between one record and the next one by simply subtracting the oldest record time from the next newest record time provided they belong to the same Asset ID.

LOCATION GROUPS

The first step was to classify each record of the REEFER data set into one of the following seven possible **Location Groups**.

XYZ: Refers to records where the trailer is located at a company XYZ's facility.

C: Refers to records where the trailer is located at a customer's facility.

T: Refers to records where the trailer is located at a carrier or transportation facility.

To XYZ: Refers to records where a trailer is moving towards a company XYZ facility.

To Customer: Refers to records where the trailer is moving towards a customer's facility.

To Transport: Refers to records where the trailer is moving towards a carrier's facility.

End of asset: Refers to the final group of records of a specific asset, when it is not possible to know the assets next destination due to a truncation of the data stream.

To make the Location Group classification possible, the first step was to assign an **XYZ**, **C**, or a **T** to the records whose coordinates matched the coordinates of a company XYZ facility, a customer's Distribution Center, or a carrier's facility. The next step was to assign the corresponding destiny (to XYZ, to Customer, to Transport) to the records that remained blank. This was possible to do since REEFER records exist in chronological order, and the destiny of the trailer (XYZ, C, or T) had already been identified in the prior step. As a third step, the value of end of asset was assigned to the remaining records.

TRIP NUMBER

Next, REEFER data records where grouped by generating a dummy Trip Number for each load. A new Trip Number was assigned to a group of records that displayed a pattern of arriving at one of company XYZ's three Pick Locations, was loaded, executed the deliveries to customers and returned back to the company's Pick Location. Below is an illustration to better explain this rationale:

$$\underbrace{\text{XYZ -> To Customer -> C -> To Customer -> C -> To XYZ ->}}_{\textbf{TRIP N}}\ \underbrace{\text{XYZ-> To Customer -> C -> To XYZ}}_{\textbf{TRIP N+1}}$$

The trip number was used to verify that only one Load ID per trip had been assigned during the data linking process. After performing this verification and cleaning any inconsistency it was possible to reliably group the remaining 181,526 REEFER data records into 1,457 Trip numbers or Load IDs, which became the initial Load sample used to run the regressions.

TEMPERATURE AVERAGE:

The third step employed was to identify when the average temperature of the trailer was within or outside of its acceptable temperature range. To do this, the upper and lower trailer temperature limits, 42°F and 31°F respectively, were included in the data table as new fields. Next, the average temperature of the trailer was calculated in a new field as the arithmetic average of the Return Temperature, Probe 1 Temperature, and Probe 2 Temperature.

FIELD VALUE CALCULATION:

After having followed the three steps described above, the data table was ready to allow for the calculation of each constructed variable. A short description of how each variable was calculated is listed below:

Customer Dwell Time: Using Excel pivot tables, customer dwell times for each Load ID were calculated by adding the time duration of the records that were classified as C, denoting that the asset was at a customer location.

Temperature out of range: Calculated by comparing the Average Temperature of each record to the upper and lower limits of the desired trailer temperature. In the new field, a number 1 is assigned to every record that strayed outside the desired temperature range, and a 0 is assigned to every record the remained within the desired temperature range.

Time with Temperature Above Range: Calculated by adding the duration of the events or records were the Average Temperature field was above the upper temperature limit established.

Time with Temperature Below Range: Calculated by adding the duration of the events or records were the Average Temperature field is below the lower temperature limit established.

Temperature Abuse Area: Temperature abuse area is a measurement of temperature abuse used by company XYZ's quality department. The measurement is the multiplication of the trailer temperature times the total minutes the product was exposed to an out-of-range temperature.

Pre-cooling Compliance: Pre-cooling load compliance means that the product should was loaded into the trailer only after the temperature of its walls reached a temperature close to 31°F. Since from the available data it is impossible to know the exact moment at which the produce is loaded into the trailer, this calculated field is just an approximation given by comparing the average temperature of the trailer just before it leaves the company's pick location with the desired 31°F temperature.

Number of Stops: Is calculated by counting the number of customers served by each load.

Load Duration: Is calculated by adding the duration of all records within a Load and excluding the time the trailer spent in the pick location facility. Doing so ensures that only true load time is captured for analysis.