NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A PREDICTIVE ANALYSIS OF THE DEPARTMENT OF DEFENSE DISTRIBUTION SYSTEM UTILIZING RANDOM FORESTS

by

Amber G. Coleman

June 2016

Thesis Advisor: Samuel E. Buttrey
Second Reader: Jonathan K. Alt

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2016	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE A PREDICTIVE ANALYSIS OF THE DEPARTMENT OF DEFENSE DISTRIBUTION SYSTEM UTILIZING RANDOM FORESTS			**5. FUNDING NUMBERS**
6. AUTHOR(S) Amber G. Coleman			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This thesis develops machine-learning models capable of predicting Department of Defense distribution system performance of United States Marine Corps ocean requisitions to the United States Pacific Command area of operations. We use historical data to develop a model for each sub-segment of the Transporter leg within the distribution pipeline and develop two different models to predict the ocean transit sub-segment based on Hawaii and non-Hawaii destinations. We develop a linear regression, regression tree and random forest model for each sub-segment and find that the weekday and month in which requisitions begin the Transporter segment are among the most significant drivers in variability. United States Transportation Command currently uses the average performance per sub-segment to estimate Transporter length, and our models, when applied to the test set, perform considerably better than the average. We conclude that the random forest models provide the best and most robust results for most sub-segments. However, we encounter several issues concerning missing values within our dataset, which we suspect artificially inflate the significance of some of our predictor variables. We recommend refining data collection processes in order to collect observations that are more accurate and applying the same methodologies in the future.

14. SUBJECT TERMS Department of Defense (DOD) distribution, linear regression, regression trees, random forests, machine learning, ocean shipments, predictive models, United States Transportation Command (USTRANSCOM), Marine Corps Logistics Command (MARCORLOGCOM)			**15. NUMBER OF PAGES** 111
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release;distribution is unlimited

A PREDICTIVE ANALYSIS OF THE DEPARTMENT OF DEFENSE DISTRIBUTION SYSTEM UTILIZING RANDOM FORESTS

Amber G. Coleman
Major, United States Marine Corps
B.S., United States Naval Academy, 2004
M.S., Syracuse University, 2012

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2016

Approved by: Samuel E. Buttrey
Thesis Advisor

Jonathan K. Alt
Second Reader

Patricia Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis develops machine-learning models capable of predicting Department of Defense distribution system performance of United States Marine Corps ocean requisitions to the United States Pacific Command area of operations. We use historical data to develop a model for each sub-segment of the Transporter leg within the distribution pipeline and develop two different models to predict the ocean transit sub-segment based on Hawaii and non-Hawaii destinations. We develop a linear regression, regression tree and random forest model for each sub-segment and find that the weekday and month in which requisitions begin the Transporter segment are among the most significant drivers in variability. United States Transportation Command currently uses the average performance per sub-segment to estimate Transporter length, and our models, when applied to the test set, perform considerably better than the average. We conclude that the random forest models provide the best and most robust results for most sub-segments. However, we encounter several issues concerning missing values within our dataset, which we suspect artificially inflate the significance of some of our predictor variables. We recommend refining data collection processes in order to collect observations that are more accurate and applying the same methodologies in the future.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
AO	Area of Operations
CCP	Consolidation and containerization point
CENTCOM	United States Central Command
CMC	Commandant of the Marine Corps
CNA	Center for Naval Analysis
CONUS	Continental United States
CWT	Customer Wait Time
DLA	Defense Logistics Agency
DMC	Distribution Management Center
DOD	Department of Defense
DORRA	Defense Logistics Agency Office of Operations Research and Resource Analysis
DPO	Distribution process owner
DSU	Deployable supply unit
EUCOM	United States European Command
FY	Fiscal year
GAO	Government Accountability Office
GLM	Generalized linear model
HL	Heavy lift
HZRD	Horizon Lines, LLC
IGC	Integrated Data Environment Global Transportation Network
IDE	Integrated Data Environment
ICP	Initial consolidation point
IDL	Integrated distribution lanes
IG	Inspector General
IPG	Issue priority group

iSDDC	Integrated Mission Support for Surface Deployment and Distribution Command
ITV	In-transit visibility
JDDE	Joint Deployment and Distribution Enterprise
LRT	Logistics response time
MAE	Mean absolute error
MAEU	Maersk Line
MARCORLOGCOM	Marine Corps Logistics Command
MARFORPAC	Marine Corps Pacific Command
MATS	Matson, Inc.
MILAIR	Military air
MSE	Mean square error
OBB	Out of bag
OD	Outsize dimension
OIF	Operation IRAQI FREEDOM
PACOM	United States Pacific Command
PMO	Priority Material Office
POD	Port of debarkation
POE	Port of embarkation
Q-Q	Quantile-quantile
QSDSS	Supply chain quality sustainability
RDD	Required delivery date
RFID	Radio frequency identification
RMSE	Root mean square error
SSA	Supply support activity
SDDB	Strategic Distribution Database
SPOD	Seaport of Debarkation
SPOE	Seaport of Embarkation
TDD	Time definite delivery
USMC	United States Marine Corps
USTRANSCOM	United States Transportation Command

EXECUTIVE SUMMARY

This thesis uses historical data and machine-learning algorithms to develop a series of models capable of predicting the length of the Transporter segment within the Department of Defense (DOD) distribution system. United States Transportation Command (USTRANSCOM) currently uses the average length of each sub-segment to estimate Transporter performance times, and we use this as our baseline to compare our models. We focus on 2015 United States Marine Corps (USMC) ocean shipments to the United States Pacific Command (PACOM) area of operations.

The distribution system consists of four main segments and is further broken down into 12 sub-segments, each of which receives a separate timestamp at completion. The Transporter leg begins when the carrier picks up a requisitioned item from a supplier and ends when the carrier delivers the item to the point of need. This segment consists of five sub-segments, which we show to be independent and model separately. Additionally, we create models for the ocean transit sub-segment to account for the large difference in distance between shipments traveling to Hawaii and those traveling to non-Hawaii destinations.

We collect and clean twelve months of data from the Strategic Distribution Database (SDDB) in preparation for analysis and encounter multiple data quality issues. We remove all unique identifiers, variables that do not apply to ocean shipments and the Transporter segment and any variable missing more than 60 percent of observations. This reduces our dataset to approximately 40 variables, which we further reduce to 20 variables. We also created variables to represent the weekday, month and quarter in which the Transporter segment began. The combination of missing observations across all variables results in only 40 percent of the dataset containing complete cases, which is enough data to build models; however, we suspect this negatively affects the accuracy of our models.

We build a linear regression, regression tree and random forest model for each sub-segment of the Transporter leg and two models for the ocean transit sub-segment.

Many of our models find the weekday and month in which the Transporter leg began to be significant drivers of variability. Upon further exploration of this result, we find these results are artificially high. We run two simple linear regressions for the Hawaii ocean transit model with two subsets of the data using transit time as the response and month as the only predictor. Model A utilizes a data subset with only complete observations and finds that month explains almost 80 percent of the variation in ocean transit time. Model B utilizes a subset including missing values and finds month explains less than 40 percent of variation in transit time. We conclude the information held by the dataset is not completely representative of the sustainment materiel that flows through the system, and this negatively affects our ability to analyze performance accurately.

When applied to our test set, most of our random forest models perform considerably better than the baseline model, and, in some cases, result in average root mean square errors of less than one day. Only in sub-segment 5 is the baseline model a more accurate predictor of performance than our random forest model; however, both models produce errors of approximately one day. We conclude that our models develop a more accurate means of estimating Transporter leg performance than the current USTRANSCOM standard; however, we have preliminary indications that the models perform poorly on 2016 data.

Although our models perform very well against the test sets, we deduce that the quality of data from which we base our models negatively affects our ability to model the system accurately. We recommend re-evaluating and updating the collection and consolidation processes associated with the SDDB. Additionally, we also recommend implementing accountability measures to ensure the system accurately captures timestamps throughout the process, as the timestamps are vital to predicting distribution system performance. Finally, we recommend employing these methodologies in the future on better quality data.

ACKNOWLEDGMENTS

To my advisors, Dr. Buttrey and LTC Alt, I appreciate all of your guidance throughout this process. I would also like to thank LTC Hiltz and his USTRANSCOM J4/J5 staff and Cam Klunder from the MARCORLOGCOM DMC for all of their assistance. Lastly, I would like to express my sincere appreciation to my husband, Joshua, for his unrelenting patience and support.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The Department of Defense (DOD) distribution system lacks an adequate method to estimate delivery dates of requisitioned materiel. According to our dataset, over half of all shipments do not meet internal delivery standards, and the Sustainment Dashboard, the current predictive tool available to some users, lacks statistical rigor. Unit commanders must make logistical decisions based on potentially inaccurate information, which equates to more risk. In this research, we develop a tool, using statistical methods and historical data, capable of providing more accurate delivery-date predictions. Equipping leaders with this information will enable them to make better decisions with limited resources while minimizing risk.

A. PURPOSE

The DOD distribution pipeline consists of a complex combination of people, resources, and policies designed to support the warfighter. Despite numerous improvements over the last 15 years, it continues to perform below expectations, a problem identified by several government agencies (Government Accountability Office [GAO] 2015a). The system consists of four legs—source, supplier, transporter, and theater—which are further divided into 12 sub-segments. When a unit requests an item through the supply system, the item typically travels through the segments depicted in Figure 1 before finally reaching the requesting unit.

Analysts use data collected from these segments to measure system performance on two internal metrics within the distribution chain—Time Definite Delivery standards (TDD) and Logistics Response Time (LRT). TDD measures consistency and dependability within the system, and LRT measures the time between order placement and receipt by the using unit (Hiltz 2015, 1). The DOD standard requires the LRT to be less than the TDD. Mahan explains that users generally accept that the system will operate at 85 percent reliability. This metric indicates what the customer actually "feels" while waiting for a requisition to arrive (Mahan et al. 2007, 17). However, this only reflects the expectation of variation and not the actual variation within the system (Hiltz

2015, 3). During calendar year 2015, our dataset indicates that close to 50 percent of all United States Marine Corps (USMC) requisitions to the United Stated Pacific Command (PACOM) area of operations (AO) did not meet TDD standards. Table 1 shows the average number of days it took to complete ocean requisitions in 2015 broken down by final destination as well as the 2015 TDD standards.

Table 1. PACOM Calendar Year 2015 Distribution Performance by Requisition Destination.

Destination	Total Number	Average (days)	TDD	Proportion On-Time
Hawaii	3191	47.70	43	0.58
Korea	5	108.80	57	0
Guam, Japan, Okinawa	8953	67.50	57	0.48
Singapore, Diego Garcia, Hong Kong, Australia, Marshall Islands, Pacific, Philippines, Thailand and all other PACOM countries	4	110	70	0.60

The Sustainment Dashboard provides decision makers information regarding late shipments with the limitations previously discussed. In many cases, leaders learn about late shipments after the requisition misses the required delivery date (RDD). This erodes confidence in the system and often leads to negative behaviors, such as hoarding of supplies and multiple ordering (Mahan et al. 2007, 17–18).

This research focuses on creating a more accurate predictive tool in order to provide leadership early notification of potentially late shipments. Alerting decision makers to potential problems earlier in the process enables them to take action before the RDD and can prevent negative effects on mission accomplishment. While the ability to deliver items quickly is important, the ability to deliver items within the promised delivery window is equally important (Slone 2004).

B. PROBLEM STATEMENT

Predicting future performance of the distribution system requires detailed analysis of multiple variables. This research seeks to address the following questions:

- What factors drive variability within the distribution system?
- Can a more accurate predictive tool be developed in order to inform decision makers of late shipments prior to shipments missing the RDD?

C. MOTIVATION

Lack of proper and timely logistics support creates unnecessary risk to unit mission accomplishment, potentially jeopardizing national security. The DOD Joint Logistics Publication explains the importance of logistics in the accomplishment of military missions.

> The relative combat power that military forces can generate against an adversary is constrained by a nation's capability to plan for, gain access to, and deliver forces and materiel to required points of application. (Chairman of Joint Chiefs of Staff [CJCS] 2013, ix)

The ability of the United States to deploy and sustain its military serves as a limiting factor on the nation's projection of power abroad. Inaccurate logistics data negatively affects command and control decision-making and forces logisticians to be reactive rather than proactive ultimately affecting support to the warfighter (Schaffer and Borns 2015). Major General John Broadmeadow, former Commanding General of Marine Corps Logistics Command (MARCORLOGCOM), explains the role of MARCORLOGCOM in supporting Marine Corps logistics.

> Marine Corps Logistics Command executes its global mission with a clear and precise objective—to ensure that Marines in harm's way have every measure of logistics support to accomplish their mission. (Wingard et al. 2015)

The results of this research will provide process owners with improved insights into the performance of their systems and will serve as a foundation for future work in the improvement of the DOD supply chain.

D. METHODOLOGY AND LIMITATIONS

The distribution system, a multibillion-dollar enterprise, supports over 6 million requisitions annually (Mahan et al. 2007). At the request of MARCORLOGCOM, this research focuses on the Transporter segment of USMC ocean requisitions to the PACOM AO. We explore the performance of each of the five Transporter sub-segments

independently. This research also explores the quality of data available to distribution customers as well as its influence on prediction accuracy. We use R, a statistical computing language, to explore and analyze the data (R Core Team 2015).

E. THESIS STRUCTURE

This study begins with gathering and cleaning all data that could potentially influence shipment performance. Once we clean and format the data, we use it to train and validate machine-learning models to develop a predictive tool capable of estimating delivery dates.

Chapter II covers background information and relevant orders, a sustainment dashboard overview and a summary of reports on distribution performance. It also provides an overview of similar problems and the methods used to solve them. We provide details concerning the datasets, data cleaning and methodology in Chapter III. Chapter IV explains the analysis behind the model. Finally, Chapter V provides a summary of research results and recommendations for future work.

II. BACKGROUND AND LITERATURE REVIEW

This literature review contains three parts. The first delivers an overview of the Department of Defense (DOD) distribution pipeline structure and operations. It provides background and context to the problem this thesis aims to solve. The second part includes reports and analysis from various government agencies that highlight several inefficiencies within the system as well as several recommendations for improvement. The last section of this literature review assesses methods used to solve similar problems. Reviewing these methods provides a basic framework from which to begin work on developing a distribution system predictive tool.

A. BACKGROUND

The DOD distribution pipeline consists of multiple sources of supply, modes of transportation, and final destinations focused on providing the right equipment, at the right time, to support the warfighter. United States Transportation Command (USTRANSCOM) oversees the Joint Deployment and Distribution Enterprise (JDDE), a collection of resources necessary to conduct joint distribution operations (Deputy Under Secretary of Defense for Acquisition Transportation and Logistics 2007). On average, it manages 1,900 air missions, 25 ships underway, and 10,000 ground shipments per week along with a workforce of 140,000 personnel operating in 75 percent of the world's countries (USTRANSCOM 2016).

Each service component depends on USTRANSCOM's management of the strategic distribution system to support its warfighters. The United States Marine Corps (USMC), the smallest component, makes up approximately 5 percent of total distribution traffic. The USMC supply system's expeditionary mission often suffers from slow response times due to distribution requirements to remote locations with low volume and frequency (Nickle 2015). The Marine Corps Logistics Command (MARCORLOGCOM) serves as the service's Distribution Process Owner (DPO). The USMC tasks MARCORLOGCOM with maintaining near real-time visibility of all assets with the ability to track, trace and expedite shipments from the point of origin to final destination

utilizing the Distribution Management Center (DMC) (Commandant of the Marine Corps [CMC] 2014). The DMC monitors daily distribution traffic throughout the USMC and is responsible for further analysis of the system's performance.

1. Organization of the DOD Supply Chain

The DOD supply chain consists of four segments—source, supplier, transporter, and theater—each with different process owners (Hiltz 2015, 5). Figure 1 illustrates the four segments of the process and its 12 sub-segments. We combine these segments to measure the *Logistics Response Time* (LRT) (Hiltz 2015, 5). This metric determines compliance with the *Time Definite Delivery* (TDD) standards. The TDD is intended to be a number of days such that 85 percent of requisitions are delivered in fewer days than the TDD standard (Mahan et al. 2007). LRT is compliant when it is less than or equal to TDD.

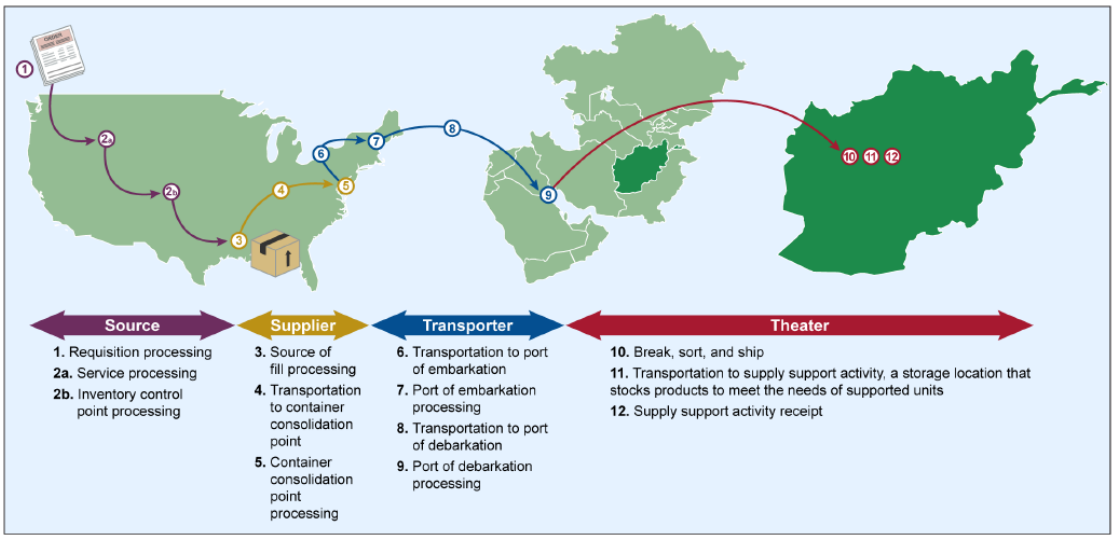

Figure 1. The DOD Global Distribution Pipeline Broken down by Segments and Sub-segments. Source: Government Accountability Office (2015a).

Integrated Distribution Lanes (IDL) extend from the supply source to the using unit and exist to enable further analysis of the system. Grouping these distribution lanes by mode of transportation and final destination results in 111 different TDDs. The JDDE

agrees upon these standards at the annual TDD conference. The LRT measures system response time, and the TDD measures reliability (Hiltz 2015, 1). Table 2 lists the fiscal year 2015 (FY15) United States Pacific Command (PACOM) TDD standards. We focus our research on this geographic location.

Shipment priority codes determine mode of transportation. Three issue priority groups (IPG) exist to accommodate three shipment speeds. An IPG 1 requisition requires the fastest mode of transportation available, IPG 2 requires faster transportation than IPG 3, but not as fast as IPG 1, and IPG 3 is the slowest mode available (Under Secretary of Defense for Acquisition, Technology and Logistics).

Table 2. PACOM FY15 Ocean Time Definite Delivery (TDD) Standards
Source: Hiltz (2014).

PACOM - Ocean

		Alaska	Hawaii	Korea	Guam, Japan, Okinawa	Singapore, Diego Garcia, Hong Kong, Australia, Marshall Islands, Pacific, Philippines, Thailand
FY15 LRT Standard		43	43	57	57	70
Segment Goals	Source	2	2	1	1	2
	Supplier	21	21	21	21	24
	Transporter	14	14	28	28	37
	Theater	6	6	7	7	7

The Integrated Data Environment (IDE)/Global Transportation Network (IGC) ties together multiple databases to provide the customer with near real-time visibility (Assistant Secretary of Defense for Logistics and Materiel Readiness 2014). The Strategic Distribution Database (SDDB) provides retrospective performance data for analysis at various levels. The Defense Logistics Agency Office of Operations Research (DORRA) collects and consolidates the SDDB, and USTRANSCOM publishes it monthly. Despite the introduction of numerous tools throughout the last 15 years, the

DOD supply chain continues to experience inefficiencies and has drawn negative attention from various government agencies (Government Accountability Office [GAO] 2015b).

2. USTRANSCOM Sustainment Dashboard

The Sustainment Dashboard, the current predictive tool, is based on performance averages and fails to consider the time necessary to complete the current sub-segment. For example, the requisition in Figure 2 is currently executing the Seaport of Embarkation (SPOE) Hold sub-segment, and the Sustainment Dashboard assumes the ocean phase begins tomorrow. It then adds the averages of the remaining sub-segments to estimate that the shipment will arrive in theater in 51 days. If this exceeds the TDD, the requisition will potentially be late.

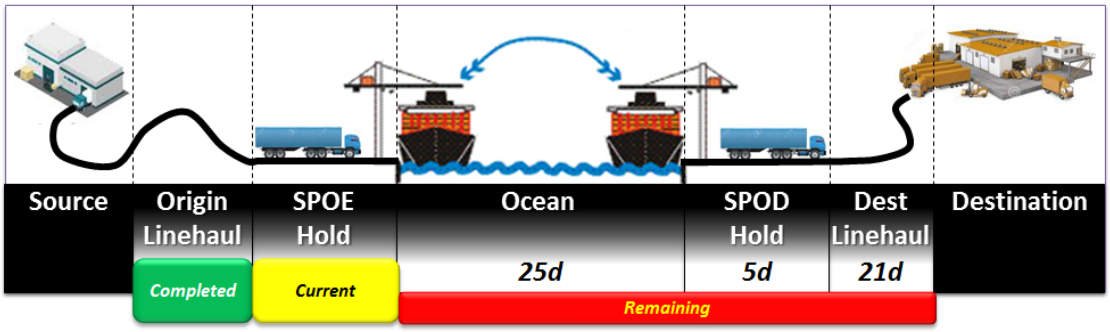

Figure 2. USTRANSCOM Predictive Model Example Source: USTRANSCOM (2015).

This model depends on two over-simplifications that can potentially lead to inaccurate predictions. First, it assumes the current sub-segment will end the following day without taking into consideration how long it has been in that sub-segment. If, on average, it takes 10 days to complete SPOE Hold and the shipment has been there only one day, the estimate will presumably be too small. Second, using average transit times for each of the sub-segments can also lead to inaccurate predictions. Savage (2009) points

out that plans based on averages often go wrong because they ignore the impact of variations, and instead, recommends replacing averages with frequency distributions.

The GAO and the RAND Corporation flagged multiple inefficiencies within the distribution system ranging from ineffective organization to lack of asset visibility. As discussed in Chapter I, these issues promote a lack of warfighter confidence, which lead to negative behaviors such as hoarding and multiple ordering, which further confound the problem and degrade efficiency. The GAO placed the DOD supply chain on the GAO high-risk program list in 1990 where it currently remains today (GAO 2015a).

B. REPORTS AND ANALYSIS

This section of the chapter reviews reports from RAND Cooperation, the DOD Inspector General (IG) and the GAO to provide more background on the distribution problems the DOD currently faces.

1. Effectively Sustaining Forces Overseas

RAND conducted a supply chain study in 2006 focusing on distribution support of Operation IRAQI FREEDOM (OIF). This study looked at staging inventory at forward deployed distribution depots in order to offset transportation costs. The authors found that weight, rather than IPG, drove transportation mode selection. Peltz et al. (2006) recommended maintaining a healthy forward stock of approximately 20,000 different items. However, more inventory makes forces less mobile and requires a larger forward deployed support infrastructure.

2. DOD Inspector General (IG) Report

In 2007, the DOD IG released a report on *Customer Wait Time* (CWT) transactions for selected Army and USMC units to analyze the CWT effect on operational availability of equipment (Inspector General 2007). CWT is the response time metric for maintenance-specific organizations. The IG chose the Army and USMC because the Army made up 76 percent of all requisitions, and because the USMC averaged 36 days per maintenance requisition based on FY05 data. The Army reported an average of 24 days, and the FY05 CWT goal was 15 days. DOD officials attributed

higher CWT averages to an increased demand due to OIF, and USMC officials attributed delays to improperly closed requisitions. The authors sampled the available data to conduct an independent analysis, resulting in a 90 percent confidence interval of 21.9 to 26.8 days, which was still greater than the FY05 goal of 15 days.

1. GAO reports

Recent GAO reports highlight an inability to track the location and status of cargo, which has led to shortages of critical equipment and supplies in both Iraq and Afghanistan (GAO 2011). In 2011, GAO attributed inefficiencies to a fragmented chain of responsibility because no single entity oversees the entire system. USTRANSCOM oversees the Source, Supplier, and Transporter segments, while the geographic combatant commanders oversee the Theater segment. GAO argues this leads to inefficiencies within the process (GAO 2011). The 2011 report also highlights limited data reliability due to missing delivery information. In its most recent report, GAO highlighted a need for improvement in both the establishment and measurement of performance metrics (GAO 2015b).

C. ATTEMPTS AT SOLVING SIMILAR PROBLEMS

The final section of this chapter includes highlights from scholarly papers reviewed prior to formulating the methodology outlined in Chapter III. The major areas we review range from using artificial neural networks (ANN) in supply chain planning to employing classification trees to reduce delivery variability to using distribution models and associate rules to determine optimal shipping combinations. These methods provide insight into solving similar problems and provide a baseline from which this thesis builds.

1. Artificial Neural Networks (ANN) in Supply Chain Planning

Chui and Lin (2004) use ANNs to model resource-oriented supply chain networks for assembly-to-order products with quick delivery lead times. The authors use three ANNs to map supply, production and delivery resources capable of meeting both customer and individual resource constraints and goals while also maximizing the global benefit to the supply chain. Decomposing the supply chain into smaller, more

manageable problems enabled complete fulfillment of all orders while significantly improving resource utilization rates throughout the supply chain.

2. DOD Source of Supply and Carrier Effects on Shipping Timelines

Sagara (2008) uses Poisson generalized linear models (GLM) to determine if source of supply and carrier impact shipping times of Navy IPG 1 requisitions processed by the Bremerton, WA Priority Material Office (PMO). He focuses on shipments to PACOM, United States Central Command (CENTCOM), United States European Command (EUCOM) and major fleet concentrations within the Continental United States (CONUS) from 2005 to 2008. His research concludes that carrier selection impacts shipping times and better performing carriers are often underutilized. Additionally, he notes statistically significant differences in processing times based on the assigned source of supply.

3. Logistics Support for the Marine Corps Distributed Laydown

The Center for Naval Analysis (CNA) reviews various aspects of the current Marine Corps Forces Pacific Command (MARFORPAC) logistics support system, including a supply support simulation model for Guam and Australia. Fredlake and Randazzo-Matsel (2013) look at the distribution of consumable items to simulate supplies issued daily by deployable supply units (DSU) utilizing military air (MILAIR) and commercial air networks. Figure 3 shows the model inputs and parameters used to determine the impact on total transportation costs and the percent of days the unit is at target inventory level. Using historical averages, the model estimates transportation time beginning when an item arrives at the port of embarkation (POE) until it is ready for pickup at the port of debarkation (POD). This covers the Transporter segment depicted in Figure 1.

The model utilizes historical distributions to determine the source of supply as well as the time and cost of delivery. It limits transportation modes to air only despite utilizing supply sources both within and outside of the area of operations (AO) which often require the use of surface assets. The model bases supply effectiveness on the percentage of days the unit is at 95 percent of its target inventory level and reorder points

are based on the estimated lead times required to maintain these levels. Using average transportation times and limiting transit modes to air does not take into account the unpredictability of ocean transit times, thus providing optimistic estimated lead times and potentially setting the conditions for supply shortfalls. Further analysis of transportation times is required in order to provide realistic transportation expectations from this model.

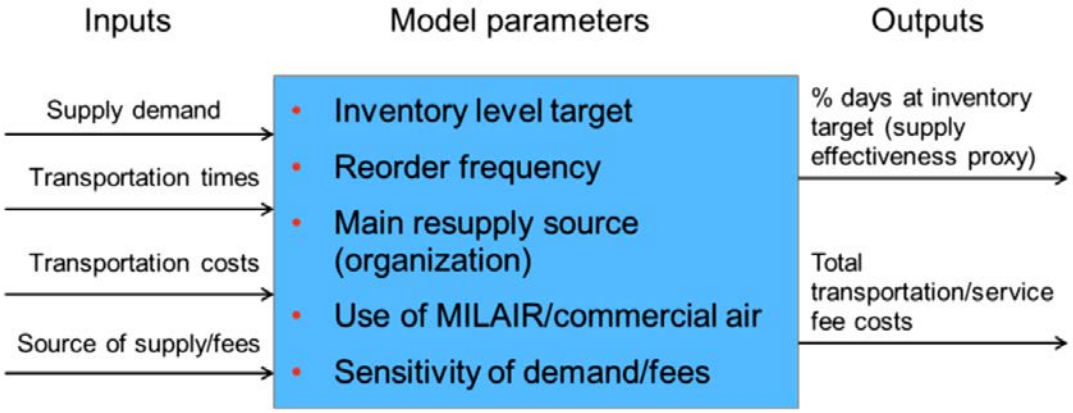

Figure 3. Center for Naval Analysis Supply Model Overview Source: Fredlake and Randazzo-Matsel (2013)

4. Logistics Data Mining to Improve Food Supply Chain Sustainability

Ting, et al. (2013) use association and probability rules to determine the optimal red wine distribution network for an Italian-based wine producer. The decision support model for supply chain quality sustainability (QSDSS) includes transit time, storage temperature, and humidity among other input variables to determine the best combinations of factors that will result in delivery of the highest quality wines. The model's first stage inputs basic logistics information to look for relationships among the combinations of shippers and receivers and outputs a ranked list of quality assurance settings. This becomes the input to the second stage, which returns an aggregated, ranked list of quality settings to determine optimal routes within the distribution network. Radio frequency identification (RFID) gathers point-to-point transactions and temperature

monitoring devices record climate data. This model uses probabilistic support rules to determine the likelihood of two events occurring in the same transaction in order to determine the best combination of shipping factors to maintain product quality during transit.

5. Using Classification Trees for Amazon Inbound Shipments

Chun (2014) uses a classification tree model based on key dates and basic shipment attributes to reduce the variation between the estimated and actual delivery dates to Amazon distribution centers. He uses the Kruskal-Wallis one-way Analysis of Variance (ANOVA) test to determine which shipping attributes reduce joint variation the most and uses these factors to produce a vector of prediction errors for various combinations of delivery dates, vendor codes, carrier codes and final destinations. He uses the resulting error distributions to generate new estimated delivery dates leading to a reduction in customer back orders, the consequence of late shipments. His use of classification trees and error distributions present a good starting point for the methodology development of the DOD distribution system model.

This literature review provides insight into solutions for related problems. The Amazon and Italian wine maker models use forms of In-Transit Visibility (ITV), which provides regular and accurate location updates, but is also very expensive and not widely used by DOD. Chui and Lin (2004) use machine learning algorithms to decompose their supply chain network, and Chun (2014) uses tree models to identify attributes that drive variability within the Amazon system. This research builds upon these concepts, among others, in order to provide a prediction tool utilizing the available databases. In Chapter III, we discuss how we use machine-learning algorithms to determine which predictors drive variability and develop models to predict late shipments.

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA COLLECTION AND PREPARATION

In preparation for analysis, we collect and format relevant data concerning the distribution system. This chapter provides a description of our data as well as the method we use to clean it. Section A gives an overview of the main dataset, and Section B provides an explanation of the process to prepare it for analysis. Section C describes the remaining variables, Section D highlights data quality issues and Section E describes the methodology of this research. We conduct all data cleaning in R, a statistical computing language (R Core Team 2015).

A. DATA

We download and combine 12 monthly iterations of the Strategic Distribution Database (SDDB), available to customers of the Joint Deployment and Distribution Enterprise (JDDE) via online resources. This database represents a comprehensive view of requisition-level data and provides JDDE customers a means to analyze the distribution system (Robbins et al. 2004). Additionally, USTRANSCOM also provided us with the 2015 Hawaii carrier schedules, which we use to compare trends in Hawaii requisitions, and further explain in Chapter IV.

The SDDB includes information about all segments, sub-segments and modes of transportation. It consists of 227 variables from various data collection systems within the JDDE. The Defense Logistics Agency (DLA) Office of Operations Research and Resource Analysis (DORRA) consolidates data monthly and forwards it to the USTRANSCOM J4/J5. The J4/J5 provides additional data and data cleaning before making the database available online to the JDDE (Hiltz 2015). We could not find openly available information concerning the methods by which DORRA consolidates the SDDB. However, RAND originally developed the methodology, which eventually became a DORRA responsibility (Boren 2016).

We begin with over 860,000 observations from United States Pacific Command (PACOM), spanning January to December 2015. Several variable name changes occurred in February 2015 that required significant data formatting and results in several missing

January observations. We filter the data to include only United States Marine Corps (USMC) requisitions shipped by ocean, leaving over 15,000 observations. We remove several variables including unique identifiers and those not applicable to ocean shipments or the Transporter segment. Additionally, we remove variables missing more than 60 percent of observations because there is not enough information stored in these variables for modeling (Kelleher et al. 2015). This process results in 41 variables from which we chose 20 to begin analysis. Additionally, we create three more variables, which we describe in the next section. We use 20 percent of the data to create a test set, comprising of 3,045 observations, which we do not use in fitting the model. This leaves 12,184 observations in the training set with which we begin our analysis.

B. DATA PROCESSING

The following list describes the steps to clean and prepare the final datasets:

1. We create an "other" option for all categorical variables with levels containing fewer than 100 observations. Levels with few observations provide little insight into drivers of variability and further complicate the model.

2. We consolidate location variables to represent geographic combatant commands instead of specific locations in order to reduce the number of categories. Hawaii destinations are the only exception because we encounter unique trends in the data, which we explain in Chapter IV.

3. We create variables to represent the weekday, month, and quarter in which the Transporter leg began.

4. We convert all blank spaces to "NA."

C. VARIABLES

Our analysis begins with 5 different response variables and 18 independent variables, some of which we determine to be insignificant. Chapter IV provides details concerning variable significance. Table 3 provides a brief description of each variable remaining in our dataset, the variable type and the number of missing values per variable.

D. DATA QUALITY

This section describes some of the data quality issues we encounter while working with the SDDB.

Table 3. Strategic Distribution Database (SDDB) Variables Retained for Analysis and Number of Missing Values per Variable.

Variable Name	Type	Description	#NA
Sub-segment 1 (response variable)	Integer	Number of days origin line haul	1428
Sub-segment 2 (response variable)	Integer	Number of days seaport of embarkation (SPOE) hold	1219
Sub-segment 3 (response variable)	Integer	Number days ocean transit	2193
Sub-segment 4 (response variable)	Integer	Number days seaport of debarkation (SPOD) hold	4464
Sub-segment 5 (response variable)	Integer	Number of days destination line haul	5019
Afloat	Binary	1 = ship-based customer, 0 = not ship-based	30
Booking method	Categorical	Booking method	785
Carrier	Categorical	Contracted carrier	765
Container	Categorical	Type of container	922
Handling	Categorical	Shipment processing requirements due to size, weight or security	2640
Initial consolidation point	Categorical	Initial Consolidation Point organization	64
Integrated distribution lane	Categorical	Assigned integrated distribution lane (IDL) short name	30
Issue priority group	Categorical	Designates shipping priority	64
Location	Categorical	Customer location	30
Month	Categorical	Month Transporter leg initiated	958
Quarter	Categorical	Quarter Transporter leg initiated	958
Service terms	Categorical	Service terms of booking	30
Shipping cost	Continuous	Cost to ship the item	500
Supply class	Categorical	Class of supply	32
Unit price	Continuous	Item cost	45
Weekday	Categorical	Weekday Transporter leg initiated	958
Weight	Continuous	Shipping weight	280

1. Missing Values

We encounter multiple missing values in this dataset even after reducing it to only a fraction of its original size. Missing values range from zero to 41 percent per variable, and we list the number of missing observations for each variable in Table 3. We provide a breakdown of missing percentages per variable in Chapter IV. Machine learning algorithms cannot train on missing values (Kelleher et al., 60). This dataset contains only 4,919 complete cases, meaning that over 60 percent of this already reduced dataset does not have the information necessary to train accurate models capable of analyzing and predicting a complex system such as the DOD distribution pipeline.

2. Erroneous Entries

Missing values are easily identifiable data quality issues within the SDDB. However, we have no way of determining the quality of data available in the SDDB and must trust that it is high enough to support our analysis.

USTRANSCOM provided a 5-year subset of the Integrated Mission Support for Surface Deployment and Distribution Command (iSDDC) dataset for this research. This information is specific to ocean shipments and serves as the source of sub-segment timestamps for the SDDB (USTRANSCOM 2015). However, the iSDDC tracks all classes of supply at the container level, while the SDDB focuses on sustainment materiel at the requisition level. The datasets do not directly compare, however, working with the iSDDC provides insight into the quality of data compiled into the SDDB.

Focusing primarily on iSDDC timestamp data, we find that over 50 percent of recorded shipments in 2015 contain either erroneous entries or missing values. We define erroneous entries as negative travel times, ocean transit times of zero days, and sub-segment lengths lasting longer than 365 days. The SDDB consolidation process omits most erroneous entries (Boren 2016). We suspect the SDDB does not accurately reflect the sustainment requisitions that pass through the system even before we remove missing values.

E. METHODOLOGY

The following sections describe the methods we use to develop predictive models. Each sub-segment measures a different activity in the transportation process and requires a separate model for accurate prediction. We assume sub-segments to be independent and provide an explanation of this assumption in Chapter IV. Due to the significant differences in distance between Hawaii and other PACOM destinations, we create subsets to represent Hawaii and non-Hawaii observations and develop two different ocean transit models. Figure 4 illustrates the organization of the six models resulting from this research. We employ three different analytical methods to develop these models. The following sections describe these methods.

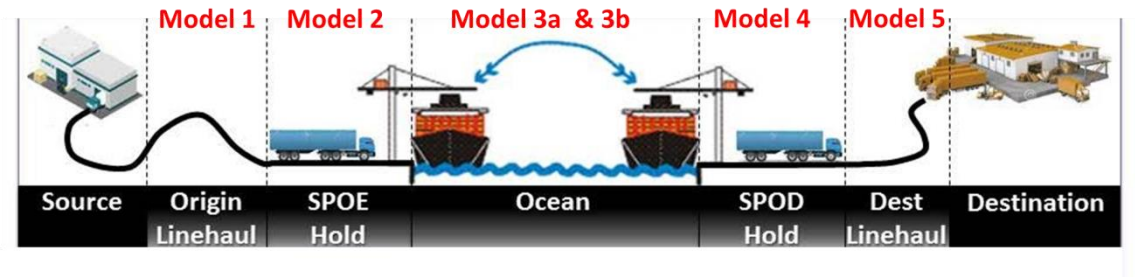

Figure 4. Predictive Model Organization by Segment. Adapted from USTRANSCOM (2015).

1. Baseline Model

As discussed in Chapter II, USTRANSCOM utilizes averages in their current prediction model, and this research seeks to improve upon performance of that model. We assume the average to be our baseline and use it to evaluate the performance of the models we discuss in Chapter IV.

2. Multivariate Linear Regression

Multivariate linear regression describes the expected value of the response variable as a linear function of independent predictor variables and fits a plane through the data in order to minimize the errors between the actual dependent variable values and

the values predicted by the model (Wackerly et al. 2008, 567). Linear regression requires errors with a normal distribution, constant variance and no unusual or overly influential observations (Faraway 2015, 73). Any violation of these assumptions can lead to problems with the model or its conclusions. Linear regression is the simplest of the techniques we employ and provides insight into the drivers of variability even if the model does not meet the required assumptions.

We initially use bidirectional stepwise regression and choose the model that minimizes Akaike Information Criterion (AIC) to avoid overfitting. The AIC provides a balance between the model fit and simplicity (Faraway 2015, 154). We then use manual variable deletion based on a 0.05 p-value threshold to further tune our model.

We use diagnostic plots to validate model assumptions. Patterns in plots of the fitted versus residual values indicate non-constant variance, which can reduce the accuracy of model inferences (Faraway 2015, 77). We use quantile-quantile (Q-Q) plots to validate the normality assumption, a lack of which reduces optimality in the estimates (Faraway 2015, 78–80).

3. Regression Trees

Regression trees use a recursive partitioning algorithm to split observations into tree nodes (Breiman et al. 1984). The model bases predictions on the average of the observations partitioned into each terminal node, and measures of impurity evaluate the overall performance of the tree, which the algorithm bases on the total sum of squares at each node (Faraway 2006, 252 and Grömpling 2013). Lower impurity values indicate better fitting models. Cost-complexity-pruning controls tree size using cross-validation thus preventing the tree from overfitting the training data, and the optimal number of splits provides the tree with the minimum cross-validated error (Faraway 2006, 252). Regression trees easily detect feature interactions and handle potential outliers better than linear regression. They split outliers into a different node thus reducing the node residual sum of squares (RSS) making trees more robust to outliers and a more effective analysis tool for this research (Breiman et al. 1984, 253). However, regression trees must still

meet the assumption of homoscedasticity of errors using the validation techniques previously described (Faraway 2006, 254).

We use variable importance rankings to gain insight into the significant drivers of variability. Variable importance is indicative of the splitting power of the variable and measures the decrease in impurity produced by the best split on a variable at each node (Breiman et al. 1984, 147)

4. Random Forests

Random forests fit several regression trees on the same dataset and average the outcome (Breiman 2001). The model chooses a random subset of the training data for each tree and a random predictor without replacement at each split. This results in reduced correlation among trees, and averaging uncorrelated trees reduces overall variation (Grömping 2013). Averaging many trees also reduces the effect of non-normality and heteroscedasticity of errors, and unlike regression trees, random forests will not over-fit the data even as the number of trees increases (James et al. 2015, 320).

We determine the number of splits by dividing the total number of predictors by three and then fit 1000 trees (Welling et al. 2016). The model omits out-of-bag (OOB) observations, approximately 37 percent of the training data, from each tree, and then uses these observations to calculate cross-validated predictions (Welling et al. 2016). We tune the model by adjusting the number of trees and random splits based on the OOB error estimates. Random forests evaluate variable importance based on the average increase in accuracy of OOB estimates as well as the total decrease in node purity resulting from splits on that feature (Louppe et al. 2013).

We utilize feature contribution plots to visualize the structure and variable interactions of our random forest models. Welling et al. (2016) find individual feature contributions to be additive within the random forest model. They sum the local increment, a scalar that describes the relationships between the predictor and response variables, which results from each split within the random forest for each predictor variable (Welling et al. 2016). We utilize *forestFloor* to plot the feature contributions of OOB observations and use different color schemes to identify feature interactions

(Welling 2016). Feature contributions enable analysis of significant variables in the model and their impacts on predictions (Palczewska et al. 2013). Variable importance assesses the average importance of each variable within the model, and the percent increase in mean square error (MSE) shows how much the MSE increases by removing the feature (Breiman 2001).

5. Model Evaluation

Because our models are error-based, we utilize root mean square error (RMSE) and mean absolute error (MAE) to evaluate performance. RMSE sums the square of the actual minus predicted values and then takes the square root of that value. The result is in the same units as the response, making it more desirable than other metrics such as MSE (Kelleher et al. 2015, 444). Because RMSE squares errors, it weights larger errors more than smaller ones. Therefore, we also use MAE as a performance metric; this is also in the same units as the response variable, but weights all errors proportionally to their size (Kelleher et al. 2015, 444). MAE will always be smaller than RMSE, but RMSE provides a more pessimistic metric making it more desirable for estimating model performance (Kelleher et al. 2015, 446).

We apply these techniques to our SDDB training dataset and evaluate their performance using our test set. Chapter IV outlines the development and evaluation of each model.

IV. MODEL ANALYSIS AND EVALUATION

This chapter covers the analysis and evaluation of the methods described in Chapter III and focuses on sub-segment 3, the ocean transit sub-segment. We cover both the Hawaii and non-Hawaii subset models in this chapter and detail the remaining sub-segment models in Appendices C through F. We conduct all analysis using R, a statistical computing language (R Core Team 2015).

F. DATA EXPLORATION

This section includes a brief overview of the 2015 Strategic Distribution Database (SDDB) dataset. We discuss descriptive statistics and the assumption of independence between all sub-segments. We use our training set, consisting of 12,184 observations, to fit all models described in the following sections.

1. Descriptive Statistics

Sub-segment length serves as the dependent variable for each model, and Figure 5 shows a boxplot of each of the sub-segment lengths for which we build a model. We model them separately because each measures a different part of the Transporter process. We discuss the independence of each sub-segment later in this chapter. Additionally, we model Hawaii and non-Hawaii transit times separately to account for the difference in travel distance. Table 4 lists the summary statistics for each sub-segment. The means we list in this table serve as our baseline models. Figure 5 shows long tails in most sub-segments, which indicate skewed distributions. This suggests the average is not an appropriate method for predicting sub-segment length. Figure 6 shows a histogram of the logarithmic transformation of each sub-segment, which we use as the our response variable.

Table 4. SDDB Descriptive Statistics of Transporter Sub-Segments in Days.

	Sub-Segment 1	Sub-Segment 2	Sub-Segment 3 Hawaii	Sub-Segment 3 Non-Hawaii	Sub-Segment 4	Sub-Segment 5
Min	0	0	1	9	0	0
1st Quartile	0	4	2	16	4	0
Median	0	6	4	17	6	0
Mean	1.5	6.1	3.8	17.5	7.8	0.3
3rd Quartile	1	8	5	19	12	0
Max	129	47	24	56	37	58
Stan. Dev	4.8	4.3	2.2	2.8	5.9	1.2

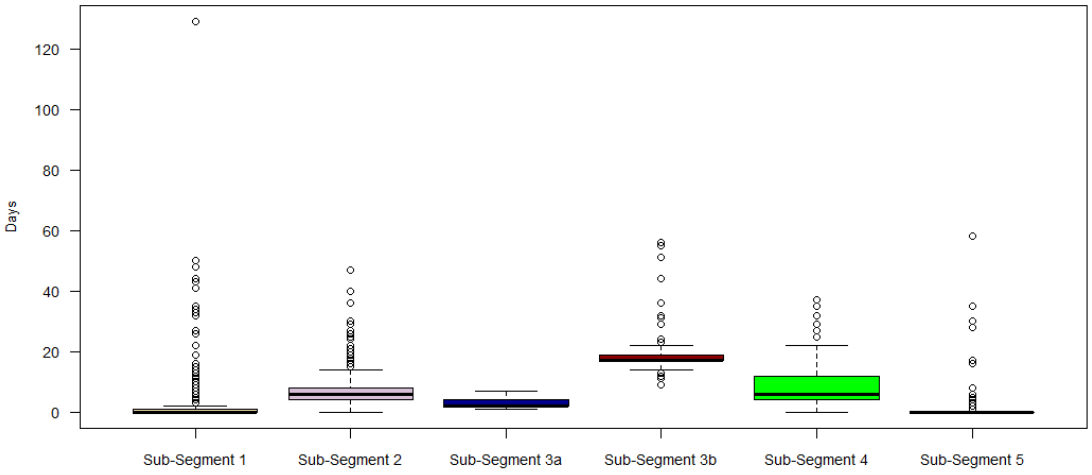

Figure 5. Length of Transporter Sub-Segments Measured in Days

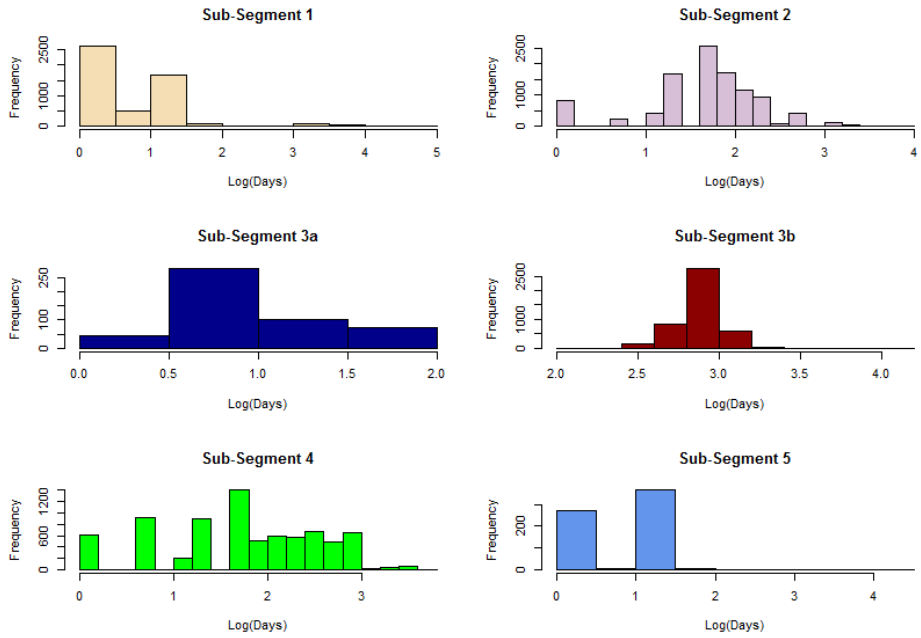

Figure 6. Logarithmic Transformation of Length in Days per Model Sub-Segment.

2. Independence Assumption

We use a pairs plot and correlation table to determine independence between each of the sub-segments, without which we could not model sub-segments separately. Figure 7 illustrates the pairs plot of each sub-segment. We observe no significant visual indications of correlation between sub-segments. We use a correlation table (Table 5) to verify these results. Based on both the pairs plot and the correlation table, we assume independence between the sub-segments and model them separately.

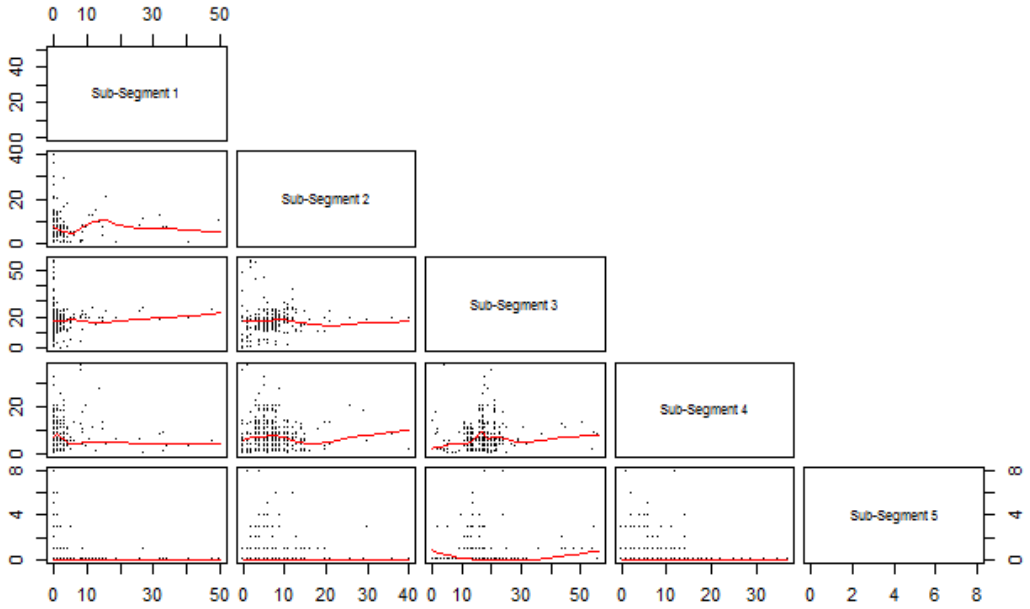

Figure 7. Pairs Plot of SDDB Transporter Leg Sub-Segments/

Table 5. SDDB Transporter Leg Sub-Segments Correlation Table.

	Segment1	Segment2	Segment3	Segment4	Segment5
Segment1	1	–0.08	0.13	–0.10	–0.08
Segment2	–0.08	1	0.20	–0.08	–0.15
Segment3	0.13	0.20	1	0.17	–0.55
Segment4	–0.10	–0.08	0.17	1	–0.22
Segment5	–0.08	–0.15	–0.55	–0.22	1

3. Data Quality

As discussed in Chapter III, we encounter data quality issues that influence the outcome of our models. Table 6 shows the proportion of missing values per variable. Table 7 shows the number of complete cases available for analysis per sub-segment model. Figure 8 shows the total number of missing values across all variables per month. When we remove incomplete cases from the dataset, we also remove all observations from units afloat. Additionally, missing values influence the significance of some predictor variables, which we discuss later in this chapter. We suspect that missing data affects our ability to accurately model each sub-segment as some sub-segment models

lose over 60 percent of observations due to low data quality. Additionally, we currently possess no way to gauge the quality of the data available within this dataset.

Table 6. Proportion of Missing Values per SDDB Variable.

	# NA	percentage missing
Sub-segment 1	1428	12%
Sub-segment 2	1219	10%
Sub-segment 3	2193	18%
Sub-segment 4	4464	37%
Sub-segment 5	5019	41%
Issue priority group	30	0%
Weekday	958	8%
Month	958	8%
Quarter	958	8%
Integrated distribution lane	30	0%
Supply class	32	0%
Carrier	765	6%
Weight	280	2%
Booking method	785	6%
Container	922	8%
Shipping cost	500	4%
Location	30	0%
Initial consolidation point	64	1%
Service terms	30	0%
Handling	2640	22%
Origin	1961	16%
Afloat	30	0%
Unit price	45	0%

Table 7. Number of Complete Cases and Percentage of Missing Cases per Sub-Segment Model.

	# complete cases	percentage missing
Sub-segment1	7722	37%
Sub-segment2	7502	38%
Sub-segment 3 Hawaii	1079	66%
Sub-segment 3 Non-Hawaii	5515	38%
Sub-segment4	5243	57%
Sub-segment5	4981	59%

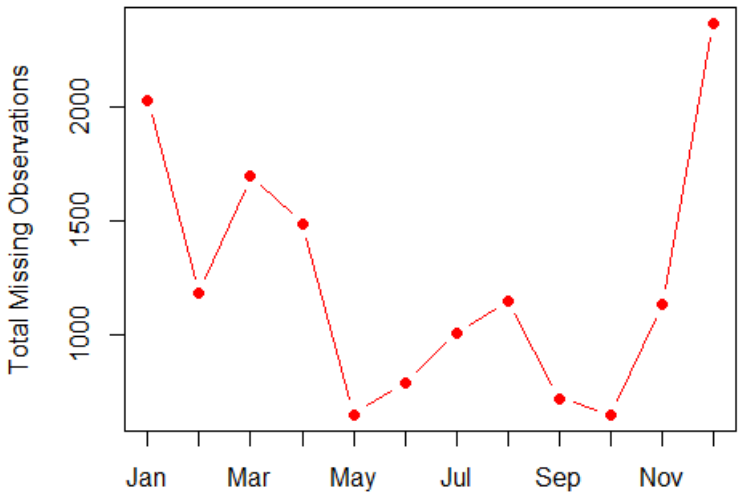

Figure 8. Total Number of Missing Observations across All Variables per Month.

G. HAWAII SEGMENT 3 MODEL ANALYSIS

In this section, we apply the techniques covered in Chapter III to the Hawaii subset of the SDDB. The model uses the length of sub-segment 3, measured in days, as the dependent variable and begins with the remaining SDDB variables described in

Chapter III as the independent variables. We present the results of each technique followed by model comparison and discussion of significant findings.

1. Baseline Model

We use the mean of the response variable as a baseline model from which to compare subsequent models. As discussed in Chapter II, the current USTRANSCOM prediction tool uses only the mean to predict shipment timelines. The mean ocean transit length, as shown in Table 6, is 3.8 days. We use the baseline model to predict sub-segment length on the test set and list the root mean square error (RMSE) and mean absolute error (MAE) in Table 12.

2. Multivariate Linear Regression Analysis

We use the logarithmic transformation of the response variable to reduce the skewed distributional effects previously discussed. Sub-setting the data into Hawaii and non-Hawaii observations and eliminating incomplete cases reduces *integrated distribution lane*, *origin*, and *afloat* to only one factor level, so we exclude them from the linear regression model. We also eliminate *location* because of uneven representation of factor levels following the removal of incomplete cases. Kaneohe Bay has 1041 observations and Pearl Harbor has only 38 observations.

Using bidirectional stepwise regression, we fit an initial model and then use manual deletion based on a 0.05 p-value threshold to develop the final linear regression model. Table 8 shows the coefficients of the significant predictor variables as well as their associated standard errors and p-values, and Table 9 shows the model goodness of fit metrics resulting from this model.

Table 8. Hawaii Linear Regression Model Coefficients.

	Estimate	Std. Error	P-value
Intercept	0.83	0.06	< 2e-16
Feb	0.20	0.13	0.12
Mar	–0.05	0.05	0.39
Apr	0.78	0.06	< 2e-16
May	0.27	0.06	0.00
Jun	–0.09	0.06	0.15
Jul	0.03	0.07	0.63
Aug	–0.70	0.07	< 2e-16
Sep	–1.38	0.07	< 2e-16
Oct	–1.08	0.08	< 2e-16
Nov	–0.44	0.06	0.00
Dec	–0.92	0.06	< 2e-16
Handling B: High sensitivity category I, heavy lift (HL)*	0.60	0.04	< 2e-16
Handling G: High sensitivity category I confidential, HL*	0.59	0.06	< 2e-16
Handling N: low sensitivity category IV, outsize dimension (OD)*	0.59	0.05	< 2e-16
Handling: Other	0.88	0.08	< 2e-16
Handling R: No special handling, OD*	0.30	0.09	0.00
Handling Z: No special handling, HL and OD*	0.29	0.03	< 2e-16
Tue	–0.06	0.05	0.28
Wed	0.70	0.03	< 2e-16
Thu	0.30	0.03	< 2e-16
Fri	0.64	0.03	< 2e-16

*Source: Defense Transportation Electronic Business Reference Data

Table 9. Hawaii Linear Regression Model Goodness-of-Fit Performance Metrics Using the Logarithmic Transformation of the Response.

Metric	Value
Residual Standard Error	0.19
R Squared	0.88
Adjusted R Square	0.88
Degrees of Freedom	1054

We use residual plots and quantile-quantile (Q-Q) plots to verify the assumptions of the model, both of which are shown in Appendix A. The model violates the homoscedasticity and normal errors assumptions, but still provides useful insight into the drivers of variability, which we will discuss in a later section.

3. Hawaii Regression Tree Model

We use the *rpart* package in R to implement regression trees as described in Chapter III and the *rpart.plot* package to plot the results (Therneau, Atkinson, Ripley 2013 and Milborrow 2015). This model uses the same response variable as the linear regression model. We grow a full regression tree and then prune it to the optimal number of splits based on the complexity parameter (cp) with minimum cross-validated error. This occurs at cp = 0.004 and results in 10 splits. We show the regression tree in Appendix A, and Table 10 lists the resulting variable importances.

4. Random Forest Analysis

We use the *randomForest* package to fit an initial model with 1000 regression trees, each with six random splits (Cutler et al. 2015). This method averages the outcomes of the trees to determine variable importance and explains variation in the response as a function of the predictors. We eliminate *afloat, origin, booking method, carrier, supply class, issue priority group, unit price,* and *shipping cost* because their presence in the model did not increase model performance. We fit our final model with only two random splits per tree. Table 11 lists the percent increase in mean square error (MSE) that would result from removing each of the remaining variables.

The random forest model yields new insights into the drivers of variation in the distribution system. All three models find *month*, *handling*, and *weekday* highly significant, but the random forest model also finds *service terms* significant.

Table 10. Hawaii Regression Tree Model Variable Importance Table.

	Variable Importance
Month	293.01
Handling	216.74
Weekday	146.45
Container	102.72
Shipping cost	43.73
Weight	30.36
Unit price	12.03
Location	10.55
Service terms	9.11
Issue priority group	8.14
Initial consolidation point	2.45
Supply class	1.04
Carrier	0.46

Table 11. Hawaii Random Forest Variable Importance Table.

	% Inc MSE
Month	116.39
Handling	81.15
Weekday	61.64
Service terms	37.9

5. Hawaii Model Evaluation

Comparing the performance of all models on estimating the ocean transit sub-segment length, we find that the regression tree model has a slightly lower MAE, however regression trees can over-fit the training data. Both the regression tree and

random forests models perform significantly better than linear regression and the baseline model, the current USTRANSCOM basis for predictive analysis. Table 12 lists the performance metrics for each Hawaii model when applied to the test set.

Table 12. Hawaii Model Test Set Performance Metrics Measured in Days.

	Root Mean Square Error	Mean Absolute Error
Baseline	1.54	1.42
Linear Regression	0.71	0.31
Regression Tree	0.19	0.03
Random Forest	0.18	0.06

All models find *month* to be a significant driver of variability and we suspect its importance is artificially inflated by removing missing observations. We ran simple linear regressions on two subsets of the data. Model A uses a complete cases subset, the same response variable and *month* as the only predictor variable, which results in an R Square of 0.79. Model B uses a subset including incomplete cases and results in an R Square of 0.37, a decrease in over 40 percent of variation explained. Poor data quality negatively affects the ability to accurately analyze the performance of the distribution system.

We utilize *forestFloor* to decompose our random forest model and better understand the relationship between our predictor and response variables (Welling et al. 2016). As discussed in Chapter III, the OOB feature contribution is the sum of all local increments per variable, and the local increment is a scalar that describes the relationship between the predictor and responses variables at each split in the forest (Welling et al. 2016). Each point on the plot in Figure 9 represents one OOB observation that falls into one of the feature categories shown on the x-axis. The vertical position represents the random forest's estimate of the effect of the variable, or its feature contribution. *ForestFloor* computes feature contributions by summing the OOB local increments and dividing by the number of times that observations fell out of the bag (Welling et al. 2015). The colors on each graph reflect the month in which the observation began the Transporter segment and enables us to visualize feature interactions such as those in

Figures 10 and 11. We jittered both the horizontal and vertical scales to make the individual points easier to discern. The smooth line represents trends associated with each group of observations across each feature level (Welling et al. 2016).

Table 11 shows that *month* is the main contributor to variance in our model and removing it would result in over 100 percent increase in MSE. Figure 9 suggests that requisitions beginning the Transporter segment in January through July experience longer ocean transit times than those beginning in August through December because the January to July shipments have a positive relationship with the response variable. Our feature contributions plots enable us to see that a requisition beginning the Transporter segment on a Thursday in January and shipped under code B handling conditions will take longer to complete the ocean transit segment due to the feature contributions of variables at the stated levels. We show these relationships in Figures 9, 10 and 11.

Figure 10 confirms the regression tree splits for *handling*, which indicate that handling codes 9 and N lead to shorter transit times. The blue shaded points represent requisitions beginning Transporter in the second half of 2015, and most indicate decreasing transit times except for potential outliers in category B.

We encounter an interesting relationship between the feature contributions of *month* and *weekday*. *Weekday* significantly contributes to variation in all models, and Figure 11 shows the interaction between the two variables. Royal blue points represent shipments beginning Transporter in December and they perform better on Mondays and Tuesdays than later in the week.

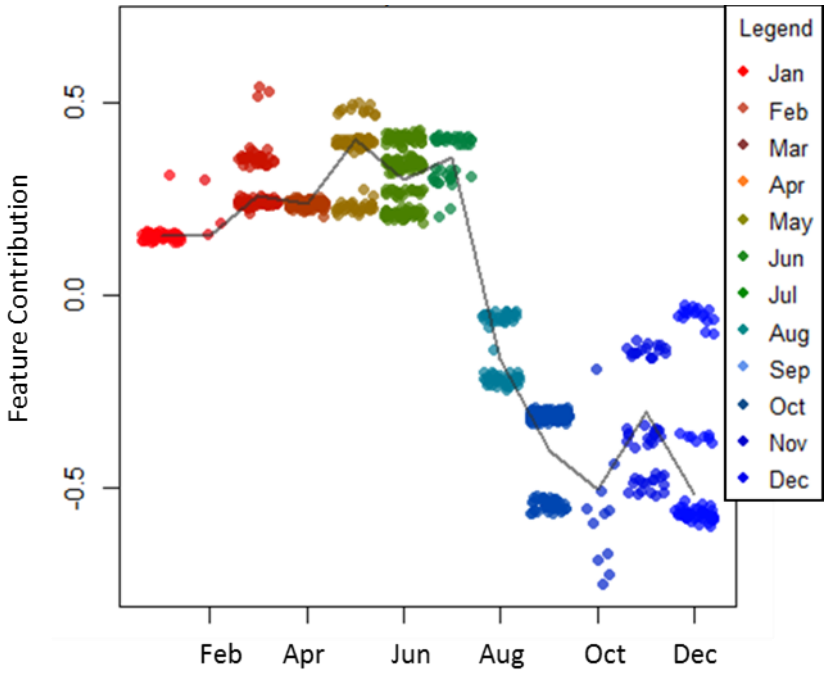

Figure 9. Random Forest *month* Feature Contribution.

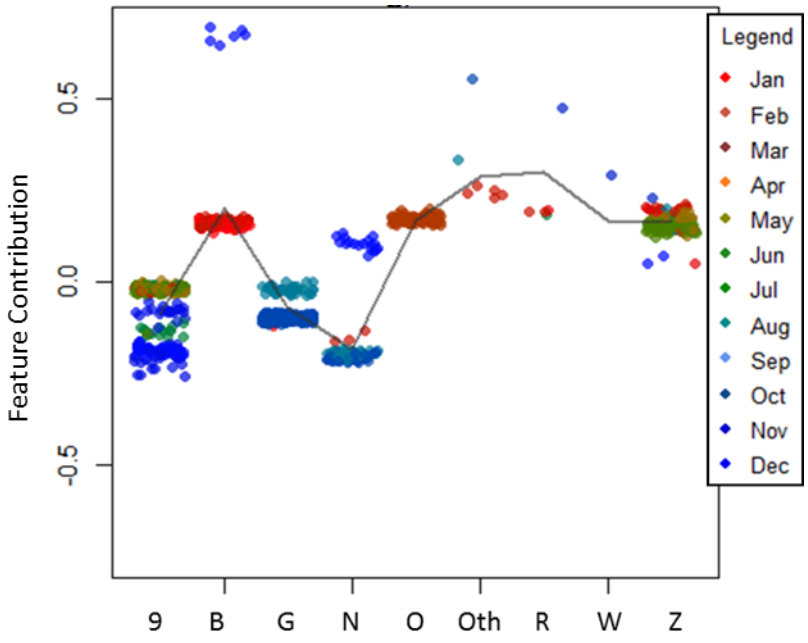

Figure 10. Hawaii Random Forest *handling* Feature Contribution.

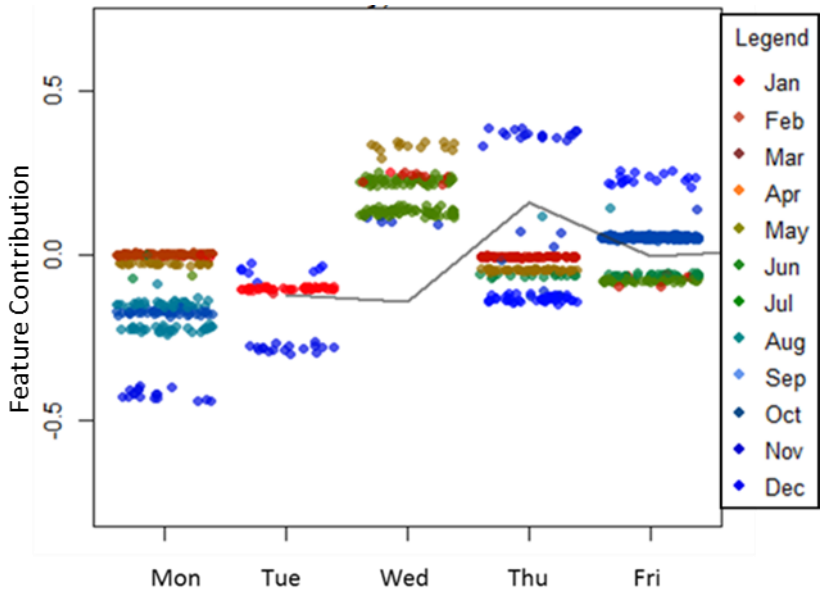

Figure 11. Hawaii Random Forest *weekday* Feature Contribution.

We create Table 13 to further explore the interaction between *month* and *weekday* and find that shipments do not begin the Transporter segment on Tuesdays during most months in 2015. We use the carrier schedules described in Chapter III to look for patterns in shipping schedules that could potentially provide insight into this relationship. Table 14 shows the schedules for carriers departing the west coast for Hawaii broken down by month and weekday. While we cannot attribute any direct causes to the Tuesday effect found in Table 13, we see an uneven distribution between the days of the week in which carriers leave port.

Lastly, the random forest model provided intuitive results concerning service terms. Shipments with multiple stops prior to their final destination take longer than those shipped via other service terms. Based on the results in Table 12, we determine that the random forest provides the most insight and thus the best predictive results for this sub-segment.

Table 13. Hawaii Shipments Broken Down by Month and Weekday the Transporter Segment Began.

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Jan	0	0	45	1	0	0	0
Feb	0	1	1	0	0	1	0
Mar	0	16	0	11	83	5	0
Apr	0	97	0	0	0	0	0
May	0	27	0	14	56	0	0
Jun	0	1	0	121	0	51	0
Jul	0	2	0	0	14	26	0
Aug	0	71	0	0	1	1	0
Sep	0	43	0	0	1	253	0
Oct	0	0	0	4	6	1	0
Nov	0	17	17	0	18	0	0
Dec	0	0	8	0	47	18	0

Table 14. 2015 Hawaii Carrier Schedule Broken Down by Month and Weekday the ship departed the Seaport of Embarkation.

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Jan	37	0	8	31	6	26	15
Feb	26	6	12	17	8	19	12
Mar	47	5	10	13	8	18	10
Apr	38	8	10	17	6	15	11
May	49	7	12	13	8	18	2
Jun	39	11	25	14	6	18	14
Jul	8	14	9	28	6	16	19
Aug	8	20	6	28	8	12	33
Sep	8	19	10	30	7	11	29
Oct	53	0	6	43	0	25	0
Nov	60	0	5	37	0	23	0
Dec	37	0	6	42	0	19	17

H. NON-HAWAII SEGMENT 3 ANALYSIS

We utilize the same modeling techniques and variables for this model, but use the non-Hawaii subset of the training data. We explain each of the four models separately,

compare their performance and provide a brief analysis of the resulting significant variables in the following sections.

1. Baseline Model

The baseline model uses only the average transit time of 17.5 days, as listed in Table 5, to predict performance. We use this outcome to gauge the improvement of our subsequent models. Table 19 lists the model performance metrics against the test set.

2. Multivariate Linear Regression Analysis

Using the same methodology and variables previously described, we fit a linear regression model for the complete cases of the non-Hawaii subset all remaining categorical variables have two or more factor levels, so we do not remove variables before fitting the initial model. Table 15 lists the significant variables in the linear regression model, and Table 16 lists the goodness of fit metrics.

We use residual and Q-Q plots to verify that the model does not meet homoscedasticity or normal errors assumptions. We show these plots in Appendix B.

3. Regression Tree Analysis

We fit a regression tree using the training set and, the minimum cross-validated error for our regression tree occurs at cp = 0.00029. This results in a tree with 75 splits, which is too large to plot. Table 17 lists the variables in order of importance in this model. Table 19 lists the evaluation metrics for this model.

4. Random Forest Analysis

We initially fit a random forest with all potential predictor variables and remove insignificant variables to develop the final prediction model. We define insignificant variables in this model as those with less than 0.01 on the variable importance table. We remove six predictor variables so the subsequent model fits 1000 trees with four random splits. This model results in 92.3 percent of variation explained when applied to the test set. Table 18 lists the model variables in order of importance. This model indicates that removing *month* will increase MSE by almost 300 percent.

Table 15. Non-Hawaii Linear Regression Model.

	Estimate	Std. Error	P-value
(Intercept)	2.55	0.01	< 2e-16
Feb	–0.04	0.01	0.00
Mar	0.18	0.01	< 2e-16
Apr	0.04	0.00	< 2e-16
May	–0.11	0.00	< 2e-16
Jun	–0.15	0.00	< 2e-16
Jul	–0.04	0.01	0.00
Aug	–0.06	0.01	< 2e-16
Sep	–0.10	0.00	< 2e-16
Oct	–0.12	0.00	< 2e-16
Nov	–0.04	0.01	0.00
Dec	–0.12	0.01	< 2e-16
Carrier: MAEU	0.03	0.01	0.00
Carrier: MATS	0.16	0.06	0.01
Carrier: OTHER	–0.13	0.03	0.00
Booking: IBS	0.14	0.02	0.00
Booking: Old Method	0.74	0.03	< 2e-16
Booking: Unknown	0.29	0.02	< 2e-16
Handling B: High sensitivity category I, HL*	0.17	0.03	0.00
Handling G: High sensitivity category I confidential, HL*	0.08	0.01	0.00
Handling N: low sensitivity category IV, OD*	0.24	0.01	< 2e-16
Handling O: Highest sensitivity category I classification secret, OD*	0.10	0.01	< 2e-16
Handling: Other	0.02	0.01	0.02
Handling R: No special handling, OD*	0.10	0.01	< 2e-16
Handling W: Highest sensitivity category I classification secret, HL and OD*	–0.02	0.01	0.00
Handling Z: No special handling, HL and OD*	0.06	0.00	< 2e-16
Location: Okinawa	0.32	0.01	< 2e-16
Location: Other	0.01	0.05	0.90

*Source: Defense Transportation Electronic Business Reference Data

Table 16. Non-Hawaii Linear Regression Model Goodness-of-Fit Performance Metrics Using the Logarithmic Transformation of the Response.

Metric	Value
Residual Standard Error	0.09
R Square	0.64
Adjusted R Square	0.64
Degrees of Freedom	6829

Table 17. Non-Hawaii Regression Tree Variable Importance Table.

	Variable Importance
Month	54.16
Location	43.10
Weekday	24.88
Handling	20.57
Carrier	20.47
Quarter	18.74
Booking	16.15
Container	12.79
Shipping cost	9.34
Origin	6.39
Weight	6.05
Unit price	4.59
Supply Class	4.14
Initial consolidation point	1.85
Issue priority group	1.49
Service terms	0.48
Integrated distribution lane	0.01

Table 18. Non-Hawaii Random Forest Variable Importance Table.

	% Inc MSE
Month	294.35
Location	198.31
Weekday	142.77
Handling	88.52
Container	76.73
Service terms	28.69
Booking	28.46
Weight	26.75
Carrier	24.58

5. Non-Hawaii Model Evaluation

Requisitions traveling to non-Hawaii destinations have a greater chance of stopping in multiple locations before reaching their endpoints thus creating a dataset with more noise. However, the models still perform relatively well against the test set. Table 19 lists the performance metrics for each model on the test set.

We find *month*, *handling*, *location* and *carrier* to be among the top predictors in each model. Again, we fit simple regression models for *month* using both a complete and an incomplete cases subset of the data. Model A, the complete cases subset, produces an R square of 0.27, and model B produces an R square of 0.07. Again, we find that the effect of *month* is artificially inflated because of low data quality. We also compare simple linear regressions of *handling*, *location* and *carrier*. We find higher R squares for all complete cases subsets; however, the *location* R Square is over 15 percent higher in the complete cases subset. This leads us to conclude that low data quality has a negative effect on our ability to accurately model the DOD distribution system.

Using *forestFloor*, we decompose our random forest model to evaluate the effects of the main contributors to variation (Welling et al. 2016). Figure 12 shows the feature contributions of each month along the y-axis, and we plot each month in a different color. We find that requisitions beginning Transporter in March have higher transit times than other months throughout the year. We did not find any significant weekday interactions.

The random forest model also identifies *location*, *carrier* and *handling* as significant variables, which confirms the linear regression findings we list in Table 15. Additionally, different handling requirements based on security or size also influence transit times. From our linear regression results, we find that handling code W has a negative relationship with sub-segment time indicating highly sensitive, outsize dimension cargo arrives faster than other handling codes. The random forest model also identifies *container*, *service terms* and *weight* as contributors to variation. Surprisingly, the method by which the shipment is booked also effects ocean transit times. Specifically, requisitions booked via the "Old Method" take longer to complete this sub-segment than requisitions booked by other means as shown in Figure 13.

Table 19. Non-Hawaii Model Test Set Performance Metrics measured in Days.

	Root Mean Square Error	Mean Absolute Error
Baseline	2.93	1.92
Linear Regression	2.27	1.24
Regression Tree	1.49	0.42
Random Forest	1.08	0.44

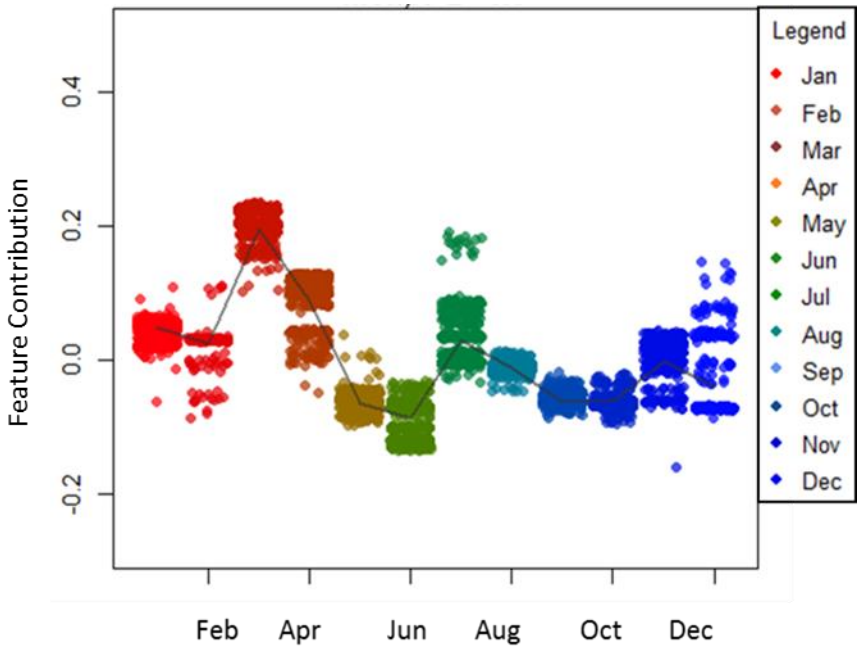

Figure 12. Non-Hawaii Random Forest *month* Feature Contribution.

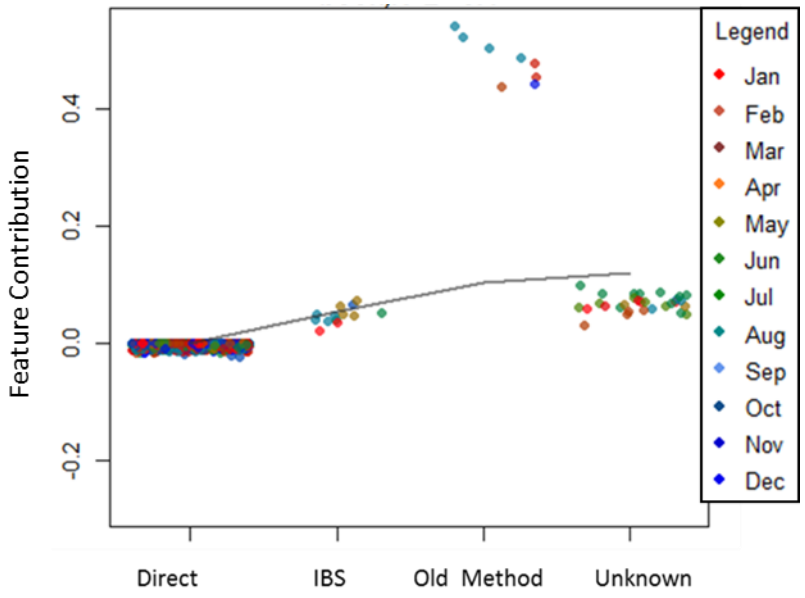

Figure 13. Non-Hawaii Random Forest *booking method* Feature Contribution.

I. TOTAL PERFORMANCE

We explain the details of the remaining sub-segment models in Appendices C through F, but list their test set performance metrics in Table 20. We convert these metrics back to days, the original sub-segment units, and highlight the lowest RMSE and MAE for predicting sub-segment lengths using the test set.

Table 20. Test Set Performance Metrics for All Transporter Sub-Segments Measured in Days.

	Root Mean Square Error				Mean Absolute Error			
	Base	**Lin. Reg.**	**Reg. Tree**	**Ran. Forest**	**Base**	**Lin. Reg.**	**Reg. Tree**	**Ran. Forest**
Sub-Segment 1	4.94	3.69	1.7	1.55	1.93	1.08	0.48	0.41
Sub-Segment 2	3.94	3.55	2.01	1.47	2.59	2.06	0.85	0.53
Sub-Segment 3a	1.54	0.71	0.19	0.18	1.42	0.31	0.03	0.06
Sub-Segment 3b	2.93	2.27	1.49	1.61	1.92	1.24	0.42	0.56
Sub-Segment 4	6.2	3.93	2.3	1.46	5.05	2.95	1.33	0.87
Sub-Segment 5	0.77	1.05	1.01	1.03	0.42	0.15	0.06	0.06

Although regression tree models perform slightly better in some instances, we recommend the random forest models because each is a collection of 1000 regression trees and provides results that are more robust. All of our models show significant improvement over the baseline models except sub-segment 5. In this case, the baseline model provides a better RMSE. As discussed in Chapter III, RMSE provides a more pessimistic evaluation, so we recommend the baseline model to predict sub-segment 5.

We suspect our results are artificially good. All literature reviewed suggest that random forests are less prone to overfitting because of aggregating the outcomes of many trees. Our models fit our data well, but the information in our dataset is not necessarily representative of what actually flows through the system. As discussed in Chapter III, SDDB consolidation filters erroneous entries, and we further filter missing observations

in preparation for our analysis. As a result, we suspect only the highest quality data remains, which is not necessarily representative of shipments that traverse the system. The performance metrics listed in Table 20 indicate that our models successfully fit this dataset. However, in this chapter we present evidence to suggest that low data quality artificially inflates the significance of some variables. Furthermore, we have preliminary indications that our models perform poorly on 2016 data.

Surprisingly, few models found IPG to be a significant driver of variability. As described in Chapter II, IPG 1 requisitions should take less time to complete the Transporter segment than IPG 3. However, the results of the model indicate no significant difference in average delivery time of IPG codes.

J. SUMMARY

In this chapter, we describe the analysis and findings of two of our six models. We provide the detailed explanations of the remaining models in the appendices. Because we assume independence among each of the sub-segments, we add the predictions resulting from each model to estimate the length of the Transporter segment. Table 21 lists the actual sub-segment lengths and predicted sub-segment lengths of five randomly selected requisitions from our test set using the random forest model. We add the sub-segment lengths to calculate the total Transporter transit time.

Table 21. Actual verses Predicted Total Transporter Time in Days.

	Actual						Predicted					
	S1	**S2**	**S3**	**S4**	**S5**	**Total**	**S1**	**S2**	**S3**	**S4**	**S5**	**Total**
1	1	9	20	4	0	34	1	8	18	3	0	29
2	1	6	21	5	0	33	1	7	20	6	0	33
3	0	6	16	5	0	27	0	6	17	2	0	25
4	0	11	17	5	0	33	0	7	16	13	0	36
5	1	7	16	4	0	28	1	6	16	4	0	27

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY AND RECOMMENDATIONS

This chapter provides a summary of the techniques we utilized and the results discussed in previous chapters. We also include recommendations and identify areas for future research.

A. SUMMARY

The goal of this research was to develop a statistical model capable of predicting late shipments based on historical performance. We created a model for each of the five sub-segments within the Transporter leg of the DOD distribution process and addressed these questions in our research:

- What factors drive variability within the distribution system?
- Can a more accurate predictive tool be developed in order to inform decision makers of late shipments prior to shipments missing the RDD?

To support our research, we utilized a subset of data from the SDDB, which we cleaned and reduced to 23 variables. We created three different models for each sub-segment including a linear regression model, regression tree model, and a random forest model. Our research found that the random forest model resulted in the lowest RMSE and MAE for most sub-segments when applied against the test set—not involved in fitting the models—and that most of our models perform better than the baseline model. Additionally, we found that the weekday and month in which requisitions begin the Transporter segment significantly influences many of the sub-segment lengths. However, preliminary trials suggest that the models perform poorly on 2016 data.

Missing values significantly degrade our ability to properly analyze the system. Kelleher et al. (2015) suggest that any variable missing 60 percent or more observations does not have enough information stored to support modeling. Only 40 percent of our training data contains complete cases observations. While this is enough information to complete a model, it is not enough information to assess performance.

B. RECOMMENDATIONS

In this section, we provide recommendations to improve the distribution system as well as recommendations for future work. RAND developed the current SDDB consolidation process over 10 years ago as an in-house analysis tool, and SDDB consolidation later became a DORRA responsibility (Boren 2016). We recommend USTRANSCOM re-evaluate the data collection and consolidation process and take over responsibility of the SDDB as the distribution process owner. Maintaining the process within USTRANSCOM will enable fluid changes and adaptations as the distribution system changes. We also recommend implementing accountability procedures to ensure proper timestamp entries for each segment and sub-segment in the distribution process, as these are the most important data for timeline prediction. Then this analysis should be repeated using random forests with a higher quality dataset.

In order to build on this research, we recommend applying the same algorithms to a multi-year dataset thus enabling better analysis of the monthly trends we highlight in Chapter IV. Additionally, we recommend developing geographic combatant-command-specific predictive tools inclusive of all modes of transportation. Lastly, we recommend a detailed analysis of data quality within the SDDB and how the quality level influences distribution system analysis.

APPENDIX A. HAWAII MODEL

In this section, we provide details on linear regression model diagnostics as well as our regression tree model.

A. LINEAR REGRESSION DIAGNOSTICS

As discussed in Chapter III, our linear regression model must meet model assumptions in order to provide accurate predictions. We use a residual versus fitted values plot and a quantile-quantile (Q-Q) plot to verify the model does not meet these assumptions. Figure 14 indicates unequal variances in the residuals, which reduces the accuracy of model inferences.

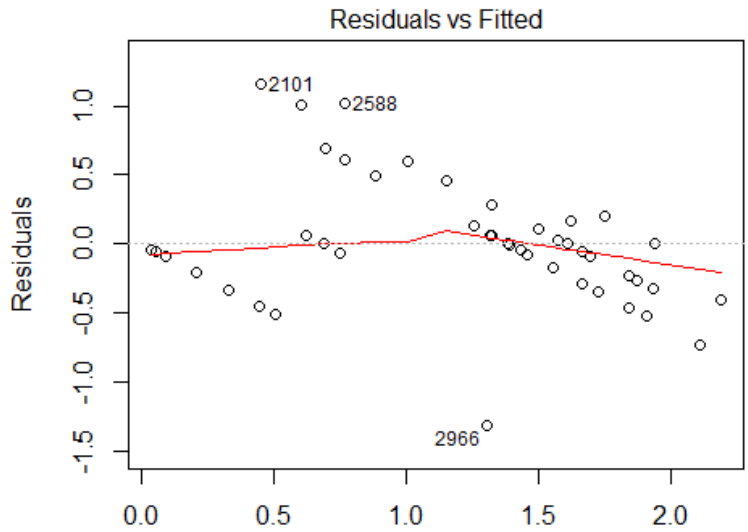

Figure 14. Hawaii Model Residuals versus Fitted Plot.

Figure 15 shows the model violates the normal errors assumption, which also reduces the accuracy of model inferences.

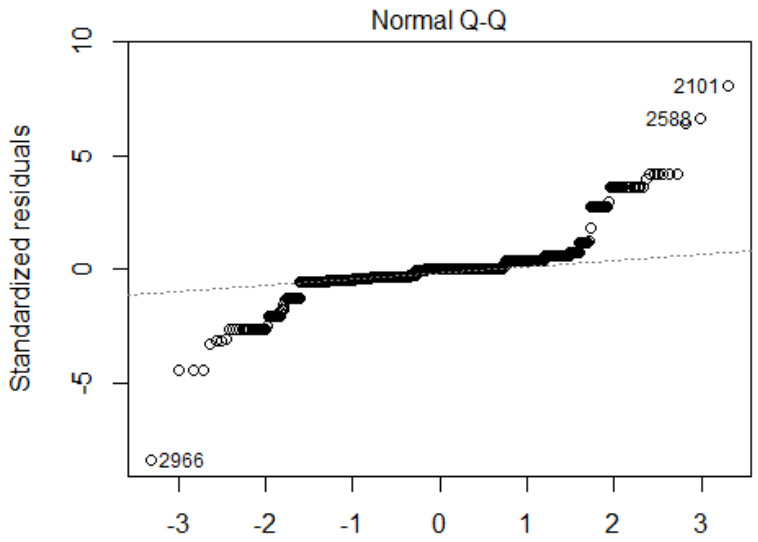

Figure 15. Hawaii Model Quantile-Quantile (Q-Q) Plot.

Residuals still show evidence of long tails even after transforming the response variable. We conclude that the structure of the linear regression model does not support prediction, but does provide insight into variation within the distribution process.

B. REGRESSION TREE MODEL

Figure 16 shows the Hawaii model regression tree. We follow each branch to the terminal node in order to predict future performance of shipments. Regression trees model variable interactions far better than linear regression. Each split beyond the main one indicates an interaction between two or more variables. We follow the branches to the terminal node, which lists the average number of days per shipment that meet branch characteristics.

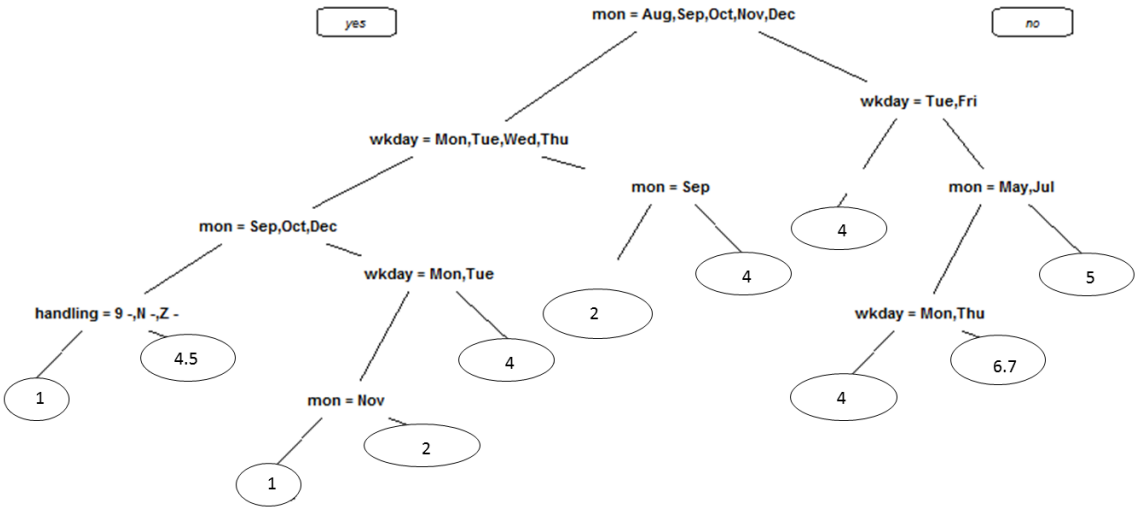

Figure 16. Hawaii Regression Tree Model.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. NON-HAWAII MODEL

In this section, we provide an overview of the non-Hawaii linear regression model diagnostics.

Figure 17 shows the model has non-constant variance, which can negatively influence model predictions, and Figure 18 indicates a long-tailed distribution, which we discuss in Chapter IV. This can negatively influence confidence intervals. We conclude that the linear regression model does not support accurate prediction, but does supplement our understanding of variation in the distribution system.

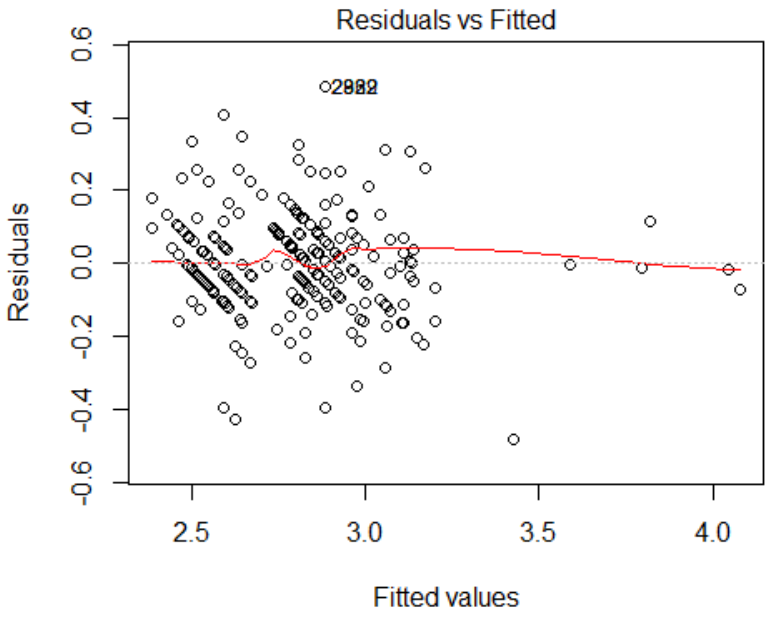

Figure 17. Non-Hawaii Residuals versus Fitted Values Plot.

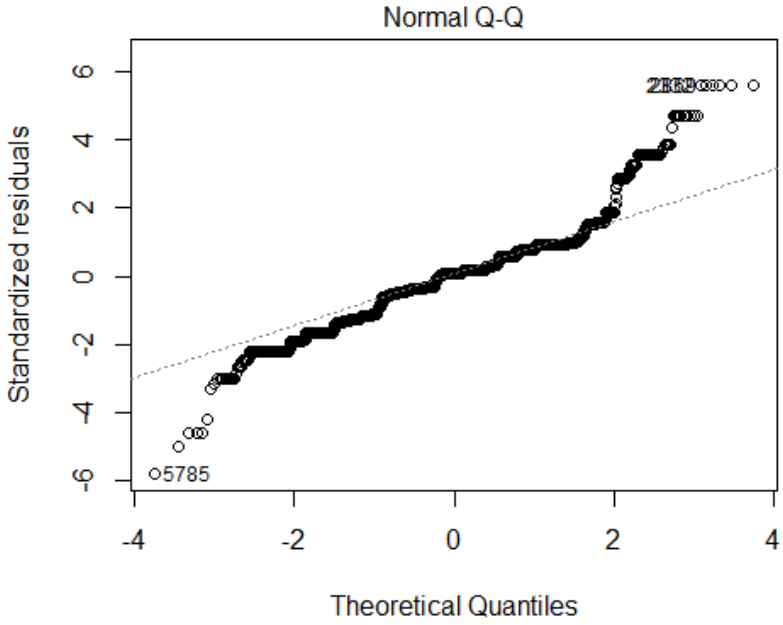

Figure 18. Non-Hawaii Quantile-Quantile (Q-Q) Plot.

APPENDIX C. SUB-SEGMENT 1 MODEL AND EVALUATION

In this section, we provide an overview of the development and evaluation of the sub-segment 1 model. We utilize the techniques outlined in Chapter III and employ the entire training set.

A. BASELINE MODEL

Origin line haul describes the time from which the carrier picks up a shipment from a supplier until it reaches the seaport of embarkation (SPOE), which takes an average of 1.5 days, as shown in Table 4. We use this average as a baseline from which to evaluate our models.

B. MULTIVARIATE LINEAR REGRESSION

We begin this model with all predictor variables described in Chapter III, and use the logarithmic transformation of the sub-segment length as our response variable. We show our results in Tables 22 and 23.

We use Figures 19 and 20 to verify model assumptions. Figure 19 confirms the presence of heteroscedasticity due in part to the discrete nature of the response, an attribute visible in the diagonal lines. Figure 20 confirms non-normal errors. We conclude that the linear regression does not support accurate prediction, but use its results to gain further insight into variation within the distribution system.

Table 22. Sub-Segment 1 Linear Regression Coefficients.

	Estimate	Std. Error	P-value
(Intercept)	0.99	0.02	< 2e-16
Tue	0.18	0.02	< 2e-16
Wed	0.09	0.02	0
Thu	0.48	0.02	< 2e-16
Fri	0.44	0.02	< 2e-16
Feb	−0.62	0.04	< 2e-16
Mar	−0.78	0.03	< 2e-16

	Estimate	Std. Error	P-value
Apr	–0.48	0.03	< 2e-16
May	–0.93	0.03	< 2e-16
Jun	–0.73	0.03	< 2e-16
Jul	-0.7	0.03	< 2e-16
Aug	–0.66	0.03	< 2e-16
Sep	–0.65	0.02	< 2e-16
Oct	–0.72	0.03	< 2e-16
Nov	–0.48	0.04	< 2e-16
Dec	–0.45	0.03	< 2e-16
Carrier: HRZD	1.07	0.1	< 2e-16
Carrier: MAEU	–0.01	0.04	0.8
Carrier: MATS	–0.54	0.02	< 2e-16
Carrier: OTHER	–0.55	0.23	0.02
Booking: IBS	1.92	0.14	< 2e-16
Booking: Old Method	–0.56	0.21	0.01
Booking: Unknown	0.14	0.13	0.29

Table 23. Sub-Segment 1 Linear Regression Goodness of Fit Metrics Using the Logarithmic Transformation of the Response.

Metric	
Residual Standard Error	0.55
R Square	0.40
Adjust R Square	0.40
Degrees of Freedom	7696

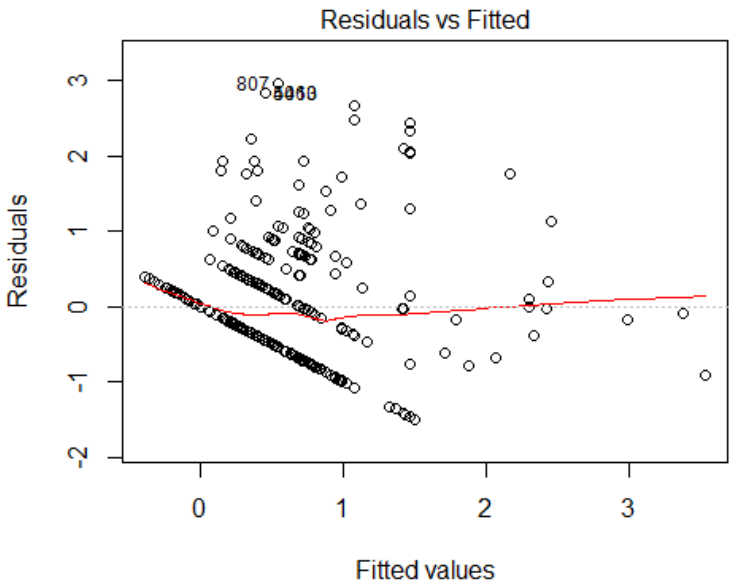

Figure 19. Sub-Segment 1 Residual versus Fitted Values Plot.

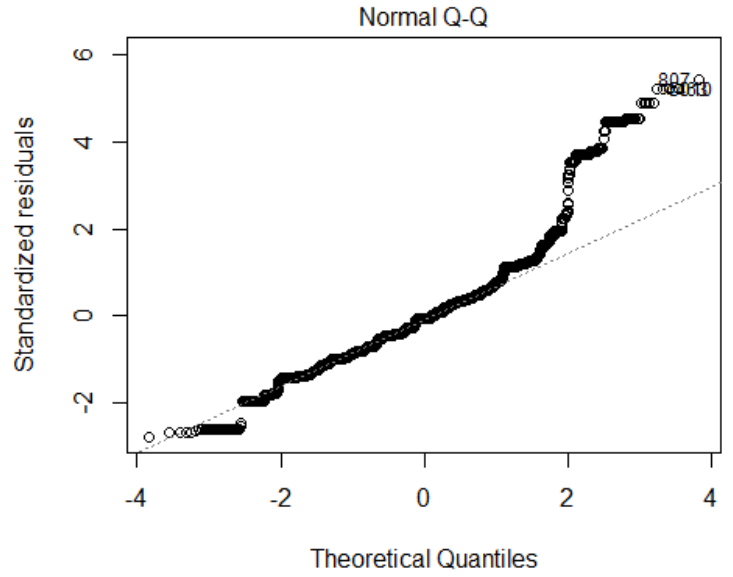

Figure 20. Sub-Segment 1 Quantile-Quantile (Q-Q) Plot.

C. REGRESSION TREE MODEL

Using the training set, we grow a full tree and prune it to the minimum cross-validated error which occurs at complexity parameter (cp) = 0.00057. This produces a

tree too large to plot. Table 24 lists the important variables in the regression tree model, and Table 26 lists the model performance metrics when applied against the test set.

Table 24. Sub-Segment 1 Regression Model Variable Importance.

	Variable Importance
Month	1173.04
Weekday	940.76
Location	465.99
Carrier	431.95
Service terms	410.05
Handling	397.10
Integrated distribution lane	382.30
Container	366.75
Weight	197.04
Shipping cost	187.38
Unit price	132.46
Supply class	48.82
Booking	39.53
Initial consolidation point	35.64
Issue priority group	28.95
origin	10.81

D. RANDOM FOREST MODEL

We remove *supply class, issue priority group, initial consolidation point, booking, origin, quarter,* and *afloat* from our model because they do not improve performance. We fit our final random forest model with 1000 trees and four random splits per tree. Table 25 lists the percent decrease in MSE from removing each variable from the model, and Table 26 shows the performance metrics of the random forest model on the test set.

Table 25. Sub-Segment 1 Random Forest Percent Increase Mean Square Error.

	%IncMSE
Weekday	300.37
Month	241.55
Carrier	34.11
Weight	73.20
Container	122.32
Shipping cost	91.40
Location	37.33
Service terms	29.57
Handling	82.73
Unit price	64.09

E. SUB-SEGMENT 1 MODEL EVALUATION

We find the random forest model provides the lowest root mean square error (RMSE) and mean absolute error (MAE). Table 26 lists the RMSE and MAE for all models.

Table 26. Sub-Segment 1 Test Set Performance Metrics Measured in Days.

	RMSE	MAE
Baseline	4.94	1.93
Linear Regression	3.69	1.08
Regression Tree	1.7	0.48
Random Forest	1.55	0.41

Our random forest model confirms the relationship between sub-segment length and weekday suggested by our linear regression model in Table 22. Linear regression results indicate a positive relationship for all weekdays except for Monday, and Figure 21 confirms this. Figure 21, which plots the feature contribution on the y-axis, highlights poor performance for shipments initiating Transporter on Thursdays in January, but also shows that the other days of the week in January perform better than most other

combinations of *month* and *weekday*. Each color in Figure 21 represents a different month.

Unlike many of our other models, *month* is not the most significant driver of variation in this random forest model. Figure 22 indicates constant performance for many months throughout the year with the exception of an increasing relationship in January and many better performing Friday requisitions in March and August. The linear regression model uses January as the base case, and Figure 22 shows requisitions beginning Transporter in January take longer to complete this sub-segment. Each color in Figure 22 represents a different weekday. This explains why all other months have a decreasing relationship with the response in the linear regression results.

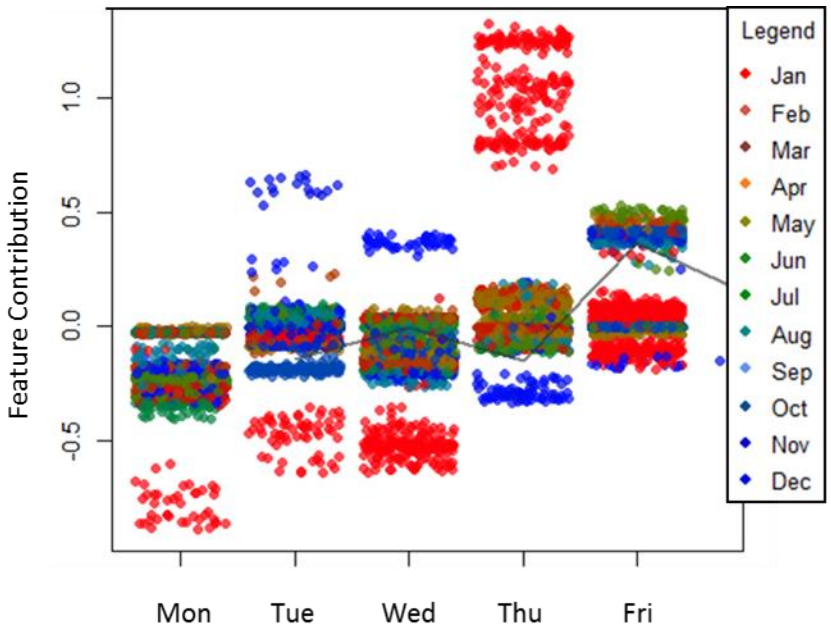

Figure 21. Sub-Segment 1 Random Forest *weekday* Feature Contribution.

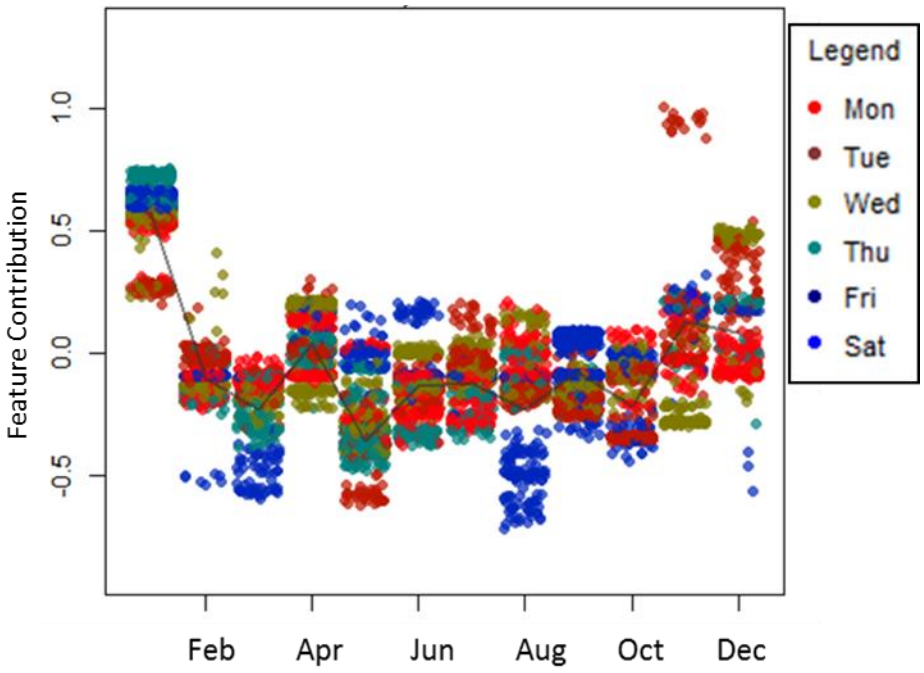

Figure 22. Sub-Segment 1 Random Forest *month* Feature Contribution.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. SUB-SEGMENT 2 MODEL AND EVALUATION

In this section, we cover the model development and analysis of sub-segment 2 using the techniques outlined in Chapter III.

A. BASELINE MODEL

Sub-segment 2 accounts for the holding time between dropping the shipment off at the seaport of embarkation (SPOE) and the beginning of the ocean transit sub-segment. This takes, on average, 6.1 days, as shown in Table 4. We use this as a baseline from which to compare our models.

B. SUB-SEGMENT 2 MULTIVARIATE LINEAR REGRESSION

We begin this model with all variables described in Chapter III and utilize the logarithmic transformation of sub-segment 2 as the response variable. Tables 27 and 28 list the results of our linear regression model.

We use a residual versus fitted values plot and a quantile-quantile (Q-Q) plot to verify model assumptions. Figure 23 indicates heteroscedasticity due in part to the discrete nature of the response, visible in the diagonal lines. Figure 24 indicates non-normal errors, both of which negatively affect the predictive capabilities of the model.

Table 27. Sub-Segment 2 Linear Regression Coefficients.

	Estimate	Std. Error	P-value
(Intercept)	1.58	0.02	< 2e-16
Tue	–0.15	0.02	0.00
Wed	–0.18	0.02	< 2e-16
Thu	–0.14	0.02	0.00
Fri	0.05	0.02	0.01
Feb	–0.06	0.04	0.09
Mar	0.40	0.03	< 2e-16
Apr	0.24	0.02	< 2e-16
May	0.33	0.02	< 2e-16
Jun	0.43	0.02	< 2e-16

	Estimate	Std. Error	P-value
Jul	0.26	0.02	< 2e-16
Aug	0.51	0.02	< 2e-16
Sep	0.15	0.02	0.00
Oct	0.27	0.03	< 2e-16
Nov	0.13	0.03	0.00
Dec	0.53	0.03	< 2e-16
Carrier: HRZD	–2.05	0.08	< 2e-16
Carrier: MAEU	–0.14	0.03	0.00
Carrier: MATS	–0.66	0.02	< 2e-16
Carrier: OTHER	-1.38	0.14	< 2e-16
Handling B: High sensitivity category I, HL*	0.66	0.04	< 2e-16
Handling G: High sensitivity category I confidential, HL*	0.47	0.03	< 2e-16
Handling N: low sensitivity category IV, OD*	0.30	0.04	< 2e-16
Handling O: Highest sensitivity category I classification secret, OD*	–0.07	0.04	0.07
Handling: Other	0.28	0.05	0.00
Handling R: No special handling, OD*	0.24	0.04	0.00
Handling W: Highest sensitivity category I classification secret, HL and OD*	0.15	0.04	0.00
Handling Z: No special handling, HL and OD*	0.24	0.02	< 2e-16

*Source: Defense Transportation Electronic Business Reference Data

Table 28. Sub-Segment 2 Linear Regression Goodness of Fit Metrics Using the Logarithmic Transformation of the Response.

Metric	Value
Residual Standard Error	0.45
R Square	0.41
Adjust R Square	0.41
Degrees of Freedom	7468

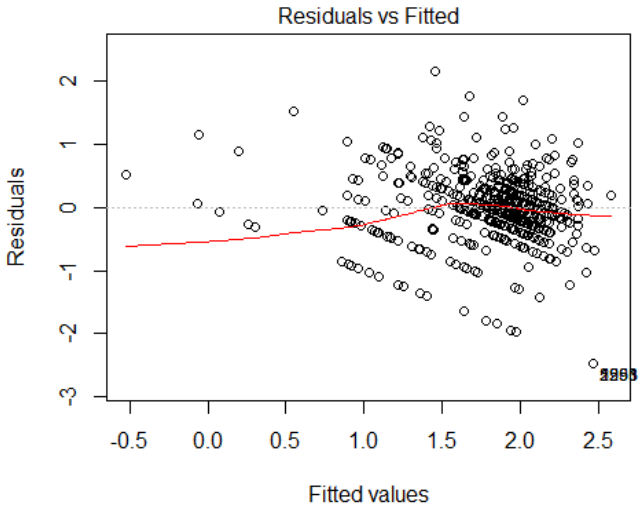

Figure 23. Sub-Segment 2 Residual versus Fitted Values Plot.

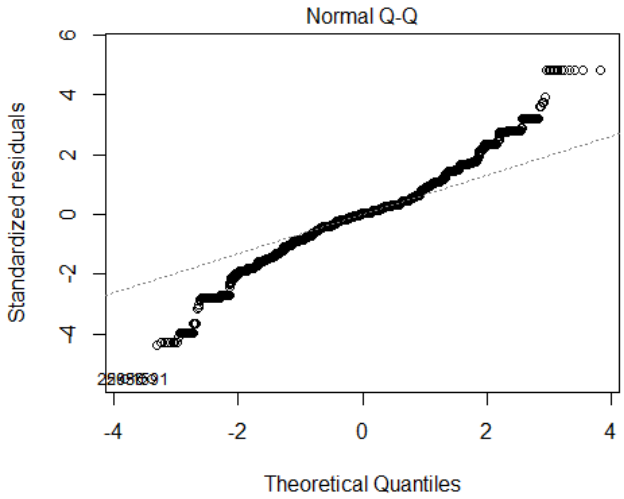

Figure 24. Sub-Segment 2 Quantile-Quantile (QQ) Plot.

C. REGRESSION TREE MODEL

Using the training set, we grow a full tree and prune it to the minimum cross-validated error which occurs at complexity parameter (cp) = 0.00011. This results in a tree too large to plot. Table 29 lists the regression variables in order of importance.

Table 29. Sub-Segment 2 Regression Tree Variable Importance.

	Importance
Carrier	755.34
Month	738.76
Service terms	717.69
Location	708.99
Weekday	621.35
Handling	620.26
Integrated distribution lane	600.99
Container	123.57
Shipping cost	77.41
Supply class	49.23
Unit price	44.55
Issue priority group	28.63
Booking method	18.59
Initial consolidation point	10.64
Origin	1.20

D. RANDOM FOREST MODEL

We remove *supply class*, *issue priority group*, *initial consolidation point*, *booking method*, *quarter*, *origin*, and *afloat* because presence does not improve the performance of the random forest model. Our final model fits 1000 regression trees with four random splits. Table 30 lists the percent increase in error that would results from removing each variable, and Table 31 lists the performance metrics when applied against the test set.

Table 30. Sub-Segment 2 Random Forest Percent Increase Mean Square Error.

	%IncMSE
Weekday	220.17
Month	285.39
Integrated distribution lane	15.32
Carrier	30.67
Weight	66.65

	%IncMSE
Container	75.70
Shipping cost	80.63
Location	25.25
Service terms	27.84
Handling	98.92

E. SUB-SEGMENT 2 MODEL EVALUATION

We find the random forest model performs better than all other models, and performs significantly better than the baseline model. Table 29 lists the root mean square error (RMSE) and mean absolute error (MAE) of each model when applied against the test set.

Table 31. Sub-Segment 2 Test Set Performance Metrics Measured in Days.

	RMSE	MAE
Baseline	3.94	2.59
Linear Regression	3.55	2.06
Regression Tree	2.01	0.85
Random Forest	1.47	0.53

The linear regression model uses January as the base case and all months, except February, have a positive relationship with the response variable. Figure 25 plots each OOB observation in a different color to represent each weekday and suggests that February has better performing shipments on Mondays and fewer poor performing *weekday* and *month* combinations relative to other months. Requisitions shipped on Fridays in July and September complete this sub-segment in less time whereas requisitions shipped on Mondays in August appear to take longer to complete this sub-segment than any other *month* and *weekday* combination. Interestingly, our regression tree finds carrier more significant than month. Figure 26 shows the relationship between carrier and month and indicates requisitions shipped by Horizon Lines, LLC (HRZD) and "other" carriers perform better than the other listed carriers, American President Lines (APLS), Maersk Line (MAEU) and Matson, Inc (MATS). MATS acquired parts of

HRZD in May 2015, and HRZD is no longer an operational ocean carrier (Horizon Lines, LLC 2014).

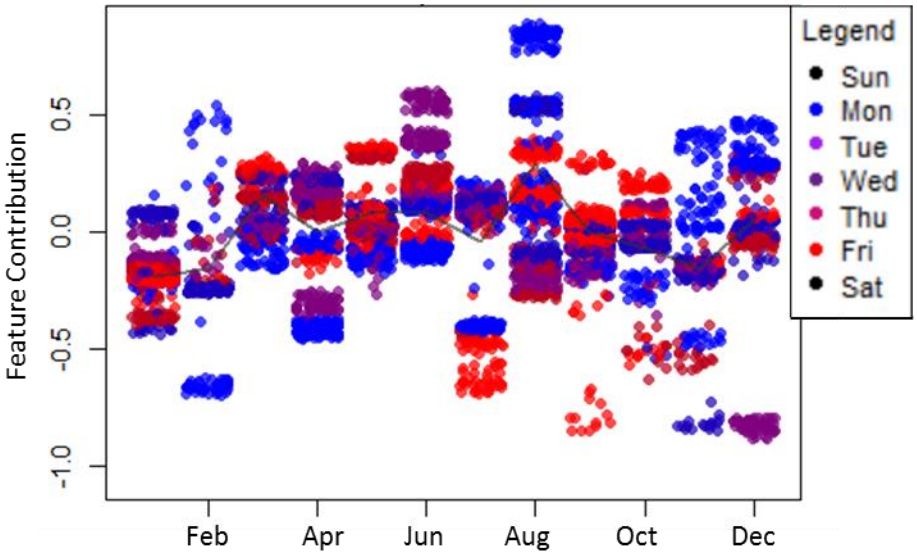

Figure 25. Sub-Segment 2 Random Forest *month* Feature Contribution.

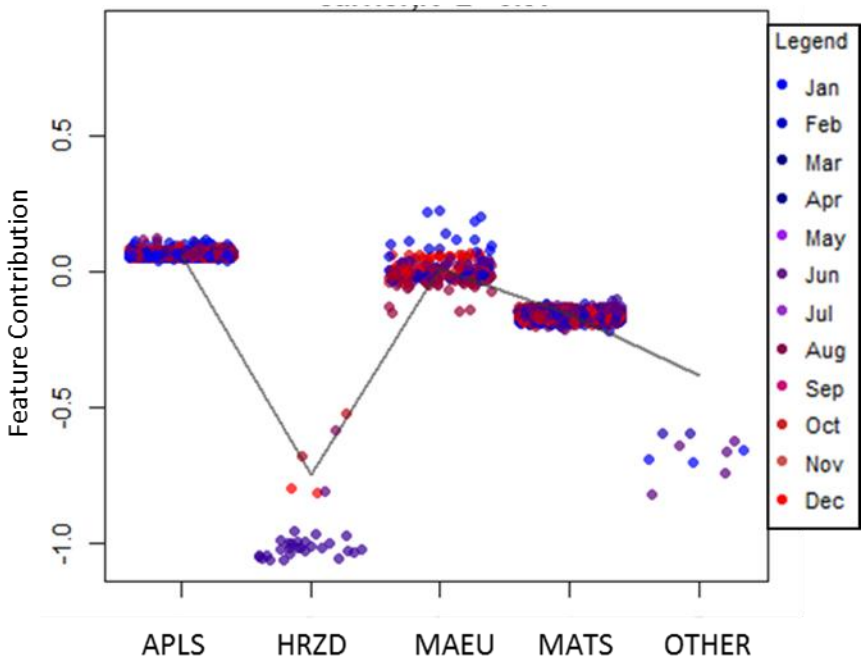

Figure 26. Sub-Segment 2 Random Forest *carrier* Feature Contribution.

APPENDIX E. SUB-SEGMENT 4 MODEL AND EVALUATION

In this section, we cover the development and analysis of the sub-segment 4 models. We employ the techniques outlined in Chapter III.

A. BASELINE MODEL

Sub-segment 4 measures the holding time at the seaport of debarkation (SPOD) between the completion of the ocean transit segment and before beginning destination line haul. We use the average, 7.8 days, as a baseline from which to compare the performance of our models.

B. MULTIVARIATE LINEAR REGRESSION MODEL

Using the predictor variables previously described, we fit a linear regression model to estimate length of sub-segment 4 and use the logarithmic transformation of sub-segment 4 as the response variable. Tables 32 and 33 list the regression coefficients and goodness of fit metrics, respectively.

We use a residual versus fitted values plot and a quantile-quantile (Q-Q) plot to verify the linear regression model assumptions described in Chapter III. Figure 27 confirms non-constant variance, and Figure 28 confirms non-normal errors, both of which negatively affect model inferences.

Table 32. Sub-Segment 4 Linear Regression Coefficients.

	Estimate	Std. Error	P-value
(Intercept)	1.15	0.06	< 2e-16
Tue	–0.13	0.02	0.00
Wed	–0.13	0.02	0.00
Thu	–0.13	0.02	0.00
Fri	–0.42	0.02	< 2e-16
Feb	0.76	0.07	< 2e-16
Mar	0.51	0.03	< 2e-16
Apr	1.28	0.03	< 2e-16
May	1.08	0.03	< 2e-16

	Estimate	Std. Error	P-value
Jun	0.52	0.03	< 2e-16
Jul	0.10	0.03	0.00
Aug	1.00	0.03	< 2e-16
Sep	0.95	0.02	< 2e-16
Oct	1.29	0.03	< 2e-16
Nov	1.53	0.04	< 2e-16
Dec	–0.03	0.04	0.48
Carrier: HRZD	–0.64	0.11	0.00
Carrier: MAEU	–0.34	0.04	0.00
Carrier: MATS	0.52	0.09	0.00
Carrier: OTHER	–1.59	0.18	< 2e-16
Location: Kaneohe	–1.09	0.09	< 2e-16
Location: Okinawa	0.12	0.03	0.00
Location: Other	–1.60	0.18	< 2e-16
Handling B: High sensitivity category I, HL*	0.16	0.08	0.04
Handling G: High sensitivity category I confidential, HL*	–0.01	0.05	0.81
Handling N: low sensitivity category IV, OD*	0.51	0.05	< 2e-16
Handling O: Highest sensitivity category I classification secret, OD*	0.14	0.07	0.05
Handling: Other	0.36	0.06	0.00
Handling R: No special handling, OD*	0.17	0.05	0.00
Handling W: Highest sensitivity category I classification secret, HL and OD*	0.07	0.06	0.24
Handling Z: No special handling, HL and OD*	0.32	0.04	0.00

*Source: Defense Transportation Electronic Business Reference Data

Table 33. Sub-Segment 4 Linear Regression Goodness of Fit Metrics Using the Logarithmic Transformation of the Response.

Metric	Value
Residual Standard Error	0.43
R Square	0.63
Adjusted R Square	0.63
Degrees of Freedom	5209

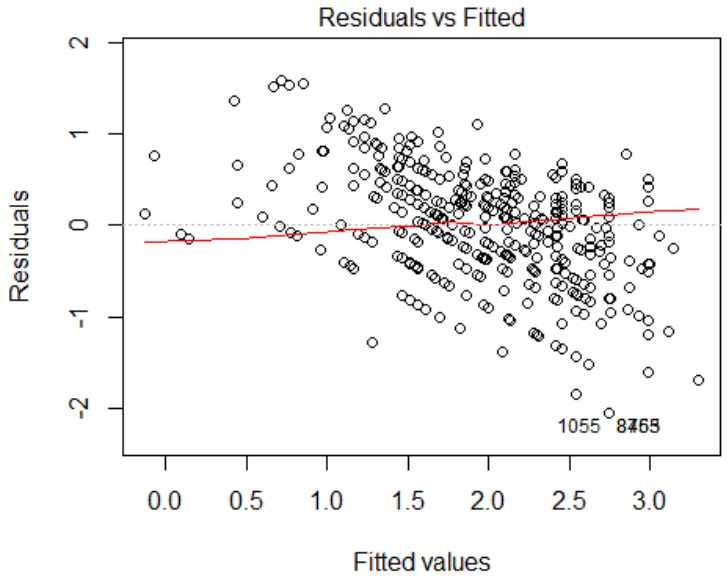

Figure 27. Sub-Segment 4 Residual versus Fitted Values Plot.

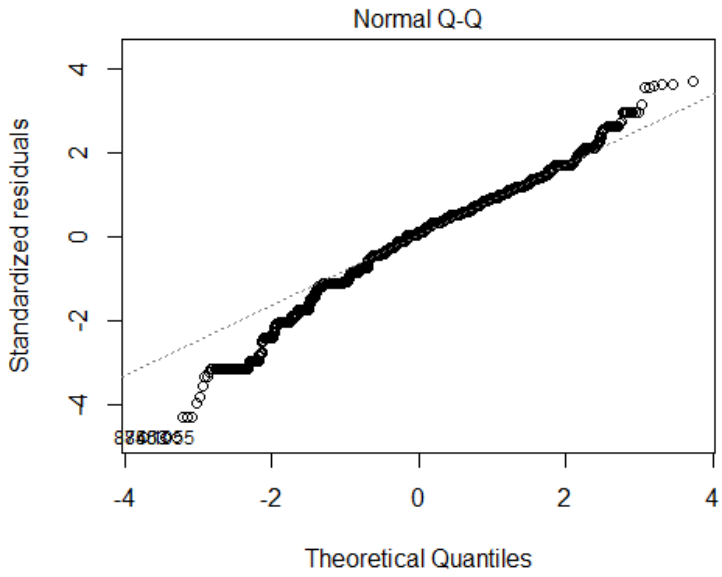

Figure 28. Sub-Segment 4 Quantile-Quantile (Q-Q) Plot.

C. REGRESSION TREE MODEL

Using our training set, we grow a full regression tree and prune it to the minimum cross-validated error which occurs at complexity parameter (cp) = 0.0002. This results in a tree with 112 splits, which is too large to plot. Table 34 lists the regression tree variable importances.

Table 34. Sub-Segment 4 Regression Tree Variable Importance.

	Importance
Month	1155.86
Handling	523.04
Location	459.38
Service terms	349.73
Integrated distribution lane	321.44
Weekday	315.19
Container	244.59
Carrier	221.63
Shipping cost	93.68
Weight	90.46
Unit price	65.81

D. RANDOM FOREST MODEL

We eliminate *supply class*, *issue priority group*, *initial consolidation point*, *booking method*, *origin*, *afloat* and *quarter* because their presence does not improve model performance. We fit a random forest model with 1000 regression trees, each with four random splits. Table 35 lists the percent increase in error resulting from removing each variable.

Table 35. Sub-Segment 4 Random Forest Percent Increase in Mean Square Error.

	%IncMSE
Month	552.82
Handling	141.99
Shipping Cost	118.67
Weight	111.55
Unit price	97.04
Location	84.68
Integrated distribution lane	25.54

E. SUB-SEGMENT 4 MODEL EVALUATION

The random forest model performs best against the test set. Table 36 lists the root mean square error (RMSE) and mean absolute error (MAE) of each model measured in days.

Table 36. Sub-Segment 4 Test Set Performance Metrics Measured in Days.

	RMSE	MAE
Baseline	4.94	1.93
Linear Regression	3.69	1.08
Regression Tree	1.7	0.48
Random Forest	1.55	0.41

Figure 29 plots each month in a different color and displays a wide range of feature contributions from month to month. Our linear regression model indicates better performance in December because of the negative relationship with the response. The model uses January as the base case for *month* and every other month has an increasing relationship with the response except for December. Our random forest model confirms this relationship. Additionally, our regression tree model indicates that removing *months* would result in a 500 percent increase in mean square error (MSE), which would raise the error from approximately one and a half days to almost eight days for this sub-segment.

Figure 30 shows various interactions between *location* and *month*. Specifically, requisitions beginning Transporter in May going to Kaneohe Bay appear to take less time than any other combination of month and location.

Kaneohe Bay performs the best in comparison to other locations in the model, and confirms the negative relationship between Kaneohe Bay requisitions and SPOD holding time we find in our linear regression results. Figure 31 shows several interactions between *location* and *handling* code and indicates better performance from Okinawa in many handling categories including 9, G, R and W. This is counterintuitive because Figure 30 shows Kaneohe Bay performs better overall, so we expect to find better Kaneohe Bay performance in one or more handling categories.

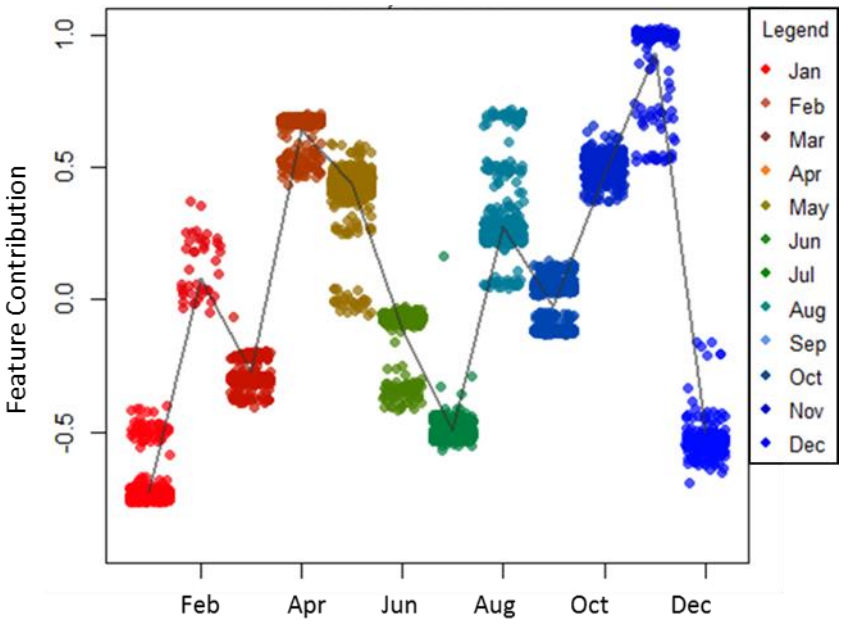

Figure 29. Sub-Segment 4 Random Forest *month* Feature Contribution.

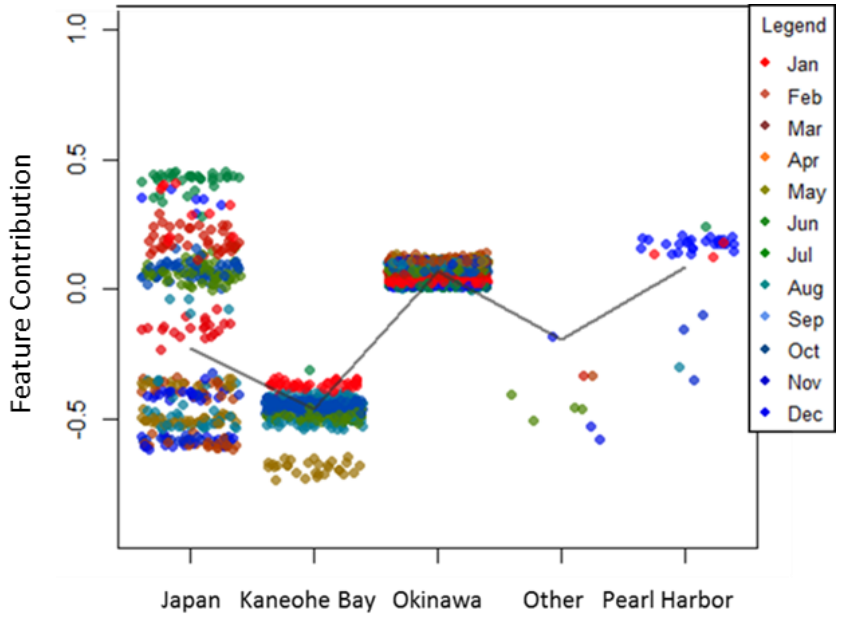

Figure 30. Sub-Segment 4 Random Forest *location* Feature Contribution.

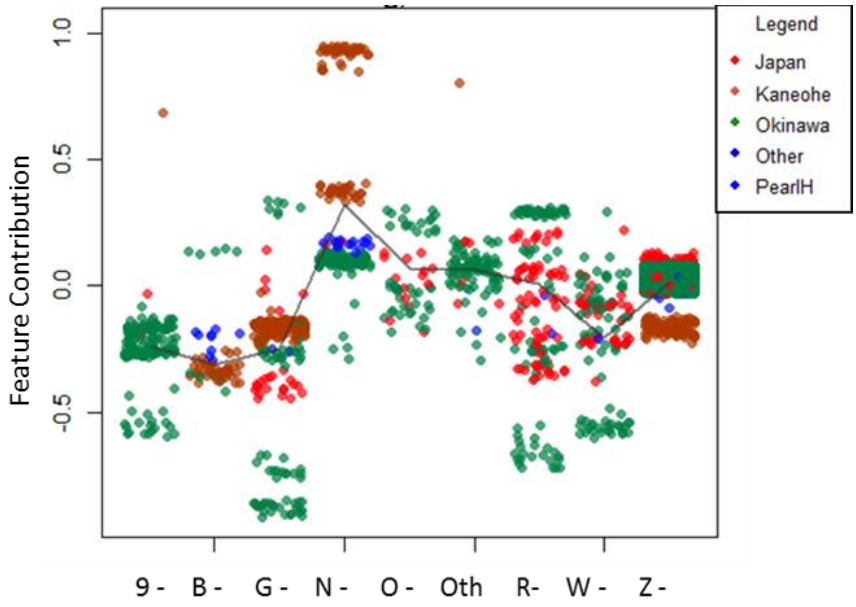

Figure 31. Sub-Segment 4 Random Forest *handling* Feature Contribution.

APPENDIX F. SUB-SEGMENT 5 MODEL AND EVALUATION

In this section, we cover the development and evaluation of the sub-segment 5 model. We use the techniques described in Chapter III.

A. BASELINE MODEL

Sub-segment 5 covers the destination line haul time, which accounts for the time until the requisition completes the Transporter segment. We use the average completion time, 0.3 days, to compare the performance of our models.

B. MULTIVARIATE LINEAR REGRESSION MODEL

We fit a linear regression model using the predictor variables described in Chapter III to determine their relationship with the sub-segment length and use the logarithmic transformation of sub-segment 5 as the response variable. Table 37 lists the regression coefficients of our model, and Table 38 lists the goodness of fit metrics.

We use a residual versus fitted values plot and a quantile-quantile (Q-Q) plot to verify the model assumptions discussed in Chapter III. Figure 32 verifies the presence of heteroscedasticity, and Figure 33 verifies the presence of non-normal errors. Heteroscedasticity and non-normal errors negatively affect model inferences.

Table 37. Sub-Segment 5 Linear Regression Model Coefficients.

	Estimate	Std. Error	P-value
(Intercept)	0.00	0.01	0.94
Carrier: HRZD	0.00	0.03	0.98
Carrier: MAEU	0.77	0.01	<2e-16
Carrier: MATS	0.59	0.01	<2e-16
Carrier: OTHER	–0.01	0.07	0.93
Handling B: High sensitivity category I, HL*	–0.49	0.03	<2e-16
Handling G: High sensitivity category I confidential, HL*	0.58	0.02	<2e-16
Handling N: low sensitivity category IV, OD*	–0.26	0.02	<2e-16

	Estimate	Std. Error	P-value
Handling O: Highest sensitivity category I classification secret, OD*	–0.01	0.03	0.84
Handling: Other	0.02	0.02	0.46
Handling R: No special handling, OD*	0.01	0.02	0.71
Handling W: Highest sensitivity category I classification secret, HL and OD*	–0.01	0.02	0.67
Handling Z: No special handling, HL and OD*	0.00	0.01	0.87

*Source: Defense Transportation Electronic Business Reference Data

Table 38. Sub-Segment 5 Linear Regression Model Goodness of Fit Metrics Using the Logarithmic Transformation of the Response.

Metric	Value
Residual Standard Error	0.17
R Square	0.78
Adjust R Square	0.78
Degrees of Freedom	4968

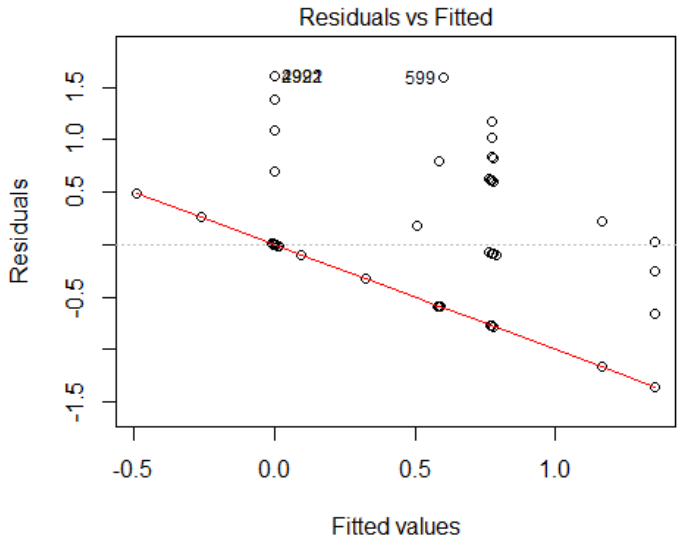

Figure 32. Sub-Segment 5 Residual versus Fitted Values Plot.

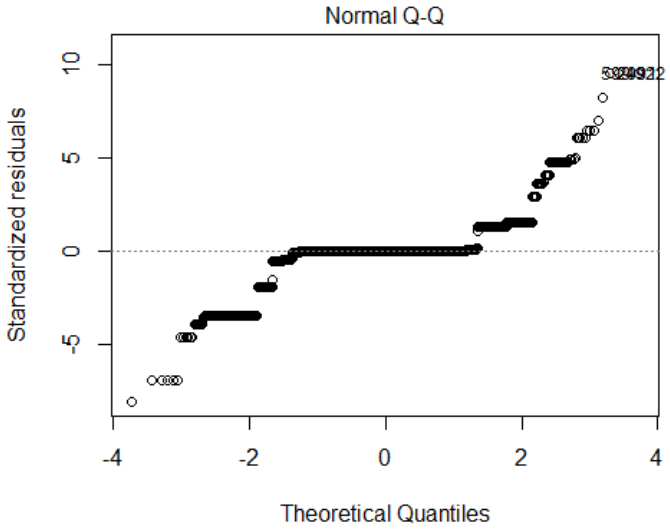

Figure 33. Sub-Segment 5 Quantile-Quantile (Q-Q) Plot.

C. REGRESSION TREE MODEL

Using our training set, we grow a full tree and prune it to the minimum cross-validated error, which occurs at complexity parameter (cp) = 0.0012. This results in a tree too large to plot. Table 39 lists the resulting variable importances.

Table 39. Sub-Segment 5 Regression Tree Variable Importance.

	Importance
Carrier	419.93
Location	375.64
Service terms	344.34
Integrated distribution lane	305.68
Handling	303.14
Month	141.52
Weekday	129.68
Container	40.96
Supply class	20.94
Weight	14.35
Booking	8.26
Shipping cost	3.05

	Importance
Unit price	2.37
Issue priority group	0.17
Origin	0.06

D. RANDOM FOREST MODEL

We remove *supply class*, *issue priority group*, *initial consolidation point*, *booking method*, *origin*, *afloat*, *weight*, *unit price*, and *shipping cost* because these variables do not improve the performance of our final model. Our random forest model fits 1000 trees each with three random splits. Table 40 lists the percent increase in mean square error (MSE) resulting from removing each variable from the model.

Table 40. Sub-Segment 5 Random Forest Percent Increase in Mean Square Error.

	%IncMSE
Weekday	33.94
Month	47.69
Integrated distribution lane	14.49
Carrier	45.75
Container	32.38
Location	34.39
Service terms	25.19
Handling	45.17

E. SUB-SEGMENT 5 MODEL EVALUATION

The random forest model has the lowest mean absolute error (MAE), but, surprisingly, the baseline model has the lowest root mean square error (RMSE). RMSE penalizes larger errors more than smaller ones, so this suggests the baseline model produces fewer large errors than the other models. The MAE weights all errors equally, and the random forest model results in the lowest MAE. Both models produce errors of

approximately one day. However, as previously discussed, RMSE provides a more pessimistic response, so we recommend the baseline model in this case. Based on the summary statistics listed in Chapter IV, this sub-segment has the lowest variation and a small difference between the mean and the median, so using the mean to predict performance presents less risk than using the mean to predict the other sub-segments. Table 41 lists the RMSE and MAE for each model when applied to our test set.

Table 41. Sub-Segment 5 Test Set Performance Metrics.

	RMSE	MAE
Baseline	0.77	0.42
Linear Regression	1.05	0.15
Regression Tree	1.01	0.06
Random Forest	1.03	0.06

Figure 34 shows Maersk Line (MAEU) and Matson Inc. (MATS) have higher destination line haul times than the other carriers, which confirms the results of our linear regression model. Additionally, Figure 35 shows all carriers take more line haul time for handling code G requisitions, which also confirms our linear regression results. This is an intuitive result as handling code G shipments are highly sensitive and require heavy lift capabilities.

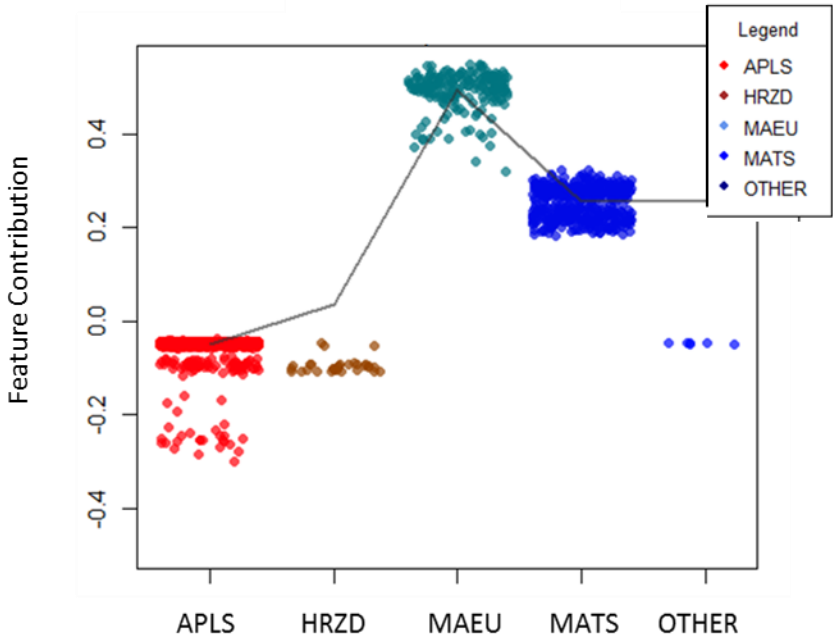

Figure 34. Sub-Segment 5 Random Forest *carrier* Feature Contribution.

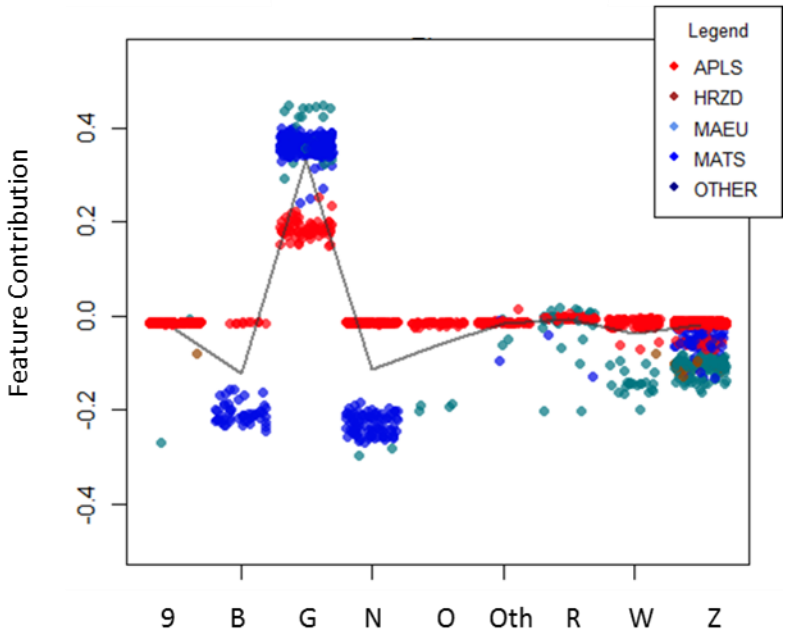

Figure 35. Sub-Segment 5 Random Forest *handling* Feature Contribution.

LIST OF REFERENCES

Assistant Secretary of Defense for Logistics and Materiel Readiness (2014) Strategy for Improving DOD Asset Visibility. Office of Assistant Secretary of Defense for Logistics and Materiel Readiness, Washington, DC.

Breiman L (2001) Random Forests. *Machine Learning.* 45:5–32.

Breiman L, Freidman JH, Olsen RA, Stone CJ (1984) *Classification and Regression Trees* (Wadsworth International Group, Belmont, CA).

Boren, P (2016) Telephone interview by author on May 25 in Monterey, Ca.

Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? — Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 7:1247-1250.

Chairman Joint Chiefs of Staff (October 16, 2013) Joint Logistics. Joint Publication 4–0. Chairman Joint Chiefs of Staff, Washington, DC.

Chui M, Lin G (2004) Collaborative supply chain using the artificial neural network approach. *Journal of Manufacturing Technology Management*. 15(8): 787–795.

Chun, M (2014) Improving inbound visibility through shipment arrival modeling. Master's thesis, Massachusetts Institute of Technology, Boston.

Commandant of the Marine Corps (2014) *USMC Marine Air Ground Task Force (MAGTF) Deployment and Distribution Policy (MDDP)* Marine Corps Order 4470.1A (Commandant of the Marine Corps, Washington, DC). Deputy Under Secretary of Defense for Acquisition Transportation and Logistics (2007) Distribution Process Owner. Department of Defense Instruction 5158.06 w/ Change 1 (Deputy Under Secretary of Defense for Acquisition Transportation and Logistics, Washington, DC).

Cutler A, Liaw A, Wiener M, Breiman L (2015) *randomForest* Breiman and Cutler's Random Forest for Classification and Regression. R package version 4.6-12 https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Defense Transportation Electronic Business Reference Data: Water-Special-Handling. Accessed March 1, 2016, http://www.ustranscom.mil/cmd/associated/dteb/files/refdata/V_WT_SP_HNDL.htm.

Deputy Under Secretary of Defense for Acquisition Transportation and Logistics (2008) Joint Deployment Process Owner. Department of Defense Instruction 5158.05 (Deputy Under Secretary of Defense for Acquisition Transportation and Logistics, Washington, DC).

Deputy Under Secretary of Defense Acquisition, Technology and Logistics (2014) DOD Supply Chain Materiel Management Procedures: Metrics and Inventory Stratification Reporting. Department of Defense Manual 4140.01-V10. Deputy Under Secretary of Defense Acquisition, Technology and Logistics, Washington, DC.

Faraway JJ (2006) *Extending the Linear Model with R Generalized Linear, Mixed Effects and Nonparametric Regression Models* (Taylor & Francis Group, Boca Raton, FL).

Faraway JJ (2015) *Linear Models with R Second Edition* (Taylor & Francis Group, Boca Raton, FL).

Faulkner WM (2014) Expeditionary logistics for the 21st Century. *Marine Corps Gazette.* 98(10):8-12.

Fredlake C, Randazzo-Matsel A (2013) Logistics Support for the Marine Corps' Distributed Laydown (Center for Naval Analysis, Washington, DC).

Government Accountability Officer (2011) DOD has made progress, but supply and distribution challenges remain in Afghanistan. GAO-12-138. (U.S. Government Accountability Office, Washington, DC).

Government Accountability Office (2015a) Improvements needed to accurately assess the performance of DOD's materiel distribution pipeline. GAO-15-226. Report, U.S. Government Accountability Office, Washington, DC.

Government Accountability Office (2015b) DOD has addressed Most Reporting Requirements and Continues to Refine Its Assets Visibility Strategy. GAO-16-88. Report, Government Accountability Office, Washington, DC.

Grömpling U (2013) Variable importance assessment in regression: linear regression versus random forest. *The American Statistician* 63(4): 308–319.

Hiltz J (2015) Time Definite Delivery: A Fully Developed and Actionable Distribution Metric. Working paper, United States Transportation Command, Scott Air Force Base.

Hiltz J (2014) Memorandum for the record effective October 1, 2014 FY15 Steering Group Approved Time Definite Delivery Standards (United States Transportation Command, Scott Air Force Base).

Horizon Lines, LLC (2014) Horizon Lines to be acquired by Matson for $0.72 per share in cash. Accessed 13 May, 2016, http://www.horizonlines.com/blog/press-releases/post/horizon-lines-to-be-acquired-by-matson-for-072-per-share-in-cash.

Inspector General (2007) Uniform Standards of Customer Wait Time. Report No. D-2007-111. (Department of Defense, Washington, DC).

James G, Witten D, Hastie T, Tibshirani, R (2013) *An Introduction to Statistical Learning with Applications in R* (Springer: New York).

Kelleher JD, Mac Namee B, D'Arcy A (2015) *Fundamentals of Machine Learning for Predictive Analytics* (MIT Press, Cambridge, MA).

Louppe G, Wehenkel L, Sutera A, Geurts P (2013) Understanding variable importances in forests of randomized trees *Advances in Neural Processing Systems*.

Mahan JM, Moon RA (2007) The Integrated Distribution Lane (IDL) Framework: Improving Joint Warfighter Sustainment *75th MORS Symposium Working Group 18 – Barchi Submission*. Military Operations Research Society, Washington, DC).

Horizon Lines, LLC. Horizon Lines to be acquired by Matson for $0.72 per share in cash. Accessed 13 May, 2016, http://www.horizonlines.com/blog/press-releases/post/horizon-lines-to-be-acquired-by-matson-for-072-per-share-in-cash.

Milborrow S (2015) *rpart.plot*: Plot rpart Models. An Enhanced Version of plot.rpart. R package version 1.5.2 http://CRAN.R-project.org/package=rpart.plot.

Nickle B (2015) Managing the Global Distribution Network. *Marine Corps Gazette*. 99 (10): 28–29.

Palczewska A, Palczewski J, Robinson RM, Neagu D (2013) Interpreting random forest classification models using a feature contribution method (arViv:1312.1121v1 [cs.LG]).

Peltz E, Girardini JJ, Robbins M, Boren P (2006) Effectively Sustaining Forces Overseas While Minimizing Costs (RAND Corporation, Santa Monica, CA).

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Robins M, Boren P, Leuschner, K (2004) The Strategic Distribution System in support of Operation Iraqi Freedom (RAND Corporation, Santa Monica, CA).

Sagara, GM (2008) Multivariate analysis of the effect of source of supply and carrier on processing and shipping times for issue priority group one requisitions. Master's thesis, Naval Postgraduate School, Monterey, CA.

Savage S (2009) *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty* (John Wiley & Sons, Inc., Hoboken, NJ).

Schaffer A, Borns N (2015) Logistics command and control. *Marine Corps Gazette*. 99 (10): DE4-DE6.

Slone, RE (2004) Leading a supply chain turnaround. *Harvard Business Review*. 82(10): 114–121.

Therneau T, Atkinson B, Ripley B (2015) *rpart*. Recursive partitioning and regression trees. R package version 4.1-10 https://cran.r-project.org/web/packages/rpart/rpart.pdf.

Ting SL, Tse YK, Ho GTS, Chung, SH, Pang G (2013) Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. *International Journal of Production Economics.* 152(2014): 200–209.

Under Secretary of Defense for Acquisition, Technology, and Logistics, Department of Defense (DOD) Time Standards for Order Processing and Delivery. Accessed January 10, 2016, http://www.acq.osd.mil/log/sci/.policy_vault.html/TDD_Standard_Script.pdf.

United States Transportation Command (2012) Distribution Process Owner Joint Deployment and Distribution Enterprise Process Flow Dashboard Supply Chain Operations Reference Model Version 7.4. United States Transportation Command, Scott Air Force Base.

United States Transportation Command (2014) FY15 Distribution Steering Group Approved Time Definite Delivery Standards. United States Transportation Command, Scott Air Force Base.

United States Transportation Command (2015) Sustainment Dashboard Executive Discussion. Unpublished Material, United States Transportation Command, Scott Air Force Base.

United States Transportation Command (2016) About USTRANSCOM. Accessed January 30, 2016, http://www.ustranscom.mil/cmd/aboutustc.cfm.

Wackerly DD, Mendenhall W, Scheaffer RL (2008) *Mathematical Statistics with Applications Seventh Edition* (Brooks/Cole, Belmont, CA).

Welling SH (2016) *forestFloor*. Visualizes Random Forests with Feature Contributions. R package version 1.9.3 https://CRAN.rproject.org/web/packages/forestFloor/forestFloor.pdf.

Welling SH, Refsgaard HHF, Brockhoff PB, Clemmensen LKH (2016) Forest floor visualizations of random forests. Cornell University Library (May 30), http://arxiv.org/abs/1605.09196v1.

Welling SH, Clemmensen LKH, Buckley ST, Hovgaard L, Brockhoff PB, Refsgaard HHF (2015) *In silico* modeling of permeation enhancement potency in Caco-2 monolayers based on molecular descriptors and random forest. *European Journal of Pharmaceutics and Biopharmaceutics*, 94, 152–159.

Wingard J, Hodge R, Stojka J (2015) Supplier relationship management. *Marine Corps Gazette*, 99(9): 35–38.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California