Multi-Objective Optimization of Container Load Plans for Modulating Inventory Flow

by

Shweta Sen

B.S., Mechanical Engineering
Pennsylvania State University, 2017

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

Master of Business Administration

and

Master of Science in Mechanical Engineering

in conjunction with the Leaders for Global Operations program

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

© Shweta Sen, 2025. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Shweta Sen
MIT Sloan School of Management and
Department of Mechanical Engineering
May 9, 2025

Certified by: Kamal Youcef-Toumi
Professor, MIT Mechanical Engineering Department
Thesis Supervisor

Certified by: Thomas Roemer
Senior Lecturer, MIT Sloan School of Management
Executive Director, Leaders for Global Operations
Thesis Supervisor

Accepted by: Nicolas Hadjiconstantinou
Chair, Mechanical Engineering Committee on Graduate Theses

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

Multi-Objective Optimization of Container Load Plans for Modulating Inventory Flow

by

Shweta Sen

Submitted to the MIT Sloan School of Management and
Department of Mechanical Engineering
on May 9, 2025, in partial fulfillment of the
requirements for the degrees of
Master of Business Administration
and
Master of Science in Mechanical Engineering

Abstract

Conventional strategies for container load planning (CLP) predominantly emphasize maximizing container utilization, which can result in suboptimally-timed inventory arrival, increased inventory holding costs, and downstream operational inefficiencies. Using a real-world case study from a global footwear and apparel retailer, this research formulates a novel multi-objective mixed-integer linear programming (MOMILP) model that jointly considers container utilization, transportation and storage costs, and timing accuracy of inventory delivery. The proposed model utilizes a branch-and-bound algorithm to evaluate numerous load configurations, assessing the impact of different load rules and weighting parameters on transportation performance metrics and inventory flow. Results highlight the cruciality of prioritizing delivery precision in transportation management decisions, demonstrating that solely maximizing volume utilization can adversely affect overall cost efficiency when downstream inventory storage and operational requirements are considered. This work also provides a process map of load planning activities and identifies targeted operational improvements, such as consolidation bypass and purchase order (PO) partitioning, that can enhance inventory flow smoothness, reduce transportation costs, and support more responsive logistics networks. Collectively, this work extends existing CLP methodologies by incorporating delivery timing and inventory storage considerations into load planning decisions, offering practical enhancements for logistics optimization.

Thesis Supervisor: Dr. Kamal Youcef-Toumi
Title: Professor, MIT Mechanical Engineering Department

Thesis Supervisor: Dr. Thomas Roemer
Title: Senior Lecturer, MIT Sloan School of Management
Executive Director, Leaders for Global Operations

Acknowledgments

I would like to express my deepest gratitude to my advisors, Dr. Thomas Roemer and Dr. Kamal Youcef-Toumi, for their invaluable guidance, feedback, and support throughout my research journey. I am sincerely thankful to my company sponsors and colleagues for their mentorship, insights, and encouragement to explore diverse research paths. I also extend heartfelt thanks to my Leaders for Global Operations (LGO) classmates for their camaraderie and collaboration over the past 2.5 years. I am equally grateful to the LGO staff for their steady support and guidance, which helped sustain and facilitate this work behind the scenes. Finally, I extend my deepest admiration and love to my parents and my partner. Their enduring support, guidance, and belief in me remain a constant source of strength and inspiration.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures **9**

List of Tables **11**

Acronyms **12**

1 Introduction **15**

1.1 The Role of CLP in Supply Chain 15

1.2 Footwear and Apparel Industry Overview 18

1.3 The Fleetform Supply Chain 19

1.3.1 Terminology and Network Map 20

1.3.2 Operational Complexities 22

1.4 Problem Motivation and Approach 25

2 Container Logistics and Process Flow **27**

2.1 Types of Containers 27

2.2 Load Rules and Load Plans 28

2.3 Load Planning Process Flow 31

2.4 Optimization Opportunities 35

2.5 Chapter Summary 36

3 Literature Review **37**

3.1 Classical Approaches to Container Load Planning 37

3.2 Freight Consolidation 49

3.3 Chapter Summary 52

4 Methods **53**

4.1 Problem Formulation 53

4.2 Modeling Approach and Setup 55

4.2.1 Branch-and-Bound Implementation 56

4.3 Model Components and Formulation 60

4.3.1 Nomenclature 60

4.3.2 Temporal Structure & Storage Costs 61

4.3.3 Container Capacity & Activation Costs 64

4.3.4 Model Components 64

4.3.5 Load Rule Modeling 69

4.3.6 Model Assumptions 70

4.4 Use Case Illustration of Model Outputs 73

4.5 Chapter Summary 76

5 Simulation Results and Key Takeaways **77**

5.1 Pareto Analysis 77

5.2 Inventory Flow Analysis 79

5.3 Load Rule Impact Analysis 82

5.4 Chapter Summary 84

6 Process Refinement and Model Extensions **87**

6.1 Revised CLP Workflow 87

6.2 Model Extension: Consolidation Bypass 89

6.3 Model Extension: PO Partitioning 92

6.4 Discussion & Considerations 93

6.5 Chapter Summary 95

7 Conclusions & Recommendations **97**

List of Figures

1-1 High-Level Schematic of Fleetform Distribution Network 21
1-2 Aggregate PO Volume Delivery to DC in 2023 23
1-3 Delivery Timing: Historical Deviation from Required Delivery Date . 24

2-1 Container Types and Relative Unit Volume Cost 28
2-2 Example of a Container Load Plan 30
2-3 Process Flow Transactions for Fleetform Container Logistics 33

4-1 Overview of Cost Minimization Methodology for Load Plan Optimization 54
4-2 Illustrative Example of Branch-and-Bound Network Diagram [27] . . 57
4-3 Temporal Nodes of Optimization Model 61
4-4 Floor Loading Setup of PO Cartons Inside Container 71
4-5 Load Plan Generation & Flow Visualization Example 74

5-1 Normalized Pareto Analysis of Delivery Precision Trade-Offs. 78
5-2 Flow Simulation of Inventory Arrival at DC Under Varying Weights. . 81
5-3 Flow Simulation of Inventory Arrival at DC Under Different Load Rules 84

6-1 Comparison of Current and Model-Enabled CLP Workflows 89
6-2 Flow Schematic & Optimization Setup for Consolidation Bypass . . . 91
6-3 Flow Simulation of Inventory Arrival at DC Using Consolidation Bypass 91
6-4 Flow Simulation of Inventory Arrival at DC Under PO Partitioning . 93

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

3.1 Comparison of Optimization Approaches for Container Loading Problems. 38

4.1 Model Nomenclature 60

5.1 Performance Metrics Comparison for Flow Simulations 80
5.2 Load Rule Impact Comparison Across Key Performance Metrics . . . 83

THIS PAGE INTENTIONALLY LEFT BLANK

Acronyms

3PL Third-Party Logistics. 20, 37, 49

CDC Central Distribution Center. 21

CDF Cross-Dock Facility. 20

CLP Container Load Planning. 15

DTC Direct-to-Consumer. 18

ERP Enterprise Resource Planning. 29, 94

FCL Full Container Load. 49

FIFO First In, First Out. 24, 30, 32, 51

GCW Goods at Consolidator Week. 31, 55, 61

LCL Less Than Container Load. 50

MILP Mixed-Integer Linear Programming. 37, 43, 55

MOO Multi-Objective Optimization. 37, 45

PO Purchase Order. 20, 34, 53

RDC Regional Distribution Center. 20

RDW Required Delivery Week. 53, 55, 61

SKU Stock Keeping Unit. 24

TMS Transportation Management System. 55, 94

VTT Vessel Transit Time. 22, 61

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

This section establishes the research context, highlighting industry trends and key operational challenges that motivate the development of a container load planning (CLP) optimization model. It also introduces the case study setting, outlines key terminology, and defines the scope and objectives of the analysis.

1.1 The Role of CLP in Supply Chain

CLP is the strategic process of assigning and arranging cargo within shipping containers to maximize space utilization, ensure load stability, and meet operational constraints such as weight distribution, handling requirements, and delivery sequencing. Efficient CLP optimizes inventory allocation, reducing the number of containers required while supporting delivery precision and cost minimization. In 2024, ocean shipping containers transported an estimated 7.4 billion metric tons of goods, [47], representing approximately 60% of the global maritime trade volume [22]. With containerized shipping becoming increasingly central to global commerce, the global shipping container market is forecasted to grow from $9 billion in 2023 to $16 billion in 2028 [42]. As a result, CLP has become a key lever for improving logistics operations, generating ripple effects that shape broader supply chain performance:

- **Transportation Costs**: Inefficient CLP can inflate transportation costs by increasing the number of containers required to move the same volume of goods. The impact of using more containers is significant because container freight rates—the cost paid by a firm to transport a container—are highly volatile and influenced by commodity market conditions, marine fuel prices, geopolitical events, trade policies, weather disruptions, vessel capacity, and logistical bottlenecks (e.g., labor strikes, piracy, equipment shortages), among other factors [29]. For instance, the historical oscillation in average freight rate for a 40-foot container is as follows: $2,409 in January 2021, $4,985 in January 2022, $955 in January 2023, and $2,136 in January 2024 [44]. Critically, the container spot market for unplanned shipments or demand surges can drive freight rates up to six times higher than contracted rates. Beyond freight costs, inefficient CLP contributes to port congestion and carbon emissions, leading to additional cost exposures such as higher tariffs, storage fees (at origin or destination ports), demurrage, and regulatory penalties [39, 30, 38]. Optimized CLP is therefore critical for minimizing per-unit transportation costs by improving container utilization, reducing reliance on spot market rates, and mitigating downstream logistical expenses associated with suboptimal shipping practices.

- **Inventory Management Policies**: CLP is crucial for aligning inventory arrivals with demand forecasts. Premature or unforecasted inventory arrivals can overwhelm downstream warehouses and distribution centers, increasing holding cost, necessitating off-site storage requirements, and causing operational inefficiencies. Late inventory arrives, in contrast, can lead to stockouts and lost sales, forcing firms to rely on costly expedited shipping or late-stage redistribution to balance inventory across multiple sales channels [1, 12]. By extension, CLP plays an integral role in shaping safety stock levels and upstream production buffers, ensuring sufficient slack between supply chain nodes to absorb variability without inflating carrying costs [2, 5]. CLP can further help synchronize inventory levels with demand cycles, thereby minimizing excess inventory.

- **Distribution Network Efficiency**: As many retailers increasingly adopt omnichannel distribution models [13, 14]—where inventory must be dynamically allocated across physical stores, fulfillment centers, and direct-to-consumer channels—inefficient CLP can disrupt network efficiency [40]. The efficiency of a distribution network is characterized by measures that include per-unit transportation cost, transit or cycle time, on-time delivery rate, backorder rate, lost sales, and customer feedback. Suboptimal CLP can misalign inventory flows, leading to stock imbalances at distribution nodes and delayed replenishment at key retail locations. The resulting distribution bottlenecks not only increase freight expenses, but also strain warehouse operations, dampening the agility required for seamless omnichannel fulfillment. Thus, precise CLP is necessary for maintaining inventory fluidity, ensuring timely deliveries, and enabling cost efficiency of the distribution network.

Within the retail environment, there are notable challenges to CLP implementation:

- **Retailer Complexities**: Retailers must navigate product heterogeneity, as variations in aspects such as style, color, size, and category add complexity to CLP. Fluctuating consumer demand across different regions adds uncertainty, making it difficult to predict optimal container allocations. Additionally, short product lifecycles, driven by seasonality and rapidly changing trends, require robust CLP strategies to minimize excess inventory and avoid stockouts.

- **Operational Limitations**: Warehouse and distribution center constraints, including limited storage capacity, can restrict the ability to pre-stage shipments for optimal container loading. Labor shortages and scheduling challenges further impact efficiency, leading to delays in packing and dispatching goods.

- **Logistical Hurdles**: Rising and often unpredictable freight rates continue to pressure retailers to maximize container utilization to minimize transportation costs. Supply chain disruptions, including port congestion, customs delays, and geopolitical instability, introduce further uncertainty, making it challenging to maintain consistent and cost-effective CLP execution.

- **Computational Constraints**: CLP involves high-dimensional optimization, where numerous constraints—such as weight distribution, stacking rules, product mixing (or co-loading), and product fragility—must be simultaneously considered. Traditional linear and static loading methodologies cannot fully capture the stochastic nature of supply-demand fluctuations, making it difficult to generate stable and efficient load plans in dynamic retail environments.

This research situates CLP as both an operational priority and an economic imperative. Through effective CLP implementation, firms can create more resilient and adaptive supply chains that reduce transportation costs, strengthen delivery precision, and smoothen inventory flow. The research and case study described in this work was conducted at an apparel and footwear retailer, which is hereafter referred to as Fleetform for anonymity. The goal of this work is three-fold: 1) provide an examination of container load planning approaches, 2) offer an optimization model and process enhancement that evaluates trade-offs between transportation cost, delivery precision, and container utilization, and 3) quantify the impact of loading constraints on logistics performance metrics. Building on existing work, this study introduces a novel optimization approach that simultaneously considers delivery precision, inventory storage costs, and inventory flow stability in load planning decisions, offering a more holistic optimization approach that evaluates the impact of CLP decisions on downstream supply chain nodes.

1.2 Footwear and Apparel Industry Overview

Transportation Management

The global footwear and apparel market, valued at roughly $1.5 trillion [41], has undergone a significant transformation in recent years, particularly in its approach to transportation management and container operations. This shift has been largely driven by the proliferation of e-commerce, the COVID-19 pandemic, and the subsequent acceleration of direct-to-consumer (DTC) models adopted by industry leaders such as Nike, Adidas, Under Armour, and Lululemon. The decentralization of distribution systems in response to modern retail demands has introduced new logistical complexities, necessitating CLP approaches that are more flexible and granular compared to traditional bulk-shipping methods.

The footwear and apparel industry employs various distribution models to meet the complex demands of modern retail environments. Traditionally, the industry relied on a linear supply chain with clear distinctions between manufacturers, wholesalers, and retailers. However, the advent of e-commerce and omnichannel retailing has blurred these lines, leading to more integrated and flexible distribution strategies. One prevalent model is the hub-and-spoke distribution system, which centralizes inventory at a primary distribution center (hub) and utilizes multiple delivery routes (spokes) to reach various destinations. This approach optimizes route planning, reduces costs, and improves inventory control. Another common model is the multi-channel distribution strategy, where companies utilize a combination of brick-and-mortar stores, e-commerce platforms, wholesale channels, and direct-to-consumer approaches. This diversification allows retailers to cater to different consumer preferences and maximize market reach. Additionally, some companies have adopted a decentralized distribution model, employing multiple smaller distribution centers located closer to end markets. This strategy can offer faster delivery times and reduced transportation costs for last-mile delivery, though it may increase overall inventory holding costs. The choice of distribution model often depends on factors such as brand positioning, product type, and target market, with many companies implementing hybrid approaches to balance efficiency and flexibility in their supply chains. Fleetform primarily utilizes a hub-and-spoke model, centralizing inventory at strategic distribution centers that serve as hubs, from which multiple spokes extend to various destinations.

1.3 The Fleetform Supply Chain

The scope of this work is linked with Fleetform's North America supply chain, and focuses on optimizing container load plans and transportation management practices to improve cost and operational efficiency. Given that this work is primarily anchored on the downstream supply chain, we assume that upstream decisions regarding incoming supply are fixed and outside the scope of our control. Furthermore, we focus exclusively on maritime containerization and use a Vietnam-to-California shipment lane as a representative route (in terms of volume of goods and profile of inventory flow) to serve as a case study for CLP optimization.

1.3.1 Terminology and Network Map

Figure 1-1 provides a high-level illustration of inventory flow within the Fleetform distribution network, with the relevant terminology described below.

- **Purchase Order (PO)**: POs are documents issued by Fleetform to factories that serve as an official request for goods or services. For a PO to be generated by Fleetform, it must contain three minimum inputs: a product code with requested quantities (for each size, style, or color pattern), a required delivery date, and destination of shipment. The required delivery date specifies when a PO must reach its shipment destination.

- **Consolidator**: As third-party logistics (3PL) facilities, consolidators aggregate manufactured goods from multiple factory locations and pack the goods into shipping containers. Consolidators are often located near shipping ports, and Fleetform engages with a multitude of consolidators globally.

- **Cross-Dock Facility (CDF)**: The primary role of the CDF is to receive, sort, and process products delivered by the consolidator; it does not operate as a storage facility or hold inventory overnight. The CDF functions exclusively as a transitional point near the cross-dock for inbound shipments, effectively serving as a pass-through facility. Importantly, the CDF is not always used for incoming inventory sortation, as many POs bypass the CDF and get directly shipped from the consolidator to regional or central distribution centers.

- **Regional Service Center (RDC)**: The RDC serves as a regional warehouse and manages the distribution of products within its designated region. The

RDC receives inventory from direct-ship POs, the CDF, or the CDC. Unlike the CDF, RDCs are designed for storage and play a key role in regional order fulfillment to stores and customers.

- **Central Distribution Center (CDC)**: The CDC is the central hub of the Fleetform supply chain, overseeing the storage and distribution of products to retail outlets, e-commerce channels, and RDCs. The CDC is crucial for maintaining inventory levels and ensuring the timely availability of products across the North American distribution network. The CDC has a significantly larger storage capacity compared to RDCs.

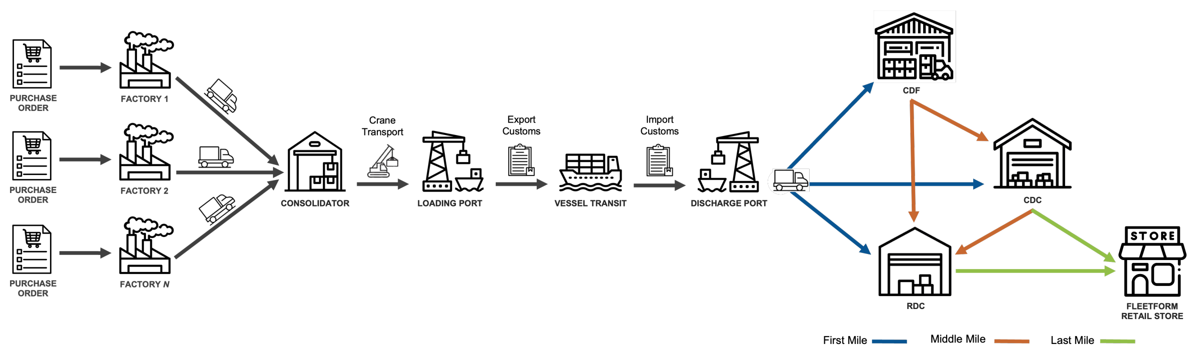


Figure 1-1: High-Level Schematic of Fleetform Distribution Network

Once a factory manufactures goods in accordance with the PO, the goods are transferred to a 3PL consolidator. The consolidator loads the goods into a container with other POs, after which the container is transferred to a vessel carrier for ocean shipment. Once the goods reach the destination, they are discharged and routed to the CDF, CDC, or RDC. The Fleetform network spans over 500 global factories, several consolidators, and 15 North American destination port locations. The route from a consolidator to destination port is termed "shipment lane." Given the multitude of locations of consolidators and desination ports, the Fleetform North America supply chain supports more than 450 distinct shipment lanes. Within this work, our analysis focuses on a single representative shipment lane from a Vietnam Consolidator to a California CDC. This particular shipment lane was selected as it represents one of the highest volume shipment lanes used in Fleetform's network and exemplifies many of the operational complexities encountered across the supply chain, providing a comprehensive case study for examining container loading optimization strategies.

1.3.2 Operational Complexities

CLP implementation at Fleetform is shaped by three principal challenges: inventory seasonality, scheduling uncertainties, and limited load planning process oversight.

Inventory Seasonality

The seasonal nature of demand and product offerings within the footwear and apparel industry pose a significant challenge in supply chain operations. Figure 1-2 illustrates the fluctuations in incoming inventory volume observed in a Vietnam-to-California shipment lane. The inventory flow exhibits a jagged profile, with notable peaks occurring approximately four weeks prior to the start of a new apparel season (Spring, Summer, Winter, Fall). Seasonal demand peaks can strain DC storage capacity, leading to overcrowding, inefficient use of space, and increased costs for temporary or overflow storage solutions. In many cases, fluctuations in inventory flow require DCs to frequently adjust labor levels to handle receiving, sorting, and dispatching goods.

Scheduling Uncertainties

In container operations, the variability in vessel transit times (VTT) and demand patterns introduces scheduling uncertainties. For example, the assigned VTT from a Vietnam Consolidator to a California DC can range from 15 to 45 days, making it difficult to accurately predict inventory arrivals and potentially disrupting downstream planning and resource allocation. As shown in Figure 1-3, which depicts deviations in delivery timing for one shipment lane, there is a significant proportion of POs that arrive at the DC well ahead of their required delivery date. Within this shipping lane, approximately 20% of POs arrive more than 60 days early, while nearly 45% arrive at least 30 days early, leading to excess inventory accumulation and increased holding costs. These premature deliveries can also contribute to DC congestion, requiring

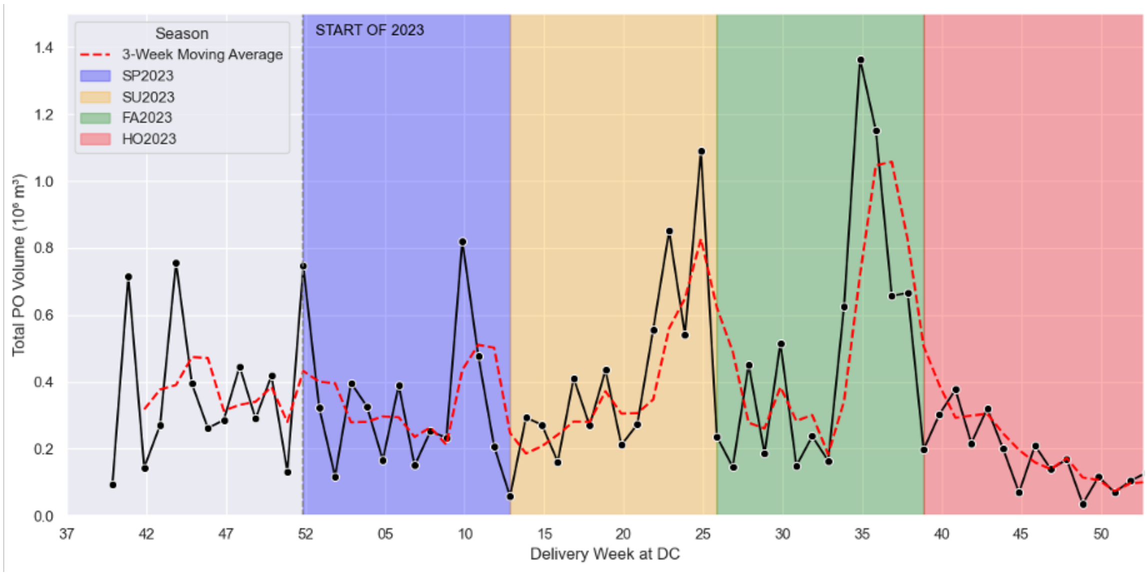


Figure 1-2: Aggregate PO Volume Delivery to DC in 2023

Note: Data reflects PO transactions from one shipment lane (from a Vietnam Consolidator to a California CDC). SU2023, SP2023, FA2023, and HO2023 denote the Summer, Spring, Fall, and Holiday product seasons of 2023, respectively.

additional labor and resources to process inventory that is not imminently needed. Conversely, roughly 15% of POs arrive 30 days or more past their required delivery date, indicating the risk of potential stockouts and unfulfilled orders. The distribution of delivery deviations suggests that early deliveries are a more prevalent issue than late deliveries, underscoring the challenge of aligning shipment schedules with demand signals. While late deliveries exist, the primary inefficiency stems from excessive early arrivals rather than widespread delays. This misalignment highlights the critical need for dynamic load planning tools and inventory inflow visibility to better synchronize deliveries with downstream inventory requirements, minimizing both early inventory buildup and late order risks.

Limited Load Planning Process Oversight

Container logistics and load planning are managed by 3PL consolidators, who are required to meet baseline container utilization metrics and comply with certain load planning guidelines provided by Fleetform (explained in greater detail in Chapter 2). As a result, Fleetform does not maintain full control over what combination of

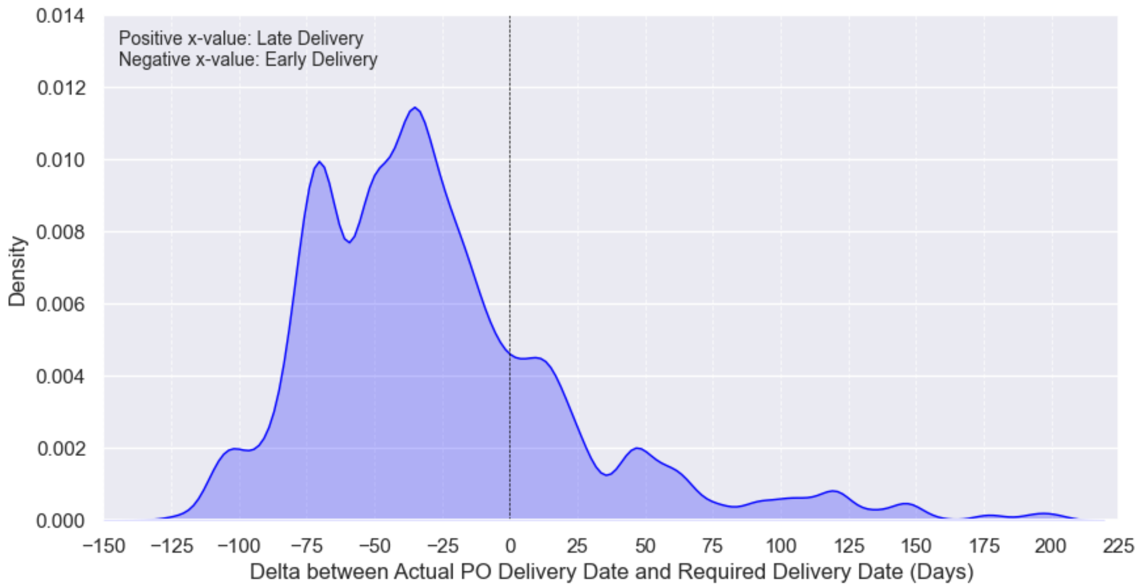


Figure 1-3: Delivery Timing: Historical Deviation from Required Delivery Date
Note: Data reflects PO delivery timing from one shipment lane (from a Vietnam Consolidator to a California CDC).

products are packed into a container, contributing to two challenges: lack of process visibility and suboptimal product mix. The loading logic used by consolidators largely relies on first in, first out (FIFO) sequencing, although load plans are often manually updated at request of Fleetform to accomodate urgent shipments. In so, Fleetform has the opportunity to have a more prescriptive hand in load planning through a model that can dynamically calculate the business impact of certain load planning decisions. Due to consolidator-managed load plans, POs are often placed into containers in lump-sum without consideration for optimal product mix. Specifically for Fleetform, 25% of the ocean containers routed to the California DC in 2023 contained just a single PO. As a result, the DC effectively receives 13 weeks (or a full season) of supply for one (SKU) in a single shipment. This single-product lump-sum delivery paradigm strains storage capacities, exacerbates early inventory costs, and front-loads inventory flow. To level-load the inventory flow, the product mix within each container could be optimized with considerations for SKU diversity.

1.4 Problem Motivation and Approach

This research aims to develop an optimization model for CLP that examines the trade-offs between prioritizing delivery precision, container utilization, and transport and storage costs to better inform load planning decisions. The goal of the model is to provide a decision support tool for logistics practitioners that enables them to conduct dynamic simulations to assess the impact of parameter changes (such as load rules, product mixes, shipment timelines, and other pertinent variables) on key business metrics. Through these simulations, the tool will provide emulation and analysis of inventory flow profiles to enhance operational visibility and facilitate more informed trade-off decisions across cost, service, and efficiency objectives.

Contributions

CLP, which explores how objects of different sizes and volumes are arranged to maximize the utilization of containers, is widely researched. However, many of the existing optimization frameworks focus solely on maximizing container utilization, which often overlooks critical supply chain factors, including inventory flow, warehousing constraints, and delivery precision. For example, achieving 100% container utilization is not always optimal if a receiving warehouse lacks the capacity to house or process incoming inventory, or if the inventory incurs holding costs due to being idle for several weeks. In such scenarios, a lower container utilization (e.g., 85%) may represent the "optimal" solution for a given supply chain when accounting for downstream nodes. The optimization model presented in this work is the first that jointly considers container utilization, delivery precision, and transport and storage costs, with an emphasis on improving the accuracy of delivery timing and smoothing inventory flow. This work also explores strategic process and policy improvements through case studies on consolidation bypass, PO partitioning, and load rule impact assessments.

Thesis Organization

This thesis is organized into six chapters. Following the introductory chapter, Chapter 2 discusses the supply chain context and operational network underpinning this research, offering an overview of Fleetform's logistics architecture and relevant processes. Chapter 3 reviews literature on CLP optimization methods and freight consolidation. Chapter 4 outlines the methods and mathematical formulation of the optimization model, providing details on model implementation, data construction, and key assumptions. Chapter 5 presents experimental results, scenario-based analyses, and operational insights. Lastly, Chapter 6 explores two strategic changes, workflow improvement recommendations, and research extension opportunities.

Note on Company Proprietary Information

To protect information that is proprietary to the corporate host, the data presented throughout this thesis may have been modified and may not represent actual values. Data labels may have been altered, converted, or removed to protect competitive information, while still conveying the findings of this work.

Chapter 2

Container Logistics and Process Flow

Container logistics, which oversees the transportation, handling, and storage of goods using standardized containers, is a critical element of global supply chain management. Containers facilitate the efficient movement of goods across various modes of transport, such as ocean freight, rail, and road transport, and play a vital role in connecting production facilities with distribution centers and end markets. This section examines the logistical context and CLP process at Fleetform, identifying operational challenges that inform the development of the optimization model and workflow design recommendations presented later in this work.

2.1 Types of Containers

Container dimensions and classifications are largely standardized in the ocean freight industry, governed by ISO regulations such as ISO 668. Among general-purpose containers, four types are most common: the 20-foot Standard Dry (20 SD), 40-foot Standard Dry (40 SD), 40-foot High Cube (40 HC), and 45-foot High Cube (45 HC). Standard Dry containers are widely used for transporting consumer goods, machinery, and textiles. High Cube containers, which offer additional vertical clearance, provide greater internal volume and are especially suited for bulky yet lightweight items such as apparel and electronics. It is important to note that containers cannot be packed to their full geometric capacity. The usable or loadable volume—typically 70% to 80% of

Container Type	Loadable Capacity (cbm)	Normalized Cost Factor Per Unit Capacity (baselined to 40 HC)
20 DRY 20' L x 8.5' H x 8' W	28	1.56
40 DRY 40' L x 8.5' H x 8' W	56	1.15
45 HIGH 45' L x 9.5' H x 8' W	73	1.11
40 HIGH 40' L x 9.5' H x 8' W	66	1.00

Figure 2-1: Container Types and Relative Unit Volume Cost

the total stated capacity—represents the amount of space within a shipping container that can be allocated for cargo, taking into account factors such as the thickness of the container walls, door frame dimensions, structural reinforcements, and any internal fixtures that reduce the available loading space within a container.

Like many organizations, Fleetform collaborates with consolidators and shipping carriers to define standardized costs for each container type. Although container pricing can fluctuate based on factors such as vessel availability, seasonal demand, and shipment lane characteristics, an average cost per unit volume can be derived by comparing typical container costs with their usable capacity. While specific cost figures are omitted in this work for confidentiality, the 40 HC container is most commonly used by Fleetform as it maintains the lowest cost per unit volume ratio. Other container types are more costly on a per-unit basis, so they are generally reserved for urgent shipments or periods of constrained carrier capacity. Figure 2-1 summarizes container types, usable volumes, and relative unit costs normalized to the 40 HC benchmark.

2.2 Load Rules and Load Plans

In container logistics, load rules define the set of operational, regulatory, and business-driven constraints that govern how goods are assigned and arranged within shipping containers. These rules serve to ensure the safe handling of cargo during transit while aligning shipments with downstream delivery requirements and operational objectives. Load rules typically address factors such as weight distribution, product compatibility, co-loading restrictions, and volume thresholds.

At Fleetform, load rules are established in collaboration with shipping carriers, 3PL consolidators, and retail stakeholders. For example, co-loading restrictions may prevent certain products (e.g., apparel and footwear) from being shipped together in the same container due to differences in handling, labeling, or customs classification. Similarly, volume-based constraints may require that containers meet a minimum utilization threshold before shipping. Time-sensitive or high-priority shipments may also be subject to expedited routing rules to guarantee timely delivery.

While load rules play a vital role in ensuring safety, compliance, and operational consistency, excessively prescriptive load rules can inadvertently introduce bottlenecks and reduce overall logistics efficiency. For example, strict enforcement of single-PO containers may result in containers being dispatched significantly below utilization capacity, thereby increasing per-unit transportation costs. Moreover, legacy load rules—often inherited from prior operational practices or longstanding customer agreements—frequently persist within logistics workflows, even when they no longer serve current strategic objectives and decrease adaptability to evolving supply chain demands. For instance, a legacy customer requirement to avoid mixing products from different factories (originally established to simplify invoice reconciliation) may impede CLP, despite improvements in digital traceability systems. Additionally, the application of load rules often necessitates substantial manual oversight, particularly when enterprise resource planning (ERP) systems lack the capability to dynamically resolve rule conflicts or accommodate exception handling. Within Fleetform's current workflow, ensuring load rule compliance with the consolidator commonly requires involvement from the internal transportation team to override default plans, reconcile conflicting rules, or approve exceptions. This manual intervention not only introduces administrative burden and elevates the risk of processing errors, but also contributes to delays in the load plan approval cycle and shipment readiness. Ultimately, the

CONTAINER LOADING PLAN / CFS OUTBOUND RECEIPT

Booking Number: [illegible] | [illegible] | **CLP No.:** [illegible]
Vessel Name: [illegible] | [illegible] | **POD:**
Container number: [illegible] | [illegible] [illegible] | **Consignee:** [illegible]
Seal number: [illegible] | [illegible] | **Loading date:** [illegible]
Cont size: [illegible] | [illegible]

Trucker:

No:	Shipping Order Ref.	[illegible]	Shipper	Division	SP#	SO #	POLine #	CTN	CBM	Gross Weight	Remark
1		[illegible]	LU	Footwear	[illegible]	[illegible]	[illegible]	454	30.95	3052.71	
2		[illegible]	LU	Footwear	[illegible]	[illegible]	[illegible]	71	4.64	465.82	
3		[illegible]	LU	Footwear	[illegible]	[illegible]	[illegible]	61	3.84	344.04	
4		[illegible]	LU	Footwear	[illegible]	[illegible]	[illegible]	51	1.81	229.78	
5		[illegible]	SHK	Apparel	[illegible]	[illegible]	[illegible]	6	0.37	40.56	
6		[illegible]	SHK	Apparel	[illegible]	[illegible]	[illegible]	9	0.45	54.28	

Figure 2-2: Example of a Container Load Plan

effectiveness of load rules depends on calibration, contextual flexibility, and automated system integration to help move from static rule sets toward adaptive load planning.

A load plan is a document that specifies the PO-level contents of each container shipment. Each load plan includes key shipment data such as the shipper (origin factory), product category (division), PO and sales order references, and the cubic volume of each PO. An example load plan is illustrated in Figure 2-2 (with select details redacted for confidentiality). Load plans are generated by 3PL consolidators, who prepare shipment configurations based on available inventory at their staging docks. The default planning logic applied by consolidators follows a (FIFO) sequence, wherein POs are loaded into containers based on their arrival time at the consolidation facility. This approach overlooks key downstream considerations such as delivery urgency, inventory needs, and destination storage constraints, often necessitating manual intervention and iterative communication loops with the Fleetform team for adjustments (discussed further in Section 2.3). Effectively, the load plan plays a central role in facilitating supply chain execution between suppliers, consolidators, carriers, and Fleetform logistics teams by documenting the full shipment composition and serving as the basis for customs declarations and downstream receiving activities.

2.3 Load Planning Process Flow

Key Players and Responsibilities

Container logistics at Fleetform involves a multi-stakeholder framework, where coordination across parties is essential to the successful execution of containerized shipments. The following roles underpin the transportation management process:

- **Fleetform Transportation Team**: This group within Fleetform works with consolidators to review and approve load plans, align on shipment timelines, communicate weekly shipment priorities based on business needs (e.g., product launches, backorders), and initiate duty processes prior to shipment arrival. The team also coordinates with carriers and consolidators to resolve booking issues or reroute cargo as needed.

- **Factories**: A goods-at-consolidator date is assigned to every PO based on its required delivery date. Factories use the goods-at-consolidator date to plan production and arrange shipment to the consolidator. For this analysis, the goods-at-consolidator date is converted to a week number, referred to as the goods-at-consolidator week (GCW).

- **Shipping Carriers**: Carriers are responsible for securing vessel space, ensuring regulatory compliance (including labeling and documentation), and coordinating the physical movement of containers from origin ports to destinations.

- **Consolidators**: 3PL consolidators aggregate goods from multiple factories, apply load rules, and generate load plans. They manage container staging, enforce load rules, and serve as intermediary between Fleetform transportation teams and shipping carriers.

- **Customs Brokers**: Brokers ensure compliance with international trade regulations, submit export and import documentation, and help prevent clearance delays at both ends of the shipment process.

- **Fleetform DC**: DCs are the final destination in the workflow scope of this research. They are responsible for receiving, unloading, and storing incoming inventory. The DC team verifies goods against shipping documentation and coordinates with middle- and last-mile logistics teams for the redistribution and rerouting of goods.

Fleetform CLP Workflow

The end-to-end transportation process from factory to destination port, as depicted in Figure 2-3, is structured across five process stages, involving multiple handoffs between factories, consolidators, carriers, brokers, and internal Fleetform teams. The workflow begins when the factory submits a booking request to the consolidator approximately 10 days before the PO is scheduled to arrive. Preliminary load plans are generated by the consolidator based on available and incoming PO inventory using (FIFO logic, factoring in loading rules provided by Fleetform and carrier capacity constraints. These load plans are then shared with the Fleetform transportation team via email for review and approval.

Upon receiving the load plan, the Fleetform transportation team reviews the proposed schedule, consulting internal departments for shipment alignment. The load plan approval process typically requires multiple iterations of back-and-forth communication to accommodate changes in product urgency, resolve PO inconsistencies, or address exceptions. Once the load plan is approved by Fleetform, the consolidator books container space with a designated carrier. If the carrier approves the booking, the factory is notified by the consolidator and will arrange trucking to transport goods to the consolidation site. Upon arriving at the consolidation site, the POs are subject to a limited free storage window, beyond which Fleetform incurs storage charges per unit volume. The consolidator will pack the POs into a container in accordance with the load plan. The consolidator also handles outbound documentation, including export declarations and destination shipping instructions. If the carrier denies the booking, Fleetform gets notified and initiates alternative arrangements, potentially engaging with spot market carriers.

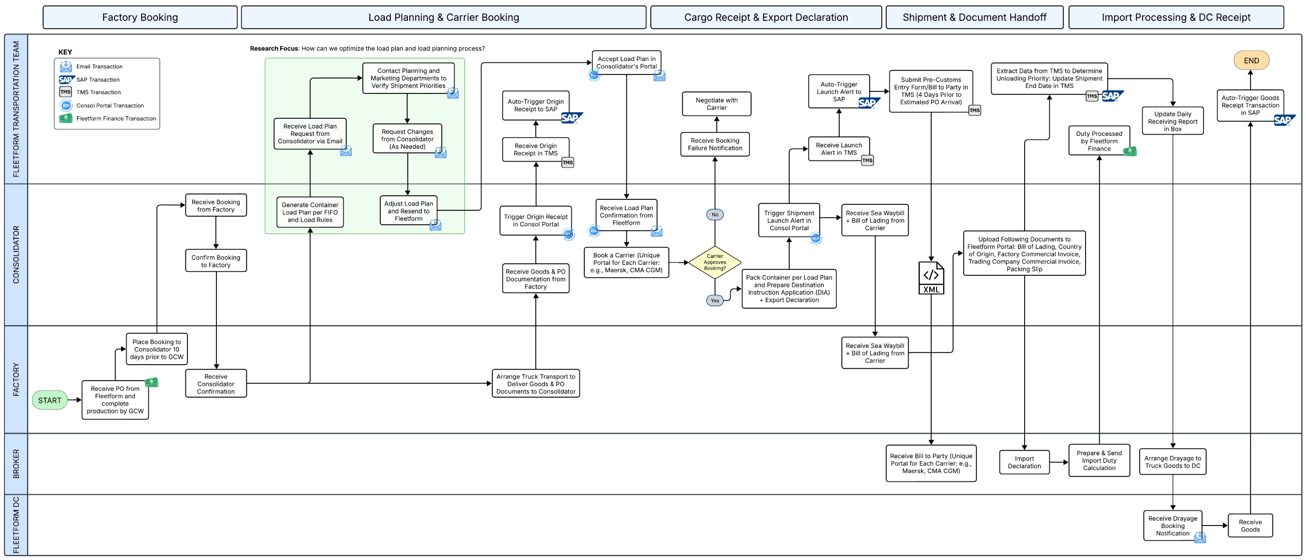


Figure 2-3: Process Flow Transactions for Fleetform Container Logistics

Once packing is complete, the carrier loads the container onto the vessel and issues a bill of lading. Approximately four days before the vessel is expected to reach the destination port, the Fleetform transportation team submits a pre-customs entry form through its broker for import declaration. After the vessel docks at the destination port, the containers are offloaded by the carrier and moved to the container yard, where they remain under a limited free demurrage period. The Fleetform DC team then coordinates inland transportation from the port, receives and inspects the shipment, verifies PO-level accuracy, and places the inventory into storage.

Bottlenecks and Process Limitations

The current CLP workflow at Fleetform reveals several operational inefficiencies stemming from decentralized decision-making, manual coordination loops, and the absence of advanced optimization logic. As illustrated in Figure 2-3, the end-to-end process comprises numerous handoffs among internal teams, external partners, and third-party consolidators, yet lacks the system integration and prescriptive logic needed to ensure consistent and efficient outcomes.

A core limitation is that load plans are generated by consolidators using static heuristics—primarily FIFO logic and proximity-based inventory pooling—rather than optimization-based strategies. While load rules enforce basic utilization thresholds and product compatibility constraints, consolidators are not incentivized to consider broader supply chain objectives for Fleetform, such as delivery urgency, inventory balance at destination, warehouse capacity, or inventory holding costs. This misalignment introduces inefficiencies in CLP and increases the likelihood of early deliveries and reduced operational recourse for Fleetform.

The most prominent bottleneck lies in the manual, iterative coordination loops between consolidators and the Fleetform transportation team. Load plans are submitted to Fleetform via email and document-sharing platforms (e.g., Box, SAP) for review and approval. This process often involves multiple back-and-forth communications to adjust for shifting priorities, correct discrepancies, or rebook space in the event of capacity constraints. These asynchronous feedback loops contribute to latency and make the workflow particularly vulnerable to disruptions such as factory delays, vessel rollovers, or market-driven volume shifts. System-level integration is also notably absent across key milestones in the CLP process. As an example, booking confirmation, shipment prioritization, and customs submission each require manual intervention and linear task execution. This lack of parallel processing introduces unnecessary dependencies and reduces the system's ability to adapt to changes in business needs or carrier availability. Furthermore, the seasonal nature of demand for apparel and footwear leaves limited flexibility to absorb disruptions.

Lastly, the current framework provides limited control over how (POs) are grouped within containers. In the absence of prescriptive planning tools, container composition is dictated largely by consolidator convenience, rather than strategic criteria. Without direct control over container composition, Fleetform faces increased risk of early deliveries, fragmented PO shipments, and suboptimal inventory allocation. Seasonal demand peaks further amplify these risks, as the rigidity, complexity, and slow responsiveness of the existing workflow reduce Fleetform's ability to absorb sudden shifts in supply or demand conditions. More broadly, the current CLP workflow reflects a system governed by decentralized decision-making and a heavy reliance on manual coordination processes. Collectively, these limitations underscore the need for a centralized, optimization-enabled CLP approach that aligns upstream decisions with downstream operational goals.

2.4 Optimization Opportunities

Container logistics is integral to global commerce, facilitating the efficient and timely movement of goods across complex supply chains. To address existing bottlenecks and inefficiencies, organizations like Fleetform can adopt targeted optimization strategies that enhance supply chain resilience and cost efficiency. The following areas represent significant opportunities for optimization:

- **Load Plan Emulator**: The development of a load plan emulator and optimization model can help determine the optimal combination of goods in each container along with the shipment schedule, enabling organizations to systematically customize load plans and dynamically analyze various loading scenarios. By implementing this approach, organizations can improve delivery precision and reduce transportation costs, thus improving logistical efficiency.

- **Network Predictability**: There is opportunity for organizations to develop more precise predictions of inventory arrival at downstream warehousing or retail nodes. This can help facilitate the improved synchronization of inventory flow with production schedules, product demand, and transportation management to reduce carrying costs and minimize the risk of early inventory and/or excess inventory. By understanding the impact of upstream transportation decisions on downstream nodes, organizations can optimize resource allocation and develop a more responsive framework for inventory management.

- **Workflow Simplification**: A more prescriptive approach to load planning allows organizations like Fleetform to reduce reliance on 3PL consolidators and minimize the need for recursive approval cycles, which can accelerate decision-making. By gaining insight into the business impact of different loading strategies and load configurations, organizations can be better equipped with tools for scenario analyses and informed load planning decisions.

2.5 Chapter Summary

Chapter 2 provides an overview of key operational elements in container logistics, including container selection, cargo consolidation, and load planning rules. It outlines how these factors shape freight movement and cost efficiency in global supply chains. The chapter then details Fleetform's end-to-end logistics workflow from factory shipment to destination delivery, highlighting how current practices, such as manual load planning and default FIFO logic, contribute to inefficiencies. This context establishes the foundation for identifying optimization opportunities in container load planning and workflow management.

Chapter 3

Literature Review

This literature review surveys the key methodologies and advancements in freight consolidation and load planning, with a focus on multi-objective optimization (MOO) frameworks. Section 3.1 examines traditional and emerging modeling approaches for load plan optimization, including heuristics, exact algorithms, and MOO techniques, highlighting their respective strengths, limitations, and utility in real-world logistics problems. Section 3.2 explores the operational dimensions of freight consolidation, with particular attention to coordination challenges, policy design, and cost-performance trade-offs in (3PL) logistics environments.

3.1 Classical Approaches to Container Load Planning

This chapter examines the most prevalent CLP approaches: heuristic, metaheuristic, mixed integer linear programming (MILP), and multi-objective optimization (MOO). Each approach is analyzed for its advantages and limitations, offering insight into the applications in academic research and industry practice. Table 3.1 provides a summary of these approaches.

Aspect	Heuristics	Metaheuristics	MILP	MOO
Objective	Find near-optimal solutions quickly	Find near-optimal solutions through guided search	Optimize a single objective function with integer and continuous variables	Optimize multiple competing objectives
Solution Space	Single or multiple near-optimal solutions	Multiple guided search solutions	Single optimal solution (integer and continuous)	Set of Pareto-optimal solutions
Flexibility	Highly flexible with domain knowledge	Highly adaptable for various problem structures	Suitable for well-defined problems with linear structure	Flexible for handling multiple conflicting objectives
Scalability	Good for large-scale problems	Scales well with appropriate tuning	Can handle large problems but may be slow with many constraints	Handles multiple objectives but can be computationally intensive
Optimality	No guarantee of optimality	No guarantee, but often close to optimal	Guarantees global optimality (if solvable)	Provides trade-off (non-dominated) solutions
Key Strength	Fast computation for large instances	Balancing exploration and exploitation	Optimal solutions with strong guarantees	Insights into trade-offs among competing objectives
Key Weakness	Prone to local optima; sensitive to initial conditions	May suffer from convergence issues and slow efficiency in high-dimensional spaces, requires careful tuning	Can become computationally expensive	Requires expert interpretation of Pareto set; can be computationally intensive
Leading Algorithmic Approaches	Greedy heuristics, constructive algorithms, domain-specific rule sets	Genetic Algorithms, Simulated Annealing, Tabu Search, Ant Colony Optimization	Branch-and-Bound, Cutting Plane Methods, Linear Programming Relaxation	NSGA-II, SPEA2, Multi-Objective Evolutionary Algorithms

Table 3.1: Comparison of Optimization Approaches for Container Loading Problems.

Heuristic Approaches

Heuristic methods constitute a class of problem-solving methodologies that employ approximations and rule-based logic to provide approximate solutions to CLPs. These approaches are computationally efficient compared to exact optimization methods such as MILP, but they do not guarantee globally optimal solutions [37]. Instead, heuristics generate practically feasible load configurations by prioritizing factors such as space utilization, stability, and ease of implementation. Heuristic strategies for CLP can be broadly categorized into greedy algorithms, wall-building techniques, and rule-based approaches:

- **Greedy Algorithms**: Greedy heuristics construct load plans by sequentially selecting and placing the most suitable box based on predefined criteria. Common variants include:
  - First Fit (FF): Items are assigned to the first available location within the container that can accommodate them. This method is computationally efficient but often results in suboptimal packing due to its inability to reconsider earlier placements.
  - Best Fit (BF): Items are placed in the tightest available space, aiming to minimize residual gaps within the container. While this reduces fragmentation, it may lead to poor weight distribution or instability.
  - Largest First / Smallest First: Items are sorted in either decreasing or increasing order based on dimensions or volume before placement. Sorting by largest first often improves stability by placing heavy items at the bottom, whereas smallest first may optimize residual space utilization.
  - Largest Area First-Fit (LAFF): Introduced by Patil and Patil [35], this heuristic prioritizes items with the largest base area, ensuring heavier items are placed first to maintain load stability.
- **Wall-Building**: Introduced by George and Robinson (1980), these algorithms fill a container by packing boxes into vertical "walls" or layers [21]. Each wall spans the height and width of a container, and subsequent walls are placed sequentially along the depth. Variants of wall-building, incorporating different box-ordering strategies, layer dimensions, or weight constraints, have served as the foundation for many subsequent algorithms. Wall-building methods

are highly efficient for structured cargo types (e.g., palletized goods), but they struggle with irregularly shaped items or multi-drop deliveries, where accessibility and flexibility are crucial.

- **Rule-Based Approaches**: Rule-based heuristics incorporate domain knowledge and practical constraints to create feasible and stable load plans. These heuristics mimic human decision-making by applying a set of predefined packing rules (e.g., "place heavier boxes at bottom"). Common rules include those of weight distribution (heavier items should be positioned at the bottom to enhance load stability), fragility constraints (fragile items should not be placed underneath heavier objects or in unstable positions), and multi-drop sequencing (items for earlier deliveries should be placed near the container opening to minimize rehandling). Bischoff and Ratcliff (1995) identified several practical loading constraints (orientation limits, weight distribution, stability requirements, complete-shipment needs, etc.) that are widely used in industry and serve as a checklist for developing robust CLP algorithms [4].

Heuristic approaches offer several distinct advantages for container load planning, including computational efficiency, ease of implementation, and domain specificity. Due to straightforward and rule-based designs, heuristic algorithms can generate feasible load configurations rapidly, making them particularly suitable for real-time applications and large-scale logistics scenarios. Additionally, heuristic methods typically require minimal parameter tuning, simplifying its adoption and deployment in practical contexts. Another key strength is its ability to readily incorporate industry-specific constraints and best practices, such as minimizing cargo shifting or accommodating particular handling requirements. Nonetheless, heuristic methods exhibit notable limitations. Given that heuristic solutions rely on locally efficient choices, a primary shortcoming is the inherent lack of global optimization. Furthermore, rule-based heuristics often depend on the scenario, performing effectively under certain conditions but failing to generalize adequately across diverse logistics settings or changing cargo characteristics. Given these limitations, heuristic methods are often combined with optimization techniques (e.g., metaheuristics or MILP-based refinements) to enhance solution quality while maintaining computational feasibility.

Metaheuristic Approaches

Metaheuristics apply higher-level iterative search strategies (inspired by phenomena such as evolution, physical annealing, or animal behaviors) to explore a broader solution space compared to single-shot heuristics. In CLP, popular metaheuristics include Genetic Algorithms, Particle Swarm Optimization, Simulated Annealing, and Tabu Search. Metaheuristic approaches start from one or more initial loading configurations (often generated by a heuristic) and then iteratively improve the solution through randomized transformations, selection, or memory-based moves. They are particularly useful for large-scale instances of the problem.

- **Genetic Algorithms (GA)**: GAs encode a loading pattern as a "chromosome" (e.g., a sequence of packing orders or positions) and apply selection, crossover, and mutation operators to evolve better solutions. The application of GAs to CLP was pioneered by Gehring and Bortfeldt in 1997, who employed a genetic approach based on a column-building packing order [20]. Since then, many GA variants have been developed. For instance, Gonçalves et al. (2011) designed a biased random-key GA with multiple populations to improve container volume utilization [24]. The strength of GAs lies in its ability to combine beneficial features from different solutions through crossover, leading to improved load planning. However, GAs require meticulous encoding of cargo packing configurations and can be computationally intensive for large-scale instances.

- **Particle Swarm Optimization (PSO)**: PSO is a swarm intelligence technique where a set of candidate solutions (particles) "fly" through the search space, guided by their own and their neighbors' best-known positions. In CLP, PSO has been used to optimize packing sequences and box orientations to minimize unused space. Each particle might represent an ordered list of box placements while randomized velocities assigned to each particle drive changes in ordering or positioning. Tlili et al. (2013) [43], along with Zhou and Liu (2017) [50], have demonstrated the collective particle learning thourgh PSO helps generate arrangements with improved container fill rate. Although PSO can converge faster than some metaheuristics, there is a risk of converging prematurely to a mediocre layout if diversity in the swarm is not maintained.

- **Simulated Annealing (SA)**: Inspired by the metallurgical annealing process, SA is a probabilistic technique that finds approximate solutions through gradual "cooling," which represents a slow decrease in the probability of accepting worse solutions as the solution space is explored. The algorithm works by progressively lowering the "temperature" of each solution from an initial positive value to zero. At each step, it randomly selects a nearby solution, evaluates its quality, and decides whether to move to it based on temperature-dependent probabilities. Better solutions are always accepted, while the probability of accepting worse solutions decreases over time, helping SA escape local optima and refine its search. In CLP, SA is often used to perturb a packing order or the position of one box or cargo item to evaluate the change on load plan. Egeblad and Pisinger (2008) [16] created an SA-based algorithm for 3D packing that iteratively improves a sequence of placements, achieving high utilization for knapsack problems. Yamazaki et al. (2000) [48] combined SA with greedy algorithms to guide the selection of different packing sequence heuristics. The SA approach excels at avoiding local optima. However, SA is highly sensitive to parameter tuning and cooling schedules, and can be computationally intensive due to slow convergence.

- **Tabu Search (TS)**: TS is an iterative improvement method that utilizes memory; it operates by moving to the best neighboring solution, even if it is worse than the current one, while using a tabu list to prevent or penalize recent moves from being reversed, thus encouraging exploration. This method uses memory structures to avoid revisiting previously explored solutions and to guide the search for new solutions in promising regions of the solution space. In CLP, typical moves include rearranging sets of boxes or swapping packing sub-blocks. Bortfeldt and Gehring (1998) laid the foundational groundwork for applying TS to CLP [6]. In 2003, Bortfeldt et al. applied TS to a single-container loading problem, where multiple search threads ran simultaneously with different starting solutions and shared information, leading to improved results on benchmark instances [7]. TS is particularly effective at intensifying the search within promising areas of the solution space while avoiding cycles, and is often hybridized with GA and SA methods. However, TS is computationally intensive, as it requires evaluating numerous potential solutions and maintaining a tabu list to avoid revisiting previously explored solutions. Additionally, TS can potentially reduce diversity in the solution set if the algorithm focuses on intensifying the search in only select promising regions and not adequately exploring other parts of the solution space.

Metaheuristic approaches are capable of exploring large solution spaces and overcoming the limitations of traditional heuristic methods. These algorithms iteratively refine initial solutions, effectively escaping local optima by diversifying its search strategies and occasionally accepting suboptimal intermediate solutions. Despite its robustness and flexibility in handling complex constraints, metaheuristics require careful parameter tuning and might be computationally demanding for large-scale instances, potentially limiting its scalability.

Mixed-Integer Linear Programming (MILP) Approaches

Mixed-Integer Linear Programming (MILP) is an optimization framework widely used for CLP modeling, particularly when precise control over positioning, orientation, and spatial constraints is required. MILP models combine linear programming with integrality constraints, making it suitable for modeling the discrete nature of packing decisions (e.g., box placements, orientations, sequencing) alongside continuous elements like weight distribution or cost minimization. In a standard MILP formulation for CLP, binary variables may represent the presence or absence of an item at a specific coordinate or relative ordering (e.g., item i placed in front of item j along a given axis to ensure non-overlap), while continuous variables capture spatial positioning or container dimensions.

MILP models for CLP have historically suffered from scalability issues due to the exponential growth in variables and constraints as the number of items increases. However, significant progress has been made in recent years. A significant advancement in CLP modeling was made by Kurpel et al. (2020), who developed new MILP formulations for various container loading problem variants. Their approach systematically enumerates feasible box placements based on discrete position-orientation pairs, while applying symmetry-breaking techniques and bounding strategies to significantly reduce the model's size and computational complexity [28]. These enhancements enabled exact or near-exact solutions for benchmark instances that were previously unsolvable due to computational limits, thus expanding the reach of exact optimization in CLP research. Another example is the work of Junqueira et al. (2012), who incorporated realistic stability and load-bearing constraints into a three-dimensional MILP container loading model [26]. Their formulation enforced support conditions under each item to guarantee physical feasibility, demonstrating how MILP can effectively encode complex loading rules for practical applications. MILP is also instrumental in hybrid or decomposed models. For instance, Eley (2005) introduced a two-phase strategy that first generates feasible packing patterns for individual containers, capturing key spatial and stability constraints, and then uses a MILP-based set partitioning model to assign these patterns across containers while minimizing load inefficiency [17]. Such pattern-based decomposition illustrates how MILP can play a central role in larger logistics optimization systems.

The key advantage of MILP in CLP lies in its guarantee of optimality. This capability is particularly valuable in small- to medium-scale planning scenarios or in high-value shipment settings, where maximizing container usage or ensuring safe and stable packing is mission-critical. MILP also allows for flexible integration of diverse constraints, including weight balancing, stacking limitations, and even product-specific handling rules, which are often difficult to accommodate in purely heuristic models. The primary drawback of MILP is its computational intractability for large-scale CLP instances. As the number of items increases, the number of binary decision variables grows quadratically. Even moderate instances may involve thousands of variables and constraints, rendering exact solution via branch-and-bound algorithms computationally expensive or infeasible. In response, studies bound MILP approaches by restricting input parameters (e.g., assuming identical box sizes, grid-based positioning, or axis-aligned orientations) or by incorporating MILP in a hybrid architecture, where it optimizes subproblems or verifies heuristic outcomes rather than solving the entire CLP directly. Generally, MILP methods remain best suited for small to moderate-scale problems or for hybrid models in larger optimization systems. Its computational intensity and limited scalability in high-dimensional, real-time applications mean that heuristic and metaheuristic methods remain dominant in operational settings, particularly for large-scale or time-constrained logistics environments.

Multi-Objective Optimization (MOO) Approaches

Multi-Objective Optimization (MOO) offers a CLP framework to model trade-offs such that competing logistics objectives can be optimized simultaneously. Typical goals include:

- Cost minimization (e.g., minimizing the number of containers, fuel consumption, or handling costs)
- Space utilization (e.g., maximizing packed volume or weight efficiency)
- Operational reliability (e.g., minimizing load/unload time, maximizing load stability, or prioritizing certain shipments).

The central goal of MOO is to generate a Pareto front—a set of non-dominated solutions—where no objective can be improved without worsening another—providing decision-makers with a spectrum of trade-offs to enhance strategic flexibility in logistics planning. Metaheuristic algorithms, particularly those based on population search methods, are especially well-suited for MOO because they naturally generate diverse solution sets. Prominent techniques include:

- **Artificial Bee Colony (ABC) Algorithm**: Phongmoo et al. (2023) developed a Pareto-based ABC algorithm in which artificial "bees" explore the solution space by mimicking foraging behavior—using employed, onlooker, and scout bee operators—to iteratively generate and refine diverse container load plans that balance packed value and volume utilization [36].

- **Non-Dominated Sorting Genetic Algorithm II (NSGA-II)**: NSGA-II has been widely applied to container loading problems due to its ability to maintain population diversity while converging toward the Pareto front. Cao et al. (2013) implemented a multi-objective NSGA-II framework to optimize volume utilization, load stability, and spatial efficiency, incorporating customized encoding and genetic operators tailored to the geometric packing constraints [10]. Fom et al. (2024) extended this approach by integrating NSGA-II with a bottom-left-fill heuristic, enabling the joint optimization of space utilization and item value while satisfying operational constraints such as orientation, overlap prevention, and weight distribution [18].

- **Multi-Objective Biased Random-Key Genetic Algorithms (BRKGA)**: Gonçalves and Resende (2012) proposed BRKGA, which handled load stability, weight balance, and bearing constraints, demonstrating the flexibility of evolutionary algorithms in handling diverse objectives [23]. In this approach, candidate solutions are represented as vectors of randomly generated keys, which are then decoded into packing sequences; the algorithm applies a biased selection process that prioritizes elite individuals while maintaining population diversity, allowing for effective exploration and exploitation of the solution space.

- **Pareto Clustering Search (PCS)**: Araújo et al. (2016) propose a hybrid metaheuristic called PCS to minimize container re-handling and ship instability across ports [3]. PCS combines Simulated Annealing (SA) for broad exploration with clustering-based local search to intensify efforts in promising regions. Solutions are encoded using rule-based representations, which streamline the search space while maintaining feasibility. Although developed for maritime stowage, the PCS framework is extensible to general CLP contexts to assess trade-offs between spatial utilization, accessibility, and structural stability.

A core benefit of MOO in CLP is the decision support it enables. Instead of prescribing a single configuration, MOO models offer a range of viable solutions, allowing logistics managers to select a load plan that best fits current operational constraints and adapt plans dynamically as shipment priorities evolve. Jugović (2020) highlights that MOO-based decision support systems can guide planners in assessing whether small increases in cost (e.g., using one additional container) might yield substantial improvements in stability or service quality [25].

Despite its advantages, MOO introduces several notable challenges in CLP. First, the computational complexity of MOO models is considerably higher than that of single-objective formulations, as they must evaluate and maintain a diverse set of non-dominated solutions across iterative generations. This becomes particularly burdensome in large-scale or real-time applications, where solution spaces grow exponentially. Second, MOO inherently results in a solution set rather than a single optimal configuration, potentially creating ambiguity in decision-making or requiring expert judgment. Bortfeldt and Wäscher (2013) emphasize that while real-world CLP scenarios often involve multiple competing objectives such as cost, stability, and service reliability, academic research has historically focused on simplified, single-objective models [8]. Recent advancements, however, have begun to bridge this gap by developing MOO frameworks capable of supporting dynamic re-planning, online optimization, and integration with inventory planning [36, 23]. These developments aim to enhance the applicability of MOO-based CLP in complex, fast-moving logistics environments by improving adaptability and alignment with operational realities.

Limitations of Existing Classifications for Real-World Logistics

Many classical CLP frameworks exhibit limitations when applied to the complexities of real-world logistics environments. The divergence between theoretical models and practical logistics applications is primarily reflected in the following three areas:

- **Neglect of Practical Constraints**: Traditional CLP research often simplifies or omits constraints that can be crucial in real shipping scenarios. For instance, early formulations treated container loading as a pure volume-maximization (3D bin packing) problem, without enforcing cargo stability, stacking strength, or orientation rules. In reality, heavy items must be placed low, fragile items cannot be stacked under weight, and some goods cannot touch others. A comprehensive survey by Bortfeldt and Wäscher (2013) observes that many proposed algorithms "do not pay enough attention to constraints encountered in practice" [8]. As a result, a model that simply packs boxes to maximize filled volume may output impractical plans.

- **Static Planning vs. Dynamic Demand**: Most classical CLP models assume a fixed set of items and deterministic parameters, solving a one-shot packing problem. However, real-world logistics is dynamic: orders can change, items may be added or removed, and container assignments might need updating in real-time. Many classical CLP models lack mechanisms for incremental or real-time re-optimization as they are typically run offline with complete data. In practice, a loading plan might need to be recomputed on-the-fly if, hypothetically, a pallet is delayed or a new high-priority shipment arrives. Traditional models fall short because re-solving a complex MILP model from scratch when a small change occurs is computationally expensive and too slow for operational timelines. Consequently, organizations often resort to manual adjustments, which can be suboptimal. The lack of robust, real-time adaptive algorithms in classical CLP means that uncertainty (e.g. varying item volumes or weights) and last-minute changes are not well-handled by most academic models as the deterministic nature of classical frameworks is a poor fit for the stochastic, dynamic environment of logistics.

- **Single-Objective Focus vs. Multi-Priority Needs**: Classical formulations typically optimize a single objective (maximize packed volume or minimize number of containers). In contrast, real logistics decisions involve multiple criteria and shifting priorities. For instance, one shipment may prioritize cost minimization (fullest packing), while another prioritizes speed or customer service (ensuring priority items definitely ship, even if some space is wasted). Classical bin packing doesn't account for such priorities – every item is treated equally, and the goal is purely efficient packing. But in practice, managers might deliberately leave some space unused to expedite a high-value item or to comply with weight limits. Moreover, shipment priorities can change, and different stakeholders value different metrics (e.g. warehouse wants fewer loads, transport wants balanced weight, customer demands no mixing of their products). Traditional models struggle to incorporate these trade-offs. Recent works have started to address multi-objective CLP – for example, the solution by Gajda et al. balances cargo value against minimizing "unnecessary move operations" during unloading by treating the latter as a soft constraint with penalties [19]. This is a departure from classical models that would either enforce unloading order rigidly or ignore it. In general, classical MILP models can be extended with multi-objective formulations, but they become even more complex and rarely scale. Thus, many classical approaches lack flexibility to prioritize certain shipments or sacrifice utilization for other goals (safety, customer preference, etc.). This limitation means that a "optimal" solution in the mathematical sense may not be truly optimal when considering broader business priorities or service level commitments.

3.2 Freight Consolidation

Freight consolidation—also referred to as shipment pooling or cargo bundling—is a foundational strategy in third-party logistics (3PL) aimed at enhancing transportation efficiency by combining multiple smaller shipments into a full container load (FCL) [31]. This practice is particularly relevant in scenarios involving fragmented order volumes and dispersed factories, enabling logistics providers to reduce transport costs, optimize container space, and lower carbon emissions. The economic rationale stems from the fixed pricing of container shipments: regardless of the fill rate, a flat fee is often charged per container, making underutilized loads inefficient both financially and environmentally. Prominent firms, such as Wal-Mart, Bosch, and Toyota, routinely employ consolidation in their 3PL-managed networks to achieve cost-effective and streamlined distribution [31].

A 3PL often operates consolidation hubs or cross-docking centers where incoming less than container load (LCL) shipments from various sources are temporarily staged and combined into outbound FCL shipments. A 3PL receives shipments from multiple suppliers, consolidates them at an intermediary facility, and then delivers as a single full-load to the destination region before distributing to final consignees, establishing a distribution system with higher cargo density and fewer trips [31]. While the operational benefits of consolidation are well recognized and employed in practice, the academic treatment of freight consolidation is relatively limited. Much of the literature has focused on classical vehicle routing, bin packing, or facility location problems, but comparatively less attention has been directed to the operational complexities and optimization opportunities that arise in 3PL consolidation hubs. Consolidation policy frameworks generally fall into three categories [33]: 1) time-based policies, where shipments occur at fixed intervals such as a weekly dispatch; 2) quantity-based policies, where shipments are triggered once cumulative volume reaches a specified threshold; and 3) hybrid time–quantity policies, where shipments occur when either a time interval or volume threshold is met.

Çetinkaya and Bookbinder (2003) offered early models that minimize total cost by balancing inventory holding and transport savings [11]. Mutlu et al.(2010) extended this to hybrid strategies, demonstrating their superior performance under demand uncertainty [32]. These frameworks help logistics providers tailor consolidation decisions to specific operational and service-level constraints.

Advantages & Limitations of Freight Consolidation

Operational benefits of consolidation include cost savings through efficient container use, fewer shipments, reduced fuel consumption per unit, and lower emissions. One industry case study reported a 12.4% reduction in freight costs and a 1.2-day reduction in delivery time when leveraging 3PL consolidation, primarily by bypassing inefficient LCL stopovers [34]. Campbell (1990) highlights that freight consolidation capitalizes on economies of scale by exploiting nonlinear transportation cost structures, enabling shippers to reduce total logistics expenditure through fuller loads [9].

Nevertheless, consolidation also presents operational challenges. Synchronizing incoming shipments from multiple sources requires precise planning, often complicated by asynchronous order readiness, variability in supplier behavior, and communication lags. As depicted in Figure 2-3, the container load planning process involves numerous touchpoints across parties—including internal logistics teams, third-party consolidators, customs brokers, and distribution center—necessitating frequent back-and-forth communication, approval loops, and manual changes via email or phone to ensure the careful alignment of shipments. These interactions often reflect shifting shipment priorities and last-minute changes, which increase planning burden and expose the system to execution risk. The risk of "fall-down"—when expected shipments are delayed or canceled—can result in partially filled containers, rebooked space, or disrupted plans. In such fall down cases, the 3PL may have to ship a LCL load or find last-minute freight to fill the gap, which is a challenging contingency to manage.

Additionally, the need for shipment compatibility, load rule adherence (e.g., no mixing of hazardous and non-hazardous goods), and customs constraints can limit inventory pooling flexibility. In practice, many consolidators apply first-in, first-out (FIFO) logic, which does not account for delivery urgency, inventory balance at the destination, or cost trade-offs. This can lead to early deliveries, overstocking at the destination, or late arrivals that affect service levels. These challenges are magnified in environments with high SKU complexity and demand volatility. Work by Tyan et al. (2003) and Ülkü (2009) emphasizes that while consolidation can be effective under stable demand conditions, its performance diminishes without dynamic planning, especially in high-SKU, volatile environments [45, 46]. The need for prioritization strategies and multi-objective optimization in consolidation planning is increasingly recognized, yet underdeveloped in mainstream logistics models.

Freight consolidation continues to serve as both a strategic and tactical lever in modern logistics, offering substantial cost savings, improved container utilization, and reduced environmental impact. However, its effectiveness is contingent on the ability to dynamically coordinate across shipment schedules, product constraints, and evolving priorities. Despite its widespread adoption, consolidation efforts are frequently hindered by static rule sets, fragmented visibility, and the complexity of coordinating among multiple stakeholders. To fully realize the potential of consolidation, logistics systems must evolve toward more adaptive, data-driven planning frameworks that integrate real-time insights with flexible operational policies.

3.3 Chapter Summary

This chapter presents a literature review on CLP and freight consolidation, emphasizing methodological approaches and implementation challenges encountered in real-world logistics. The chapter begins by outlining classical optimization techniques—including heuristic, metaheuristic, MILP, and MOO—and evaluates their respective strengths and limitations across key dimensions such as scalability, flexibility, and optimality. While heuristic and rule-based strategies offer computational efficiency, they often lack generalizability and global optimality. Metaheuristic techniques demonstrate improved search capabilities, but require careful tuning and may suffer from convergence challenges. The chapter also highlights the limitations of traditional CLP formulations in addressing multiple priorities, uncertain demand, and trade-offs across cost, utilization, and delivery precision. Recent efforts in MOO provide a promising direction, but practical deployment remains hindered by computational complexity and the need for decision-maker input. Overall, the review underscores the need for adaptive, hybrid approaches that balance performance, interpretability, and operational relevance.

Chapter 4

Methods

This section presents the CLP optimization framework developed to balance container utilization, delivery precision, and transport and storage costs. The proposed model formulates the problem as a cost-minimization objective that optimally assigns a set of purchase orders (POs) to containers over a multi-week planning horizon. The model integrates constraints related to container capacity, inventory storage policies, and shipment scheduling, enabling a trade-off between cost efficiency and service-level performance.

4.1 Problem Formulation

The optimization model evaluates the performance of a transportation logistics system across three key dimensions:

1. **Delivery Precision** measures the temporal accuracy of shipments by tracking whether POs are delivered on their required delivery week (RDW). Early and late deliveries are penalized using associated cost parameters.

2. **Transportation Costs** represents the total cost impact of moving and storing goods within the logistics network, encompassing:

    - *Direct Transport Costs*: Fixed freight cost incurred for each container used.

- *Storage Costs*: Fees applied for holding inventory at either the origin consolidation facility or at the destination port. These costs are influenced by the volume of goods stored and the duration of storage.

3. **Container Utilization** captures volumetric efficiency by measuring the proportion of container space occupied by POs. A higher utilization rate reduces the cost penalties associated with capacity under-utilization and lowers the per-unit shipping cost of goods.

An overview of the cost-minimization objective function is provided in Figure 4-1. Decision variables govern container selection along with PO assignments and volume distribution across shipment weeks, ensuring an optimized load plan that balances transportation costs with logistics efficiency (container capacity utilization and service reliability (delivery precision).

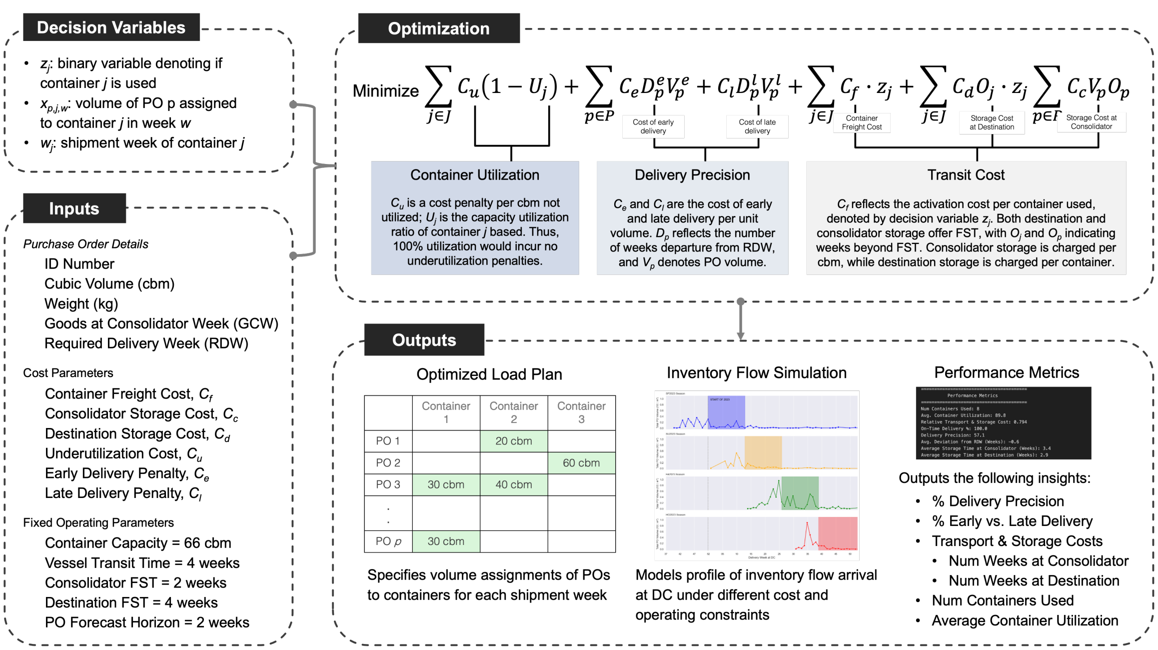


Figure 4-1: Overview of Cost Minimization Methodology for Load Plan Optimization

Dataset Construction & Processing

The dataset used in this analysis captures PO transactions for a single shipping lane from a Vietnam Consolidator to a California DC, covering all products for the 2023 footwear and apparel season. From a geographical and operational perspective, the model focuses exclusively on the Vietnam-to-California shipping lane, as it represents one of Fleetform's highest-volume trade routes and maintains a representative flow profile of inventory delivery, making it a strong candidate for modeling broader supply chain dynamics. The dataset consists of 225,880 PO line items, with each entry specifying product code, order quantities, and cubic volume. It was constructed by stitching together data from three key sources at Fleetform:

- Sales Orders (SOs): Provided the required delivery date for each PO. This was rolled into a week number to form required delivery week (RDW), which is the week on which the PO is to arrive at the destination port.
- ERP System: Contained the volume and product category of each PO.
- Transportation Management System (TMS): Supplied the goods-at-consolidator date, indicating when goods became available for shipment. This was rolled into a week number to form goods at consolidator week (GCW).

4.2 Modeling Approach and Setup

In this work, the container loading optimization problem is formulated as a multi-objective mixed-integer linear programming (MILP) model that seeks to optimize three competing objectives: maximizing container utilization, maximizing delivery precision, and minimizing transportation costs. A weighted-sum approach is employed to aggregate these objectives into a single scalar objective function, allowing a trade-off analysis between operational priorities.

4.2.1 Branch-and-Bound Implementation

The branch-and-bound algorithm is a widely used method for solving combinatorial optimization problems, particularly those involving integer programming [49]. The algorithm divides the bin-packing problem into smaller sub-problems (branching) and evaluates each using rigorous bounds (bounding), discarding infeasible or suboptimal branches (pruning). Starting from the root node containing all possible assignments, the algorithm creates child nodes by branching on each configuration of the decision variables. Each node represents a partial solution, with the tree structure naturally encoding the hierarchical nature of container loading decisions. The branch-and-bound implementation is particularly effective for CLP due to the hierarchical structure of the decisions and the natural bounds provided by the constraints. The branch-and-bound algorithm provides optimality through three mechanisms:

1. **Enumeration Control**: The algorithm systematically explores all viable container loading combinations while maintaining a comprehensive tree structure of possible solutions.

2. **Bound Management**: The algorithm tracks upper and lower bounds for each objective and updates bounds continuously as superior solutions are identified.

3. **Pruning**: The algorithm eliminates suboptimal branches through bound comparisons while exploring the solution space. This systematically eliminates large portions of the solution space that cannot yield optimal solutions, while focusing computational effort on the most promising solution paths.

The foundational work on branch-and-bound algorithms for cargo loading operations was introduced by Kolesar (1967). In figure 4-2, Kolesar demonstrates how the branch-and-bound algorithm systematically explores different combinations of items (0-7) that can be included in a knapsack problem, with the arc lengths representing item weights. The network structure shows all possible paths from the starting node to different total weights (0-133), in which unlabeled arcs have zero weight, allowing the algorithm to efficiently explore and prune suboptimal solutions while finding the

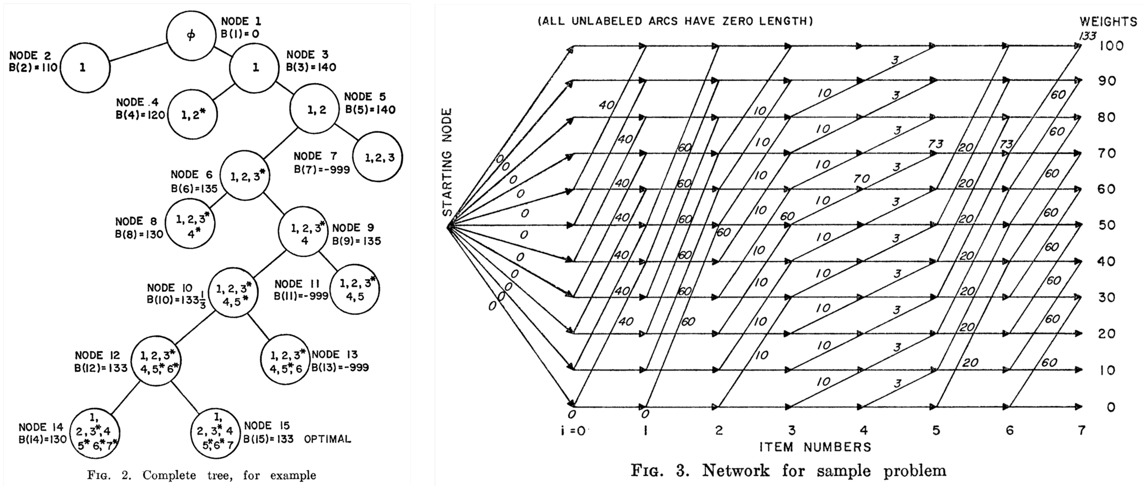


Figure 4-2: Illustrative Example of Branch-and-Bound Network Diagram [27]

optimal combination of items that maximizes the total weight without exceeding capacity constraints.

Branching

Branching is performed on fractional values of decision variables that arise in the linear relaxation of the MILP problem. If the variable is binary (e.g., z_j), two subproblems (branches) are created:

$$\begin{array}{ccc} & \text{Root Node: } z_j & \\ \swarrow & & \searrow \\ z_j = 0 & & z_j = 1 \end{array}$$

in which setting to $z_j = 0$ (left branch) indicates that container j is not utilized, while setting $z_j = 1$ (right branch) indicates that container j is utilized. For continuous variables, branching involves splitting the search space by enforcing upper and lower bounds on feasible integer values. For example, if the cubic volume of a given PO j is 45.6 cbm, the two subproblems (branches) resemble:

$$\begin{array}{ccc} & \text{Root Node: } x_{p,j} = 45.6 \text{ cbm} & \\ \swarrow & & \searrow \\ x_{p,j} \leq 45 \text{ cbm} & & x_{p,j} \geq 46 \text{ cbm} \end{array}$$

in which PO p is restricted to at most 45 cbm in container j (left branch) or PO p needs to occupy at least 46 cbm in container j (right branch). Thus, each branch corresponds to a more constrained version of the feasible region, systematically partitioning the solution space to ensure that all possible assignments are explored while excluding infeasible or suboptimal paths.

Bounding

Bounding is conducted by solving a linear relaxation of the current subproblem, where binary decision variables are allowed to take continuous (fractional) values between 0 and 1. The solution to this relaxed formulation yields a:

- **Lower Bound:** The best possible objective value that can be achieved within the subproblem.
- **Upper Bound:** The value of the best-known feasible solution (if one exists).

The bounding step leverages the weighted objective function:

$$Z = \alpha \cdot (1 \text{ - Utilization}) + \beta \cdot (1 \text{ - Delivery Precision}) + (1 - \alpha - \beta) \cdot \text{Transportation Cost}$$

in which α and β are weights for each dimension of the objective function used to evaluate partial solutions. Bounding leverages a weighted composite objective function that balances three competing metrics: container utilization, delivery precision, and transportation cost. The lower bound is calculated by solving the LP-relaxation of the multi-objective MILP model, in which binary decision variables—such as z_j (indicating whether container j is used)—are relaxed to take on fractional values. The upper bound is updated dynamically as feasible integer solutions are identified during the search process.

In this work, the CBC (COIN-OR Branch and Cut) solver is employed to implement a branch-and-bound framework. It begins by solving the LP-relaxation of the root node to obtain a lower bound. If the solution includes fractional values for binary decision variables—such as $z_7 = 0.48$ (indicating partial use of container 7)—the solver branches on this variable to explore both possibilities ($z_7 = 0$ and $z_7 = 1$), each corresponding to excluding or including the container in the solution. Similarly, if a PO assignment indicator variable x_{pjw} is fractional (e.g., $x_{41,3,10} = 0.5$), this triggers a branch to determine whether PO 41 should be assigned to container 3 in week 10 or not. At each branch, a new LP-relaxation is solved recursively to produce an updated lower bound. This LP-relaxed lower bound allows for rapid screening of subproblems that are unlikely to outperform current solutions. Additionally, domain-specific constraints—such as earliest shipment eligibility based on GCW and co-loading restrictions—are embedded within the relaxed model to ensure operational relevance and feasibility. Bounding plays a crucial role by providing lower and upper estimates on objective function values at each node. These bounds guide the search process by identifying promising regions of the solution space and enabling early elimination of suboptimal branches.

Pruning

Pruning enhances computational efficiency through the early elimination of branches that cannot contribute to an optimal solution. Three pruning criteria are employed:

1. **Infeasibility**: Branches that violate container capacity or temporal constraints are discarded.

2. **Bound Dominance**: Nodes with bounds worse than the best-known solution across any objective are excluded.

3. **Optimality**: Nodes that cannot improve upon the best-known solution or contribute to non-dominated solutions along the Pareto frontier are terminated.

4.3 Model Components and Formulation

4.3.1 Nomenclature

Table 4.1 provides a summary of the notation used throughout the optimization model. This nomenclature defines the sets, indices, decision variables, and parameters that structure the CLP formulation.

Notation	Description
Sets & Indices	
p	Purchase order (PO) index, where $p \in P$
j	Container index, where $j \in J$
w	Time index (week number)
Decision Variables	
x_{pjw}	Volume (cbm) of PO p assigned to container j in shipment week w
w_j	Shipment week for container j
z_j	Binary variable indicating whether container j is used
Parameters	
GCW_p	Earliest possible shipment week for PO p
RDW_p	Required delivery week for PO p
V_p	Volume (CBM) of PO p
C_u	Cost per cbm of underutilization
C_e	Early delivery cost penalty
C_l	Late delivery cost penalty
C_f	Container freight cost
C_d	Storage cost per container per week at the destination port
C_c	Storage cost per CBM per week at the origin consolidator
FST_c	Free storage time at the consolidator (2 weeks)
FST_d	Free storage time at the destination (4 weeks)
VTT	Vessel transit time from Vietnam Consolidator to California Port (4 weeks)

Table 4.1: Model Nomenclature

4.3.2 Temporal Structure & Storage Costs

The optimization model evaluates a sequence of interconnected temporal nodes (Figure 4-3) to determine the most efficient container shipment schedule. The decision-making process centers on selecting the optimal shipping week (w) for each container while minimizing transportation cost and maintaining high delivery precision and container utilization. The temporal constraints in the model establish clean decision boundaries that are based on the manufacturing and transportation setup used by Fleetform:

- **Goods at Consolidator Week (GCW)**: The week that the PO arrives to the consolidator from the factory. It denotes the earliest week that PO p can ship out of the consolidation facility.
- **Shipment Week (w, Decision Variable)**: The model determines the optimal week to ship container j from the origin (Vietnam) port.
- **Vessel Transit Time (VTT)**: A fixed 4-week shipping duration from the Vietnam consolidator to the California destination port.
- **Required Delivery Week (RDW)**: The week on which POs are expected to arrive to the destination port.

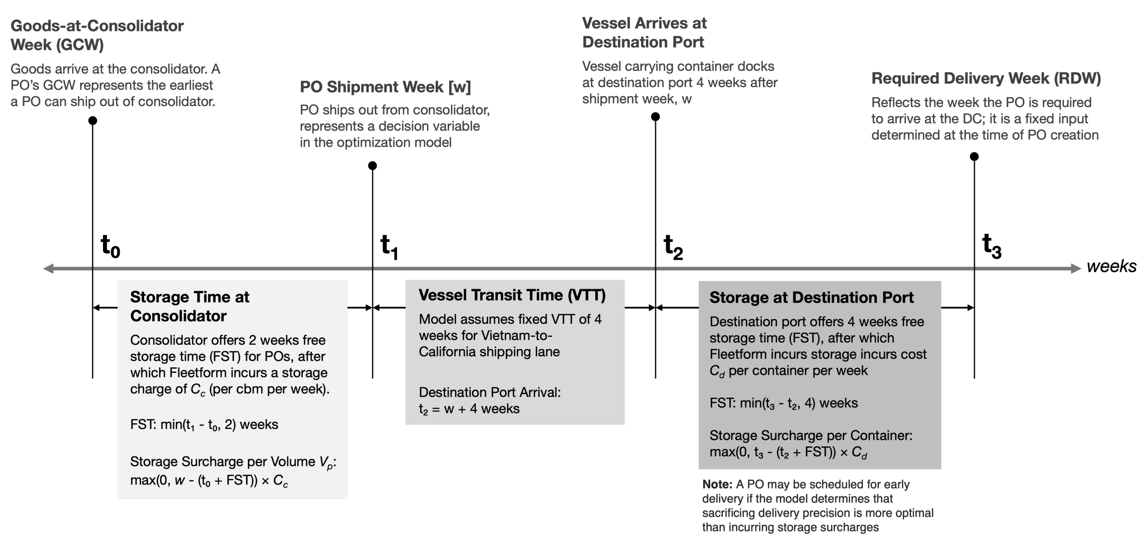


Figure 4-3: Temporal Nodes of Optimization Model

Storage Constraints & Cost Implications

Due in part to seasonality and long lead times in the footwear and apparel industry, merchandise is typically manufactured months in advance of scheduled customer deliveries [15], necessitating strategic inventory storage points throughout the supply chain. Given that the optimization model maintains a port-to-port purview, POs may be stored at either the origin consolidator facility in Vietnam or at the destination port in California. Storage costs are incurred when POs exceed the allocated free storage time (FST) at either location. Given that certain costing data is confidential to Fleetform, approximate cost structures are reported for model illustration. The linearized storage cost formulations reflect:

- **Storage at Origin Consolidator:** Each PO p receives 2 weeks of free storage at the origin consolidator before charges apply. While precise cost details cannot be reported, beyond 2 weeks, a cost of approximately \$2 per CBM per week is incurred. The storage cost at the consolidator for each PO can be described by:

$$\text{Storage Cost}_{Consolidator} = \sum_{p \in P} \max(0, (w_j - GCW_p - FST_c) \cdot V_p \cdot C_s) \tag{4.1}$$

where:

  - w_j = Shipment week for container j (applies to all POs assigned to j) [1]
  - GCW_p = Goods at Consolidator Week for PO p (earliest week it can ship)
  - FST_c = Free Storage Time at the consolidator (2 weeks),
  - V_p = Volume (cbm) of PO p
  - C_s = Storage cost per cbm per week (\$2).

To linearize the storage cost setup, a variable O_p is introduced to calculate "over-storage," representing the number of weeks a PO remains at the origin

[1] A separate constraint in the optimization model ensures that each PO p inherits the shipment week w_j of its assigned container, thereby maintaining consistency between PO-container shipment schedules.

consolidator beyond its free storage period.

$$\text{Storage Cost}_{Consolidator} = \sum_{p \in P} O_p \cdot V_p \cdot C_s \tag{4.2}$$

in which over-storage, O_p is defined as:

$$O_p = w_j - GCW_p - FST_c \tag{4.3}$$

subject to the constraint:

$$O_p \geq 0 \tag{4.4}$$

- **Storage at Destination Port:** Each container receives 4 weeks of free storage at the destination port before charges apply. While precise cost details cannot be reported, beyond 4 weeks, a cost of roughly \$250 per container per week is incurred. The storage cost at destination for each container can be described by:

$$\text{Storage Cost}_{Destination} = \sum_{j \in J} max(0, (RDW_j - (w_j + VTT) - FST_d) \cdot C_d) \tag{4.5}$$

where:

  - RDW_j = Required Delivery Week for container j,
  - w_j = Shipment week of container j
  - VTT = Vessel Transit Time (4 weeks)
  - FST_d = Free Storage Time at the destination port (4 weeks)
  - C_d = Storage cost per container per week (\$250)

To linearize the storage cost setup, a variable O_j is introduced to calculate "over-storage," which represents the number of weeks the container remains at the destination port beyond its free storage period.

$$\text{Storage Cost}_{Destination} = \sum_{j \in J} O_j \cdot C_d \tag{4.6}$$

in which over-storage, O_j, is defined as:

$$O_j = RDW_j - (w_j + VTT) - FST_d \tag{4.7}$$

subject to the constraint:

$$O_j \geq 0 \tag{4.8}$$

The formulation correspondingly ensures that storage costs are only incurred if the container remains at the destination port beyond the free 4-week period. This ensures that storage costs are only incurred if a PO remains at the consolidator beyond the free 2-week period, while also respecting the constraint that all POs in container j must have the same shipment week w_j.

4.3.3 Container Capacity & Activation Costs

The optimization model incorporates several key logistics and cost components that reflect real-world shipping operations at Fleetform. Each 40-ft container has a maximum capacity of 76 cbm, of which roughly 66 cbm is considered usable. Each time a container is used for shipping, a fixed freight cost of C_f is incurred. The model further accounts for container availability constraints: the maximum number of containers available each week corresponds to the forecasted annual volume demand divided by 52 weeks. If the number of containers in a given week is insufficient, the PO volume is rolled over to the subsequent week for shipment.

4.3.4 Model Components

The optimization framework integrates binary decision variables for container usage, continuous variables for PO volume allocation, and integer variables for container shipment scheduling. The model is structured as a cost-minimization MILP problem that generates optimized load plans balancing container utilization, delivery precision, and transportation cost.

Decision Variables

- $x_{pjw} \rightarrow$ Continuous variable representing the volume (cbm) of PO p assigned to container j in week w.

- $w_j \rightarrow$ Integer variable denoting the shipment week for each container j within a single calendar year, constrained to $w_j \in \{1, 2, \ldots, 52\}$.

- $z_j \in \{0, 1\} \rightarrow$ Binary variable indicating whether container j is utilized (1 if used, 0 otherwise).

Objective Function

The optimization framework balances three key dimensions of logistics performance: container utilization, delivery precision, and transportation cost. Given that we aim to minimize transportation cost while maximizing container utilization and delivery precision, complementary metrics are employed to formulate the objective function as a cost minimization problem. Correspondingly, container utilization and delivery precision are optimized by their complements: underutilization ($1-$ utilization) penalizes inefficient container usage, while delivery imprecision ($1-$ precision) penalizes deviations from the required delivery schedule. The resulting objective function for the container loading optimization is expressed as:

$$
\begin{aligned}
\text{Minimize } Z = \quad & \underbrace{C_u \sum_{j \in J} (1 - U_j)}_{\text{Container Underutilization Cost}} \\
& + \underbrace{\sum_{p \in P} \left(\underbrace{C_e D_p^e V_p^e}_{\substack{\text{Early Delivery} \\ \text{Penalty}}} + \underbrace{C_l D_p^l V_p^l}_{\substack{\text{Late Delivery} \\ \text{Penalty}}} \right)}_{\text{Delivery Imprecision Cost}} \\
& + \underbrace{\left[\underbrace{\sum_{j \in J} C_f \cdot z_j}_{\substack{\text{Container} \\ \text{Freight Cost}}} + \underbrace{\sum_{j \in J} C_d O_j z_j}_{\substack{\text{Storage Cost at} \\ \text{Destination Port}}} + \underbrace{\sum_{p \in P} C_c V_p O_p}_{\substack{\text{Storage Cost at} \\ \text{Origin Consolidator}}} \right]}_{\text{Transportation Cost}}
\end{aligned}
\tag{4.9}
$$

The formulation reflects three key dimensions in logistics performance:

- **Container Utilization** measures the efficiency of space usage within a container, ensuring that available capacity is consumed to reduce fixed transportation costs. The capacity utilization of a container, U_j is defined as:

$$U_j = \frac{\sum_{p,w} x_{p,j,w}}{\lambda_j} \tag{4.10}$$

  where:

  - $x_{p,j,w}$: Volume (cbm) of PO p assigned to container j in week w.
  - λ_j: Loadable volume capacity of container j. Within this model, assume 40-ft High containers provide 66 cbm of loadable volume.

  To discourage inefficient packing, the model penalizes underutilized containers through a per cbm cost of underutilization (C_u). This penalty reflects the wasted capacity in partially filled containers, incentivizing shipment consolidation to maximize space utilization.

- **Delivery Precision** measures the accuracy of shipments arriving on their required delivery week (RDW) by comparing the estimated delivery week of a PO p to its RDW. In the optimization model, delivery deviations are separately penalized for early (D_p^e) and late (D_p^l) delivery.

$$D_p^e = \max(0, RDW_p - (w_j + VTT)) \tag{4.11}$$

$$D_p^l = \max(0, (w_j + \mu_j) - RDW_p) \tag{4.12}$$

where:

- w_j: Shipment week of container j
- VTT: Port-to-port vessel transit time (4 weeks)

The early delivery penalty applies if a PO arrives before its RDW; the late deviation penalty applies when a PO arrives after its RDW. In a MILP solver, the max() function introduces non-linearity. As a result, the formulation is piecewise linearized using auxiliary binary variables y_p^e and y_p^l. To activate D_p^e and D_p^l only when necessary, we introduce Big-M constraints:

$$D_p^e \geq RDW_p - (w_j + VTT) - M(1 - y_p^e) \tag{4.13}$$

$$D_p^e \leq M y_p^e \tag{4.14}$$

$$D_p^l \geq (w_j + VTT) - RDW_p - M(1 - y_p^l) \tag{4.15}$$

$$D_p^l \leq M y_p^l \tag{4.16}$$

Since the penalty for early or late delivery depends on the volume of the affected PO, the final penalty term in the objective function for delivery precision includes the cubic volume of the PO, V_p. The early and late delivery cost penalties, C_e and C_l, are thus multiplied on a per-cbm basis by V_p and delivery deviation, D_p.

- **Transportation Cost** Transportation cost is the summation of three cost components: container freight cost, storage cost at origin consolidator, and storage cost at destination port (which capture detention and demurrage fees). Given that the storage costs depend on shipment timing, the formulation for consolidation storage cost and destination storage cost were previously derived in Equations (4.2) and (4.6), respectively. The container freight cost, C_f, is the cost incurred for activating and utilizing a shipping container, represented by binary decision variable, z_j. In the model, total freight cost is set to $6,000 per container to account for cost exposures beyond base commodity freight rates: bunker adjustment factor (BAF) for fluctuations in fuel price, terminal handling charges (THC) for container (un)loading at the ports, wharfage fees for containers using port facilities, customs and clearance fees, container inspection fees, bill of lading fees, and third-party handling fees. Correspondingly, the total freight cost per container can be expressed as $C_f = \text{Base Freight} + \text{Port Handling} + \text{Customs Fees} + \text{Third-Party Handling} + \text{Surcharges}$.

Constraints

The model enforces the following constraints:

- **Container Volume Capacity**: The sum of PO volumes assigned to container j must be less than or equal to the loadable capacity, λ_j, of container j.

$$\sum_{p \in P} x_{p,j} \leq \lambda_j \quad \forall j \in J \tag{4.17}$$

- **Container Weight Capacity**: The combined weight (kg) of POs assigned to container j must not exceed the maximum weight limit, μ_j, of container j.

$$\sum_{p \in P} \omega_p x_{p,j} \leq \mu_j, \quad \forall j \in J \tag{4.18}$$

  where ω_p represents the weight per cbm of PO p.

- **PO Volume Conservation**: All POs $p \in P$ must be fully assigned across all container-week combinations as a reconciliation step to ensure no portion of the

PO remains unshipped. This also prevents the MILP solver from choosing not to load a PO simply to minimize costs.

$$\sum_{j \in J} \sum_{w \in W} x_{p,j,w} = V_p, \quad \forall p \in P \tag{4.19}$$

- **Shipment Timing Constraint (GAC Week)**: A PO cannot be assigned to a container before its Goods at Consolidator Week, GCW_p, which represents the first week the PO p is available for shipment from the origin consolidator.

$$x_{p,j,w} = 0, \quad \forall p \in P, j \in J, w_j < GAC_p \tag{4.20}$$

- **PO-Container Shipment Week Consistency**: All POs assigned to the same container j inherit the shipment week of container j. This constraint ensures that all POs assigned to the same container share the same shipment week, w.

$$w_j = w, \quad \forall p \in P, j \in J, w \in W \text{ if } x_{p,j,w} > 0 \tag{4.21}$$

4.3.5 Load Rule Modeling

Fleetform has an established set of load rules that consolidators must follow to ensure safe and compliant cargo loading practices. Load rules function as constraints in logistics optimization models, encompassing considerations that include co-loading restrictions, volume-based handling requirements, and specific container allocation criteria. Within the CLP framework, these rules are implemented as hard constraints that must be satisfied for any feasible solution. The formulation presented within this-nwork incorporates three key load rules utilized by Fleetform: co-loading restrictions based on product category, first-sail scheduling for large POs, and dedicated container assignments for large POs.

- **Product Category Co-Loading Restriction**: POs belonging to different product categories (e.g., apparel, footwear, accessories) cannot be co-loaded into the same container. P_z is the set of POs belonging to category z, and $y_{p,j,w}$ is a binary variable indicating whether PO p is loading into container j in week w. This constraint ensures that at most one category is assigned per container.

$$\sum_{p \in P_z} y_{p,j,w} \leq 1, \quad \forall j \in J, w \in W, z \in Z \tag{4.22}$$

- **First-Sail Requirement**: POs with a volume greater than 40 cbm must be shipped on their first available sailing week, corresponding to their GAC Week.

$$x_{p,j,w} = 0, \quad \forall p \in P, j \in J, w > GAC_p \text{ if } V_p > 40 \tag{4.23}$$

- **Single PO-to-Container Allocation Requirement**: POs with a volume between 45 and 66 cbm must be assigned exclusively to a standalone container, ensuring that no other POs are co-loaded in that container. The summation ensures that only one PO is loaded into container j in week w.

$$\sum_{p' \in P} x_{p',j,w} = x_{p,j,w}, \quad \forall j \in J, w \in W, p \in P \text{ if } 45 \leq V_p \leq 66 \tag{4.24}$$

By integrating these load rules into the optimization model, Fleetform can assess their impact on critical performance metrics, such as delivery precision, container utilization, and transportation cost. Within the model, load rules can be toggled on and off, allowing Fleetform to analyze their individual and combined effects on logistics performance

4.3.6 Model Assumptions

The optimization model is predicated on several key assumptions that simplify real-world logistics while maintaining a high level of fidelity to the Fleetform logistics operation. These assumptions govern transit times, container specifications, loading practices, cost structures, weight distribution, and demand signals.

- **Transit Time**: A fixed "on water" vessel transit time of four weeks is assumed based on historical averages for the Vietnam-to-California shipping lane. Variability due to seasonal port congestion, weather disruptions, or carrier schedule deviations are not considered in this analysis.

- **Container Specifications**: The model assumes that only 40-ft High Cube (40 HC) containers, each with a 66 cbm usable capacity, are used for shipment. Alternative container sizes (e.g., 40-ft Dry or 20-ft High) are not considered in this analysis. Further, the model assumes weekly container allowances without dynamic adjustments for demand fluctuations or spot market conditions.

- **Floor-Loading**: All POs are packed into cartons and floor-loaded into containers without the use of pallets. Cartons are stacked floor-to-ceiling to help maximize volume utilization, as shown in Figure 4-4. This analysis assumes that cartons are stacked uniformly within the container, minimizing void spaces, and that all POs can be freely stacked atop one another without weight-based stacking sequences or fragility constraints.

Figure 4-4: Floor Loading Setup of PO Cartons Inside Container

- **Carton Packing**: The model assumes uniform carton packing, where each carton has a standardized volume of 1 cbm (1m × 1m × 1m). This assumption simplifies space allocation by ensuring a direct correlation between the number of cartons and total cubic volume (cbm) assigned to each container.

- **Cost Uniformity**: Handling, storage, and transportation costs are assumed to be fixed over the time horizon of this study for the Vietnam-to-California shipping lane.

- **Weight Distribution**: The model does not explicitly account for weight-based constraints when loading POs into containers. The basis for this assumption is that the maximum payload of Fleetform's POs to form a full container load (FCL) is significantly below the payload capacity of a 40 HC container. Specifically, the average payload capacity of a 40 HC container is 24,600 kg, while the maximum payload of an FCL for Fleetform's footwear and apparel is approximately 3,000 kg. As a result, the CLP is considered volume-constrained rather than weight-constrained.

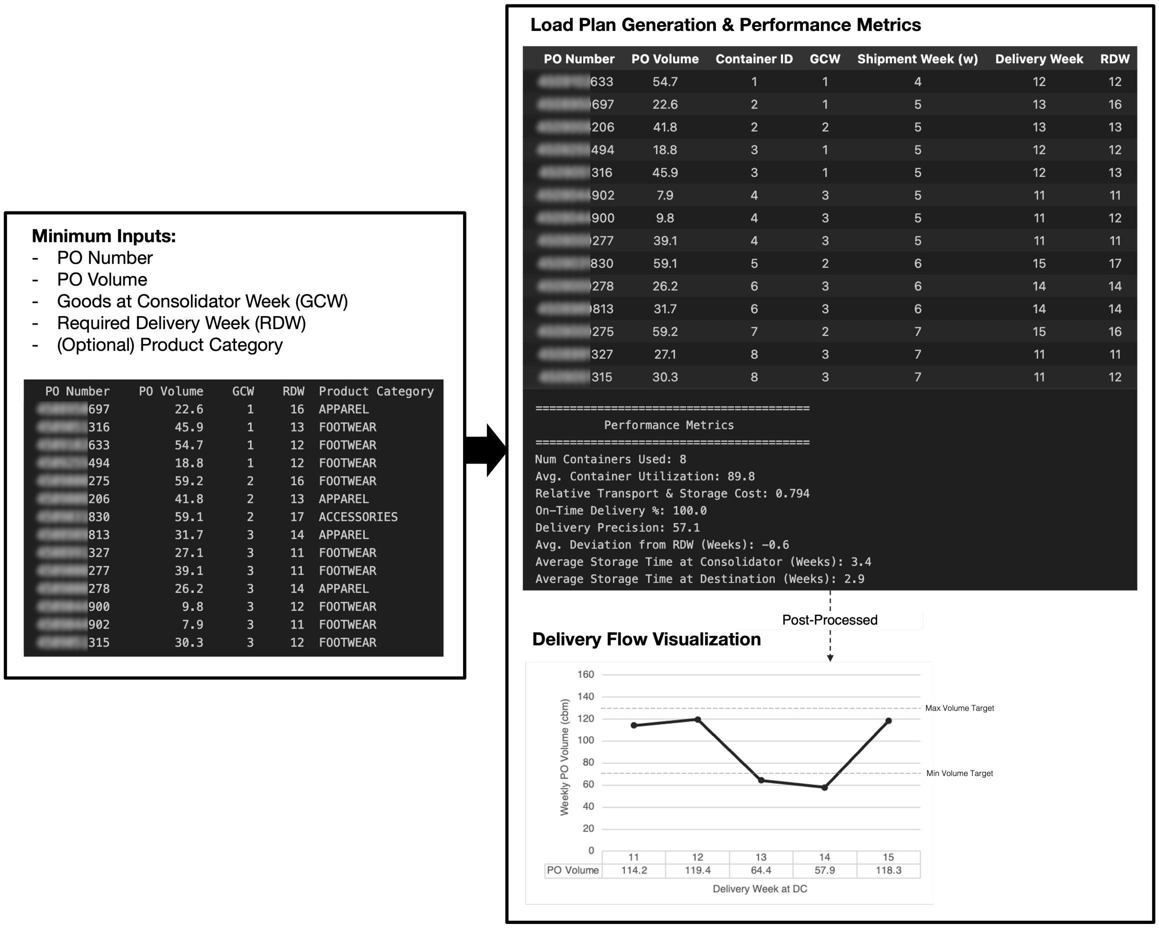
- **Demand Signal**: The required delivery week (RDW) of a PO serves as the final demand signal, indicating the week on which POs need to arrive at the destination port. No secondary demand signals (e.g., downstream customer demand fluctuations) are incorporated into the model.

- **Port-to-Port Modeling**: The Fleetform DC is co-located with the destination port in California, allowing for efficient container unloading and inventory transfer. On average, it takes 2-3 days to move inventory from containers into the warehouse. As a result, arrival at the port is treated as the final milestone in the shipping process, serving as a proxy for warehouse arrival within the same week. This assumption implies that container unloading and warehouse intake delays are negligible and do not impact shipment scheduling, ensuring that port arrival reliably represents warehouse arrival in the optimization model. Further, inventory holding costs at the DC are roughly similar to those at the destination port, allowing port storage costs to serve as a proxy for DC holding costs.

- **Weekly Shipment Frequency**: The model assumes that container shipments occur on a weekly basis, meaning that all shipments are scheduled and optimized at the week level rather than on specific dates. This assumption aligns with the weekly aggregation of Goods at Consolidator Week (GCW) and Required Delivery Week (RDW), ensuring consistency in the temporal structure of the optimization model. In practice, outbound container shipments occur 2–3 times per week. However, a weekly shipment schedule is adopted to simplify computation by using integer-based week values instead of continuous date-based scheduling. This simplification is particularly beneficial when running large-scale analyses across a full year, reducing computational complexity while maintaining a realistic approximation of shipment patterns. Although the model currently operates on a weekly scale, it can be extended to accommodate finer temporal resolution (e.g., daily scheduling) if required.

4.4 Use Case Illustration of Model Outputs

To highlight the CLP optimization model in a practical lens, Figure 4-5 presents an example load plan generated over a two-week planning horizon using a small data subset. Under typical operating conditions, logistics practitioners are equipped with the list of POs that are scheduled to arrive at a consolidator within the following two weeks. In the example shown, the input dataset includes a subset of 14 POs that are expected to arrive at the consolidator (denoted by GCW) in weeks 1 through 3. Practitioners also have access to PO attributes such such as PO volume and RDW, which are required model inputs. This is the stage where practitioners can also adjust RDWs to reflect changes in PO delivery priorities. The example shown includes product category as an optional attribute to illustrate a co-loading constraint (preventing the mixing of different product categories within the same container). Additional data fields, such as launch indicators or SKU-level flags, can also be incorporated as PO attributes to enable more customized load rule configurations. Under default workflows, the consolidator generates the load plan largely based on FIFO logic. As a result, POs in this example would be slotted into containers according to their GCW, unless specific overrides are requested by Fleetform to prioritize urgent shipments.



PO Number	PO Volume	GCW	RDW	Product Category
[redacted]697	22.6	1	16	APPAREL
[redacted]316	45.9	1	13	FOOTWEAR
[redacted]633	54.7	1	12	FOOTWEAR
[redacted]494	18.8	1	12	FOOTWEAR
[redacted]275	59.2	2	16	FOOTWEAR
[redacted]206	41.8	2	13	APPAREL
[redacted]830	59.1	2	17	ACCESSORIES
[redacted]813	31.7	3	14	APPAREL
[redacted]327	27.1	3	11	FOOTWEAR
[redacted]277	39.1	3	11	FOOTWEAR
[redacted]278	26.2	3	14	APPAREL
[redacted]900	9.8	3	12	FOOTWEAR
[redacted]902	7.9	3	11	FOOTWEAR
[redacted]315	30.3	3	12	FOOTWEAR

Load Plan Generation & Performance Metrics

PO Number	PO Volume	Container ID	GCW	Shipment Week (w)	Delivery Week	RDW
[redacted]633	54.7	1	1	4	12	12
[redacted]697	22.6	2	1	5	13	16
[redacted]206	41.8	2	2	5	13	13
[redacted]494	18.8	3	1	5	12	12
[redacted]316	45.9	3	1	5	12	13
[redacted]902	7.9	4	3	5	11	11
[redacted]900	9.8	4	3	5	11	12
[redacted]277	39.1	4	3	5	11	11
[redacted]830	59.1	5	2	6	15	17
[redacted]278	26.2	6	3	6	14	14
[redacted]813	31.7	6	3	6	14	14
[redacted]275	59.2	7	2	7	15	16
[redacted]327	27.1	8	3	7	11	11
[redacted]315	30.3	8	3	7	11	12

```
========================================
         Performance Metrics
========================================
Num Containers Used: 8
Avg. Container Utilization: 89.8
Relative Transport & Storage Cost: 0.794
On-Time Delivery %: 100.0
Delivery Precision: 57.1
Avg. Deviation from RDW (Weeks): -0.6
Average Storage Time at Consolidator (Weeks): 3.4
Average Storage Time at Destination (Weeks): 2.9
```


Figure 4-5: Load Plan Generation & Flow Visualization Example

After the model ingests the data, it computes an optimized load plan based on the weights chosen by the user for each element of the objective function. In this instance, the weight configuration prioritizes transport and storage cost (0.40), followed by delivery precision (0.35) and utilization (0.25). The resulting load plan specifies how POs are assigned to containers while adhering to operational constraints. In this example, the optimized load plan recommends co-loading POs -697 and -206 into Container 2. Given that the product category co-loading restriction is activated as a constraint, each container holds product belonging to only one product category.

For each load plan simulation, the model calculates key performance metrics to help practitioners evaluate alignment with business targets. In this scenario, the average container utilization is 89.8% across the 2-week planning horizon. Hypothetically, if

Fleetform had a minimum utilization target of 90.0%, they would observe that the load plan falls below their business metric and can correspondingly adjust the model weights to increase the prioritization on container utilization. Similarly, they can observe that the POs they can observe that the average deviation from RDW is -0.6 weeks, indicating that deliveries are arriving slightly ahead of schedule. They can correspondingly tune the delivery precision weight to reduce the deviation. In the example shown, the average storage time at consolidator is 3.4 weeks, which exceeds the contractual 2-week free storage period. From a policy perspective, if such trends persist, logistics teams may opt to pursue alternative storage solutions or engage in contract renegotiations with consolidators to explore alternative fee structures and inventory staging options. Ultimately, the performance metrics offer a granular lens on week-level alignment with planned delivery goals, which can be used to inform longer-term business strategies.

Following load plan generation, a post-processing step aggregates PO volumes by delivery week to visualize the inventory arrival profile at the destination DC. Practitioners can implement target volume bands to assess whether any weeks exhibit volume excursions (above or below threshold). Such fluctuations are operationally meaningful: volume peaks may necessitate surge staffing, extended overtime, and/or temporary warehousing, whereas troughs may result in underutilized resources and excess capacity. In the example provided, weeks 13 and 14 display dips in target PO volumes (relative to a simulated target volume band), suggesting irregularities in shipment pacing and volume. Such findings can be used to assess operational consistency and may indicate the need to recalibrate model weights to better align with volume smoothing objectives. Beyond immediate load plan adjustments, these insights can also prompt strategic policy reconsiderations, such as redesigning the system buffer between GCW and RDW, or revisiting upstream and downstream storage fee structures.

While this example contains a highly limited sample of 14 POs, it underscores the extensibility of the model across large-scale operational environments, where practitioners may routinely manage thousands of POs across a season. The model's ability to dynamically adjust and evaluate load plans under customizable priorities and constraints reinforces its value as a decision-support tool for load plan optimization and delivery flow smoothing.

4.5 Chapter Summary

This chapter discusses the methodological framework for CLP optimization through a multi-objective MILP model. The proposed formulation integrates three competing objectives—maximizing container utilization, maximizing delivery precision, and minimizing transportation and storage costs—through a weighted composite objective function. Real-world operational constraints such as container capacity, shipment timing, and storage policies are embedded to ensure practical relevance and emulation of Fleetform's logistics workflow. The model is implemented using a branch-and-bound algorithm supported by a CBC solver and incorporates domain-specific load planning rules to align with industry practices. Lastly, an illustrative walkthrough of model outputs is included to demonstrate how simulation results can inform load planning decisions and strategic trade-off management.

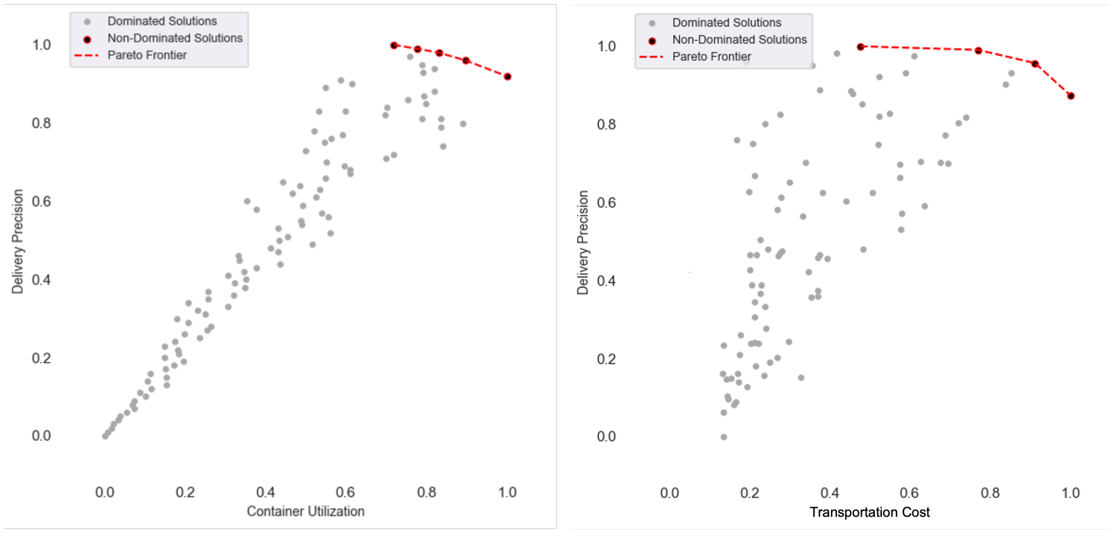
Chapter 5

Simulation Results and Key Takeaways

The following presents the results of simulation experiments conducted using the multi-objective CLP model. The analysis evaluates how different weightings of container utilization, delivery precision, and transportation costs affect logistics performance. Through trade-off visualizations, inventory flow comparisons, and load rule impact analysis, the chapter highlights the key operational insights enabled by the model.

5.1 Pareto Analysis

The Pareto analysis shared in Figure 5-1 provides insights into the trade-offs between delivery precision, container utilization, and transportation costs in the optimization model. Plot 5-1a illustrates the relationship between delivery precision and container utilization, while plot 5-1b examines delivery precision and transportation costs. In both cases, non-dominated solutions (shown in red) define the set of optimal trade-offs where improving one objective would result in the deterioration of another. Dominated solutions (shown in gray) indicate suboptimal configurations that can be improved in at least one dimension without negatively impacting others. This analysis was executed across 200 weight combinations, demonstrating that traditional container loading practices–focused solely on maximizing container utilization—may not yield the best outcome when considering broader supply chain implications.



(a) Precision vs. Utilization (b) Precision vs. Transportation Cost

Figure 5-1: Normalized Pareto Analysis of Delivery Precision Trade-Offs.
(a) against container utilization and (b) against transportation costs

In plot 5-1a, the Pareto frontier indicates a general trend in which higher container utilization is associated with improved delivery precision. However, the dispersion of points around the frontier suggests that this relationship is neither strictly linear nor consistent across all configurations. This suggests that consolidating shipments effectively (i.e., minimizing underutilization) can under certain circumstances, enhance delivery timing by reducing variability in container activation. However, the frontier shows diminishing returns: beyond a utilization threshold of approximately 75%, further improvements in container efficiency diminish delivery precision. This aligns with practical constraints, where opting to maximize container fill rate can reduce shipment frequency, making it more challenging to align with target delivery windows.

Plot 5-1b) illustrates that under certain conditions, efforts to minimize transportation costs can lead to increased variability in delivery timing, highlighting the trade-off between cost efficiency and service reliability. While many non-dominated solutions show that higher delivery precision tends to be associated with increased cost, the relationship is not monotonic. Notably, beyond a cost threshold of roughly 60% (relative to baseline), additional cost reductions often correspond to sharp declines in precision. This suggests that strict cost minimization strategies (e.g., delaying shipments to minimize the number of containers needed) may introduce inefficiencies in meeting desired service levels.

The Pareto analysis highlights two notable insights. First, maximizing container utilization does not necessarily result in better delivery precision. The upper right region of the utilization-precision curve suggests that stuffing a container to nearly 100% utilization can reduce shipment flexibility, ultimately leading to weaker delivery precision and suboptimal outcomes. Contrary to conventional CLP approaches that emphasize maximizing pack efficiency, we observe that optimal solutions for delivery precision can emerge at approximately 80% utilization. Second, achieving high delivery precision does not always correlate with the most expensive solutions. When considering inventory holding costs and unplanned storage fees, we observe that prioritizing delivery precision can, in some cases, lead to overall cost benefits. Ultimately, these trade-offs highlight the importance of selecting a balanced logistics strategy that ensures operational efficiency while maintaining service level commitments.

5.2 Inventory Flow Analysis

Figure 5-2 presents simulated inventory arrival flows at the DC under a range of objective function weight combinations for container utilization, delivery precision, and transportation cost. The historical flow (black line) serves as a baseline, while the modeled flows (green lines) simulate the profile of inventory delivery under different load plan configurations. Red-circled markers denote volume excursions, defined as weekly inbound volumes exceeding one standard deviation above the mean of the optimized flow. This visualization provides a diagnostic lens to compare load planning priorities and how they affect shipment timing and operational consistency.

The results reveal that excessive emphasis on container utilization (Sub-Figure a) produces a highly volatile arrival pattern. By aggressively consolidating shipments to maximize fill rates (96.7% utilization), the model tends to front-load POs, resulting in shipments that arrive significantly earlier than their required delivery week (RDW).

This behavior leads to large weekly spikes in inbound volume, with an average deviation from RDW of -11.2, indicating that POs are arriving to the DC roughly 3 months early. Although this approach lowers total cost (0.636 relative to baseline), it comes at the expense of delivery precision (16.3%). Conversely, when delivery precision is prioritized (Sub-Figure b), the resulting inventory flow is more evenly distributed and while flow variability persists, the profile is less jagged and exhibits fewer extreme fluctuations. When transportation cost is prioritized (Sub-Figure c), the flow pattern resembles that of Sub-Figure a, characterized by pronounced fluctuations. From a metrics standpoint, Scenarios a and c yield comparable container utilization levels. However, Scenario c increases the average amount of time a PO spends at consolidator and destination storage. This extended storage dwell time reduces the deviation from RDW from -11.2 (Scenario a) to -7.9 weeks (Scenario c), which is equivalent to nearly 2 months of early delivery. The mixed prioritization cases (Sub-Figures e-g) demonstrate that a balanced approach can smooth inventory arrivals such that the week-to-week change is more gradual. Table 5.1 provides an overview of key performance metrics for each variant of the simulation.

Metrics	**Historical**	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**
Num Containers Used	2,155	1,909	2,017	1,934	1,946	1,923	1,999	1,962	2,034
Avg. Container Utilization (%)	85.6	96.7	91.5	95.4	94.8	95.9	92.2	94.0	90.7
Transport & Storage Cost (Relative to Baseline)	1.000	0.636	0.546	0.520	0.732	0.681	0.668	0.584	0.651
On-Time Delivery (%)	89.8	74.0	97.3	83.1	92.6	74.8	90.2	88.5	93.8
Delivery Precision (%)	3.7	16.3	77.6	43.5	67.8	26.6	71.5	58.2	75.9
Avg. Deviation from RDW (Weeks)	-4.7	-11.2	-0.8	-7.9	-2.2	-6.9	-1.3	-6.4	-1.7
Avg. Storage Time at Consolidator (Weeks)	1.7	1.9	5.5	2.4	4.8	2.7	5.3	3.9	4.7
Avg. Storage Time at Destination (Weeks)	3.8	0.2	4.7	4.1	3.9	5.3	4.4	4.1	4.0

Table 5.1: Performance Metrics Comparison for Flow Simulations
Scenario Weight Parameters: a: Utilization Only ($\alpha = 1.0$), b: Delivery Precision Only ($\beta = 1.0$), c: Transportation Cost Only ($\alpha = \beta = 0.0$), d: Balanced ($\alpha = 0.33$, β=0.33), e: Mixed ($\alpha = 0.50$, $\beta = 0.25$), f: Mixed ($\alpha = 0.25$, $\beta = 0.50$), g: ($\alpha = 0.25$, $\beta = 0.25$), h: ($\alpha = 0.14$, β=0.62).

The flow analysis is crucial for effective supply chain planning, as erratic inventory arrivals can create downstream inefficiencies such as warehouse congestion, higher storage costs, and increased labor requirements for unloading and processing shipments.

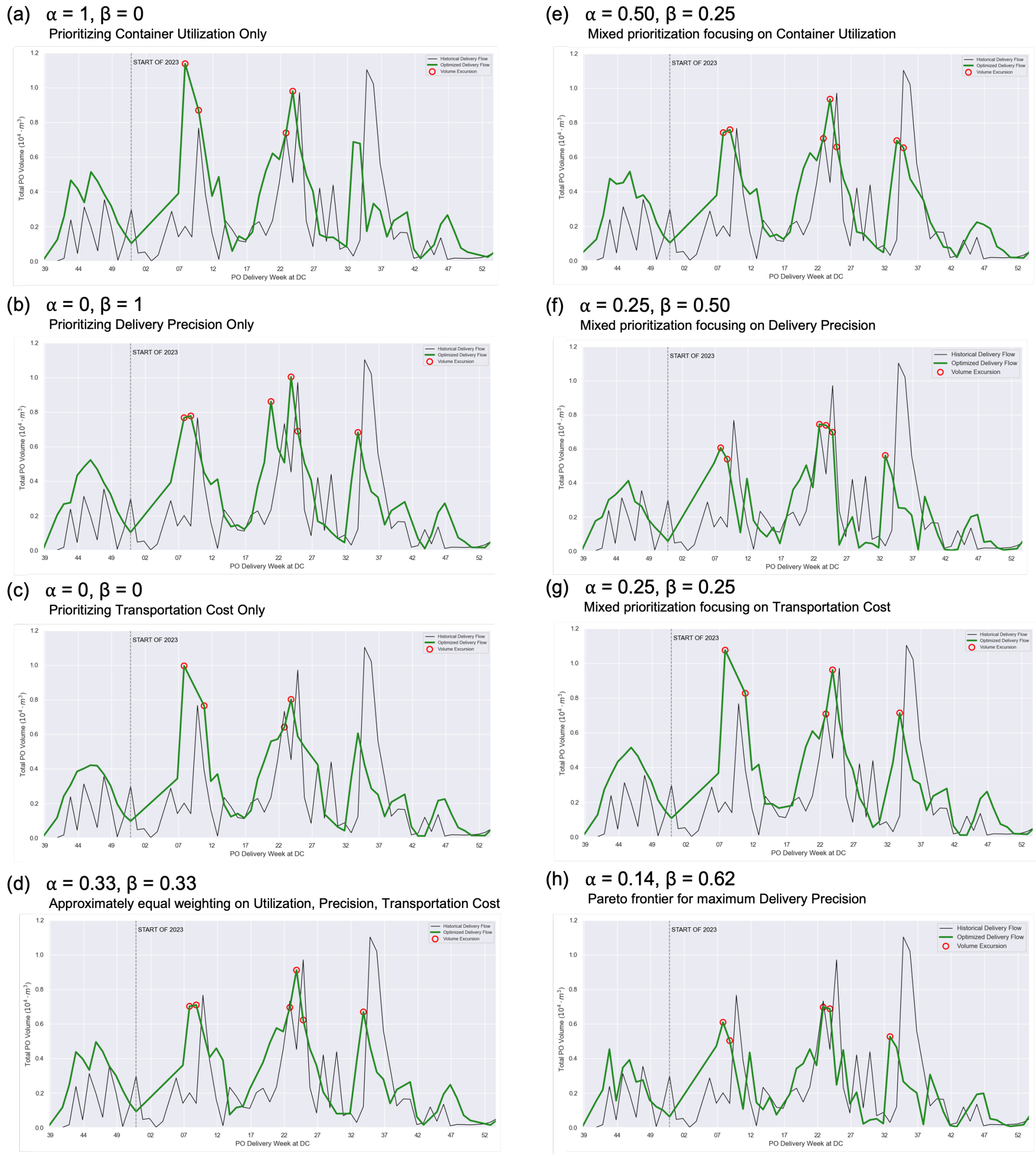


Figure 5-2: Flow Simulation of Inventory Arrival at DC Under Varying Weights. Weight parameters for Container Utilization, Delivery Precision, and Transportation Cost are denoted by α, β, and $(1-\alpha-\beta)$, respectively.

The results highlight that pure cost minimization or utilization maximization may not align with operational stability and that a balanced weighting approach can offer better predictability and control over inventory movement into the DC. It is important to recognize that across all scenarios, the inventory flow exhibits fluctuations and irregular peaks; strategies for further smoothing and level-loading flow through operational interventions are explored in Chapter 6.

5.3 Load Rule Impact Analysis

As described in Section 4.3.5, operational load rules were encoded as model constraints to evaluate their effects on delivery flow and logistics system performance. Figure 5-3 illustrates the simulated weekly PO volume arriving at the DC under four different load rule scenarios, benchmarked against historical delivery flow. Table 5.2 summarizes key performance metrics for each scenario.

- The **product co-loading rule** prohibits the mixing of different product categories (e.g., apparel, footwear, accessories) within the same container. While this rule may improve sorting accuracy or simplify customs classification, the optimization model significantly increases average storage time at the origin (4.3 weeks) due to the increased wait time required to pool homogenous POs. Although container utilization remains relatively high at 91.6%, the flow profile exhibits sustained peaks similar in magnitude to the baseline, suggesting limited improvement in delivery smoothness. Delivery precision improves notably (63.0%), though at the cost of extended inventory holding periods.

- The **first sail rule**, which mandates that POs over 40 cbm ship in their first eligible week (GCW), generates the most volatile flow profile. The constraint forces a front-loading effect that eliminates the buffering available through the consolidator storage window, leading to sharp, concentrated delivery spikes and the lowest delivery precision (1.6%). Although on-time delivery appears high (93.5%), the outcome is primarily driven by prioritizing immediate shipments rather than accurate alignment with RDWs. While utilization remains moderate at 87.6%—a result of certain POs being individually large enough to fill a container—the lack of pooling leads to the highest number of containers used (2,073) and heightened transportation costs (0.927 relative to baseline). Among the load rule scenarios tested, the first sail requirement exerts the most pronounced negative impact on delivery precision and overall logistics efficiency.

- The **single PO allocation rule** allows containers to carry only one PO if its volume falls between 45 and 66 cbm. While the load rule is useful in scenarios requiring shipment segregation—such as product launches, promotional campaigns, or high-value SKUs demanding enhanced traceability—it significantly compromises operational efficiency. The restriction that each container carry only one PO (when within a certain volume threshold) results in a marked decline in container utilization (70.8%). The resulting loading structure is highly fragmented, contributing to a high logistics cost.

- When all three load rules are simultaneously activated in the **combined load rules** scenario, the resulting delivery profile roughly emulates patterns from the historical baseline. The number of containers used (2,118) aligns closely with the baseline (2,155), along with the average delivery deviation from RDW (-4.9 weeks versus -4.7 weeks baseline) and logistics cost (0.945 relative to baseline). This approximate convergence in delivery timing and cost relative to the baseline suggests that the aggregate effective of multiple load constraints may moderate the extremes introduced by individual rules. However, it is important to note that the first sail requirement exerts a disproportionately strong impact on overall performance metrics in the combined scenario, particularly in triggering earlier deliveries, increased container usage, and diminished delivery precision.

Metrics	**Historical**	**Product Co-Loading** (a)	**First Sail Requirement** (b)	**Single PO Allocation** (c)	**Combined Load Rules** (d)
Num Containers Used	2,155	1,992	2,073	2,009	2,118
Avg. Container Utilization (%)	85.6	91.6	88.9	91.6	87.1
Transport & Storage Cost (Relative to Baseline)	1.000	0.694	0.927	0.841	0.945
On-Time Delivery (%)	89.8	88.1	93.5	87.2	90.6
Delivery Precision (%)	3.7	63.0	1.6	52.4	15.9
Avg. Deviation from RDW (Weeks)	-4.7	-2.0	-7.2	-3.1	-4.9
Avg. Storage Time at Consolidator (Weeks)	1.7	4.3	0.3	2.5	2.2
Avg. Storage Time at Destination (Weeks)	3.8	4.0	4.8	3.9	4.3

Table 5.2: Load Rule Impact Comparison Across Key Performance Metrics

The load rule impact analysis underscores the trade-offs between operational rigidity and logistical efficiency and illustrates the cost implications of rigid shipment

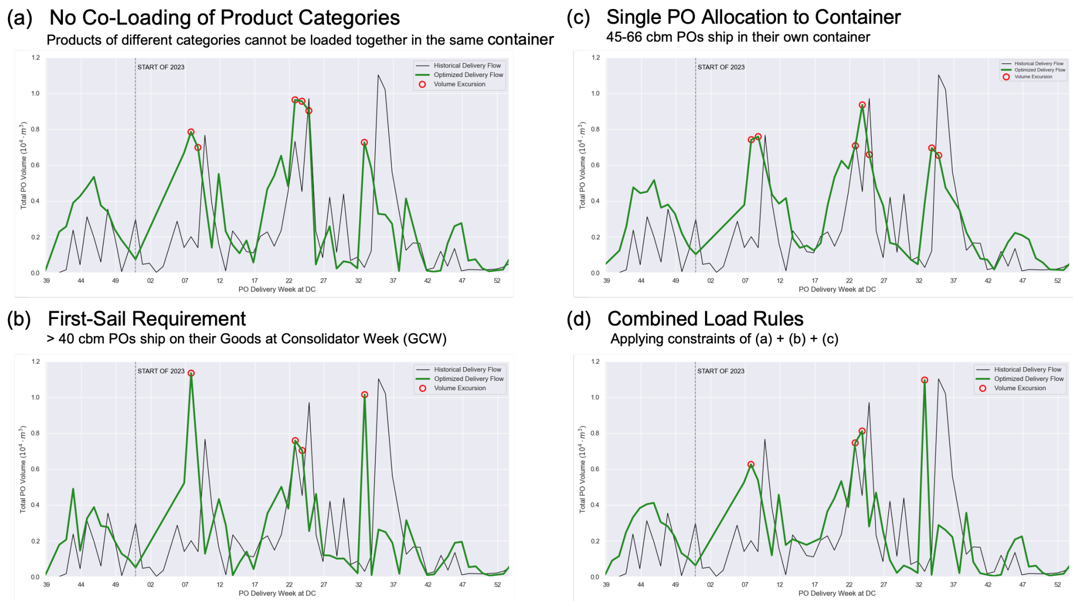


Figure 5-3: Flow Simulation of Inventory Arrival at DC Under Different Load Rules

constraints. As indicated in Tables 5.1 and 5.2, the absence of load rules consistently improves delivery precision and relative transportation costs across all simulations tested. This indicates that the easing or elimination of rigid shipment restrictions allows for more efficient load configurations and better alignment with delivery windows, without incurring significant cost penalties. Among the load rules tested, the first-sail requirement—which mandates immediate shipment of large-volume POs—emerges as the dominant constraint disrupting flow regularity and inflating early inventory delivery penalties. This outcome suggests that strict rules prioritizing immediacy over coordination can undermine flow stability and introduce cost inefficiencies. Taken together, these findings illustrate the importance of carefully calibrating load rules to balance cost, logistics efficiency, and service performance.

5.4 Chapter Summary

This chapter evaluates simulation results derived from the multi-objective CLP optimization model, revealing how trade-offs between container utilization, delivery precision, and transportation cost vary under different planning strategies. The Pareto analysis demonstrates that delivery precision peaks at approximately 84% utilization, beyond which service reliability in terms of delivery precision declines. Inventory flow simulations further illustrate that prioritizing cost or utilization alone produces erratic shipment patterns, while balanced or precision-focused strategies yield smoother inflows without incurring significant cost penalties. The load rule analysis highlights the operational risks of rigid shipment policies, such as the first-sail rule, which can increase cost and degrade delivery precision. Collectively, the simulations help validate the model as a potential decision support tool, offering logistics practitioners the ability to quantify trade-offs, explore rule variations, and design more effective load planning strategies.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6

Process Refinement and Model Extensions

This section examines enhancements to the optimization framework that can improve its practical applicability in logistics settings and reflect real-world implementation considerations. It presents a revised CLP workflow that enables more prescriptive load planning, followed by simulations of consolidation bypass and PO partitioning strategies. These extensions offer quantitative insights into how different logistics strategies affect delivery precision, transportation cost, and inventory flow, providing actionable levers for improving CLP workflows and service levels.

6.1 Revised CLP Workflow

Figure 6-1 presents a comparative analysis of the current CLP workflow and the proposed model-enabled workflow, illustrating the structural improvements introduced by the optimization framework. As highlighted in the Section 4.4 use case, the optimization model enables Fleetform to take a more prescriptive role in load planning by directly determining container assignments based on real-time PO availability at the consolidator, shipment priorities, and inventory management directives. The revised workflow shifts decision-making agency upstream to Fleetform, removing the need for the consolidator to independently assign POs to containers using FIFO methods, which often results in premature deliveries to the DC.

Under the proposed framework in Figure 6-1, Fleetform gains direct control over inventory-holding decisions, allowing it to dictate which POs should be stored beyond the free storage time (FST) and which should be prioritized for shipment. This added control enables Fleetform to optimize key trade-offs between transportation cost, delivery precision, and container utilization, rather than heavily relying on the consolidator's default age-based shipment logic. A key advantage of the model-enabled workflow is the significant reduction in manual coordination and email-based back-and-forth between Fleetform and the consolidator. In the current process, consolidators pack containers using FIFO, unless explicitly instructed by Fleetform, leading to misalignment between shipment timing and downstream demand requirements. PO priorities can also change dynamically in response to external factors beyond the consolidator's control, resulting in reactive interventions and manual overrides that introduce inefficiencies and potentially cost-ineffective decisions. For inventory to be held at origin beyond FST, consolidators must obtain approval from Fleetform to authorize storage surcharges, which introduces administrative overheads. The approval process requires time and resource approvals, creating a disincentive for consolidators to store inventory longer than necessary. As a result, consolidators default to shipping inventory before the FST period expires, often leading to suboptimal shipment timing. At the same time, Fleetform currently lacks full visibility into the cumulative cost impact of such storage decisions, limiting its ability to make data-driven trade-offs between storage costs and shipment timing.

The proposed workflow eliminates these inefficiencies by automating the load plan creation process, allowing Fleetform to feed a pre-validated load plan directly to the consolidator. This removes the iterative review cycle in which Fleetform must adjust or override consolidator-generated plans. By embedding container utilization, storage costs, and delivery precision into the optimization model, the new workflow enables Fleetform to proactively manage outbound freight decisions, reduce premature deliveries, and enforce strategic inventory-holding decisions at the consolidator level.

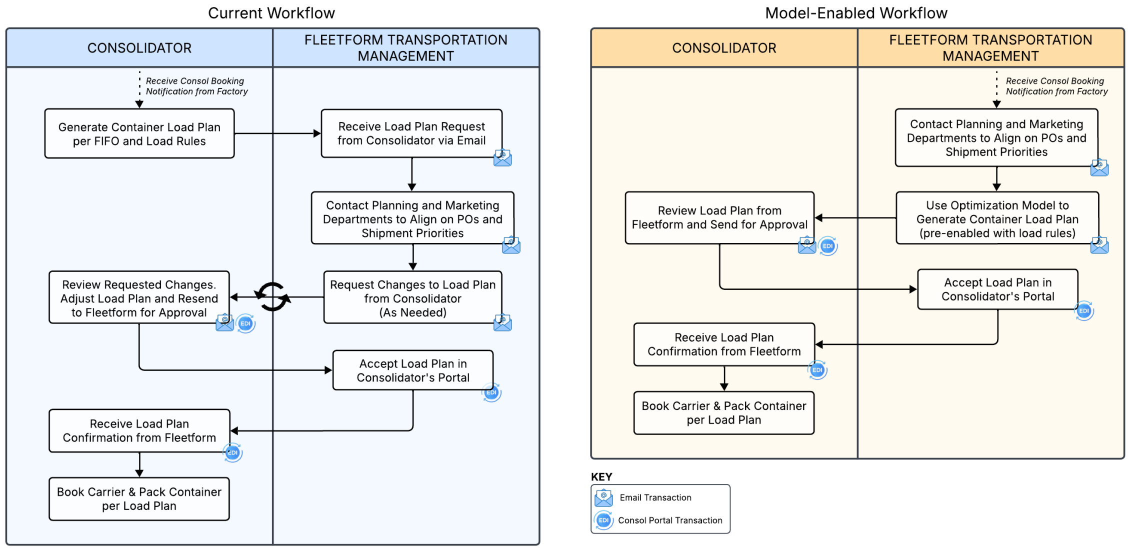


Figure 6-1: Comparison of Current and Model-Enabled CLP Workflows

6.2 Model Extension: Consolidation Bypass

In traditional retail logistics networks, shipments from multiple suppliers are typically aggregated at 3PL consolidation facilities. While this approach supports efficient container utilization, it introduces additional handling steps, intermediate processing, and the risk of delays. As global supply chains become increasingly complex and speed-to-market grows in strategic importance, consolidation bypass, also known as factory-direct shipping, has emerged as a viable alternative that allows containers to be shipped directly from factories to destination markets, thereby eliminating the need for intermediary consolidation centers. Consolidation bypass has gained traction among firms seeking to streamline logistics operations, lower total landed costs, and improve delivery precision. By removing an entire processing node from the supply chain, the strategy reduces handling complexity and minimizes touchpoints, which in turn lowers the risk of disruption and reduces transportation cost. In particular, the consolidation bypass model identifies POs that meet specific volume criteria (e.g., 55 - 66 cbm) and routes them directly from the factory to the DC. Compared to traditional consolidation routes, this strategy can reduce average port-to-port transit time by approximately one week, while avoiding storage and processing fees incurred at consolidation hubs.

The consolidation bypass strategy is particularly advantageous for products with short lifecycles, such as fashion items, entertainment goods, and promotional merchandise, where speed to market is crucial. Similarly, seasonal or perishable goods also benefit from consolidation bypass, as their time-sensitive nature requires precise delivery windows to maximize selling periods. Companies with stable and predictable demand patterns are especially well-positioned to leverage consolidation bypass, as they can better forecast volumes and optimize container utilization directly from factories. The success of consolidator bypass in such cases often manifests in the ability to reduce transit time, minimize handling costs, and maintain product integrity while meeting tight delivery windows.

The optimization approach for consolidation bypass tested within this work employs a two-stage decision process. In the first stage, POs are assessed within a rolling two-week planning horizon. POs with volumes between 55 and 66 cubic meters are automatically designated for factory-direct shipment. Under this paradigm, the shipment week is set equal to the GCW, reflecting the operational constraint that factories lack inventory storage capacity and must ship as soon as goods are ready. In the second stage, the remaining POs within the two-week planning horizon are processed through the default consolidation process used by the CLP optimization model.

The cost structure differs substantially between the two transportation strategies. Consolidated shipments incur various intermediate costs—including consolidator service fees, temporary storage, and handling charges—which are baked into the container freight cost, c_f. In contrast, consolidation bypass shipments eliminate the intermediate layers. As a result, the freight cost decreases by a flat discount, $\$D$, per cbm shipped. Additionally, the vessel transit time (VTT) for bypass shipments is modeled as three weeks, in contrast to four weeks for consolidated cargo, due to the elimination of delays associated with planning and executing inland truck transfers from factory to consolidator. Figure 6-2 illustrates a high-level schematic of the consolidation bypass workflow along with the associated simulation setup.

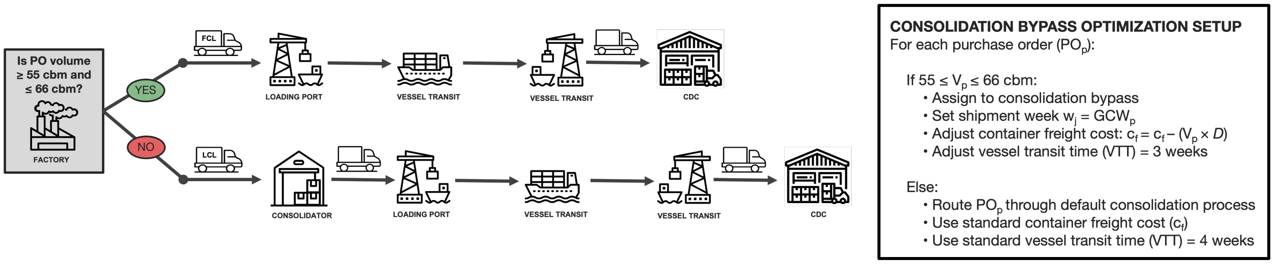


Figure 6-2: Flow Schematic & Optimization Setup for Consolidation Bypass

The simulation results, shown in Figure 6-3, highlight the impact of consolidation bypass on inventory flow dynamics. Approximately 54% of POs met the requirements for consolidation bypass, while the remaining 46% followed the default consolidation route. This bifurcation in flow pathways significantly altered delivery timelines, with inventory arriving an average of 7.1 weeks earlier through the bypass route compared to the default consolidation process. The simulation further yielded a 6% reduction in transportation cost due to discounted processing fees when bypassing consolidation services. While the consolidated bypass approach enables an expedited delivery strategy, it also introduces challenges related to excess and early inventory management at DCs, particularly due to the lack of storage options at origin. As a result, it is imperative to calibrate GCW more closely with the RDW to minimize the risk of early or idle inventory.

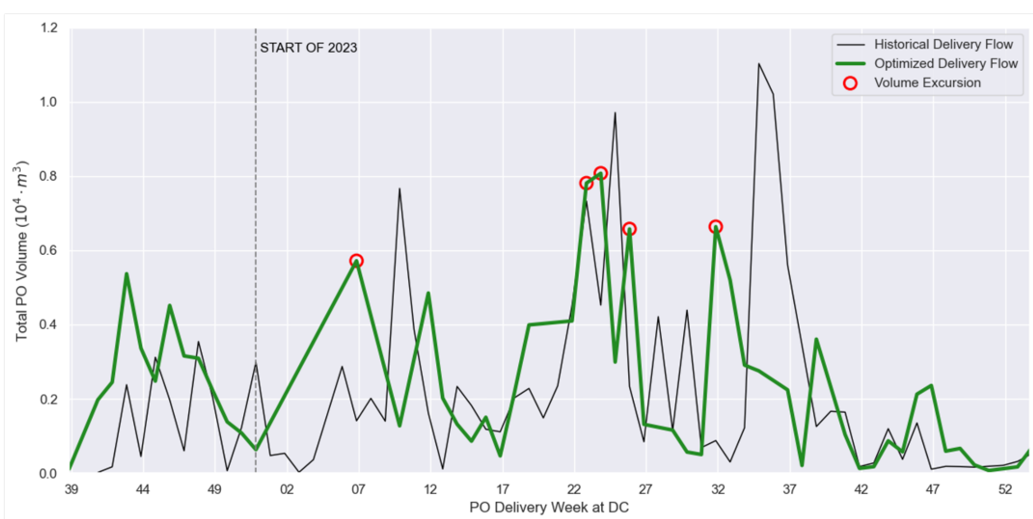


Figure 6-3: Flow Simulation of Inventory Arrival at DC Using Consolidation Bypass

6.3 Model Extension: PO Partitioning

In the current logistics setup, each PO is assigned a single required delivery week (RDW), which can be inefficient for managing seasonal demand fluctuations. A single PO may contain inventory (such as 12,000 pairs of shoes) for an entire season, although only a fraction of that quantity is needed in the initial weeks of the demand cycle. This setup results in Fleetform holding excess inventory well beyond the near-term needs of the demand cycle. To mitigate this imbalance, a PO partitioning approach is introduced in which the quantity of each PO is divided into three equal portions. The first portion is assigned to the original RDW, while the second and final portions have modified delivery requirements of RDW + 4 weeks and RDW + 8 weeks, respectively. By distributing inventory inflow over a 12-week season, this PO "slow-release" strategy reduces excess stock accumulation at the DC and aligns deliveries more closely with actual demand patterns, thereby alleviating inventory storage requirements. To illustrate the simulation setup, a PO containing 12,000 pairs with an original RDW of Week 4 would be re-segmented as follows: 4,000 pairs with an RDW of Week 4, 4,000 pairs with an RDW of Week 8, and 4,000 pairs with an RDW of Week 12.

As shown in Figure 6-4, the PO partitioning simulation demonstrates a leveling of inventory intake at the DC by distributing deliveries for each PO over a 12-week season. The approach also helps organizations like Fleetform become more responsive to demand fluctuations (since product is staged across multiple rounds of intake) and reduces the risk of overstocking in the early weeks of the season. While PO partitioning has benefits in maintaining smoother inventory flow, there are implementation challenges related to import and customs management. Given that customs processes for FLeetform are conducted at the PO level, splitting POs into multiple portions can complicate workflows. As a result, transportation management system changes are necessary to efficiently handle split POs.

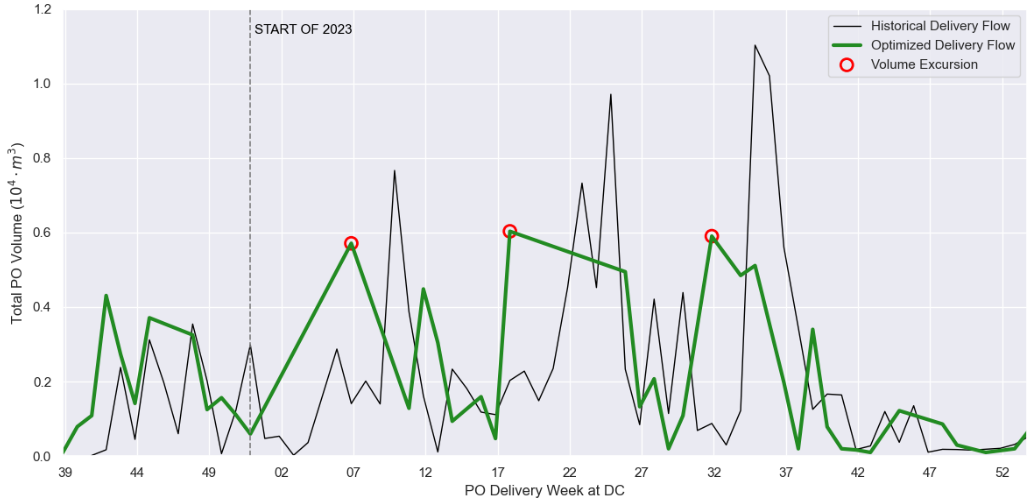


Figure 6-4: Flow Simulation of Inventory Arrival at DC Under PO Partitioning

6.4 Discussion & Considerations

The multi-objective optimization model presented in this research contributes insights to CLP decision-making, but its current form has certain limitations and simplifying assumptions that do not fully capture real-world supply chain complexities. Future research can expand the operational utility of the model through refinement in the following areas:

- **Stochastic Transit Time Modeling & Dynamic Pricing**: Develop stochastic representations of transit time uncertainties, incorporating disruptions such as severe weather, port congestion, labor disputes, and multi-modal transportation delays. Introduce dynamic pricing models (particularly for spot market planning) to replace static container pricing.

- **Synchronized Logistics Scheduling**: Integrate production schedules with logistics planning to minimize inventory holding times and reduce associated costs. The current process isolates GCW from RDW, so systemic process changes are necessary to align production schedules more precisely with demand cycles.

- **Expanded Load Rules**: Incorporate more comprehensive operational constraints into the model, including detailed rules for product co-loading compatibility, specific customs regulations, and customer-centric delivery preferences. Employing advanced rule-based decision frameworks and constraint handling methodologies would provide greater realism and practical utility, better reflecting the nuanced complexities of global logistics operations.

- **Multi-Echelon & Multi-Lane Optimization**: Extend the model's scope beyond the distribution center endpoint to encompass downstream inventory management across first-mile, middle-mile, and last-mile operations. Incorporate constraints and objectives relevant to each logistics stage, such as warehouse capacity, regional inventory balancing, and timely fulfillment at retail and customer locations. Extend analysis beyond a single shipment lane by incorporating multiple transportation routes to provide a comprehensive picture of network-wide logistics operations.

- **Digital Twin Capability Building**: Create comphrensive emulator of load plans and establish model as decision-support tool through integration with enterprise resource planning (ERP) and transportation management systems (TMS). This integration can provide logistics planners with real-time, data-driven insights and improved scenario-testing capabilities.

- **Sustainability Integration**: Incorporate environmental performance indicators (e.g., carbon emissions, fuel consumption) into the multi-objective framework to align with growing industry emphasis on sustainable supply chain practices while maintaining the core objectives of cost efficiency, container utilization, and delivery precision.

The trade-offs between container utilization, delivery precision, and transportation cost emphasize the need for a more holistic approach to CLP that considers a fuller extent of operational impact, rather than focusing solely on volumetric efficiency.

6.5 Chapter Summary

This chapter presents a set of exploratory improvements to the container load planning process and extends the optimization model to test the effect of consolidation bypass and PO partitioning strategies on logistics performance. These use case applications demonstrate that process flow changes can help address operational challenges such as early deliveries, inventory spikes, and inefficient shipment sequencing. The chapter further discusses of model limitations and future opportunities for incorporating dynamic scheduling, multi-echelon planning, and system integration to strengthen decision support for logistics practitioners.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 7

Conclusions & Recommendations

Conclusions

Cargo containers play a crucial role in global supply chains, serving as the backbone of international trade. This research presents a multi-objective optimization model that helps balance three key operational goals: minimizing cost, improving delivery timing, and making strategic use of container space. Unlike many traditional approaches that focus solely on maximizing container utilization, the proposed model allows practitioners to explore trade-offs between Pareto front solutions and to select operating points that align with their specific organizational priorities and constraints.

One of the core findings is that maximizing container fill rates does not always lead to the lowest overall cost. In fact, doing so without considering the delivery timing and storage needs of inventory can create downstream bottlenecks and added holding costs. The model shows that a moderate fill rate coupled with better delivery precision leads to more stable inventory flow and reduced total logistics cost. This work also shows that high delivery precision can be achieved without a significant cost penalty. Ultimately, the model provides a flexible decision support tool that enables more informed trade-offs between cost, delivery timing, and utilization. Practitioners can tailor optimization priorities to specific business goals, ensuring that container loading strategies support not only cost efficiency but also broader supply chain resilience and warehouse management objectives.

Recommendations

Building on the findings of this research, several opportunities exist to advance both the practical utility and academic development of CLP optimization:

- **Workflow Redesign for Centralized Decision-Making**: The revised CLP workflow presented in this work highlights the value of centralizing planning decisions to reduce manual coordination and misaligned priorities between retail organizations and consolidators. Organizations can evaluate opportunities to reconfigure their logistics workflows to shift load planning authority upstream—toward the shipper or centralized control teams—rather than relying on decentralized channels and potentially suboptimal decision-making by third-party consolidators. Such shifts can enables more strategic control over PO assignment, shipment timing, and inventory holding, which can help reduce premature deliveries, administrative overhead, and downstream distribution center inefficiencies. Importantly, it also facilitates a more holistic planning approach that considers delivery timing, inventory storage costs, co-loading rules, and service-level requirements rather than solely optimizing for container fill rates. This integrated perspective enhances trade-off analysis between cost efficiency and operational performance.

- **Strategic Process Enhancements**: Extensions such as consolidation bypass and PO partitioning demonstrate significant potential to improve network speed and balance inventory flow. However, these strategies require structural changes to customs processing, ERP integration, and internal data infrastructure. There is a compelling case for exploring how digital twin platforms and event-driven triggers can be integrated with optimization models to better support advanced planning mechanisms.

- **Model Expansion**: Future research should expand the optimization framework to account for multi-echelon networks and multi-lane shipment configurations, which would allow for more comprehensive end-to-end planning across first-mile, middle-mile, and last-mile operations. Incorporating stochastic elements—such as variable transit times, uncertain demand, and dynamic order arrivals—would more closely reflect real-world uncertainty. In parallel, there is opportunity for embedding sustainability metrics, such as emissions per container or carbon intensity of shipping modes, into CLP frameworks to align logistics optimization with emergent environmental goals.

Through these improvements, CLP optimization can offer greater practical value to logistics practitioners while also supporting the development of more adaptive, efficient, and resilient global supply chain networks.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] AlixPartners. *The Hidden Costs of Excess Inventory.* 2023. URL: https://www.alixpartners.com/insights/102jo2h/the-hidden-costs-of-excess-inventory/.

[2] Henrik Andersson et al. "Industrial Aspects and Literature Survey: Combined Inventory Management and Routing". In: *Computers & Operations Research* 37.9 (2010), pp. 1515–1536. DOI: 10.1016/j.cor.2009.11.009.

[3] Eliseu Junio Araújo et al. "Pareto clustering search applied for 3D container ship loading plan problem". In: *Expert Systems with Applications* 44 (2016), pp. 50–57. DOI: 10.1016/j.eswa.2015.09.005.

[4] E.E. Bischoff and M.S.W. Ratcliff. "Issues in the development of approaches to container loading". In: *Omega* 23.4 (1995), pp. 377–390. DOI: 10.1016/0305-0483(95)00015-G.

[5] Dennis E. Blumenfeld et al. "Analyzing Trade-offs Between Transportation, Inventory, and Production Costs on Freight Networks". In: *Transportation Research Part B: Methodological* 19.5 (1985), pp. 361–380. DOI: 10.1016/0191-2615(85)90051-7.

[6] A. Bortfeldt and H. Gehring. "Applying Tabu Search to Container Loading Problems". In: *Lecture Notes in Computer Science.* 1998. DOI: 10.1007/978-3-642-58891-4_84.

[7] A. Bortfeldt, H. Gehring, and D. Mack. "A Parallel Tabu Search Algorithm for Solving the Container Loading Problem". In: *Parallel Computing* 29 (2003), pp. 641–662. DOI: 10.1016/S0167-8191(03)00047-4.

[8] A. Bortfeldt and G. Wäscher. "Constraints in container loading – A state-of-the-art review". In: *European Journal of Operational Research* 229.1 (2013), pp. 1–20. DOI: 10.1016/j.ejor.2012.12.006.

[9] James F. Campbell. "Freight consolidation and routing with transportation economies of scale". In: *Transportation Research Part B: Methodological* 24.5 (1990), pp. 345–361. DOI: 10.1016/0191-2615(90)90008-M.

[10] Fei Cao. "Study on Container Loading Problem Based on NSGA-α Algorithm". In: *Advanced Materials Research* 753–755 (Aug. 2013), pp. 2954–2958. DOI: 10.4028/www.scientific.net/AMR.753-755.2954.

[11] Sıla Çetinkaya and James H. Bookbinder. "Stochastic models for the dispatch of consolidated shipments". In: *Transportation Research Part B: Methodological* 37.8 (2003), pp. 747–768. DOI: 10.1016/S0191-2615(02)00060-7.

[12] Anna Collins, Dimitrios Letsios, and Gueorgui Mihaylov. "Algorithms for Shipping Container Delivery Scheduling". In: *arXiv* (2023). URL: https://arxiv.org/abs/2306.17789.

[13] McKinsey & Company. *Beyond the Distribution Center*. June 2022. URL: https://www.mckinsey.com/industries/retail/our-insights/beyond-the-distribution-center.

[14] McKinsey & Company. *Retail's Need for Speed: Unlocking Value in Omnichannel Delivery*. Accessed: March 3, 2025. URL: https://www.mckinsey.com/industries/retail/our-insights/retails-need-for-speed-unlocking-value-in-omnichannel-delivery.

[15] McKinsey Company. *Measuring the Fashion World*. 2025. URL: https://www.mckinsey.com/industries/retail/our-insights/measuring-the-fashion-world.

[16] Jens Egeblad and David Pisinger. "Heuristic Approaches for the Two- and Three-Dimensional Knapsack Packing Problem". In: *Computers & Operations Research* (2008). DOI: 10.1016/j.cor.2007.12.004.

[17] Michael Eley. "A Bottleneck Assignment Approach to the Multiple Container Loading Problem". In: *Container Terminals and Automated Transport Systems*. Ed. by Hans-Otto Günther and Kap Hwan Kim. Berlin, Heidelberg: Springer, 2005, pp. 359–374. DOI: 10.1007/3-540-26686-0_16.

[18] Dung D. Fom et al. *A Hybrid Multi-Objective Optimization Approach for Efficient 3D Container Loading Problem*. Aug. 2024. DOI: 10.21203/rs.3.rs-4854532/v1.

[19] Mikele Gajda et al. "An optimization approach for a complex real-life container loading problem". In: *European Journal of Operational Research* 307.1 (2022), pp. 261–277. DOI: 10.1016/j.omega.2021.102559.

[20] H. Gehring and A. Bortfeldt. "A Genetic Algorithm for Solving the Container Loading Problem". In: *European Journal of Operational Research* (1997). DOI: 10.1016/S0969-6016(97)00033-6.

[21] J.A. George and D.F. Robinson. "A heuristic for packing boxes into a container". In: *Computers & Operations Research* 7.3 (1980), pp. 147–156. DOI: 10.1016/0305-0548(80)90001-5.

[22] Rose George. *Ninety Percent of Everything: Inside Shipping, the Invisible Industry That Puts Clothes on Your Back, Gas in Your Car, and Food on Your Plate*. New York: Metropolitan Books, 2013.

[23] José Fernando Gonçalves and Mauricio G.C. Resende. “A biased random key genetic algorithm for 2D and 3D bin packing problems”. In: *International Journal of Production Economics* 145.2 (2013), pp. 500–510. DOI: 10.1016/j.ijpe.2013.04.019.

[24] José Fernando Gonçalves and Mauricio G.C. Resende. “A Parallel Multi-Population Biased Random-Key Genetic Algorithm for a Container Loading Problem”. In: *Computers & Operations Research* 39 (2011). DOI: 10.1016/j.cor.2011.03.009.

[25] A. Jugović. “Economic impact of container-loading problem”. In: *Transactions on Maritime Science* 9.2 (2020), pp. 271–278. DOI: http://dx.doi.org/10.7225/toms.v09.n02.010.

[26] L. Junqueira, R. Morabito, and D. Yamashita. “Three-dimensional container loading models with cargo stability and load bearing constraints”. In: *Computers & Operations Research* 39.1 (2012), pp. 74–85. DOI: 10.1016/j.cor.2010.07.017.

[27] Peter J. Kolesar. “A Branch and Bound Algorithm for the Knapsack Problem”. In: *Management Science* 13.9 (1967), pp. 723–735. DOI: 10.1287/mnsc.13.9.723.

[28] D.V. Kurpel et al. “Exact solutions of several types of container loading problems”. In: *European Journal of Operational Research* 284.1 (2020), pp. 87–107. DOI: 10.1016/j.ejor.2019.12.012.

[29] Yui-yip Lau et al. “Maritime transport resilience: A systematic literature review on the current state of the art, research agenda and future research directions”. In: *Ocean & Coastal Management* 251 (May 2024), p. 107086. DOI: 10.1016/j.ocecoaman.2024.107086.

[30] Nang Laik Ma and Kar Way Tan. “Reducing Carbon Emission through Container Shipment Consolidation and Optimization”. In: *Journal of Traffic and Transportation Engineering* 7 (2019), pp. 111–121. DOI: 10.17265/2328-2142.

[31] Noha A. Mostafa and Omar Eldebaiky. “A Sustainable Two-Echelon Logistics Model with Shipment Consolidation”. In: *Logistics* 7.1 (2023), p. 18. DOI: 10.3390/logistics7010018.

[32] Fatih Mutlu, Sıla Çetinkaya, and James H. Bookbinder. “An analytical model for computing the optimal time-and-quantity-based policy for consolidated shipments”. In: *IIE Transactions* 42.5 (2010), pp. 367–377. DOI: 10.1080/07408170903462368.

[33] Christine Nguyen, Maged Dessouky, and Alejandro Toriello. “Consolidation Strategies for the Delivery of Perishable Products”. In: *Transportation Research Part E: Logistics and Transportation Review* 69 (2014), pp. 108–121. DOI: 10.1016/j.tre.2014.05.018.

[34] ODW Logistics. *Freight Consolidation Benefits and Case Studies*. https://www.odwlogistics.com/resources/freight-consolidation-benefits-case-studies. 2025.

[35] J. T. Patil and M. E. Patil. "Cargo space optimization for container". In: *International Conference on Global Trends in Signal Processing, Information Computing and Communication*. IEEE, 2016, pp. 310–314. DOI: 10.1109/ICGTSPICC.2016.7955271.

[36] S. Phongmoo et al. "Artificial Bee Colony Algorithm with Pareto-Based Approach for Multi-Objective Three-Dimensional Single Container Loading Problems". In: *Applied Sciences* 13.11 (2023), p. 6601. DOI: 10.3390/app13116601.

[37] David Pisinger. "Heuristics for the container loading problem". In: *European Journal of Operational Research* 141.2 (Sept. 2002), pp. 382–392. DOI: 10.1016/S0377-2217(02)00132-7.

[38] Siim Sild. *Supply Chain Issues: Port Congestion*. 2024. URL: https://fractory.com/port-congestion-explained/.

[39] Business for Social Responsibility. *Calculate and Manage Transport Supply Chain Greenhouse Gas Emissions*. 2015. URL: https://www.bsr.org/reports/BSR_CCWG_Calculate_Manage_Emissions_2015.pdf.

[40] Dongping Song. "A Literature Review, Container Shipping Supply Chain: Planning Problems and Research Opportunities". In: *Logistics* 5.2 (2021), p. 41. DOI: 10.3390/logistics5020041.

[41] Statista. *Apparel supply chain worldwide - Statistics & Facts*. 2024. URL: https://www.statista.com/topics/8824/apparel-supply-chain-worldwide/#topicOverview.

[42] Statista. *Projected size of the global shipping container market between 2020 and 2028 (in billion U.S. dollars) [Graph]*. Apr. 2022. URL: https://www-statista-com.libproxy.mit.edu/statistics/1097059/global-shipping-containers-market-size/.

[43] T. Tlili, S. Faiz, and S. Krichen. "A Particle Swarm Optimization for Solving the One Dimensional Container Loading Problem". In: *2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*. 2013, pp. 1–4. DOI: 10.1109/ICMSAO.2013.6552590.

[44] Trading Economics. *Containerized Freight Index*. 2024. URL: https://tradingeconomics.com/commodity/containerized-freight-index.

[45] Jonah C Tyan, Fu-Kwun Wang, and Timon C Du. "An evaluation of freight consolidation policies in global third party logistics". In: *Omega* 31.1 (2003), pp. 55–62. DOI: 10.1016/S0305-0483(02)00094-4.

[46] Muhammed Ali Ülkü. "Analysis of Shipment Consolidation in the Logistics Supply Chain". PhD thesis. University of Waterloo, 2009.

[47] United Nations Conference on Trade and Development. *Review of Maritime Transport*. Report. UNCTAD, 2024. URL: https://unctad.org/system/files/official-document/rmt2024_en.pdf.

[48] Hiroyuki Yamazaki et al. "The 3D-Packing by Meta Data Structure and Packing Heuristics". In: *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* E83-A.4 (2000), pp. 639–646. URL: https://globals.ieice.org/en_transactions/fundamentals/10.1587/e83-a_4_639/_p.

[49] L. Zerfa and M.E.A. Chergui. "An efficient branch-and-bound algorithm to optimize a function over a nondominated set". In: *International Transactions in Operational Research* (2024). DOI: 10.1111/itor.13547.

[50] Q. Zhou and X. Liu. "A Swarm Optimization Algorithm for Practical Container Loading Problem". In: *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*. 2017, pp. 5690–5695. DOI: 10.1109/IECON.2017.8216987.