Delivery Estimate Accuracy: Understanding and Reducing Virtual-Physical Mismatches and Missorts in Fulfillment Centers

by

Rong (Jenny) Yao

BASc, Industrial Engineering
University of Toronto, 2016

Submitted to the MIT Sloan School of Management and
Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degrees of

Master of Business Administration

and

Master of Science in Civil and Environmental Engineering

in conjunction with the Leaders for Global Operations program
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 2024

Author .....
MIT Sloan School of Management and
Department of Civil and Environmental Engineering
May 10, 2024

Certified by .....
Y. Karen Zheng, Thesis Supervisor
Associate Professor, Operations Management

Certified by .....
John Williams, Thesis Supervisor
Professor of Civil and Environmental Engineering

Accepted by .....
Heidi Nepf
Professor of Civil and Environmental Engineering

Accepted by .....
Maura Herson
Assistant Dean, MBA Program, MIT Sloan School of Management

Delivery Estimate Accuracy: Understanding and Reducing Virtual-Physical Mismatches and Missorts in Fulfillment Centers

by

Rong (Jenny) Yao

Submitted to the MIT Sloan School of Management and
Department of Civil and Environmental Engineering
on May 10, 2024, in partial fulfillment of the
requirements for the degrees of
Master of Business Administration
and
Master of Science in Civil and Environmental Engineering

Abstract

Delivery Estimate Accuracy (DEA) is the Amazon Operations metric that measures the percentage of items that attempted delivery on or before the Promised Delivery Date (PDD). There are significant costs and customer experience impacts when packages are not delivered on time, resulting in a DEA miss. Specifically, there are two types of DEA misses that are less well-understood than others and make up a large proportion of the overall misses - Virtual-Physical Mismatch (VPM) and Missort. This project focuses on understanding and reducing the number of VPM and Missort misses in Fulfillment Centers, with the scope being Amazon's Traditional Non-Sort Fulfillment Centers in the US.

Thesis Supervisor: Y. Karen Zheng
Title: Associate Professor, Operations Management

Thesis Supervisor: John Williams
Title: Professor of Civil and Environmental Engineering

Acknowledgments

First, I want to thank my supervisor Ryan Morrison for the guidance and wisdom he shared with me throughout the project to make this a success. I also appreciate Ben Anderson, Ben Whidden, and Travis Dorion for their invaluable support.

Or Dan, your mentorship has been a guiding light for me throughout this process. I also couldn't have done it without the generous support from Brock Jensen, Neil McCoy, Brian Wensel, Alex Liu, and Kosay Abu-Ghaban. A big thank you to Chuck Cummings, Carolyn Cade, Nora Martin, and the entire LGO internship committee for making this experience possible. Shout out to the outbound team - Cameron Winn, Denney Domnik, Dylan Downie, Justin Barker, Kerrie McIntyre, Steve Smith, and Luke Borron - for their valuable assistance throughout this journey. I'm also grateful to have worked with Courtney Salindong, Taylor Wisnasky, and Drew Williams - our collaboration has been instrumental in the success of this project.

I also want to express my heartfelt gratitude to Professor Karen Zheng for all the insights and thoughtful feedback she shared with me throughout the year, and to Professor John Williams for all his help and encouragement.

Finally, I want to thank my dear friends and family for their unwavering belief in me as always. No more "I can't, I have to write my thesis", let's go enjoy this beautiful life and this wonderful world.

Note on Proprietary Information

In the interest of protecting Amazon's competitive and proprietary information, tables, figures, and processes presented throughout this thesis are used solely for the purpose of illustration. The data presented throughout this thesis has been modified and does not represent actual values. Data labels have been altered, converted, and/or removed in order to protect competitive information, while still conveying the findings of this project.

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

Acronyms

AM Area Managers. 51

CPT Critical Pull Time. 53

DCS Dynamic Container Stacking. 60

DEA Delivery Estimate Accuracy. 15, 51, 63

DS Delivery Station. 17, 19, 63

FC Fulfillment Center. 15, 19, 63

OM Operations Managers. 51

PDD Promised Delivery Date. 15

RCA Root Cause Attribution. 6, 33, 64

RFID Radio Frequency Identification. 27, 61

SC Sort Center. 17, 19, 63

SLAM Scan, Label, Apply, and Manifest. 15, 20

SME Subject Matter Experts. 28, 34, 42, 51

USNS U.S. Traditional Non-Sortable. 16, 64

VPM Virtual Physical Mismatch. 16, 25, 63

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

1.1 Background on Amazon

Amazon.com, Inc. is an American multinational company based in Seattle, WA, founded in 1994. As one of the world's largest companies by market capitalization, it has business operations in e-commerce, cloud computing, online advertising, digital streaming, and AI. Amazon's e-commerce platform is country-specific and offers a wide range of products including apparel, beauty products, household products, kitchen items, jewelry, musical instruments, sporting goods, automotive parts, and toys. In the U.S., delivery of orders made on the online e-commerce platform is supported by Amazon's fulfillment process in partnership with third-party organizations.

1.2 Problem and Objectives

An internal metric at Amazon - Delivery Estimate Accuracy (DEA) - measures the percentage of items that attempted delivery on or before the Promised Delivery Date (PDD), the date on which Amazon as a platform has "promised" to deliver the item(s) by. As the first stage of a sophisticated fulfillment process, the Fulfillment Center (FC) has significant influence over the success of delivery because an FC DEA miss can cause downstream recovery costs, concessions costs, and opportunity costs as a result of unsatisfactory customer experience.

Post-Scan, Label, Apply, and Manifest (SLAM) FC DEA is further broken down into

subcategories, one of which is Package Flow Exceptions, which is composed of Virtual Physical Mismatch (VPM) and missort. VPM is a measure of how accurately the physical movements of packages match their virtual movements, and missort occurs when packages end up going on a completely different route from the original route intended.

This project focuses on First Mile U.S. Traditional Non-Sortable (USNS) Fulfillment Centers, where a significant proportion of post-SLAM DEA misses were caused by VPMs and missorts in 2022, resulting in millions of DEA misses. In comparison to other types of DEA misses, VPM and missort misses are harder to detect and correct in real-time since they typically become apparent one to three days after trailer departure, when feedback is received from downstream processes. The purpose of this project is to develop a deeper understanding of VPM and missort at the FC level, identify root causes, and study potential solutions to reduce these misses.

1.3 Project Approach

A combination of qualitative and quantitative methods were used to provide a holistic perspective of the issue. Qualitative methods, such as expert interviews and observing front-line associates, were used to gain a deep understanding of the process from the executors' perspectives. Meanwhile, quantitative methods, including data collection and analysis, were used to provide statistical insights based on system records.

The analysis was based on data and processes at SLC2, a USNS site in West Jordan, UT that opened on August 9, 2020. Timestamped data of packages as they moved across different steps of the outbound process are recorded by automatic or manual scanning. This data was searched and obtained from the internal database. Data from pilots that were run on the floor were mostly collected manually via counting and timing.

Data modeling and analysis were done in R and via QuickSight - an internal data manipulation and visualization software. In R, multiple linear regression models were built, with extracted .csv files as input. In QuickSight, data tables were pulled via SQL and visualized in a dashboard.

1.4 Thesis Overview

The content of this thesis is based on results from a six-month research internship conducted from February to August 2023. The primary site studied was SLC2, an FC based near Salt Lake City, UT. Insights from additional sites of different types (e.g. Sort Center (SC), Delivery Station (DS)) and from different locations (e.g. Seattle, Los Angeles, Chicago) were also included.

This thesis is broken down into six chapters: Chapter 1 introduces the program at a high level by providing some background information on the company, the problem being studied, the overall objective, and the approach taken. Chapter 2 provides more details on the fulfillment process, how it is conducted at SLC2, the quality metrics involved, and the literature review conducted. Chapter 3 focuses on results from exploratory analysis, including noticeable trends, breakdowns, ratios, and correlations. Chapter 4 lays out the seven root causes that were discovered, including description and supporting evidence for each. Given the root causes from the previous chapter, Chapter 5 then explains the seven solutions that were explored, studied, or implemented. Finally, Chapter 6 summarizes the key findings from this research project and suggests further areas for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

Background

2.1 Amazon's Fulfillment Process

At a high level, Amazon's US operation is divided into three main stages: First Mile, Middle Mile, and Last Mile.

First Mile is where the process starts: when a customer places an order online, details of the order are received in a warehouse known as the Fulfillment Center (FC). The FC is responsible for intaking and maintaining inventory, picking the item(s) from inventory, packing accordingly, and shipping the packed order to the next leg of the journey.

In the Middle Mile, each package sent by the FC is received by a Sort Center (SC) for sorting into the appropriate routes. Once sorted, these packages are loaded and shipped again to the Last Mile.

The Last Mile of the operation is responsible for delivering packages to the doorsteps of the end customer. This component is characterized by its close proximity to the final delivery destination. Packages sent by SCs in the Middle Mile are received by Delivery Station (DS)s in the Last Mile. Here, packages are received, sorted, and loaded one last time before being sent out on delivery trucks for daily delivery.

As this entire fulfillment process is sequential, errors made earlier in the process will impact downstream components. As such, the focus of this project is on First Mile, more specifically, on the operational processes within FCs. FC sites at Amazon vary by type and generation. The type of site can be influenced by factors such as the technology used

(e.g. robotic sorting vs manual sorting) and the size of items processed (e.g. small items vs furniture). Within the same site type, each FC will also have a generation number - newer generations are typically built later and have incorporated improvements from previous generations. SLC2 - the site studied in more detail in this project - is a manual sort site for medium-sized items.

2.2 Overview of SLC2's Outbound Process

At SLC2, operations and teams are split into Inbound and Outbound. Inbound operations focus on the intake and storage of inventory while Outbound operations focus on moving items out of the building once an order is received. The Outbound Process has five main stages, which are:

1. Pick
2. Pack
3. Scan, Label, Apply, and Manifest (SLAM)
4. Ship
5. Dock

Pick is the first step in the process. When an order is received, an associate ("picker") walks or drives along the shelves within the FC to locate the relevant items and place them in a cart or cage. At this point, the item's barcode is scanned by the picker to indicate that it has left the storage area and moved onto Pack.

At Pack, items are taken out of the cart or cage and placed into Amazon boxes. Multiple items from the same order may be packaged into a single box. Both the items and the box are scanned so the system can track placement.

Once packed, boxes are placed onto conveyor belts to be labeled through a process known as SLAM. Being SLAMed means receiving a label that includes the delivery address as well as instructions for routing - note that customer info is not disclosed before this step. Following SLAM, boxes move onto Ship where the conveyor belt brings the package to a sorter.

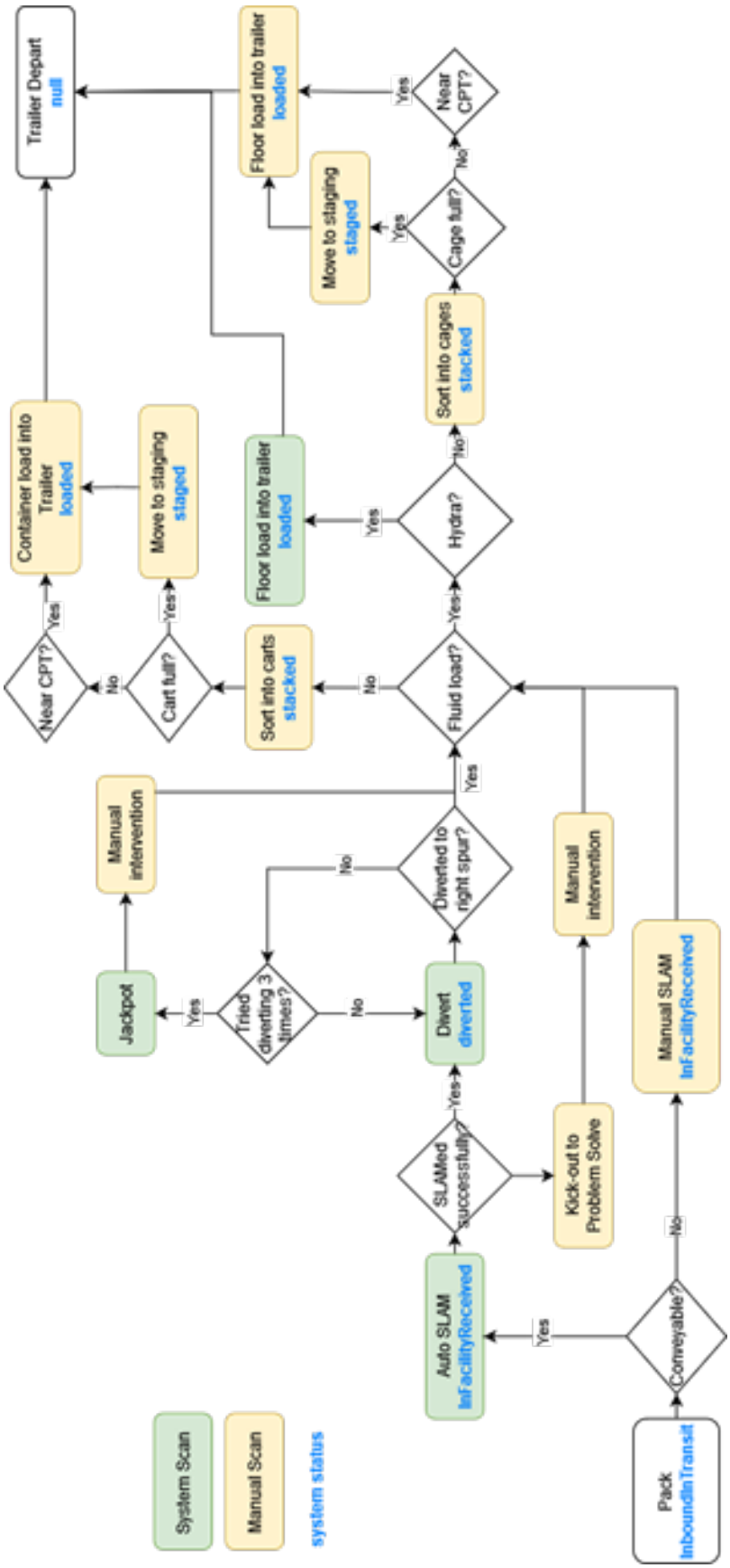


Figure 2-1: Flowchart of package process path at SLC2 dock

The dock at SLC2 performs one of the most complex processes within the fulfillment center (Figure 2-1). On average, 85% of the packages were processed through conveyor belts, eventually diverting to one of 18 lanes. After the implementation of an auto-SLAM and re-circulation project on May 9, 2023, a package that did not divert the first time will enter re-circulation and get two more tries before eventually going to the jackpot. The remaining 15% of the packages were deemed Not-Conveyable (Non-Con) due to their dimensions and/or weight limits and therefore were manually processed. Table 2.1 outlines the ways a package can be processed.

Type of Load	**Container**	**Non-Con?**	**Process Path**
Fluid	Trailer	No	SLAM -> Divert -> Load -> Depart
Fluid	Trailer	Yes	SLAM -> Stack -> Load -> Depart
Fluid	Cage -> Trailer	No	SLAM -> Divert -> Stack -> Stage -> Load (package) -> Depart
Fluid	Cage -> Trailer	Yes	SLAM -> Stack -> Stage -> Load (package) -> Depart
Containerized	Cart -> Trailer	No	SLAM -> Divert -> Stack -> Stage -> Load (cart) -> Depart
Containerized	Cart -> Trailer	Yes	SLAM -> Stack -> Stage -> Load (cart) -> Depart

Table 2.1: Process path for different packages

As mentioned in Table 2.1, there are two types of loads: fluid and containerized. Packages are considered fluid-loaded if they are eventually moved into the trailer as individual boxes and get stacked with the other boxes, as shown in Figure 2-2. There are two ways this can happen: packages can either be routed to the trailer directly using conveyor belts, or they can be routed first to a cage and then the entire cage will be offloaded into the trailer. Cages are reusable carts (shown in Figure 2-3) that are used to hold and move packages temporarily as they are being sorted. Note that cages that have been emptied are returned to the loading area and do not leave the fulfillment center.

In contact, containerized loads are packages that are loaded into trailers and shipped out in carts. Instead of being stacked with the other packages in a trailer, packages are first sorted into carts based on their next destinations, and then entire carts are loaded into the

Figure 2-2: Packages fluid-loaded into a trailer. Image source: supplychainbrain.com/-articles/29664-robots-edge-closer-to-unloading-trucks-in-amazon-era-milestone

Figure 2-3: Example of a cage used in warehouses for temporary storage. Image source: nationalcart.com/products/aluminum-deck-order-pick-carts/

trailer. This is usually done for ease of sorting in downstream processes when the entire cart can be loaded onto the next trailer. Carts (as shown in 2-4) do leave the fulfillment center and are a shared resource across First Mile, Middle Mile, and Last Mile.

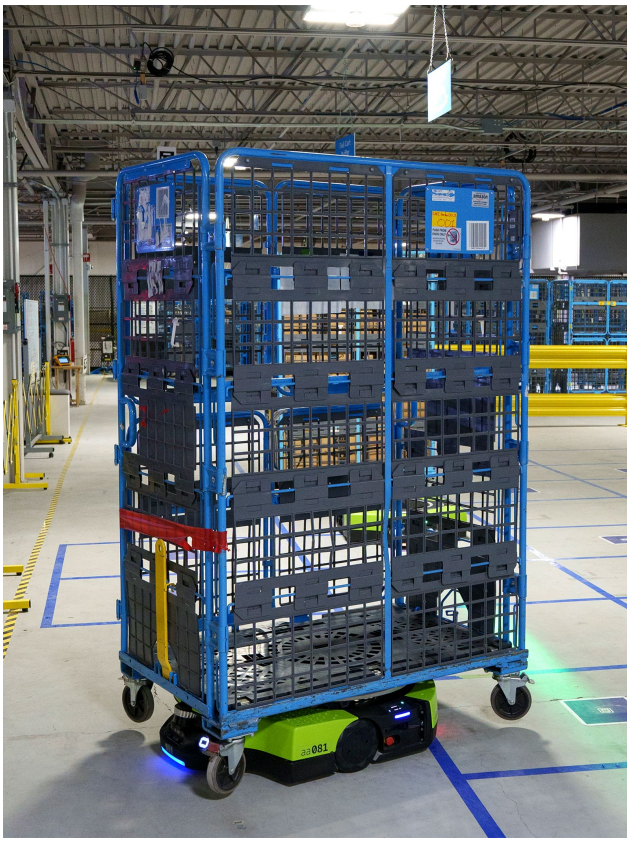


Figure 2-4: Example of a cart (in blue) used across First, Middle, and Last Mile. Image source: robotsguide.com/robots/proteus

The outbound process path starts with SLAM for all packages. Divert means a package was scanned and the system transported it to the next destination automatically using the conveyor belt. The next destination could be a trailer, a cage, or a cart. Stack means a package is put into a container, which could again be a trailer, a cage, or a cart. Stage means the cage or cart is loaded and moved into the staging area, ready to be loaded into a trailer. Load refers to the loading of the trailer with either individual packages or entire cages. Finally, once loading is complete, the trailer departs.

2.3 Delivery Estimate Accuracy (DEA)

2.3.1 Virtual-Physical Mismatch (VPM)

Virtual Physical Mismatch (VPM) is a measure of how accurate we are at loading packages physically to match their virtual movement. A VPM miss occurs when there is a mismatch between the virtual scan received by the package and its actual location. For example, a package could be scanned to a container but not physically placed into the container, thus creating a mismatch. On the other hand, a package could also be placed into a container but did not receive a scan to inform the system that is it there. In the first case, the container may eventually depart assuming the package was within, and discover the discrepancy as the trailer gets unloaded at the next facility. In the second case, if the package ends up in the right facility and gets scanned there, it will continue on its originally planned path. VPMs can happen at the container level as well. An entire cart of packages can become VPMs if the cart's scans do not match its physical movements. Therefore, VPMs usually occur when a package or a container gets scanned and moved from one location to another, which happens during divert, stack, stage, and load.

A large VPM miss can cause package delays in reaching the next step in the process. It also causes a snowball effect and poor customer experience. The downstream transportation team relies on accurate VPM information to process the volume in time and plan labor accordingly - without proper accuracy, they can't plan accordingly and might run out of work for certain periods of time while being overrun at other times.

2.3.2 Missort

A missort occurs when a package ends up going on a completely different path and therefore reaches a different destination than what was intended. This could occur at any stage in the process where the package is being sorted, including when the package is diverted, stacked, staged, or moved into a trailer. For example, if two containers going to facilities A and B are placed right next to each other and a package intended for facility A was mistakenly placed in the container for facility B, then this package becomes a missort. This could occur if the

system made a wrong diversion or an associate manually moved the package to the wrong location. In the case of a manual move, hand scanners usually play a warning sound effect when an associate tries to scan a package into the incorrect container. At this point, the associate could choose to correct this action by scanning the package to the correct container or ignoring the warning altogether. Similar to VPMs, missorts can occur at both the package and the container level. Again, entire carts full of packages can be moved into the wrong trailer, causing a container-level missort.

2.4 Literature Review

In order to gain a comprehensive understanding of the topic, literature review was conducted on sortation, quality control, and error reduction in supply chains. Several previous research theses and industry-wide publications were examined. Relevant and useful findings are outlined in this section, from the perspectives of Process, Technology, and People.

2.4.1 Process

In a 2006 study done at Amazon's Fulfillment Centers, the company's two most commonly used operational improvement frameworks - DMAIC and DMADV - were described. "DMAIC stands for Define, Measure, Analyze, Implement, and Control. This framework is one that is often used for improving a process that is already in place" [5] while "DMADV stands for Define, Measure, Analyze, Develop, and Verify. The purpose of this framework is to introduce new processes or new equipment to solve a problem in the fulfillment center" [5]. Both frameworks were considered when implementing changes and pilots later in the project.

Separately, an analysis of the outbound process defects at an Amazon Fulfillment Center in 2016 identified a few challenges with assessing missort defects at the time. Specifically, "the sortation process at ONT2 has several different performance metrics, including the concessions report, a missort audit...and a sort center missort report but these metrics may not be accurate representations of the defects because they could be lagging indicators, they are generally lacking in clarity, and/or defects may not be allocated to the correct reason code by the customer or the customer service representative" [3]. Since then, data availability

has improved to support more accurate measurements of defects. Nonetheless, the history of the previous approaches and methods was a helpful starting point for further research on this topic.

2.4.2 Technology

Much of the discussion on technology-enabled solutions to minimize sortation defects was found to be surrounding the use of Radio Frequency Identification (RFID) technology. As mentioned in a 2006 study, "the consumer retail industry has traditionally relied on the scanning of bar codes to keep track of inventory levels, consumption patterns, and the movement of products" [5] but "RFID technology offers a new, more precise method of tracking inventory without need for line-of-sight scanning that bar codes require" [5]. It is important to note that "one of the main reasons that RFID has not been used in mainstream consumer retail products is because of its traditionally high costs" [5]. It was interesting to see that the use of RFID technology in Fulfillment Centers was proposed almost a decade ago, and yet, industry-wide adoption has still not occurred. Fortunately, the cost of adopting RFID has been decreasing over the past 18 years, thus this option can be re-evaluated in this research. This re-evaluation is done with benefits beyond reducing VPMs and missorts in mind - "the benefits of RFID in this system are not only the decreased time to locate an item but also the increased accuracy of the information [5].

2.4.3 People

Finally, research was done on the impact that staff have on overall quality.

In the previously mentioned 2016 study at an Amazon Fulfillment Center, human errors were identified as an inevitable factor during long shifts - "although manual errors such as package misplacement may seem trivial and unlikely, associates work ten-hour shifts, with just an hour break total during that time. Fatigue, the repetitive nature of the work, and frequent process and layout changes can easily lead to mistakes. The individual root causes of the human errors that result in missorts were difficult to quantify for this project, so the focus was primarily on improving the processes that might lead to missorts" [3]. Furthermore,

"the Outbound Dock Subject Matter Experts (SME) hypothesized that the reason for the dock sortation defect rate was primarily due to human error with palletized packages" [3]. Expanding beyond the workers, interactions with the environment and management were also identified in the root-cause analysis: "for dock packages, the primary contributors, based on data and observations, to missorts appear to be the people and the environment, with some contribution from management's inability to effectively measure associate performance" [3].

Supporting these observations, a very recent study quantified this relationship as "defects impact was calculated using error reduction factors assigned according to the level of human interaction in the respective solution" [1]. A possible solution was mentioned in a 2015 study, stating that "it is important to create an incentive system that supports productivity and quality improvements...this could include establishing a direct link between Associates' compensation and their team's productivity and work quality [2].

Chapter 3

Exploratory Analysis

3.1 Descriptive Statistics

Before jumping into a more detailed analysis, some basic statistics were gathered to understand current state processes. To do so, operational data at SLC2 over five months spanning spring to summer 2023 was extracted and prepared.

Multiple millions of packages were processed and shipped through SLC2 during this time. Of these, less than 2% were DEA misses. The proportion of errors that together made up these DEA misses varied greatly from week to week: Package Flow Exception, which includes VPM and missort misses, made up anywhere between 15% to 80% of all DEA misses in a given week as seen in Figure 3-1.

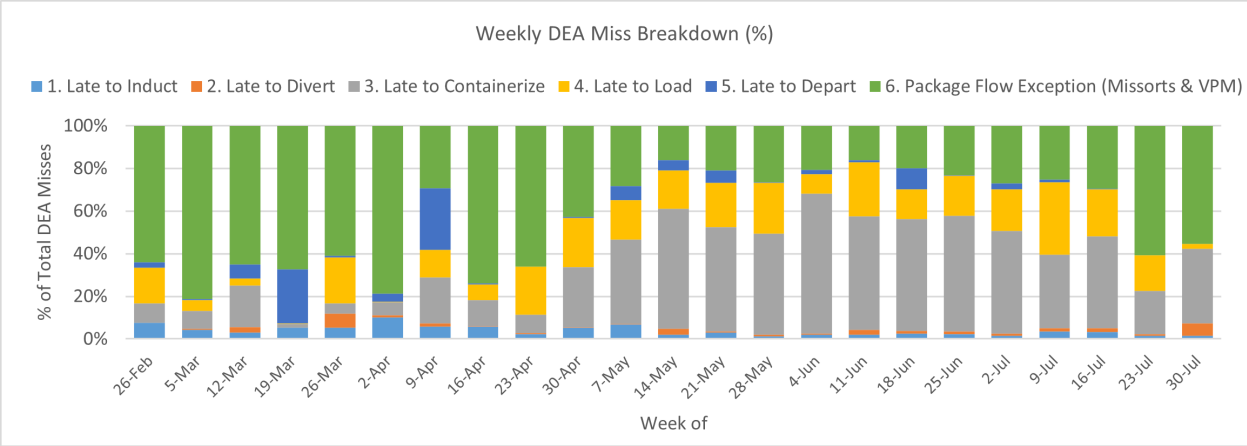


Figure 3-1: Weekly DEA misses - Package Flow Exception in green

The average daily VPM rate was much higher and more volatile than the average daily missort rate (Figure 3-2).

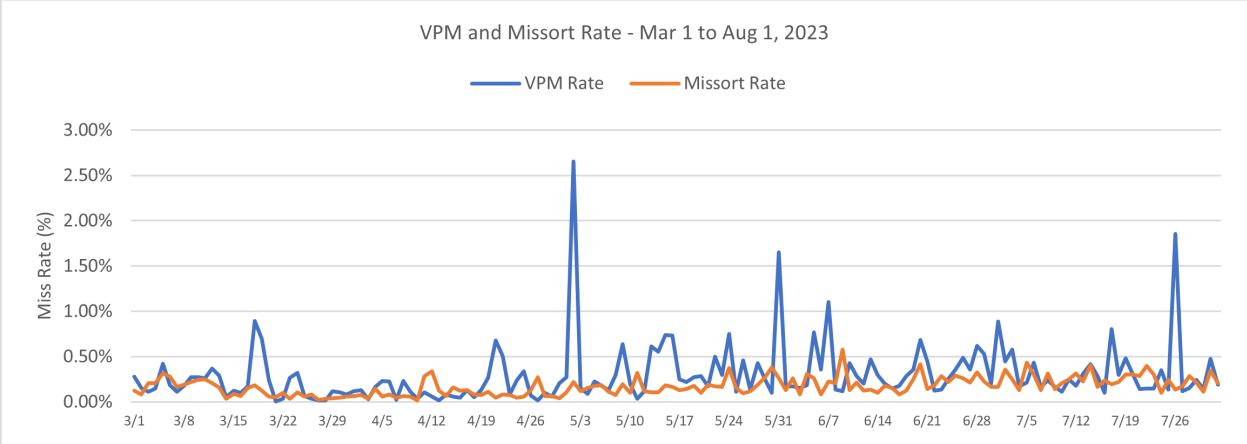


Figure 3-2: Daily VPM and missort rates

Furthermore, fluid loads, which are packages going directly from SLAM to the trailer, had a slightly lower VPM rate and missort rate than Containerized Loads, which are packages that were stacked into a container at some point after SLAM and before loading into the trailer, as outlined in Table 3.1. Interestingly, VPM and missort misses make up a smaller portion of total DEA misses for containerized loads than for fluid loads.

Type of Load	**DEA Miss Rate**	**VPM Rate**	**Missort Rate**
Containerized	2.58%	0.33%	0.19%
Fluid Loaded	0.64%	0.29%	0.17%

Table 3.1: Comparison of miss rates between load types

The DEA miss rate, VPM rate, and missort rate were also examined by ship option, type of box, and stacking method. In addition, the failure rate at each step - by individual SLAM station, sorter, destination, diverter, chute, stacking area, and stacking associate was also assessed in an attempt to find patterns. These breakdowns and miss rates were specific to the configuration of SLC2 from March 14 to April 14. This configuration has since changed dramatically due to major layout and process changes associated with two operational projects that took place after this analysis.

3.2 Correlations

A few simple correlations were tested to find patterns. Of the different factors tested, the daily volume (average number of packages processed) was found to have a moderate to strong correlation with the daily misses, as seen in Figure 3-3. For this analysis, additional data from other FCs were included.

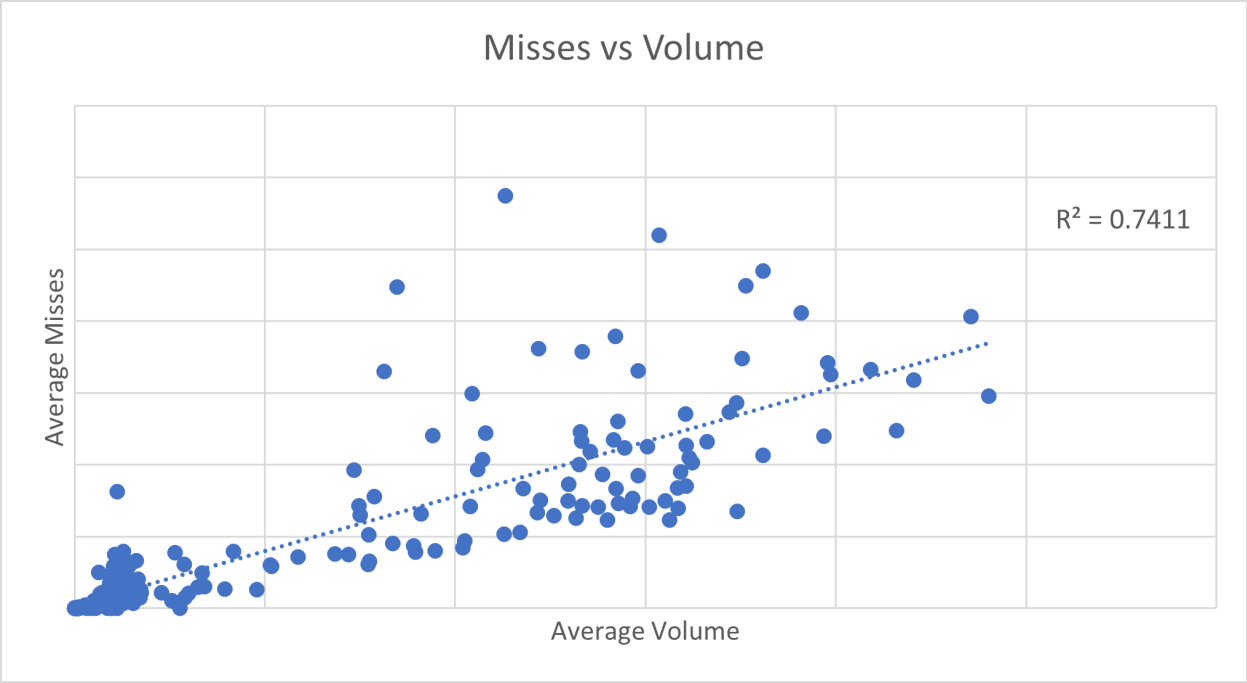
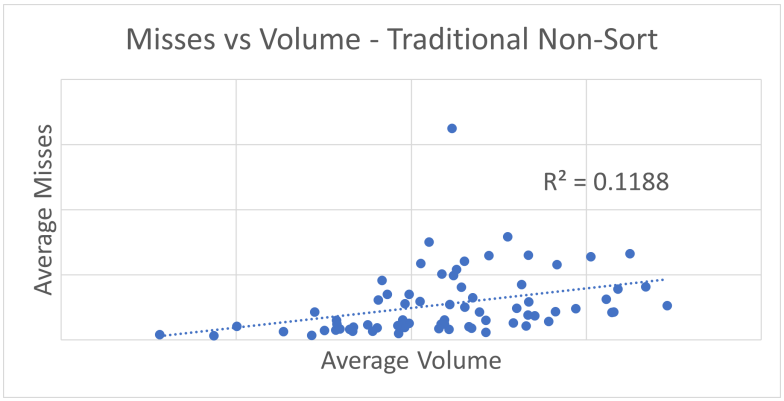
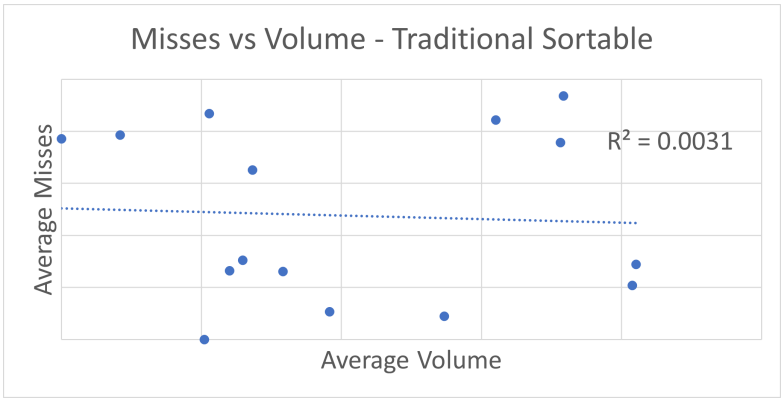


Figure 3-3: Daily average volume vs DEA misses across 268 FCs (axis values removed for confidentiality)

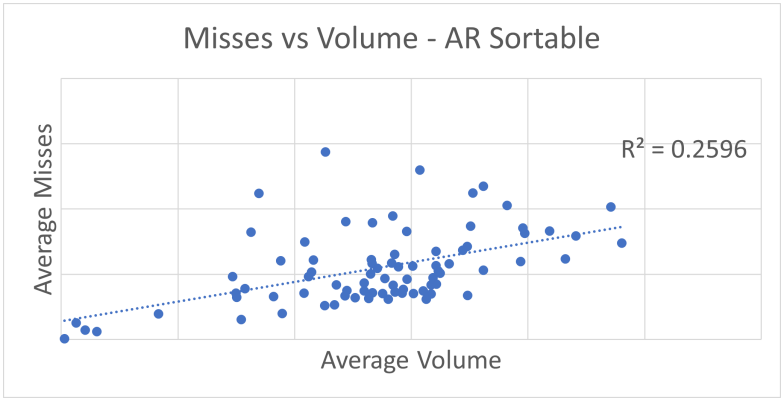
The fact that more packages processed meant more potential for misses may not be a surprising discovery. However, a closer look at the relationship by site type (Figure 3-4) reveals that this correlation is the strongest for AR Sortable sites, followed by low-to-medium correlation for Traditional Non-Sortable and AR Non-Sortable sites, and has little correlation with Traditional Sortable sites. One of the potential explanations for this could be due to the number of data points available for each site type - there are only a few sites in the Traditional Sortable and AR Non-Sortable categories. Although a clear conclusion cannot be drawn from these results, the differentiation observed in these graphs is useful to keep in mind in further research - factors that contribute to DEA misses could differ across site types, generations, layout, staffing, and other aspects of operations.



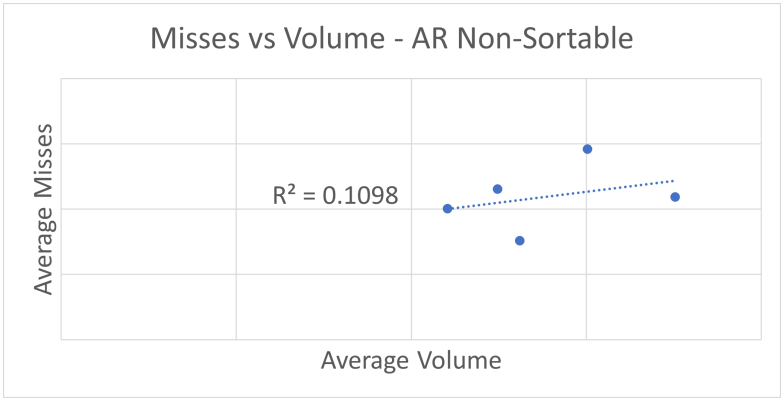
(a) Traditional Non-Sortable



(b) Traditional Sortable



(c) AR Sortable



(d) AR Non-Sortable

Figure 3-4: Misses vs Volume for different types of FCs

Chapter 4

Root Cause Discovery

4.1 Root Cause Overview

In order to solve issues associated with VPMs and missorts, an understanding of what causes each miss, as well as which factors contribute to misses, is necessary. In this research project, this was done by leveraging an internal framework to categorize types of misses, interviewing key stakeholders to collect potential factors, and then building a multiple linear regression model for each type of miss to understand which factors are important. This chapter will outline the Root Cause Attribution (RCA) methodology and provide details on each root cause of interest.

4.1.1 The Root Cause Attribution (RCA) Methodology

A methodology was developed by the Amazon Sort Tech team to attribute FC Post-SLAM DEA misses to root causes and suggest actions accordingly. The categorization of causes is determined by a series of logic written in SQL coupled with operational data from sensors and scans. While this method may not capture real-life cases perfectly, it provides a starting point for teams to track and conduct deep dives.

The RCA categorization information is available within an internal dashboard. Causes are broken into three levels: Level 0, Level 1, and Level 2. Both VPM and missort misses are categorized under the same Level 0 called Package Flow Exception. Level 1 differentiates

between package-level misses and container-level misses. Level 2 is the most detailed level, which has more details regarding what the cause is.

Leveraging this logic, an analysis of the top VPM and missort causes at SLC2 was done for the March 1 to August 1, 2023 period. The top three VPM causes (covering over 90% of total VPM) and top four missort causes (covering over 90% of total missort) were identified (Table 4.1) and became the focus of the project.

Cause	**Type**	**Level 1**	**Level 2**
A	VPM	Package	Other VPM
B	VPM	Package	Non-Containerized Package Untraceable
C	VPM	Package	Containerized Package Untraceable
D	Missort	Package	Other Missorts
E	Missort	Container	Loaded in Wrong Trailer
F	Missort	Package	Fluid Loaded in Wrong Trailer - Through Sorter
G	Missort	Package	Stacked in Wrong Container - By Associate

Table 4.1: Top VPM and Missort Causes

4.1.2 Potential Factors

The goal is to develop a multiple linear regression model for each of the above causes. The purpose of the models is to identify factors (independent variables) that are correlated with the number of misses on a given day at SLC2 (dependent variable). Note the likelihood that a specific package will become a VPM or missort given its attributes were not modeled, therefore package-level factors such as dwell time and cycle time were not included. Models were built upon data over a 122-day period, with potential factors sourced from dock Subject Matter Experts (SME) across five USNS sites. Factors are listed in Table 4.2.

Some interesting observations and patterns include:

- **Day of week**: Tuesday and Wednesday were found to have higher correlations with the number of misses, likely related to shift management.

- **Container type "null"**: these packages were considered containerized but did not have a container type in the system. Thousands of units in this cause became VPM

#	**Factor**	**Description**
1	volume	Number of packages shipped on this day
2	JPH	Dock rate: Jobs-per-Hour (packages processed per hour)
3	containerized	Number of containerized packages shipped on this day
4	vc1_4	Number of dock associates with experience levels 1 to 4 (less experienced) working on this day
5	vc5_plus	Number of dock associates with experience levels 5 and above (more experienced) working on this day
6	post_slam_package_scan_rate	Post-SLAM scan rate (%) on this day
7	actual_dock_hours	Number of hours clocked in for dock work on this day
8	fluid_loaded	Number of fluid-loaded packages shipped on this day
9	day_of_week	Day of week (Mon-Sun)
10	stacking_locations	Number of stacking locations active on this day
11	container_*(type)*	Number of packages that had each container type (trailer, cage, cart, other, null)

Table 4.2: Factors considered in the building of regression models

misses and missorts. Based on sampling individual cases, they tend to be a mix of packages from custom-made boxes, fluid loads, and packages that went to the jackpot.

- **Container type "other"**: these were mainly composed of shuttles (large cardboard box containers) with some categorized as bags; about a hundred units became VPM misses and missorts.

4.2 Root Cause Models

4.2.1 Procedures

A multiple linear regression model was developed to understand each type of cause as well as overall VPM and missorts using the following procedures:

First, data was pulled and cleansed from March 1 to August 1, 2023. This includes the overall number of VPMs misses, the overall number of missort misses, and the number of

misses within each category of causes - these serve as dependent variables. Values associated with each factor, such as the total units, are also collected and mapped to each day as independent variables.

Then, a multicollinearity check is done for each of the models using a correlation matrix. Pairs of factors that have correlation coefficients (>0.9 or <-0.9) between the two independent variables are identified and one of the factors is eliminated.

Next, the data is randomly split into training data (70%) and testing data (30%) to ensure an evaluation of each model. The training data is then used to construct the linear regression model through a series of rounds. After each round, the significance (stars or p-values in R) of each variable is checked and the variable with the least significance (highest p-value) is removed. One variable is removed each round until all remaining variables are significant. Categorical variables, such as the day of the week and container type, are preserved if most of the categories are significant.

Finally, once the model is established, its goodness of fit to the training data is examined based on the R^2 value. The model's ability to generalize and predict is assessed using the testing data and evaluated based on the out-of-sample R^2 value.

4.3 VPM Models

A model was built for overall VPM misses and each of the top three VPM misses. An overview of the results is presented in Table 4.3, with (+) denoting a positive correlation, (-) denoting a negative correlation, and non-significant otherwise. At a glance, "container_null" seems to be a common factor across multiple causes. As mentioned, these were mostly custom-made boxes, fluid loads, and packages that went to the jackpot. A potential explanation could be that custom-made boxes follow a more manual process and packages in the jackpot were often problematic packages that required human attention.

For the overall VPM model (VPM total), variables were removed from the model in the order outlined in Table 4.4. Four of the variables were removed due to their high correlation with other independent variables.

#	**Factor**	**VPM Total**	**A: Other VPM**	**B: Non-Cont. Package**	**C: Cont. Package**
1	volume		+		
2	JPH				
3	containerized				
4	vc1_4				−
5	vc5_plus				
6	post_slam_package _scan_rate				
7	actual_dock_hours				
8	fluid_loaded				
9	day_of_week	+/−	+	+/−	
10	stacking_locations				
11	container_*(type)*	+ (cart)	− (null)	+ (null)	+ (null)
	R^2	24.19%	16.92%	42.29%	27.26%
	out-of-sample R^2	Poor	Poor	32.80%	18.30%

Table 4.3: Summary of correlations in VPM models

Stage	**Variable Removed**
Multicollinearity Check	actual_dock_hours containerized container_pallet container_trailer
Round 1	post_slam_package_scan_rate
Round 2	stacking_locations
Round 3	vc1_4
Round 4	volume
Round 5	fluid_loaded
Round 6	container_null
Round 7	vc5_plus
Round 8	JPH
Round 9	container_other

Table 4.4: Factors removed from VPM models

The result is the following regression equation:

$$y = -23.47 + 66.59x_1 + 2.94x_2 + 13.39x_3 - 2.97x_4 + 84.59x_5 + 97.55x_6 + 0.61x_7 + \epsilon$$

Where:

$$\begin{aligned}
y &= \text{number of VPM misses} \\
x_1 &= 1 \text{ if it is Monday, 0 otherwise} \\
x_2 &= 1 \text{ if it is Saturday, 0 otherwise} \\
x_3 &= 1 \text{ if it is Sunday, 0 otherwise} \\
x_4 &= 1 \text{ if it is Thursday, 0 otherwise} \\
x_5 &= 1 \text{ if it is Tuesday, 0 otherwise} \\
x_6 &= 1 \text{ if it is Wednesday, 0 otherwise} \\
x_7 &= \text{number of units processed through carts}
\end{aligned}$$

$$\begin{aligned}
R^2 &= 0.2419 \\
\text{Out-of-sample } R^2 &= -0.1003
\end{aligned}$$

4.3.1 A: VPM - Other

This is a catch-all category. Potential issues that fit into this category include packages that had missing scans, went missing after going through manual troubleshooting, and all other issues that do not fit into other categories. Due to its "catch-all" nature, this cause is a blend of multiple root causes, therefore both the goodness of fit and predictive ability of the regression model were relatively poor.

During the analysis period at SLC2, this is composed of roughly 2/3 fluid load and 1/3 containerized load. It was found that the day of the week (+/-) and the number of units with container type “null” (-) were the most significant factors.

Variables were removed from the model in the order outlined in Table 4.5.

Stage	**Variable Removed**
Multicollinearity Check	actual_dock_hours containerized container_pallet container_trailer
Round 1	fluid_loaded
Round 2	post_slam_package_scan_rate
Round 3	container_other
Round 4	vc1_4
Round 5	stacking_locations
Round 6	vc5_plus

Table 4.5: Factors removed from model for Cause A

The result is the following equation:

$$y = -33.59 + 0.60x_1 + 54.27x_2 + 16.63x_3 + 6.53x_4 + 5.77x_5 + 79.13x_6 + 46.10x_7 - 0.54x_8 + \epsilon$$

Where:

$$y = \text{number of VPM - Other misses}$$

$$x_1 = \text{total number of units (volume)}$$

$$x_2 = 1 \text{ if it is Monday, 0 otherwise}$$

$$x_3 = 1 \text{ if it is Saturday, 0 otherwise}$$

$$x_4 = 1 \text{ if it is Sunday, 0 otherwise}$$

$$x_5 = 1 \text{ if it is Thursday, 0 otherwise}$$

$$x_6 = 1 \text{ if it is Tuesday, 0 otherwise}$$

$$x_7 = \text{if it is Wednesday, 0 otherwise}$$

$$x_8 = \text{number of units with null container value}$$

$$R^2 = 0.1692$$

$$\text{Out-of-sample } R^2 = -0.4363$$

4.3.2 B: VPM - Non-Containerized Package Untraceable

These packages departed as untraceable but were never recorded as containerized at any point. Therefore, the system categorized them as non-containerized packages that went missing. This could mean that they were fluid loads that were not scanned to a trailer or containerized loads that did not get scanned to a container. The solution recommended by Sort Tech is to monitor and manage package scan accuracy such that container information is recorded.

The model suggests that the number of misses is influenced by the number of containers with the type "null" (+) and the day of the week (+/-). The positive correlation with the number of units with container value being null aligns with the possibility that these were fluid loads, thus they were never containerized and the system did not record a container value. Another possibility is that these are packages that require more human intervention (e.g. custom boxes, problematic packages in the jackpot), which means they do not follow the standard path. Deviations from the standard path leave more room for bypassing scanning or the containerization step. For the day of the week, the equation seems to suggest that, with all else equal, Saturday and Tuesday could have fewer VPM misses of this type than other days of the week. Interestingly, shift changes happen on Sundays and Wednesdays.

Variables were removed from the model in the order outlined in Table 4.6.

Stage	**Variable Removed**
Multicollinearity Check	actual_dock_hours containerized container_pallet container_trailer
Round 1	vc5_plus
Round 2	stacking_locations
Round 3	post_slam_package_scan_rate
Round 4	container_other
Round 5	JPH
Round 6	vc1_4
Round 7	fluid_loaded
Round 8	container_cart
Round 9	volume

Table 4.6: Factors removed from model for Cause B

The result is the following equation:

$$y = -4.84 + 0.65x_1 + 8.25x_2 - 4.99x_3 + 5.64x_4 + 1.39x_5 - 1.84x_6 + 28.53x_7 + \epsilon$$

Where:

y = number of VPM - Non-Containerized Package Untraceable misses

x_1 = number of units with null container value

x_2 = 1 if it is Monday, 0 otherwise

x_3 = 1 if it is Saturday, 0 otherwise

x_4 = 1 if it is Sunday, 0 otherwise

x_5 = 1 if it is Thursday, 0 otherwise

x_6 = 1 if it is Tuesday, 0 otherwise

x_7 = 1 if it is Wednesday, 0 otherwise

$R^2 = 0.4229$

Out-of-sample $R^2 = 0.3280$

4.3.3 C: VPM - Containerized Package Untraceable

These misses are similar to B: VPM - Non-Containerized Package Untraceable, but these packages departed as untraceable after being recorded as containerized at one point in the process. A closer investigation of 100 random samples using an internal troubleshooting web app revealed that 70% of these were last seen in a cage and 22% in a cart, which seems to suggest that cages, the reusable containers that do not leave the fulfillment center, contribute to this category more than carts do. The recommended action was to retrain all container loaders on the importance of scanning containers as they are being loaded. The model identifies the number of containers with type "null" (+) and the number of working dock associates at VC 1 to 4 (-) as determining factors. The model suggests an interesting relationship - the more inexperienced associates working on that day, the fewer VPM misses

of this type. One possible explanation that some of the Subject Matter Experts (SME) have mentioned was that newer associates tend to follow procedures more closely.

Variables were removed from the model in the order outlined in Table 4.7.

Stage	**Variable Removed**
Multicollinearity Check	actual_dock_hours containerized container_pallet container_trailer
Round 1	fluid_loaded
Round 2	container_cart
Round 3	volume
Round 4	container_other
Round 5	JPH
Round 6	day_of_week
Round 7	stacking_locations
Round 8	vc5_plus
Round 9	post_slam_package_scan_rate

Table 4.7: Factors removed from model for Cause C

The result is the following equation:

$$y = 2.84 - 0.36x_1 + 0.62x_2 + \epsilon$$

Where:

y = number of VPM - Containerized Package Untraceable misses

x_1 = number of working associates with VC 1 to 4

x_2 = number of units with null container value

$R^2 = 0.2726$

Out-of-sample $R^2 = 0.1830$

4.4 Missort Models

Similar to VPMs, a model was built for overall missort misses and each of the top four types of missort misses. An overview of the results is outlined in Table 4.8, with (+) denoting a positive correlation, (-) denoting a negative correlation, and non-significant otherwise.

#	**Factor**	**Missort Total**	**D: Other Missorts**	**E: Container Loaded in Wrong Trailer**	**F: Fluid Loaded in Wrong Trailer**	**G: Stacked in Wrong Container**
1	volume		+			
2	JPH		+		−	
3	containerized					+
4	vc1_4			−		
5	vc5_plus	−				
6	post_slam_package _scan_rate	−	−			
7	actual_dock_hours	+		+		
8	fluid_loaded					
9	day_of_week					
10	stacking_locations				−	
11	container_*(type)*				+ (other)	
	R^2	61.46%	37.62%	50.27%	31.56%	25.68%
	out-of-sample R^2	21.43%	Poor	11.34%	19.76%	26.13%

Table 4.8: Summary of correlations in missort models

For the overall missort model (missort total), variables were removed from the model in the order outlined by Table 4.9. Three significant factors were identified by the model. Logging more hours on the dock could suggest more work or more volume, so it makes sense that more missorts could occur. The number of experienced associates is negatively correlated with the total missorts, which is similar to the hypothesis mentioned in C: VPM - Containerized Package Untraceable for newer associates. Scan accuracy has a negative correlation, suggesting that the higher the scan rate, the lower the number of missorts.

Stage	Variable Removed
Multicollinearity Check	volume containerized container_pallet container_trailer
Round 1	container_null
Round 2	container_other
Round 3	container_cart
Round 4	stacking_locations
Round 5	vc1_4
Round 6	fluid_loaded
Round 7	day_of_week
Round 8	JPH

Table 4.9: Factors removed from missort models

The result is the following equation:

$$y = 847.00 + 0.68x_1 - 0.28x_2 - 849.00x_3 + \epsilon$$

Where:

$$y = \text{number of missort misses}$$
$$x_1 = \text{number of dock hours logged (actual)}$$
$$x_2 = \text{number of working associates with VC5+}$$
$$x_3 = \text{post slam package scan rate}$$

$$R^2 = 0.6146$$
$$\text{Out-of-sample } R^2 = 0.2143$$

4.4.1 D: Missort - Other Missorts

This is the catch-all category for missort, and no information is available to identify if the missort occurred at the container level or at the package level. Again, as a blended category

that contained multiple root causes, the model showed poor predictive performance for Other Missorts. It identified the volume (+), JPH (+), and Post-SLAM scan rate (-) as predictive factors.

Variables were removed from the model in the order outlined in Table 4.10.

Stage	**Variable Removed**
Multicollinearity Check	actual_dock_hours containerized container_pallet container_trailer
Round 1	day_of_week
Round 2	container_other
Round 3	fluid_loaded
Round 4	container_cart
Round 5	container_null
Round 6	vc1_4
Round 7	vc5_plus
Round 8	stacking_locations

Table 4.10: Factors removed from model for Cause D

The result is the following equation:

$$y = 763.96 + 0.60x_1 + 1.58x_2 - 836.20x_3 + \epsilon$$

Where:

$$y = \text{number of Missort - Other Missorts misses}$$
$$x_1 = \text{total number of units (volume)}$$
$$x_2 = \text{dock jobs per hour (JPH)}$$
$$x_3 = \text{post slam package scan rate}$$

$$R^2 = 0.3762$$
$$\text{Out-of-sample } R^2 = -0.0325$$

4.4.2 E: Missort - Loaded in Wrong Trailer

These packages were missorted due to the entire container being physically loaded onto the wrong trailer. This is the only cause within the top seven causes that occur at the container level, as all other misses occur at the package level.

The containers that were loaded into the wrong trailers are mostly cages and carts that are used to batch packages together before moving them to the next step. This batching process usually happens for destinations that have low to medium volume so that packages can accumulate in containers over time. Destinations with high volumes are usually fluid-loaded with packages being fed from the conveyor built to the trailer. The downside of batching packages into containers is that each error will cause multiple missorts. On average, container-level defects resulted in 9 to 25 missorted packages each.

The model identifies dock hours (+) and the number of working dock associates with experience levels 1 to 4 (-) as predictive factors. The relationship with dock hours is similar to that of missort total - more hours suggests more volume and more chances to miss. The negative relationship with newer associates is similar to C: VPM - Containerized Package Untraceable - newer associates may follow instructions and standards more closely.

Variables were removed from the model in the order outlined in Table 4.11.

Stage	**Variable Removed**
Multicollinearity Check	volume containerized container_pallet container_trailer
Round 1	day_of_week
Round 2	container_cart
Round 3	post_slam_package_scan_rate
Round 4	JPH
Round 5	container_other
Round 6	vc5_plus
Round 7	container_null
Round 8	stacking_locations
Round 9	fluid_loaded

Table 4.11: Factors removed from model for Cause E

The result is the following equation:

$$y = -13.83 + 0.64x_1 - 0.35x_2 + \epsilon$$

Where:

$$\begin{aligned} y &= \text{number of Missort - Loaded in Wrong Trailer misses} \\ x_1 &= \text{number of dock hours logged (actual)} \\ x_2 &= \text{number of working associates with VC 1 to 4} \end{aligned}$$

$$\begin{aligned} R^2 &= 0.5027 \\ \text{Out-of-sample } R^2 &= 0.1134 \end{aligned}$$

4.4.3 F: Missort - Fluid Loaded in Wrong Trailer - Through Sorter

These are fluid-loaded packages that were diverted by a sorter. As the package moves along on the conveyor, the sorter does an automatic scan and then diverts the packages to the path that will take them to the trailer they will depart in.

After observing and talking to associates working to move packages from the conveyor belt to the trail, a few possible explanations for the missort were identified: packages can be left on the conveyor in between trailer swaps, so these packages will end up on the next trailer that will likely be going to an unintended destination. The standard practice to mitigate this issue is to hand scan each package from the conveyor belt right after a trailer swap to clear the queue. Another possible explanation is that packages are sometimes down-stacked (moved from the conveyor to the floor) if there is a problem with the package or if the trailer becomes full. These packages may also be accidentally loaded onto the next trailer if not scanned carefully.

The model for this cause identified JPH (-), number of stacking locations (-), and number of units that had a container type "other" (+) as predictive factors. The number of stacking locations is a surprising outcome since fluid-loaded packages do not have stacking locations.

Given the negative correlation, one possibility is that fewer stacking locations on a day means less containerized load and more fluid loads, thus more fluid-loaded missorts.

Variables were removed from the model in the order outlined in Table 4.12.

Stage	**Variable Removed**
Multicollinearity Check	volume containerized container_pallet container_trailer
Round 1	container_null
Round 2	volume
Round 3	post_slam_package_scan_rate
Round 4	container_cart
Round 5	vc1_4
Round 6	fluid_loaded
Round 7	vc5_plus
Round 8	day_of_week

Table 4.12: Factors removed from model for Cause F

The result is the following equation:

$$y = 58.24 - 0.06x_1 - 0.40x_2 + 0.61x_3 + \epsilon$$

Where:

y = number of Missort - Fluid Loaded in Wrong Trailer Through Sorter

x_1 = dock jobs per hour (JPH)

x_2 = number of stacking locations

x_3 = number of units with container value of "other"

$R^2 = 0.3156$

Out-of-sample $R^2 = 0.1976$

4.4.4 G: Missort - Stacked in Wrong Container

These packages were manually stacked by an associate and resulted in a package-level missort. This could happen if the dock has a poor lane layout, such as missing labels, having outdated labels, and putting similar destinations closely together. This could also occur if associates are not scanning package-to-package, which means that they scan batches of packages at a time (instead of one at a time) and then move the entire batch to the same container. Since this is a more complex task and more prone to mistakes, scanning one package at a time is the standard being taught to new associates when they join.

The model identifies only one significant factor - the number of containerized packages (+) processed and shipped in a day. While this is the only model that only identified one correlation among the potential factors, this positive relationship makes sense because more packages being containerized could mean more chances to missort into the wrong container. On the other hand, this also suggests that changing other factors may not be helpful if the goal is to reduce the number of packages stacked in wrong containers.

Variables were removed from the model in the order outlined in Table 4.13.

Stage	**Variable Removed**
Multicollinearity Check	volume actual_dock_hours container_pallet container_trailer
Round 1	day_of_week
Round 2	container_other
Round 3	stacking_locations
Round 4	container_null
Round 5	post_slam_package_scan_rate
Round 6	vc5_plus
Round 7	container_cart
Round 8	fluid_loaded
Round 9	JPH
Round 10	vc1_4

Table 4.13: Factors removed from model for Cause G

The result is the following equation:

$$y = -2.97 + 0.60x_1$$

Where:

$$x_1 = \text{number of containerized units}$$

$$R^2 = 0.2568$$

$$\text{Out-of-sample } R^2 = 0.2613$$

Chapter 5

Solutions

Based on the top seven VPM and missort causes and factors identified in Chapter 4, a series of interviews and brainstorming sessions were conducted between February and August 2023, where over 20 stakeholders were engaged, including Operations Managers (OM), Area Managers (AM), dock Subject Matter Experts (SME), Delivery Estimate Accuracy (DEA) owners, Sort Tech, Process Tech, Process Engineering, and Systems Teams. The purpose was to brainstorm, identify, and assess potential solutions. As a result, a total of seven solutions were explored, some of which were piloted while others were studied more theoretically. An overview of the solutions is summarized in Table 5.1, including their estimated entitlement (as a percentage of the total cost of the problem), effort level, and impact level. More details on each solution are provided in this chapter.

#	**Solution**	**% of Entitlement**	**Efforts**	**Impact**
1	Increase Scan Rate	1%-10%	Low	Medium
2	Clean Label Off Carts	15%	Medium	Medium
3	Deep Dive Dashboard	30%	Low	High
4	Label Stacking Locations	3%	Low	Low
5	Scanning at Dock Doors	40%	Medium	High
6	Dynamic Container Stacking	54%	Medium	High
7	RFID for Packages	55%	High	High

Table 5.1: Overview of top 7 solutions studied

5.1 Short-Term Solutions

5.1.1 Solution 1: Increase Scan Rate

Scan rate measures the success rate of scan points for each package and container across the fulfillment network as scan failures have a significant impact on DEA. A missed package or container scan will lead to the lost virtual tracing of the item and therefore cause downstream impacts.

As identified by the regression models in Chapter 4, scan rate has been found to have a negative correlation with Missort total and D: Missort - Other Missorts. The more accurately items are scanned, the fewer DEA misses will occur. Low scan rate is caused by either sorter health issues (package-level) or associate execution (package-level and container-level). Sorter health issues are mostly related to how accurately packages are being scanned and diverted on the conveyor. Low accuracy could be attributed to software issues, mechanism malfunctions, outdated figuration (e.g. location of a trailer or container has moved but has not been reflected in the system), or bad placement of packages (e.g. packages placed side-by-side or closely together could cause a mis-scan).

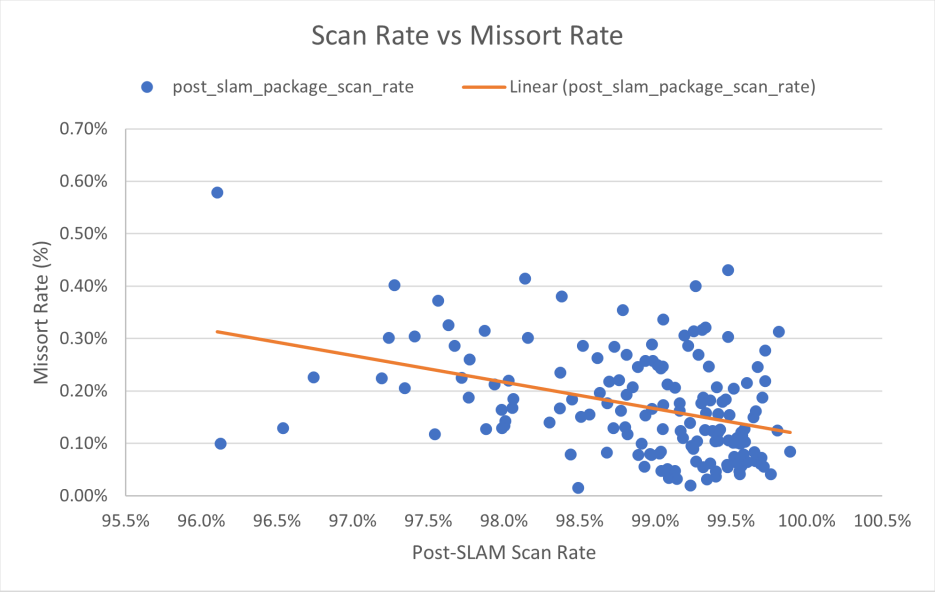


Figure 5-1: Daily scan rate vs missort rates at SLC2 from Mar 1 to Aug 1

As shown in Figure 5-1, daily scan rates have a moderate negative linear correlation with

the total missort rate. This result is in line with the outcomes of models for Missort - Total and Missort - Other Missorts. Based on high-level calculations using the Missort - Total model, with all else being equal, increasing the scan rate rate could potentially result in savings of $7,564 to $75,638 in savings each year for SLC2 (Table 5.2).

Increase in Scan Rate	Estimated Annual Savings
0.13%	$7,564
0.26%	$15,128
0.39%	$22,691
0.52%	$30,255
0.65%	$37,819
0.78%	$45,383
0.91%	$52,947
1.04%	$60,511
1.17%	$68,074
1.30%	$75,638

Table 5.2: Estimated annual savings for SLC2 for increasing scan rates in increments of 0.13%

From an execution perspective, increasing scan rate relies on training associates in their scanning activities, this includes reminding associates to scan at every step at the start of each shift, monitoring for failed moves in the system, and coaching associates in real-time.

5.1.2 Solution 2: Clean Labels Off Carts

While gathering insights on E: Missort - Loaded in Wrong Trailer, multiple dock AMs brought up the fact that carts can often be missorted based on the old label, especially near Critical Pull Time (CPT) when carts are rush loaded, resulting in Cause E - Container Loaded in Wrong Trailer. Cart labels contain critical information regarding the next destination of the entire container. They are printed and inserted into the side of the cart once the container is full of the expected packages. In standard practice, these labels should be removed by the final receiving facility (usually a Sort Center or a Delivery Station) once all packages have been removed. However, this step is often missed, thus carts are often re-used while the old labels are still present, often causing the entire cart to be missorted to the old destination, resulting in missorted packages.

On August 8, 2023, 73 carts on the SLC2 dock were audited - 97% of carts had old labels and 73% had 2+ old labels. As seen in Figure 5-2, from August 8 to 11, a pilot was run where all carts on the dock were cleared of old labels once a day, and it prevented 14%-21% of carts from being missorted based on old labels. The average cart missort rate was 0.08% during the 4-day pilot, compared to 0.15% during the 4 days prior and 0.12% during the 10 days prior.

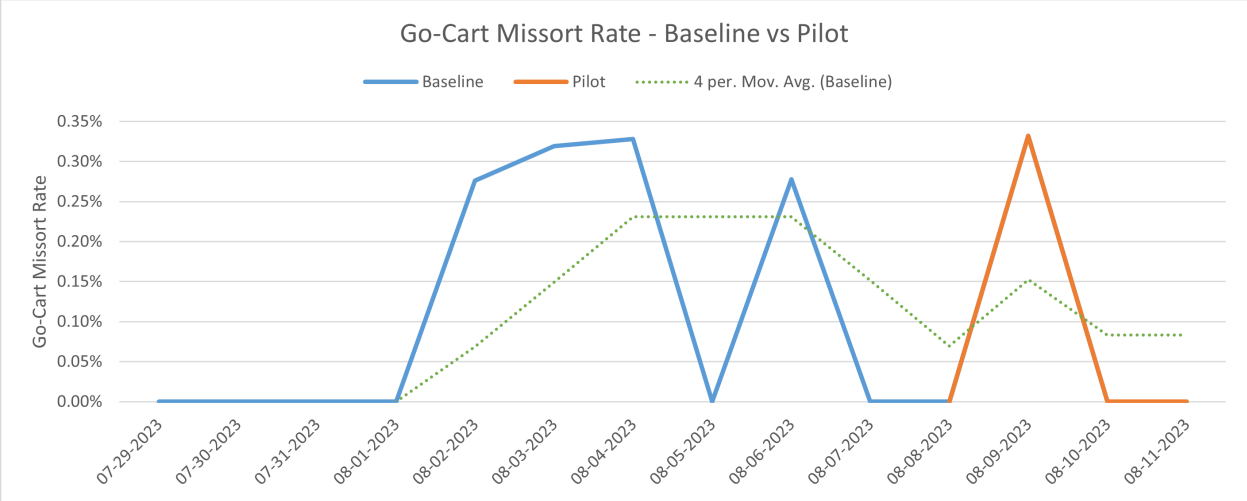


Figure 5-2: Pilot results on the effectiveness of removing old labels from carts

As discussed with dock AMs, the label-removing task makes the most sense as a part of standard work procedures when each cart gets unloaded from the trailer. This additional step would add about 13 seconds of work per cart. However, removing old labels makes it easier to identify new labels on a cart, thus knowing where to take the cart next and saving some time at the point of moving carts from staging areas into trailers.

5.1.3 Solution 3: Deep Dive Dashboard

This solution aims to make an impact on multiple causes identified in Chapter 4. Oftentimes, it is difficult to track changes in DEA misses and identify where problems have occurred. For example, it would be useful to know when misses occurred, what types of misses have been trending, which shifts perform better, and which stacking locations tend to cause problems. For F: Missort - Fluid Loaded in Wrong Trailer - Through Sorter in particular, a natural question to ask would be "which sorters are these packages coming from?"

To make it easier and faster to help dock AMs understand what, when, and where VPMs and missorts happen, as well as to provide a quick action plan, the VPM and Missort Deep Dive Dashboard was designed and developed using Amazon's internal dashboard software QuickSight. This dashboard pulls in data from Sort Tech databases and is refreshed on a daily basis. The dashboard is split into four main sections. Section A (Figure 5-3) provides a quick overview of the recent performance and immediate actions that can be taken to improve. It shows the number and percentage of misses over the past two weeks to help identify spikes and trends. It also ranks shifts (day, night, front-half of the week, back-half of the week etc.) and categories of causes that had the most misses over the past two weeks.

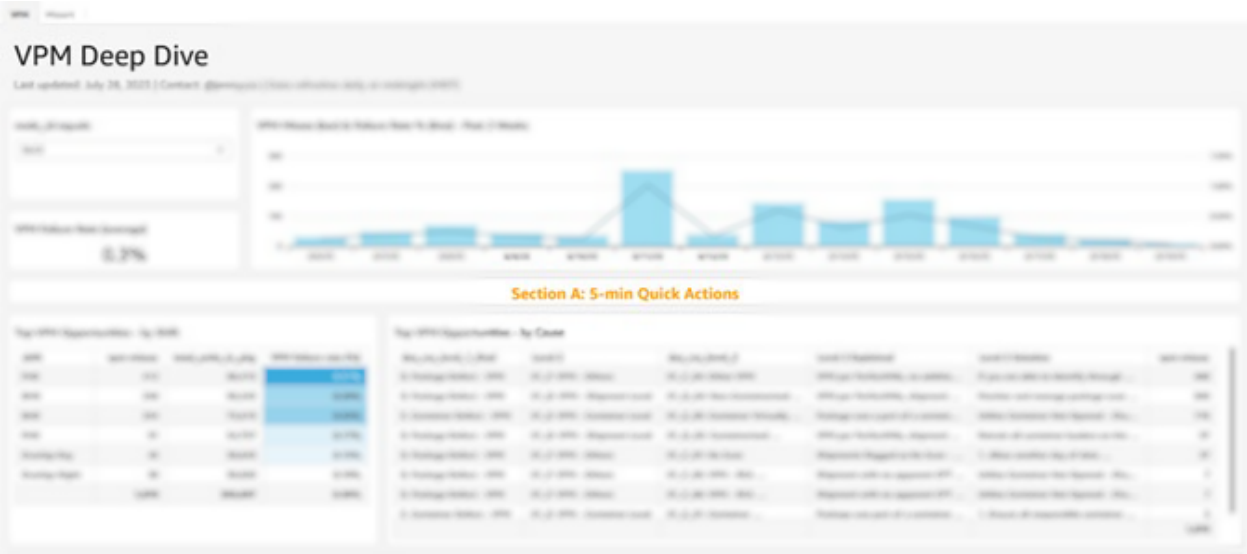


Figure 5-3: Sample screenshot from Section A of the VPM Dashboard (blurred for privacy considerations)

Section B (Figure 5-4) provides more details on the packages that became misses. This includes the load type (containerized or fluid), the type of box it was in, where it was intended to go (route), and if it experienced any order changes.

Section C further traces the missed packages by breaking down their path, including which SLAM station they came from and which scanner (sorter) they went through. Section D puts the information in the previous three sections together in a tabular format for single-package tracking, and includes timestamps of each movement.

This dashboard is meant to be interactive in that all of the graphs and charts can be used to filter each other. This makes investigating misses easier, as the user can dig into particular

Figure 5-4: Sample screenshot from Section B of the VPM Dashboard (blurred for privacy considerations)

cases by filtering down using different attributes. Previously, investigating missing packages was done by tracking a single package ID across multiple web apps internal to Amazon. This dashboard not only makes this task faster and easier, but it allows the user to see aggregated data on the misses that have occurred.

The VPM and Missort Deep Dive Dashboard was piloted with 5 sites over the course of six weeks from June 4 to August 5. Each week, subject to availability, a review of the top causes, actions, and shifts to focus on was conducted for each site. As a result, four out of five sites experienced fewer VPM misses (42% reduction on average) during weeks they had a dashboard review compared to weeks when they did not. Similarly, for missorts, four out of five sites experienced fewer misses (31% reduction on average). More than five dock SMEs have expressed that this dashboard was helpful – it revealed information they had not seen before and allowed for both high-level understanding and deep dives in one interface.

This initiative would directly impact dock AMs and indirectly impact associates by putting more emphasis on following standard work procedures. This would help associates maintain work habits that will improve their quality and performance.

5.1.4 Solution 4: Label Stacking Locations

As discussed in G: Missort - Stacked in Wrong Container, poor lane layout could cause more packages to be sorted into the wrong carts or cages. Although the number of stacking

locations was found to have correlations with this cause, the number of stacking locations is usually based on the volume planned for the day, therefore it cannot be changed easily. Instead, to reduce the impact of this cause, labels can be displayed and maintained at stacking locations throughout the day.

A stacking location is an area on the dock floor where packages received are sorted before they are staged and moved into a trailer. These locations are labeled within the floor plan layout and are agnostic of the actual destination where packages will go. To help associates find the right stacking location(s) for a package, destination labels are often used on the dock.

These destination labels contain four-digit codes (e.g. SLC2) that indicate the facility that packages from this stacking location will be sorted to next. This should match the 4-digit code that can be found on the SLAM label. Since demand and routing changes occur all the time, destinations can be assigned to different stacking locations depending on the shift and time of day. Therefore, destination labels are constantly being created, moved, and destroyed. This can often create confusion on the dock. Non-intuitive lane layout - such as having out-of-place labels, multiple labels for one stacking location (Figure 5-5), and illegible labels (Figure 5-6) - is one of the main reasons why associates may stack a package into the wrong container.

Figure 5-5: Example of a stacking location with multiple labels

Figure 5-6: Example of an illegible label where the ink has worn off

A pilot test was conducted on SLC2 dock between July 5 and July 16, 2023, where new, ergonomic destination labels (Figure 5-7) were implemented and maintained on four stacking locations with containers. New destination labels were printed with block letters to increase readability. They are also differentiated by color, shape (on the left), and background (Figure 5-8) such that they can be more memorable and recognizable in the long run. Lastly, new labels are laminated such that they can be removed and re-taped to a new stacking location as needed without being easily damaged.

During the 12-day testing period, the average error rate for stacking was 0.027%, down from the average rate of 0.052% in the 12 days before the pilot, and also lower than the average rate of 0.035% for the 12 days after the pilot (Figure 5-9). Note that the new destination labels were still used after the pilot but were no longer maintained diligently as in the testing period.

During the pilot, more than eight dock associates had proactively expressed that they found these labels helpful, and/or that they had been wanting something to this effect for a while. This provides evidence that the change can improve the associates' stacking experience and performance.

Figure 5-7: Example of a new label implemented during testing

Figure 5-8: New destination labels used during testing

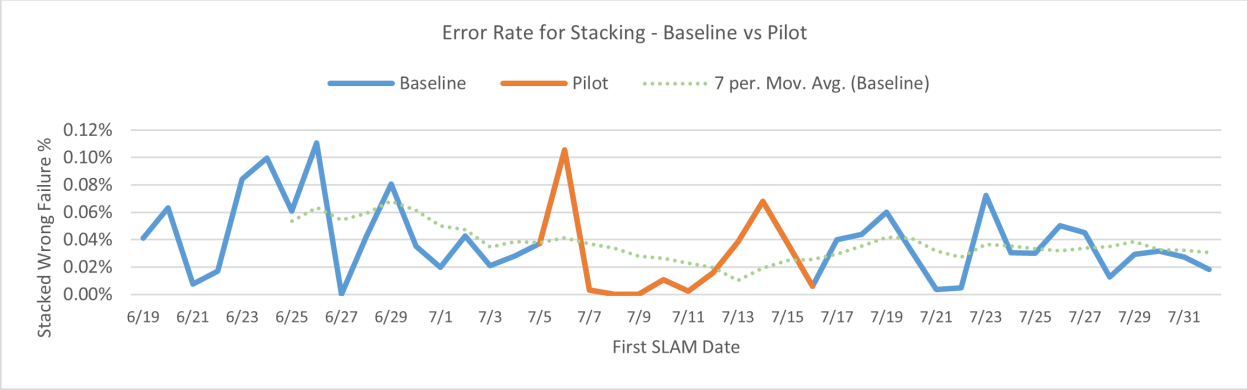


Figure 5-9: Error rate for stacking before, during, and after the pilot

5.2 Medium-Term Solutions

5.2.1 Solution 5: Scanning at Dock Doors

Another way to increase scan rates, targeting Missort total and D: Missort - Other Missorts, is to implement automated scans. Adding an extra scan at the point of dock-loading is a potential solution aiming to prevent packages from being fluid-loaded incorrectly. One way to do so would be to install a camera above the conveyor during loading such that every package that enters the trailer gets an additional confirmation scan. This would prevent fluid-loaded packages from being virtually loaded into a trailer but physically going onto a different trailer or being left on the floor.

# of dock doors to track	6
Scanners	$3,500 each
Support / Binding	$200 each
Software and integration	$400
Total set up cost	**$24,600**

Table 5.3: Cost estimates obtained from vendor for a pilot

A high-level quote (Table 5.3) was obtained through contact with a vendor who has experience implementing similar mechanisms with other logistics organizations. The initial scope of six dock doors will take four to six weeks to implement. Based on high-level calculations using concession costs for VPMs and missorts, and the number of fluid-loaded packages that become VPMs and missorts each year, this solution is estimated to yield potential savings of over 10x its cost. This solution would not impact the way associates currently load packages from the conveyor belt to the trailer. However, associates may need to learn additional exception-handling processes.

5.2.2 Solution 6: Dynamic Container Stacking

Dynamic Container Stacking (DCS) is an intuitive visual guidance solution that makes the manual container-building process easier. The idea is to use visual cues to eliminate memorization of stacking locations and thus reduce errors. More specifically, when an

associate scans a package, the interactive user interface will highlight the correct stacking location for the package to move to, giving the associate a clear path to follow. G: Missort - Stacked in Wrong Container would benefit from this solution as it would make package sorting less error-prone.

It can also improve palletizing rates, reduce post-SLAM misses, reduce failed moves, and improve the associate experience. DCS was piloted at SLC2 in 2022 and was projected to have a payback period of just 3.5 weeks. The future planned stacking mechanism of using a ProGlove (Figure 5-10) and/or light indicators above each stacking location can drastically reduce the likelihood of a stacking missort. In addition to financial benefits, this solution can also help shorten the time it takes for associates to find the right containers, therefore reducing carrying time, which reduces the likelihood of musculoskeletal injuries.

Figure 5-10: Illustrative image of ProGlove. Image source: theloadstar.com/if-the-glove-fits-wearable-devices-set-to-improve-warehouse-operations/

5.3 Long-Term Solutions

5.3.1 Solution 7: Radio Frequency Identification for Packages

Adding Radio Frequency Identification (RFID) tags at the package level can enable accurate tracking of packages without the need for automatic or manual scanning. In theory, this

could eliminate all VPM misses since virtual and physical movements can be tracked in real-time 24/7 and never get out of sync. Unfortunately, this theory has not been validated yet. Widespread implementation has been proposed for over 20 years but has not been implemented due to its traditionally high costs.

According to a McKinsey article from 2021, "the average cost of an RFID tag has fallen by 80% to about $0.04 in the last decade... prices of RFID readers have dropped by nearly 50%" [4]. According to a study of RFID in Amazon Customer Fulfillment Centers in 2006 [5], it was identified that "the best place to conduct Amazon's first implementation of RFID is at the outbound sorting dock" with package missorts being the main defect in question. By implementing RFID, packages can pass through the last step of the outbound process (loading into a trailer) and be validated passively without the need for additional manual intervention. This would theoretically detect VPMs and identify missorts in real-time. A high-level quote (Table 5.4) was obtained through contact with a vendor who has experience implementing RFID in the industry. Note that the cost of a tag at $0.07 is still quite high. Based on projected savings from reducing VPM and missort of fluid-loaded packages, the estimated annual savings are only 1.28x the annual costs. If the cost per tag could reach $0.04, this estimated annual savings can increase to 2.08x the annual cost.

Variable cost	
Cost of tag (each)	$0.07
Set up cost	
# of dock doors to track	6
Reader (hardware & configuration for each door)	$4,000
Software and integration	$60,000
Total set up cost	**$84,000.00**

Table 5.4: Cost estimates obtained from vendor for an RFID pilot

The initial use case of implementing RFID scans at dock doors would have minimal impact on the associate experience, as it would be a passive step. Associates may need to get used to responding to alerts from the scanning system, such as how to re-route a missorted package to its expected locations via a defined process.

Chapter 6

Conclusion and Future Work

6.1 Summary

In summary, this thesis outlined the problem, methods, and solutions studied for improving Amazon's fulfillment process to minimize Delivery Estimate Accuracy (DEA) misses. Specifically, the project focused on Virtual Physical Mismatch (VPM) and missort misses that originate at the Fulfillment Center (FC) level. Despite conducting most of the study at SLC2 - a Traditional Non-Sortable FC - different types and generations of FCs, as well as their interconnectivity with downstream processes at Sort Center (SC)s and Delivery Station (DS)s, were also considered throughout.

Several qualitative and quantitative methods were used in this study, including interviewing stakeholders, performing and observing frontline work, performing data analysis, designing pilots, and running experiments. This provided a comprehensive understanding of the outbound processes, a foundation for when and where the errors may occur, and where to go for more information. A review of internal and external literature was also conducted early on to obtain more knowledge on VPM, missort, and other research in this domain.

Leveraging an internal root cause framework as the starting point, a list of potential causes and factors was identified. This list was then used to build seven multiple linear regression models, one for each of the top causes to solve. These models were able to provide a more detailed understanding of what factors are associated with each cause.

Correspondingly, seven potential solutions were also identified and studied. Each solution -

whether short-term, medium-term, or long-term - had the potential to address one or more top causes. Efforts and impact of each solution were estimated, along with its potential savings as a percentage of the total cost of the problem. The effort-to-impact ratio is informative in prioritizing the next steps.

6.2 Limitations and Assumptions

The analysis from this research project has a few known limitations and assumptions:

DEA Root Cause Attribution Logic - Root Cause Attribution (RCA) logic for post-SLAM FC VPM and missorts currently only applies to cases where a DEA miss occurred. An estimated 1/3 of the true misses are recovered downstream and another 1/3 are attributed to other parts of the supply chain, so these misses have not been analyzed. In May 2023, a new project named Missorts RCA was commissioned by Sort Tech in order to address this limitation. A similar initiative for VPM is highly recommended.

Control of Experiments – Pilots were conducted in a dynamic environment where multiple factors could be at play at any moment. For example, the effects of Project Boomerang construction and go-live (April to May) and Prime Week (July 9 to 15) had major impacts on dock processes. Changes in staffing could also influence test results.

Data Quality - Data used to conduct the project analysis was pulled from multiple sources, including different databases and dashboards. At the package level, the data have sometimes been found to be different and disjointed between sources. To mitigate this risk, Sort Tech data was used as the primary basis for analysis with other sources as support.

6.3 Application to Other Sites

Based on stakeholder engagement with different sites, it was observed that although there are many common standard operating procedures, each U.S. Traditional Non-Sortable (USNS) sites operate in its own way depending on factors such as its site generation, its leadership decisions, its volume and type of packages, and whether or not it has certain features such as automated SLAM, recirculation set up, tail sorter, and certain conveyors. Therefore, a

similar analysis is recommended for each site to find:

- **Its top VPM and missort causes**: most sites will share "other VPM" and "other missort" as the top causes but have different second and third causes depending on their features and processes.
- **The factors that may be associated with their top causes**: additional factors and parameters may be included and existing ones may be removed depending on the site. A site-specific model would help identify and quantify the changes required.
- **The impact of explored solutions on their site**: given that each of the explored solutions only targets one or a few of the top causes, an assessment of the impact of each of the seven solutions explored above is recommended for prioritization.

6.4 Recommendations for Future Research

In order to prevent the annual loss due to VPM and missorts, sites like SLC2 should prioritize implementing solutions that are low/medium in effort but medium/high in impact:

1. **Focus on improving scan rate**: low effort, medium impact; spend minimal funding to solve 1% to 10% of the VPM and missort problems.
2. **Continue leveraging the Deep Dive Dashboard**: low effort, high impact; spend minimum funding to solve 30% of the VPM and missort problems by providing dock operations the insights they need to improve; easily expandable to other sites.
3. **Implement scanning at dock doors**: medium effort, high impact; spend $25K to set up cameras that can solve roughly 40% of the VPM and missort problems.

In addition, refining the RCA logic is crucial for FCs to better trust and understand their misses. Consolidating tools for DEA for VPM and missort deep dives will make this process easy and actionable, perhaps by leveraging the Deep Dive Dashboard or similar initiatives as the starting point. Developing and communicating robust insights to FCs will set the dock team up for success. Advanced technology can produce high-impact results with the support of a responsible, well-informed team.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] Nayeli Arellano Martinez. "Visual Sort Marker Digitization in Sort Center Operations". Accepted: 2023-10-18T17:09:56Z. Thesis. Massachusetts Institute of Technology, June 2023. URL: https://dspace.mit.edu/handle/1721.1/152487.

[2] Vrajesh Y. Modi. "Application of flexible labor and standard work in fulfillment center produce operations". Accepted: 2015-09-29T18:58:55Z. Thesis. Massachusetts Institute of Technology, 2015. URL: https://dspace.mit.edu/handle/1721.1/99025.

[3] Elizabeth Ann Ohrt. "Analysis of outbound process defects at Amazon's ONT2 Fulfillment Center". Accepted: 2016-09-27T15:14:54Z. Thesis. Massachusetts Institute of Technology, 2016. URL: https://dspace.mit.edu/handle/1721.1/104393.

[4] *RFID in retail | McKinsey*. URL: https://www.mckinsey.com/industries/retail/our-insights/rfids-renaissance-in-retail.

[5] Howard H. (Howard Hao) Shen. "A Radio Frequency Identification (RFID) evaluation strategy for customer fulfillment centers". Accepted: 2007-04-20T15:56:23Z Journal Abbreviation: RFID evaluation strategy for customer fulfillment centers. Thesis. Massachusetts Institute of Technology, 2006. URL: https://dspace.mit.edu/handle/1721.1/37247.