

Elucidating Import Container Flows: A Simulation Study of Port of New York/New Jersey

by

Kevin Power

Bachelor of Technology in Manufacturing Engineering Technology, McMaster University, 2022

and

Yassine Lahlou-Kamal

Bachelor of Applied Science in Chemical Engineering, University of Toronto, 2020

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

© 2025 Kevin Power and Yassine Lahlou-Kamal. All rights reserved.

The authors hereby grant to MIT permission to reproduce and to distribute publicly paper and electronic copies of this capstone document in whole or in part in any medium now known or hereafter created.

Signature of Author: _____
Kevin Power
Department of Supply Chain Management
May 9th, 2025

Signature of Author: _____
Yassine Lahlou-Kamal
Department of Supply Chain Management
May 9th, 2025

Certified by: _____
Elenna R. Dugundji, Ph.D.
Research Scientist
Capstone Advisor

Certified by: _____
Thomas Koch, Ph.D.
Postdoctoral Associate
Capstone Co-Advisor

Accepted by: _____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Elucidating Import Container Flows: A Simulation Study of Port of New York/New Jersey
by
Kevin Power
and
Yassine Lahlou-Kamal
Submitted to the Program in Supply Chain Management
on May 9th, 2025 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

This study develops a discrete-event simulation model using open-source software to analyze import container flows through the Port of New York/New Jersey. The simulation integrates parameters from extensive data analysis of vessel arrivals, container dwell times, and intermodal transfers. Data is sourced from multiple public sources including vessel GPS data and import records released by Customs and Border Patrol. The model is calibrated with real-world import data to ensure accuracy. A fine-tuned BERT model is used to predict Harmonized System (HS) codes using unstructured shipping manifest text, achieving over 90% classification accuracy at the two-digit level. This classification enables commodity-specific analysis of dwell time. Dwell time analysis results show that for dry containers, Gaussian KDE reduces mean absolute error by 39.5% relative to the normal distribution. For refrigerated containers, the Fourier Series model yields a 24.8% reduction. Scenario testing allows us to quantify the effects of changes in key variables such as container arrival volumes, resource allocation, and truck-rail modal split – performance metrics like throughput, truck congestion, and container dwell times are captured. Scenario testing reveals that increasing the outbound rail share from 15% to 25% reduces truck congestion by 11.5% and reduces median dwell times by 1.52% for dry containers and 2.55% for reefers. Extending gate hours by two hours per day leads to a 4.42% drop-in median dwell time for dry containers and 6.28% for reefers. The model provides a scalable, data-driven framework for evaluating operational and policy interventions. Further research can expand on short-haul rail utilization, extended gate hours, or inland infrastructure investments.

Capstone Advisor: Elenna R. Dugundji, Ph.D.
Title: Research Scientist

Capstone Co-Advisor: Thomas Koch, Ph.D.
Title: Research Scientist

Table of Contents

1. Introduction	4
1.1. Motivations	4
1.2. Problem Statement & Key Questions	4
1.3. Project Goals & Expected Outcomes	5
2. Literature Review	5
2.1. Current Comprehensive Literature Reviews	5
2.2. Terminal Operations in the era of Megaships	6
2.3. Traffic Flow Management within Port Gates	7
2.4. Simulation in Port and Intermodal Logistics	7
2.5. Intermodal Transport and Short-Haul Viability	8
2.6. Simulation for Rail Integration and Port Integration	9
2.7. The NY/NJ–Lehigh Valley Corridor	10
2.8. Unpacking the "Chicken-and-Egg Dilemma" in Rail Adoption.....	11
3. Methodology	12
3.1. Methodology Selection	12
3.2. Data Sources and Preparation	12
3.3. Model Construction	14
3.3.1. Model Architecture and Logic	14
3.3.2. Assumptions and Constraints	15
3.3.3. Validation and Calibration.....	16
3.3.4. Key Performance Metrics	17
3.3.5. User Interface and Simulation Workflow	18
4. Results and Discussion	19
4.1. Dwell Time Analysis	19
4.2. Baseline scenario analysis	20
4.3. Impact of Increased Rail Adoption on Port Performance.....	23
4.4. Commodity Categorization.....	25
5. Conclusion	30
6. Future Research	31
References	33
Appendix A.....	36
Understanding the Port of New York/New Jersey	36
Understanding Inbound Container Processes at the Port of NY/NJ.....	37

1. Introduction

Seaports are pivotal to the U.S. economy, managing over one-quarter of the nation's Gross Domestic Product (GDP) and facilitating the movement of more than \$6 billion worth of goods daily¹. The efficiency of port operations is indispensable, not only for millions of jobs, but also for maintaining stable consumer prices and the broader economic health of the nation. Inefficiencies in port operations propagate beyond the ports themselves, significantly impacting both retailers and consumers. During the supply chain crisis of 2021–2022, port congestion and delays led to a 17% surge in shipping costs, which ultimately were passed on to consumers through increased prices for goods.²

1.1. Motivations

Many factors contribute to inefficiencies within ports. A key issue is the lack of coordinated collaboration among stakeholders. Each container shipment involves numerous actors; from ocean carriers and terminal operators to trucking companies and beneficial cargo owners, each optimizing for their own objectives rather than for the performance of the overall system. This fragmented decision-making leads to suboptimal operations.

These dynamics create a system-level coordination problem. The issue is not just modal imbalance or infrastructure gaps, but that decisions are made in silos, without visibility into their broader consequences. Efforts to address these challenges seek to improve transparency and alignment through data sharing among public and private actors.

However, data sharing alone is not sufficient. Without tools to turn that data into actionable insights, coordination remains difficult. Our work responds to this gap by building a simulation model of import container flows through the Port of New York and New Jersey (NY/NJ) using publicly available data. The model allows users to experiment with changes in volume, capacity, and modal mix, and observe their effects on throughput, yard utilization, and container dwell times.

The model also draws on Harmonized System (HS) code-level information to capture the diversity of container contents. Different types of goods flow through the port in different ways, yet this variation is often buried in unstructured data. In the process of calibrating our simulation, we extract insights that would otherwise remain inaccessible, making it possible for stakeholders to better understand flow patterns and potential levers for intervention. This enables more informed, granular decisions without requiring each actor to develop advanced analytical tools in-house.

In this way, our simulation serves as a decision-support tool. It shows stakeholders how their local choices affect system-wide outcomes, helping to bridge the gap between siloed decision-making and coordinated action. It also provides a framework to explore segmentation and targeted policies, such as rail incentives or dwell-time penalties, based on the actual structure of container flows.

1.2. Problem Statement & Key Questions

Despite their critical role in the economy, U.S. seaports continue to suffer from chronic inefficiencies that cascade across the supply chain. These inefficiencies are not only costly, but they also undermine resilience and adaptability in the face of disruption.

The core problem we address is the absence of system-level decision-support tools that help stakeholders evaluate the impact of their actions within a complex, interdependent environment. Without such tools, it is difficult to identify high-leverage interventions or to coordinate effectively across modes and actors.

¹ FreightWaves Staff. (2019, March 20). *Study finds U.S. ports contributed \$5.4 trillion to GDP*. FreightWaves ([URL](#))

² Atkinson, H., & Bowman, N. (2024, October 2). *Strike at U.S. ports brings debate over automation front and center*. SupplyChainBrain ([URL](#))

This leads us to the following key questions:

- How do changes in container volume, port resource allocation, and modal mix affect congestion, throughput, and dwell time at the Port of NY/NJ?
- How can we use simulation to make the downstream effects of stakeholder decisions more visible?
- Can we leverage container-level insights (e.g., HS code patterns) to inform more targeted policies and operational strategies?

By exploring these questions through a discrete-event simulation model, we aim to provide a framework that helps port stakeholders (public and private) design, evaluate, and refine interventions with a clear view of system-wide consequences.

1.3. Project Goals & Expected Outcomes

The goal of this project is to build a discrete-event simulation model that helps stakeholders evaluate the system-level consequences of decisions made at the Port of New York and New Jersey. Rather than optimizing for any single stakeholder, the model enables exploration of cross-cutting interventions, such as rail incentives, equipment constraints, or shifts in container volumes, that ripple across the port ecosystem. In developing this model, we aim to:

- Quantify the impact of key variables (e.g., volume surges, yard resource changes, modal splits) on congestion, throughput, and dwell times
- Support experimentation with what-if scenarios, helping stakeholders anticipate unintended effects
- Provide a decision-support framework to test and compare potential policy or operational changes
- Make complex dynamics accessible through visualization, enabling both public agencies and private actors to align around shared outcomes

This tool is not a forecasting engine, it is a system sandbox. It allows for structured experimentation to better understand how port performance can be influenced by coordinated changes across infrastructure, policy, and behavior. The longer-term goal is to build a scalable framework that can be extended to other U.S. ports and inform national-level supply chain coordination efforts.

2. Literature Review

To effectively improve the flow of import containers at the Port of New York and New Jersey, it is essential to ground our research in the existing scholarly discourse. This literature review synthesizes key insights that inform the development of our discrete-event simulation model.

We begin with operational aspects of marine terminals, focusing on how the rise of mega-ships and evolving yard practices affect throughput and container dwell times. We then examine strategies for managing truck traffic and gate congestion inside the port. From there, we evaluate the role of simulation in capturing the complexities of port operations and intermodal transitions. Building on that foundation, we explore intermodal transport planning, particularly the viability of short-haul rail in dense port regions. We conclude with a focus on the unique dynamics between the Port of NY/NJ and inland logistics hubs such as the Lehigh Valley.

By organizing the literature in this sequence, from internal operations to external intermodal considerations, we identify practical gaps and establish a foundation for modeling strategies aimed at improving system-wide efficiency, sustainability, and coordination.

2.1. Current Comprehensive Literature Reviews

In 2024, (Abu-Aisha, Audy, & Ouhimmou, 2024) conducted a systematic literature review examining Sea-Rail Intermodal Transportation Systems (SRITS). Their findings reveal that 72% of the analyzed research

articles on SRITS were published between 2015 and 2023, while the remaining 28% were published between 2000 and 2014. The study highlights that most research has centered on optimization models at the operational level, leaving a notable gap in understanding dynamics of sea-rail intermodal operations. The researchers also observe a lack of literature employing simulation techniques to address SRITS-related challenges. Furthermore, the feasibility of integrating short-haul rail systems to mitigate port congestion is absent from the reviewed studies.

(Lange, Schweintek, & Jahn, 2017) conducted an extensive literature review to classify methods aimed at reducing congestion at container terminals. Their analysis addresses the impact of increasing container ship sizes on peak truck activity, the environmental consequences of idling trucks, and financial burdens faced by drayage truck drivers.

2.2. Terminal Operations in the era of Megaships

A container must go through several terminal operations to move from a vessel to a truck or train. Terminal operations is a well-studied area within academic literature; however, this is an evolving field due to technological advances in maritime shipping, namely the advent of megaships. The increasing size of container ships is a major factor driving change in terminal operations, as they require more containers to be handled in shorter periods of time, demanding improved planning and efficiency from container terminals. Ultra Large Container Vessels (ULCVs), with capacities up to 24,000 TEUs, can bring up to 7,000-8,000 containers per call, triple the volume compared to two decades ago (Carboni, Deflorio, Caballini, & Cangelosi, 2024).

To stay competitive, terminals must guarantee an acceptable throughput and an adequate level of service. This puts significant pressure on infrastructure upgrades and prompts changes in operations, both at the quayside and the landside. Additionally, this necessitates continually optimizing management practices, especially considering the substantial infrastructure investments required. (Musso & Sciomachen, 2020).

In recent years, advancements in quayside equipment have shifted the bottleneck in terminal operations from quayside processes to yard-side processes (Ambrosino & Xie, Optimization approaches for defining storage strategies in maritime container terminals, 2023). Numerous studies have explored yard optimization algorithms, aiming to reduce terminal throughput times in response to the increasing container volumes per vessel. Some strategies focus on clustering; (Wang, Zhou, & Wang, 2022) propose a cluster-based yard template design method that reserves storage space based on the container destination and origin, grouping containers with similar characteristics to improve utilization and reduce reshuffling operations. (Zhou, Wang, & Li, Container reshuffling considered space allocation problem in container terminals, 2020) investigates the impact of container reshuffling on space allocation in container terminals, proposing a flexible storage strategy where space boundaries adjust based on vessel occupation to optimize space utilization.

Further exploration of yard optimization strategies includes mathematical or heuristic optimization; (Ambrosino & Xie, Optimization approaches for defining storage strategies in maritime container terminals, 2023) compare yard storage strategies for varying transport demands using a mixed-integer linear programming model and a heuristic approach.

(Ambrosino & Xie, Machine Learning-Based Optimization Models for Defining Storage Rules in Maritime Container Yards, 2024) explore the use of machine learning and optimization to develop efficient storage strategies for export yards, combining spectral clustering for grouping containers with mixed-integer programming for assigning groups to bay locations.

(Wang, Zhao, & Mi, 2023) present a novel model for container space allocation that aims to maximize the "attractiveness" of yard spaces, considering factors such as operational efficiency and container flow dynamics to optimize yard operations.

2.3. Traffic Flow Management within Port Gates

Beyond managing container flows, managing commercial vehicle traffic flow on internal roads is critical for reducing the negative impacts of congestion, both within the terminal and beyond its gates (Carboni, Deflorio, Caballini, & Cangelosi, 2024). Congestion in and around terminal areas can lead to significant traffic problems (Ambrosino & Peirano, Truck Arrival Management at Marine Container Terminals, 2016). Recent research has focused on managing commercial vehicle traffic flow within ports to reduce congestion and its associated negative impacts:

- (Yıldırım & Kawasaki, 2023) employ an optimised truck management strategy to reduce congestion at container terminal gates. Using the Arena simulation tool, they develop a discrete-event microsimulation model to simulate truck arrivals, gate queuing, and service times, analyzing truck waiting times.
- (Du, Zhang, & Zhang, 2023) leverage Anylogic software to simulate and model internal port traffic, evaluating the effects of traffic organization and streamline design on traffic flow efficiency.
- (Carboni, Deflorio, Caballini, & Cangelosi, 2024) present a methodological approach for modeling and simulating road traffic flows within container terminals using the micro-simulation tool Aimsun. They evaluate the impact of different operational scenarios, including disturbances like yard closures, on the efficiency and sustainability of the terminal.

2.4. Simulation in Port and Intermodal Logistics

The studies above demonstrate the widespread adoption of simulation as a robust method for modeling complex logistics systems. Simulation models provide a comprehensive representation of a logistics system, encompassing nonlinearity, uncertainty, and complexity (Jackson, Saenz, & Ivanov, 2023). By experimenting with various control strategies in a simulated environment, businesses can conduct scenario planning, fine-tune their processes, and identify the most effective approach for their unique supply chain needs (Phandis, Sheffi, & Caplice, 2022). Simulation not only provides valuable decision-making insights but also enhances the decision-maker's understanding of the underlying operational principles of the logistics system. (Wang & Wang, 2013).

There are various approaches to developing simulation models. According to a critical analysis of case studies by (Maina & Mwangangi, 2020), system dynamics models are best suited for problems involving continuous processes that require feedback to model behavior changes, while discrete event simulation (DES) is more appropriate for systems reliant on detailed and empirical data for linear processes. Simulating container flows through a port involves linear, process-driven operations such as container unloading, customs inspections, yard storage, and intermodal transfers. Accurately modeling these processes requires detailed and empirical data, making a discrete event simulation model the most suitable choice for simulating terminal operations.

Recent research applying discrete event simulation (DES) to port operations demonstrates a variety of approaches, assumptions, and software tools. The following studies explore this topic and highlight their methods and contributions to understanding terminal operations:

- (Davies, Islam, Gheith, & Eltawil, 2024) focuses on optimising truck appointment systems (TAS) in container terminals using a combination of DES and mixed integer programming. The simulation model is developed using AnyLogic software. Validation is performed by comparing model outputs to results from similar studies in the literature and using parameters from well-established sources.
- (Cavada, Cortés, & Rey, 2023) examines container management strategies for logistics services container terminals, a specific type of inland terminal primarily serving hinterland markets. The researchers employ a discrete-event simulation approach built using the Python programming language and the SimPy library.

- (Musso & Sciomachen, 2020) investigates the impact of megaships on container terminal performance, focusing on dwell times, throughput, and berthing times. They model DES using Witness software. The study explores different scenarios based on varying sizes of container ships and aim to achieve a balanced modal split between rail and road transport for container exits.
- (Zhou, Zhao, & Li, Simulation optimization iteration approach on traffic integrated yard allocation problem in transshipment terminals, 2021) propose a simulation optimization iteration approach, where a mixed integer programming model interacts with a DES model for traffic movement. They do not mention the software used for DES modelling.

2.5. Intermodal Transport and Short-Haul Viability

Numerous studies highlight the potential of short-haul rail services and intermodal strategies to alleviate congestion at ports and improve the overall efficiency and sustainability of freight transportation systems. For instance, (Abu-Aisha, Audy, & Ouhimmou, 2024) emphasize the effectiveness of simulation techniques in diagnosing and resolving bottlenecks in sea-rail systems. Their case study of the Port of Trois-Rivières illustrates how optimizing railcar movements and increasing the share of rail transport can enhance port capacity, reduce cargo dwell time, and streamline freight flows. Similarly, (Perrykkad, Ernst, & Krishnamoorthy, 2020) underscore the importance of simulation models that incorporate both road and rail modalities. Such integrated models help stakeholders evaluate the feasibility of intermodal solutions, especially in urban areas where infrastructure constraints, demand fluctuations, and modal interactions must be carefully managed.

Despite this, short-haul rail faces significant challenges related to cost, infrastructure, and demand consistency. Addressing these barriers often involves operational interventions at the terminal level. (Maguire, Ivey, Golias, & Lipinski, 2010), for example, propose both strategic capacity expansions and more tactical measures—such as extended gate hours and appointment systems—to reduce truck queues, smooth demand peaks, and minimize unpredictability in yard operations. These actions, while not directly resolving systemic cost or infrastructure issues, mitigate immediate congestion problems and create a more stable operating environment. (Giuliano & O'Brien, 2007) add that such strategies are most effective when paired with incentives for drayage operators, adjustments in labor practices, and off-peak deliveries. This integrated, stakeholder-focused approach suggests that while short-haul rail must contend with broader economic and infrastructural hurdles, targeted operational improvements at terminals can help lay the groundwork for more efficient intermodal networks.

Economic feasibility remains central to the adoption of short-haul rail. (Kordnejad, 2016) analyze intermodal terminals in locations such as the Duss Munich Reim terminal and the Port of Valencia, demonstrating that cost-competitiveness requires optimizing loading space, reducing transshipment costs, and incorporating total cost analyses—often including state-funded infrastructure investments. They also highlight the potential of automated shunting processes, particularly for small-scale linear terminals, to lower costs and increase efficiency, thereby enhancing the financial attractiveness of short-haul rail solutions.

As the industry seeks more adaptive and integrated approaches, new concepts like synchronomodality have emerged. (Acero, 2021) introduces synchronomodality as a concept that goes beyond traditional intermodal transport by emphasizing real-time information sharing and dynamic selection of transport modes based on real-time conditions. Although promising, the empirical evidence for synchronomodality's effectiveness—especially in short-haul contexts—remains limited. More granular insights into factors affecting port efficiency can be found in the work of (De Armas Jacomino, 2021), who examine how container dwell time influences performance. Their research identifies multiple factors affecting dwell time, including container type and status, proximity of industrial hubs, rail frequency, pre-inspection processes, and the availability of tracking technologies. By addressing these factors, ports can reduce dwell times and enhance the viability of incorporating rail into their operational mix.

Despite these advancements, the literature still presents notable research gaps. (Macharis & Bontekoning, 2004), along with (Caris, Macharis, & Janssens, Planning problems in intermodal freight transport: Accomplishments and prospects, 2008) and (Caris, Macharis, & Janssens, Decision support in intermodal transport: A new research agenda, 2013), point out the underdevelopment of simulation methods tailored specifically to intermodal freight planning. Data availability and variability pose challenges in short-haul scenarios, where granular, real-time information is crucial for accurate modeling. (Tongzon, 1995) reminds us of the need for holistic frameworks that consider infrastructure, technology, management practices, and external economic conditions to assess port performance thoroughly. However, the lack of comprehensive data-driven strategies and comparative analyses across multiple ports leaves room for further inquiry.

Future research should thus concentrate on developing robust simulation frameworks that capture the complexity of short-haul intermodal operations, crafting detailed cost-benefit analyses that incorporate economic, environmental, and social dimensions, and examining the role of policy and regulation in creating favorable conditions for short-haul rail. By filling these gaps and building on existing operational and conceptual insights, stakeholders can foster more efficient, sustainable, and competitive freight transportation networks.

2.6. Simulation for Rail Integration and Port Integration

Numerous studies demonstrate the use of simulation as a powerful tool to model the complexities of port operations and assess the impact of various strategies, including increased rail adoption. (Guimarans, Harabor, & Van Hentenryck, 2015) use a combination of data analytics and discrete-event simulation to examine rail operations at Port Botany in Australia and challenge prevailing perceptions about its capacity constraints. They analyze six months of operational data, focusing on metrics like train timeliness, utilization, terminal usage, and yard congestion. Contrary to prevailing beliefs, they found that rail resources were significantly underutilized, and existing infrastructure had ample capacity to handle projected growth in container volumes. Their simulation model focused on the "last mile" rail corridor, demonstrating that operational improvements, rather than costly infrastructure investments, were key to unlocking the port's potential. They suggest strategies like flexible servicing windows, improved staging practices, and dedicated train services to optimize rail usage.

Simulation models are particularly valuable in capturing the dynamic interactions between various components of the port ecosystem and identifying potential bottlenecks. (Vidović, Krstić, & Zrnić, 2011) propose a combined approach using a deterministic p-hub location model and simulation to address the challenge of intermodal terminal location selection in Serbia. The p-hub model helps determine optimal locations for intermodal terminals, while the simulation model evaluates the chosen locations based on economic, time, and environmental criteria. This approach demonstrates the value of integrating optimization techniques with simulation to analyze the impact of infrastructure decisions on the overall performance of the intermodal transportation system. They emphasize the importance of considering factors like transport costs, time savings, and environmental benefits in evaluating intermodal terminal locations.

The calibration of simulation models with limited or aggregated data presents a common challenge, requiring careful consideration of assumptions and methodologies. (Abu-Aisha, Audy, & Ouhimmou, 2024) use a discrete-event simulation to investigate bottlenecks in the sea-rail connection at the Port of Trois-Rivières. Their model captures the movement of wagons between the train yard and the loading/unloading points, focusing on factors like train arrival and departure times, wagon loading and unloading durations, and the number of wagons per train. To calibrate their model, they collected real-world data from the port, including ship arrival patterns, cargo volume, and operational parameters. Recognizing the limitations of relying solely on historical data, they conducted multiple simulation replications to account for uncertainties and obtain robust results. Their analysis led to the identification of key bottlenecks in the system and the evaluation of

different scenarios to improve port efficiency. Their study underscores the value of simulation in identifying system vulnerabilities and testing the efficacy of various improvement strategies.

Despite these valuable insights, the sources also reveal some research gaps that necessitate further exploration. Notably, there is a need for more studies focusing on the calibration and validation of port simulation models using limited data, exploring innovative approaches for data aggregation and imputation, and incorporating techniques like sensitivity analysis to evaluate the robustness of model outputs. Further research should also focus on:

- Developing standardized frameworks for evaluating the environmental impacts of different intermodal strategies, including the assessment of factors beyond emissions, such as noise pollution and land use.
- Examining the role of emerging technologies, like automation and intelligent transportation systems, in enhancing the efficiency and sustainability of rail operations within ports.
- Investigating the impact of policy interventions, such as incentives for rail adoption and congestion pricing for trucks, on the dynamics of intermodal transportation systems.

Addressing these research gaps will enhance the understanding of the complex interplay between rail integration, port congestion, and environmental sustainability, enabling more informed decision-making and fostering the development of robust and sustainable port ecosystems.

2.7. The NY/NJ–Lehigh Valley Corridor

The Port of NY/NJ's heavy reliance on trucking presents unique challenges and opportunities for promoting rail-based intermodal transportation. The study by (Maguire, Ivey, Golias, & Lipinski, 2010) emphasizes the critical importance of efficient gate operations at intermodal marine container terminals, highlighting that inefficient gate operations can cause congestion and safety issues in surrounding areas, impacting carriers, shippers, and terminal operators. The authors note that intermodal terminals are often located in densely populated areas, like the NY/NJ region, where physical expansion is costly and impractical. This underscores the need for operational strategies to improve efficiency and reduce truck-related congestion, highlighting the urgency of exploring alternative modes of transportation like rail to mitigate the strain on the region's road infrastructure and address the environmental impacts associated with heavy truck traffic.

Lehigh Valley, with its existing logistics infrastructure and strategic location, emerges as a potential destination for rail-based solutions aimed at diverting cargo from truck-intensive routes. (Oliver Wyman and Leachman and Associates for the Pacific Harbor Line, 2023) on short-haul intermodal service, while focusing on the San Pedro Bay ports, provides valuable insights that can be applied to the NY/NJ context. They highlight the success of rail shuttle-inland port services at other North American ports like Charleston, Virginia, Savannah, and Vancouver. These services, often integrated with logistics parks offering warehousing and transloading capabilities, demonstrate the viability of shifting freight from truck to rail for both local and long-haul destinations. The report underscores the importance of public-private partnerships, incentives, and supportive policies to ensure the success of such initiatives. The authors also stress the need for a comprehensive benefit-cost analysis, considering factors like reduced truck traffic, emissions reduction, economic growth, and supply chain resilience, to justify public funding for inland port development. Applying these concepts to the NY/NJ region, Lehigh Valley, with its proximity to the port and existing warehousing capacity, could serve as a strategic inland port, facilitating the transfer of containers from trucks to rail, reducing congestion in the port's immediate vicinity, and offering shippers a more efficient and cost-effective alternative to direct trucking.

However, there is a notable lack of specific studies that directly examine the intermodal dynamics between the Port of NY/NJ and Lehigh Valley. While (Kordnejad, 2016) explores the feasibility of short-haul

intermodal rail in general terms, and (Craig et al., 2013) discuss policy interventions for promoting intermodal transportation, these studies do not specifically address the NY/NJ-Lehigh Valley corridor. Further research is needed to understand:

- The specific demand patterns for freight moving between the Port of NY/NJ and Lehigh Valley, including the types of cargo, shipment volumes, and destination markets.
- The operational and economic feasibility of establishing a dedicated rail shuttle service between the port and Lehigh Valley, considering factors like infrastructure requirements, train schedules, and cost competitiveness with trucking.
- The potential environmental benefits of such a service, including reduced emissions, noise pollution, and road congestion, as well as the challenges and opportunities associated with incorporating sustainability considerations into the design and implementation of the intermodal system.

Filling these research gaps will provide valuable insights for policymakers, port authorities, and logistics providers, enabling them to make informed decisions regarding investments in rail infrastructure, development of intermodal services, and implementation of supportive policies to promote a more balanced and sustainable freight transportation system in the NY/NJ region.

2.8. Unpacking the "Chicken-and-Egg Dilemma" in Rail Adoption

Many of the studies grapple with the inherent challenges in promoting rail adoption, particularly the interconnectedness of demand and service quality, which often presents a "chicken-and-egg" scenario. (Kordnejad, 2016) explores the factors influencing the feasibility of short-haul intermodal rail freight systems in the greater Stockholm-Mälaren region of Sweden, recognizing the pivotal role of stakeholder perspectives and the need to satisfy both policy objectives and commercial interests. The author emphasizes the role of "cost-leadership" as a primary driver for shipper adoption of rail but also acknowledges the importance of other service attributes such as accessibility, flexibility, and reliability. (Kordnejad, 2016) argues that achieving cost-competitiveness for rail often requires high and balanced loading space utilization, which in turn depends on the aggregation of sufficient demand from multiple shippers. This creates a cyclical challenge: attracting sufficient demand to justify frequent and reliable rail services, while simultaneously needing those services to entice shippers away from the convenience of road transport.

A recurring theme in these studies is the need for data-driven solutions to analyze demand patterns, optimize service offerings, and address bottlenecks that hinder rail adoption. (Guimarans, Harabor, & Van Hentenryck, 2015) provide a compelling example of this approach by using data analytics and simulation to challenge the prevailing notion of rail capacity constraints at Port Botany. Their analysis of operational data reveals that rail resources are underutilized, and existing infrastructure has ample capacity to handle anticipated growth in container volumes. They propose operational improvements such as flexible servicing windows, optimized staging practices, and "dedicated" train services to enhance rail utilization and attract shippers by offering more competitive and reliable services. This data-driven approach demonstrates the importance of moving beyond assumptions and anecdotal evidence to inform decision-making and identify practical solutions for resolving the demand-service interdependency challenge.

The sources also highlight the critical role of policy interventions, incentives, and partnerships in overcoming barriers to rail adoption and fostering a more balanced and sustainable transportation system. (Craig, Blanco, & Sheffi, 2013) in their analysis of intermodal freight transportation, emphasize the potential of policy measures to accelerate the shift from road to rail. They suggest the creation of "carbon market areas," where emissions savings from intermodal transport could be monetized, incentivizing shippers to choose rail and promoting investment in intermodal infrastructure. Additionally, they argue that policies aimed at internalizing the external costs of trucking, such as congestion pricing and stricter emissions regulations, could create a more level playing field for rail and encourage a shift towards more

environmentally friendly transportation modes. (Kordnejad, 2016) further underscores this point by highlighting the success of short-haul intermodal rail services in regions with strong policy support and regulations that promote a modal shift. They cite examples like the E&S system in Japan and the Innovatrain system in Switzerland, where favorable policy environments have fostered the growth of competitive and sustainable rail-based logistics systems.

3. Methodology

3.1. Methodology Selection

The DES model aims to quantify how operational changes, such as increased rail adoption or policy interventions, affect key performance metrics like dwell time, gate congestion, and throughput. The methodology follows a two-part structure:

- Part I focuses on building the simulation model: sourcing and preparing relevant data, designing simulation logic, and calibrating parameters specific to NY/NJ.
- Part II uses the completed model to evaluate operational and policy scenarios under different assumptions, providing a structured environment for experimentation and decision support.

The DES approach is well-suited to the port environment, where operations are dynamic, resource-constrained, and driven by events such as vessel arrivals, crane operations, yard transfers, and inland departures. Compared to continuous models, DES better captures the stochastic nature of queueing, scheduling, and resource utilization in a port system. Simulation also allows for detailed “what-if” analyses, showing how system-level outcomes respond to changes in inputs like volume, infrastructure, or policy.

All modeling is done in Python, leveraging open-source library SimPy for event-based logic. The choice of SimPy as the main DES modelling software provides a reliable and highly customizable framework with an open-source tool. Developing the framework with open-source tools was a deliberate choice to enable future researchers and stakeholders to build upon it freely, without the burden of licensing costs.

3.2. Data Sources and Preparation

Building a realistic simulation requires diverse data sources that capture operational patterns, infrastructure constraints, and demand characteristics:

Vessel Arrivals and Container Volumes

- **AIS Vessel Data:** Collected via radio frequency, this dataset provides detailed vessel-level spatiotemporal data (arrival times, anchorage durations, berthing intervals) that inform the stochastic arrival distributions of ships. It is analyzed in detail by (Aristov et al., 2024), offering insights into ship arrival patterns at the Port of NY/NJ. We incorporate these distributions to simulate varying arrival patterns over the study period.
- **ImportGenius Shipping Manifests:** Publicly accessible maritime import manifests detail container-level data—vessel name, container ID, commodity description, and consignee information. Each entry corresponds to a shipment that may span multiple containers. By “exploding” rows to container-level granularity and linking vessel names to AIS data, we generate high-fidelity estimates of container arrivals per voyage. Seasonal fluctuations, commodity distributions, and total inbound flows emerge from this integration. Container numbers were fed into APM terminal container tracking tool via API to get time-stamps for dwell time analysis.

Commodity Classification

The ImportGenius commodity descriptions present significant challenges due to high cardinality and unstructured text, where product descriptions are unique, free-form, and often ambiguous. This variability makes direct aggregation or manual classification into Harmonized System (HS) codes infeasible. To address this, we employ a fine-tuned BERT-based classification model, which processes and standardizes these descriptions into structured, machine-readable HS codes.

Utilizing the ImportGenius platform, we scraped over 500,000 import records for the country of India, each containing an 8-digit HS code. We then simplified these codes to their first two digits, resulting in a dataset of 500,000 entries with short-text product descriptions paired with corresponding two-digit HS codes.

The pipeline begins with data preprocessing, where raw product descriptions are normalized by converting them to lowercase and removing stop words and punctuation. Tokenization is applied using a pre-trained BERT tokenizer, converting text into tokenized input sequences with a fixed-length representation.

Prior to training the classification model, the dataset is split into training (80%), validation (10%), and test (10%) sets, stratified by HS codes to preserve class distribution. HS codes are then encoded into numerical labels, and the text data is tokenized into input tensors. The model is fine-tuned using a BERT architecture adapted for sequence classification, with optimization strategies including gradient accumulation, learning rate scheduling, and mixed-precision training (FP16). The training process is managed through Hugging Face's Trainer API, incorporating early stopping to mitigate overfitting and checkpointing at regular intervals to save the best-performing model.

After training, the fine-tuned model is evaluated on an unseen dataset, ensuring its robustness and generalizability. The structured classification output provides cleaned, standardized product categorizations, facilitating downstream analysis of trade flows, risk assessment, and logistics planning. By transforming highly variable textual data into structured commodity classifications, this approach enhances the ability to track commodity movements, identify trade patterns, and improve efficiency in supply chain decision-making.

Geolocation and Inland Transport Modes

By geocoding consignee addresses, we identify containers destined for specific hinterland regions, such as the Lehigh Valley. This geographic dimension supports scenario analysis in which we test shifting proportions of containers from truck to rail for short-haul distribution. Linking each container to a mode (truck vs. rail) and destination region enables evaluating how modal shifts influence congestion, resource utilization, and environmental outcomes.

Infrastructure data

- **Port Infrastructure and Equipment:** Data from the Port Authority of NY/NJ (PANYNJ) provide terminal-specific parameters, including berth configurations, yard capacity by container type (e.g., reefers, hazardous materials), and equipment availability (cranes, straddle carriers, yard hustlers).
- **RFID Gate Data and Rail Schedules:** Truck arrival/departure timestamps (RFID data) calibrate gate service times and queue lengths. Published rail schedules and corresponding constraints (train frequency, capacity, load/unload rates) are integrated to simulate intermodal transfers and assess the viability of increased rail adoption.

Port Infrastructure and Operational Constraints

We incorporate terminal-level data from the Port Authority of NY/NJ, including:

- Berth configurations and lengths
- Yard capacity by container type (standard, reefer, hazardous)
- Crane availability and productivity rates
- Gate capacity and processing windows

- Rail service frequency and train capacity

Truck gate RFID data is used to calibrate queue lengths and service times, while published rail schedules inform modal timing and constraints.

3.3. Model Construction

The simulation model developed for this study is a discrete-event system designed to emulate the end-to-end flow of import containers through a marine terminal, from vessel arrival to inland departure. The primary goal is to help users understand how operational decisions and infrastructure constraints interact to influence throughput, container dwell time, yard congestion, and gate delays. While it is not intended to serve as a full digital twin, the model is calibrated to reflect realistic dynamics at terminals like APM in the Port of New York and New Jersey. The simulation explicitly represents the following components:

- Vessels, with scheduled or randomly generated arrivals.
- Containers, tracked individually by size, type, commodity class, and inland destination.
- Terminal operations, including vessel berthing, unloading, storage, and gate processing.
- Resources, such as cranes, yard space, truck gates, and train capacity.
- Modal dispatch decisions, which determine whether a container exits the port by truck or rail.

Each event in the system, whether a ship berthing, a crane offloading a container, or a truck arriving at the gate, triggers changes in state that ripple through the system. This structure enables the simulation to capture the nonlinear and interdependent effects that drive congestion and inefficiency in real port operations.

The model is built with flexibility in mind: it supports extensive scenario testing through parameter changes, preset configurations, and input toggles. Rather than hardcoding rules for a single use case, the framework is modular and generalizable, enabling adaptation to different terminal layouts, rail strategies, and policy interventions. This structure allows researchers and stakeholders to experiment with various operational strategies and quantify their impacts with high granularity and transparency.

3.3.1. Model Architecture and Logic

The simulation model is built using SimPy, a process-oriented discrete-event simulation library in Python. It captures the core operational modules of a container terminal through a series of event-driven processes and resource constraints. The model architecture is centered around clarity, modularity, and extensibility, allowing new users to quickly understand and modify individual components without needing to grasp the entire system upfront.

At the core of the model are four main classes:

- **Container:** Each container object records all major lifecycle events, including vessel arrival, yard entry, and inland departure. The class stores container attributes such as container type (e.g., dry, reefer, hazardous), size (TEU or FEU), assigned inland transport mode, and timestamps for key operations. This structure allows the simulation to track dwell times and identify which parts of the terminal are most congested.
- **Vessel:** Vessel instances create and store containers, define unloading behavior, and trigger berth and crane usage. Each vessel has a defined arrival time, number of containers, and unloading duration, which can either be defined through a schedule or generated stochastically if no schedule is provided.
- **Yard:** The yard object tracks capacity utilization and availability of storage space. It monitors how long containers remain in storage, updates yard congestion levels, and flags when the yard is full or under pressure.
- **Simulation:** The Simulation class sets up the environment, loads the appropriate configuration, and runs the main event loop. It handles the initialization of all entities, allocation of resources, and scenario-specific behavior such as rail capacity, train schedules, or gate restrictions.

Simulation Logic and Key Processes

The simulation advances through a series of processes defined in the `simulation_processes.py` file. Each process interacts with SimPy's environment and shared resources. These include:

- **vessel_arrival**: Adds a vessel to the queue based on a schedule or random arrival. If a berth is available, the vessel is assigned and begins offloading.
- **crane_unload**: Models crane behavior and unloads containers from the vessel into the yard, constrained by the number of available cranes and crane productivity.
- **container_departure**: Assigns inland transport mode and manages dispatch through truck gates or rail. Departure is subject to gate hours and queue lengths.
- **train_departure_process**: Simulates periodic rail departures with capacity constraints. Containers scheduled for rail wait for the next departure window.
- **monitor**: Continuously tracks yard occupancy, queue lengths, gate delays, and resource utilization to generate performance metrics.

Config File and Presets

All simulation parameters are stored in a dedicated `config.py` file. This includes yard capacity, crane counts, container compositions, train schedules, gate opening hours, and vessel arrival patterns. The presence of a centralized configuration structure enables two major advantages:

1. **Ease of onboarding**: New users do not need detailed knowledge of every process in the port. They can simply select a preloaded preset (e.g., APM terminal in Port of NY/NJ), and override only the parameters relevant to their use case.
2. **Scenario testing**: The config file makes it easy to test alternate policy or operational decisions. For example, to simulate a modal shift, a user can increase the `rail_share` value or change the train frequency. No modifications to the model logic are needed.

This design emerged from direct conversations with port authorities, terminal operators, and faculty advisors, all of whom stressed that a barrier to adoption was expecting a single analyst to understand every operational detail. By using parameter presets, the model enables partial understanding while still generating meaningful and accurate outputs.

Schedule vs No-Schedule Modes

The model supports two modes of operation:

- **Scheduled mode**: Vessel arrival times and container manifests are read from a CSV file. This mode is ideal when real data is available.
- **Unscheduled mode**: The simulation randomly generates vessels and their associated containers based on configurable probability distributions. This mode is helpful for early prototyping, stress-testing, or exploring future demand scenarios without historical data.

This dual-mode structure ensures the tool is adaptable across different levels of data availability and use cases.

3.3.2. Assumptions and Constraints

The simulation simplifies several operational complexities in order to balance realism with usability. Each assumption was introduced deliberately, based on calibration discussions with terminal operators, expert interviews, or specific constraints observed in the available data. These simplifications allow the model to remain intelligible, adaptable, and suitable for scenario-based decision support.

Terminal-Level Simplifications

- Yard operations do not include container reshuffling or precise slot allocation. Instead, the model uses aggregate yard capacity by container type (dry, reefer, hazardous) and tracks overall occupancy.
- Crane productivity is modeled as a stochastic service time per container, calibrated based on historical unloading durations rather than modeling individual crane movement paths.

Gate Operations

- Gate hours are treated as fixed daily windows (e.g., 7am–6pm) with no variation across days.
- Trucks are assumed to queue outside the gate until opening hours begin, and processing times are drawn from a triangular distribution based on terminal operator input.
- Driver scheduling behavior, appointment systems, and last-mile congestion are not modeled in this version. The focus is on terminal-side delays and capacity constraints.

Inland Transportation Mode

- Containers are assigned to truck or rail at the time of yard release based on a fixed probability (e.g., 15% rail, 85% truck by default).
- Rail services depart at regular intervals (e.g., once every 12 hours), and their capacity is fixed per train.
- There is no interdependence between demand and service availability. That is, additional rail demand does not dynamically increase train frequency or vice versa.

Vessel and Container Flow

- In schedule mode, vessel arrivals and container volumes are taken directly from user-provided CSV files. These may reflect real-world data or synthetic schedules for scenario testing.
- In no-schedule mode, vessels are generated stochastically using a configurable interarrival time distribution and container count distribution.

3.3.3. Validation and Calibration

The simulation model was calibrated to reflect realistic operating conditions at APM Terminal within the Port of New York and New Jersey. Our approach combined structural validation with parameter calibration using historical data and expert input. The goal was not to replicate every operational nuance but to ensure that system-wide behaviors align with observed dynamics at the port.

Structural validation was conducted through step-by-step walkthroughs with academic advisors and terminal stakeholders. These reviews confirmed that key processes operate in ways consistent with how real terminals behave. Particular attention was paid to bottleneck propagation: how constraints in one part of the terminal affect downstream performance.

Calibration was anchored in a container-level dataset of over 50,000 records from December 2024, including vessel identifiers, cargo classifications, and approximate dwell times. Rather than tuning every internal timestamp, we calibrated parameters to ensure that the distribution of total dwell time matched historical patterns. This system-level matching reflects our goal of building a decision-support tool rather than a predictive digital twin.

The final baseline configuration reflects a blend of empirical data and operational insight:

Table 1: Baseline Input Parameters and Justifications for the Port Simulation Model

Parameter	Value	Justification
Yard Capacity	50,000 TEUs	Matches published capacity for APM Terminal (PANYNJ infrastructure data).
Initial Yard Occupancy	70%	Reflects typical steady-state conditions observed during site visits and stakeholder interviews.

Rail Utilization	15% of outbound containers	Mirrors current intermodal split at Port of NY/NJ; allows baseline comparison for modal shift scenarios.
Truck Gate Capacity	120 gates	Derived from throughput of ~5,000 trucks/day over 11 hours of operation.
Gate Processing Time	Triangular(0.1, 0.13, 0.3) in hours	Based on operator interviews; reflects variability in gate transactions while enabling queue modeling.
Active Berths	4	Matches APM Terminal berth availability during typical operation.
Cranes per Berth	4	Reflects common crane allocation strategy and informs unloading rates.
Crane Unloading Rate	~30–35 containers/hour	Reverse-engineered from historical vessel turnaround durations.

The vessel arrival schedule is a critical driver of port dynamics. For our baseline scenario, vessel schedules were carefully designed as representative scenarios informed by historical AIS vessel data at APM Terminal. These arrivals include realistic variations, such as early or late arrivals, to reflect inherent uncertainties and variability that terminal operators must manage daily.

We conducted controlled test runs under edge conditions (e.g., near-capacity yard states, peak vessel clusters, delayed train departures) to ensure the simulation responded with plausible bottleneck dynamics. System indicators such as yard congestion and truck queues reacted in ways consistent with real terminal behavior.

While this version of the model does not yet include granular reshuffling logic or dynamic appointment systems, it offers a stable, credible platform for exploring policy and operational changes. Future work may refine internal sub-processes and extend calibration to additional metrics such as crane cycle times or intra-terminal travel distances.

This calibrated baseline serves as the control environment for all scenario analyses, ensuring that results are anchored in conditions that port operators would recognize as operationally valid.

3.3.4. Key Performance Metrics

To assess system behavior and evaluate scenario performance, the simulation collects time-stamped metrics throughout the run. These metrics reflect key aspects of terminal congestion, resource utilization, and flow dynamics, and are tracked using background processes that run concurrently with operational events. The simulation monitors these variables at each simulation step and aggregates them for downstream analysis.

Metrics are stored in two main data structures:

- **metrics**: captures global system indicators such as total yard occupancy, queue lengths, and gate status.
- **yard_metrics**: stores per-yard occupancy over time, enabling differentiated analysis across container types (e.g., reefer, dry, hazardous).

A dedicated monitoring process (`monitor()`) collects the following key metrics at one-hour intervals:

Table 2: Key System-Level Performance Metrics Tracked During Simulation

Metric	Description
Total Yard Occupancy	Total number of containers stored across all yards at each time step.
Truck Queue Length	Count of containers waiting to depart by truck but not yet processed.

Rail Queue Length	Count of containers assigned to rail, staged for departure but not yet loaded.
Gate Status	Whether truck gates are currently open or closed based on the simulation clock.

A second process (`monitor_yard_occupancy()`) tracks yard-specific metrics:

Table 3: Yard-Level Occupancy Metrics by Container Type

Metric	Description
Yard Occupancy (per type)	Number of containers stored in each yard by container type (e.g., dry, reefer).

These time series allow users to analyze how specific interventions, such as increasing crane capacity, extending gate hours, or boosting rail frequency, affect yard congestion and queue dynamics over time.

In addition to these system-level indicators, the simulation tracks each container’s full lifecycle using the `Container` class, capturing timestamps for the following events:

- Vessel scheduled arrival
- Actual vessel arrival
- Vessel berthing
- Container offload to yard
- Start of inland transport queue
- Time loaded for inland departure
- Final departure from port

These timestamps are stored in the main simulation output `DataFrame` (via the `create_dataframe()` function). This granularity enables fine-grained post-simulation analysis, bottleneck identification, and root-cause tracing for outliers.

3.3.5. User Interface and Simulation Workflow

To ensure broad accessibility and usability, the simulation is packaged in a browser-based interface built using Streamlit. This interface was designed with non-technical users in mind. Its purpose is to enable users to run complex port simulations and explore the implications of different operational decisions without needing to write or modify code.

At launch, users can select from a set of pre-configured terminal presets. The user can then modify specific fields as needed to reflect new assumptions or desired what-if scenarios.

Once parameters are selected, the user can launch the simulation and track its progress through a real-time progress bar. When the simulation is complete, the interface automatically renders key performance metrics. These include interactive time-series plots of yard occupancy, rail and truck queue lengths, and gate status over time. The full container-level output can also be downloaded as a CSV file. This enables post-simulation analysis in Excel, Python, or other tools.

By exposing all relevant functionality through an intuitive graphical interface, the model lowers the barrier to entry for new users while preserving full analytical rigor. It also supports collaborative experimentation, allowing multiple stakeholders to explore how operational strategies affect outcomes in a shared, transparent environment.

4. Results and Discussion

4.1. Dwell Time Analysis

The initial analysis leverages a variety of statistical models to evaluate container dwell time distributions for both dry and reefer containers. These models, Gaussian Kernel Density Estimation (KDE), Circular KDE, Fourier Series, and Normal Distribution, are assessed for their accuracy and performance based on real-world container movement data. Performance metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are calculated for each model using a 20% test dataset.

For dry containers, the Gaussian KDE model with bandwidth of 0.1 demonstrated the most significant improvement over the baseline normal distribution, achieving a 39.5% reduction in MAE. Circular KDE, which accounts for cyclical periodicity in data, yielded a slightly higher error than the normal distribution, with an MAE increase of 3.6%. The Fourier Series model underperformed compared to other methods, showing an 18.7% increase in MAE over the baseline.

For reefer containers, the Fourier Series model outperformed all other approaches, achieving a 24.8% improvement in MAE compared to the normal distribution baseline. Gaussian KDE also performed well, with an 18.8% reduction in MAE. Circular KDE, while showing modest improvement over the baseline, did not perform as well as Fourier or Gaussian KDE models.

These initial results indicate the strengths and weaknesses of various statistical models in representing container dwell times. While Gaussian KDE demonstrates robust performance across container types, the Fourier Series model's strong performance for reefer containers highlights the potential for customized modeling approaches tailored to specific container attributes.

Table 4 summarizes the performance metrics for each model, highlighting the strengths of Gaussian KDE and Fourier models across container types. Figure 1 visualizes these distributions, providing a comparative view of model accuracy.

Table 4: Summary of Model Performance Metrics

Model	Container Type	MAE	RMSE	Improvement vs Normal (%)
KDE	Dry	0.0492	0.1221	39.5
Circular KDE	Dry	0.0842	0.1621	-3.6
Fourier	Dry	0.0965	0.1545	-18.7
Normal	Dry	0.0813	0.1716	
KDE	Reefer	0.1095	0.1383	18.8
Circular KDE	Reefer	0.1271	0.1611	5.7
Fourier	Reefer	0.1013	0.1355	24.8
Normal	Reefer	0.1349	0.1721	

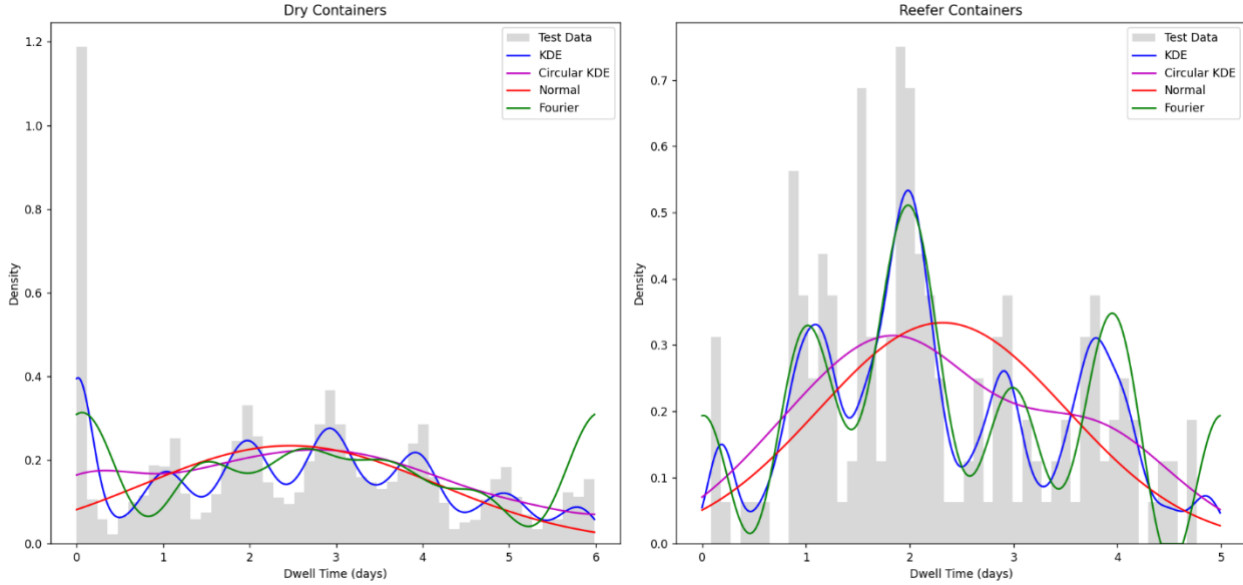


Figure 1: Plot of Dwell Time Distribution Models. Note: the y-axis represents probability density, where the total area under the histogram is normalized to one, and the area under each probability density function integrates to one

4.2. Baseline scenario analysis

To establish a performance benchmark, we begin by analyzing the baseline scenario under current operational conditions at the APM Terminal. The simulation tracks each container's journey across key operational checkpoints and aggregates system-level metrics over time. This section presents findings related to yard occupancy, inland transport queues, and detailed container dwell times.

At simulation start, the yard is pre-loaded with 35,000 containers, representing a 70% yard- utilization rate. These initial containers are not tracked individually, but their presence plays a crucial role in shaping realistic operational conditions. They create immediate pressure on gate operations, contribute to congestion, and increase the likelihood of queues—closely mirroring real-world dynamics.

Yard Occupancy Dynamics

Yard occupancy increases steadily as vessels are unloaded. The slope of this growth aligns closely with vessel arrival patterns and crane productivity rates. The number of containers in the yard peaks shortly after hour 100 (Day 4), corresponding to the point when the last vessel has completed unloading.

From that point on, a seasonal pattern emerges—yard levels fluctuate based on:

- Gate operating hours (06:00–17:00), during which truck dispatches gradually reduce yard volume.
- Train departures scheduled every 6 hours, which cause sharper downward shifts in occupancy, especially when a train departs at high yard utilization.
- Off-hours accumulation, where container outflows pause but vessel unloading continues, driving short-term spikes in yard congestion.

These patterns show that yard occupancy is highly sensitive not just to the volume of arriving containers, but also to the timing, synchronization, and capacity of outbound flows.

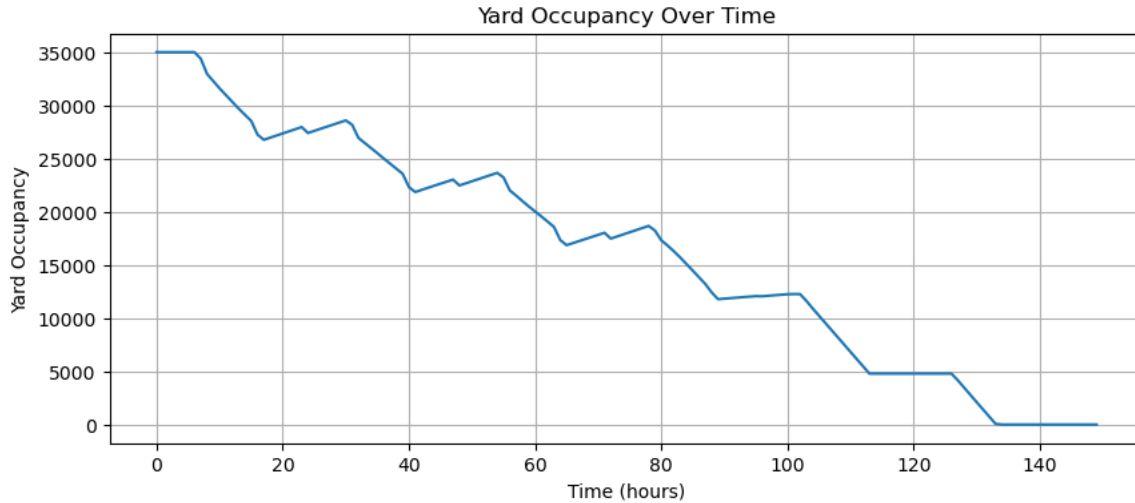


Figure 2: Yard Occupancy Over Time

Inland Transport Queues

A closer look at truck and rail queues provides deeper insight into inland dispatch efficiency.

- **Rail Queue:** The rail queue builds initially as containers are unloaded and wait for scheduled train departures. However, by hour 100, the queue reaches zero and remains empty for the rest of the simulation. This indicates that, given current adoption levels (15% of containers via rail) and a train capacity of 500 TEUs per departure, rail is underutilized for this specific set of inputs. The existing infrastructure could handle significantly more volume without additional congestion or delay.
- **Truck Queue:** Truck queue length follows a similar buildup during unloading but takes until hour 130 to reach zero. This 30-hour delay (despite high gate capacity) suggests that truck dispatch is less efficient than rail under current system configurations. The longer persistence of the truck queue points to potential benefits from shifting more volume to rail, even without infrastructure changes.

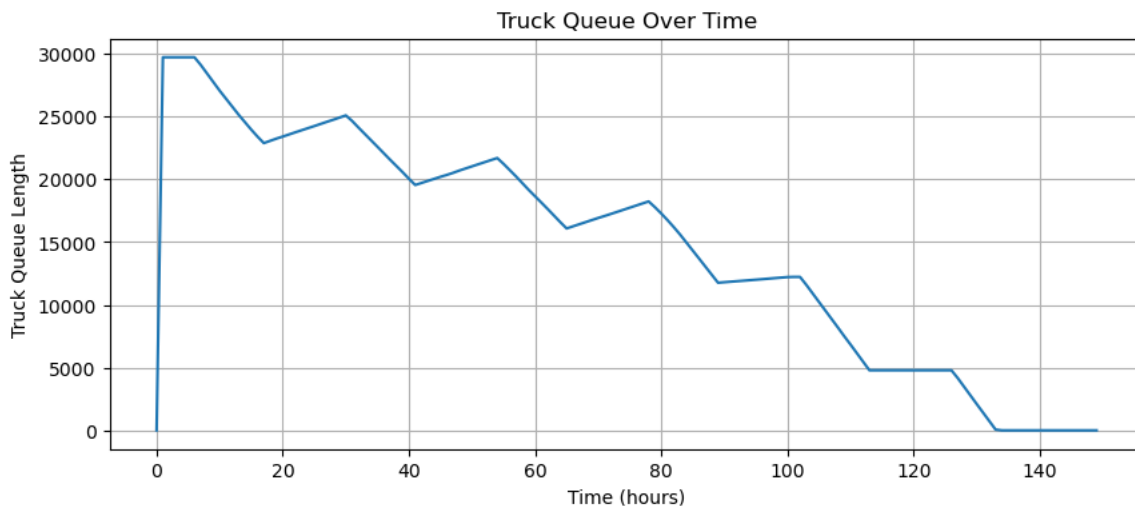


Figure 3: Truck Queue Over Time

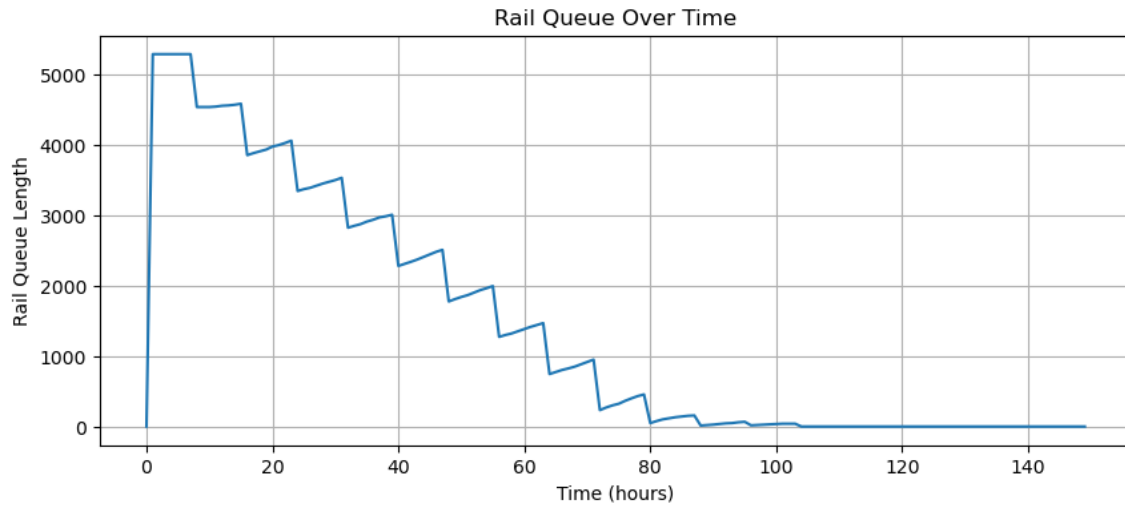


Figure 4: Rail Queue Over Time

Together, these trends suggest two important conclusions:

- The system, as modeled, has untapped rail capacity.
- Increasing rail's mode share (even with no changes to train frequency or capacity) has the potential to relieve pressure on gate operations and reduce average dwell times.

Container Dwell Time Components

To assess individual container experiences, we analyzed dwell times across five components. These granular checkpoints (tracked for every container) offer transparency into where delays accumulate.

Table 5: Breakdown of Container Dwell Time by Operational Stage

	Mean (hrs)	Std Dev	25th pct	Median	75th pct	Max
Vessel Arrival Delay	2.02	1.86	0.04	2.23	2.66	4.80
Berth Delay	11.40	14.28	0.00	0.00	27.62	37.37
Unloading Time	18.67	13.02	8.12	16.15	27.45	53.92
Yard Waiting Time	58.86	19.92	52.63	61.50	72.63	117.64
Total Duration	90.96	17.16	90.43	94.98	98.47	125.16

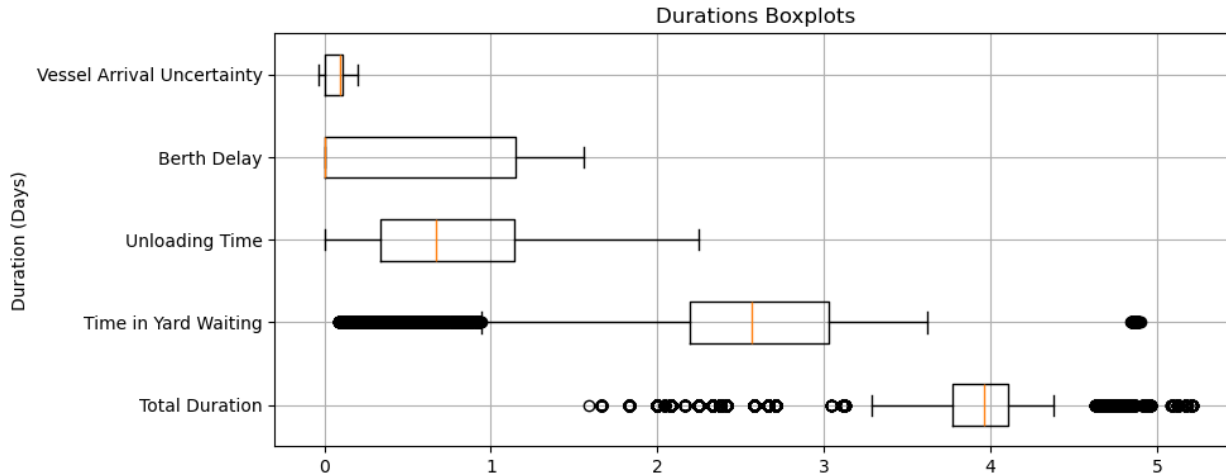


Figure 5: Boxplots of Dwell Time Components (in days)

The largest contributor to total dwell time is yard waiting time, averaging nearly 59 hours, or roughly 65% of the total. This confirms that most inefficiency arises after containers have been unloaded, emphasizing the critical role of inland transportation, gate performance, and scheduling in driving throughput.

Notably, berth delay shows significant variance, but a median of zero, implying that while many vessels berth immediately, a smaller subset faces significant delay due to berth competition. This variance has downstream effects on congestion and train/truck alignment.

These findings validate the use of simulation as a strategic decision-support tool. Even without exploring alternative policies or expanded infrastructure, the baseline scenario already reveals underutilized resources and actionable improvement levers.

4.3. Impact of Increased Rail Adoption on Port Performance

To evaluate the potential performance gains from shifting a larger share of inland container movements from truck to rail, we ran a series of simulation scenarios varying the rail adoption rate from 5% to 40%. In addition to changing the mode share, we also explored the effects of increasing rail capacity through two levers: increasing the number of trains per day and increasing the capacity of each train. In every scenario, all other operational inputs were held constant.

The key performance metric analyzed was average container dwell time, which captures the total time a container spends in the port system, from vessel arrival to departure via truck or rail.

Table 6: Average Dwell Time (in hours) of Varying Rail Usage, Departure Frequency and Train Capacity

Rail Usage	Baseline	Increasing train frequency	Increasing train capacity
	train departure freq: 4/day Train capacity: 500 TEUs	train departure freq: 6/day Train capacity: 500 TEUs	train departure freq: 4/day Train capacity: 750 TEUs
5%	106.37	106.29	106.32
10%	97.89	96.79	96.65
15%	95.82	91.90	90.96
20%	96.15	89.27	87.20
25%	99.77	89.18	85.98
30%	109.41	94.19	89.53
40%	140.23	112.72	104.07

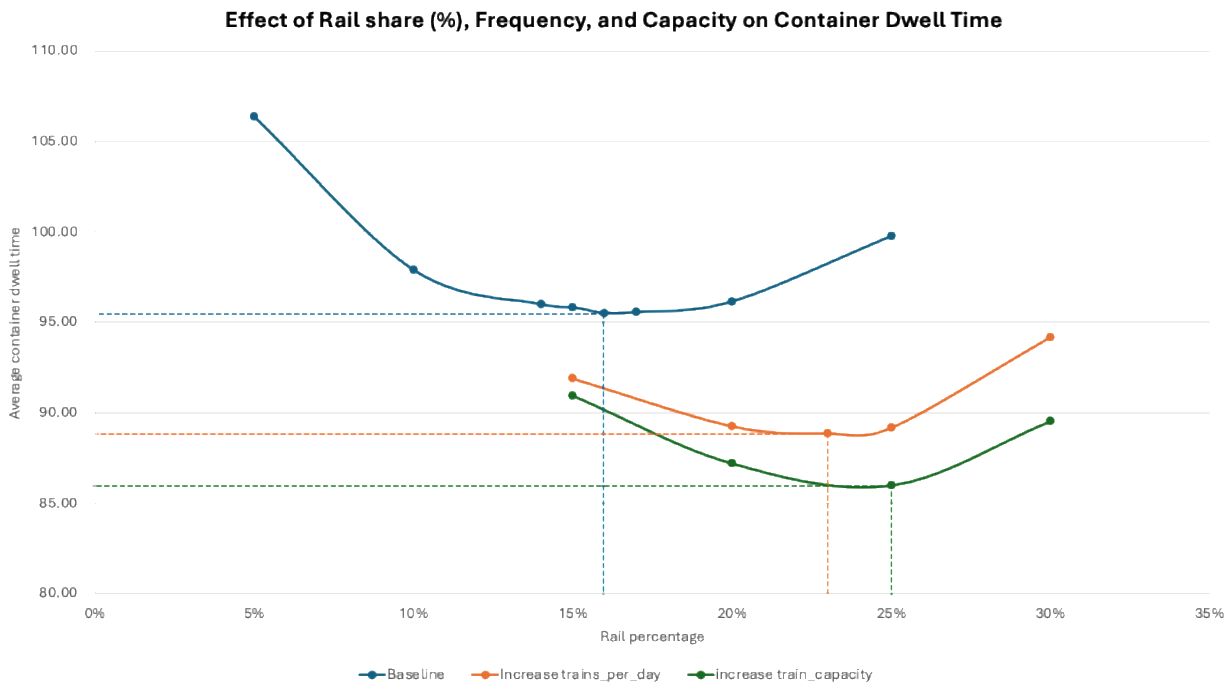


Figure 6: Effect of Rail Share (%), Frequency, and Capacity on Average Container Dwell Time

The downward shift in yard waiting time demonstrates the value of increasing rail throughput. Containers are cleared from the yard faster, which not only reduces dwell time but also eases gate congestion, improves resource utilization, and enhances overall system resilience.

Shifting Containers to Rail Without Infrastructure Changes

When rail adoption increases from 5% to 15%, average dwell time drops by over 10 hours, even with no change in rail frequency or capacity. This confirms the insight from the baseline scenario: rail is underutilized, and simply routing more volume through existing rail operations improves system performance.

However, the benefit plateaus between 15–20%, and at 25% or higher, dwell time begins to increase again. This is a clear signal: while rail has latent capacity, it is not unlimited. At higher adoption levels, congestion shifts from truck gates to rail queues, and rail becomes the new bottleneck—especially when frequency and capacity remain unchanged.

Increasing Number of Trains per Day

Raising daily train departures from 4 to 6 shifts the inflection point. At 15% rail adoption, dwell time drops by nearly 4 additional hours compared to baseline rail frequency. The performance gain grows further at higher adoption levels, with the optimal benefit at around 23–25% rail, where dwell time drops below 89 hours.

But even with this improvement, diminishing returns appear above 30%. The rail network still becomes saturated unless other constraints (such as loading times or departure scheduling) are also addressed.

Increasing Train Capacity

Increasing train capacity from 500 to 750 TEUs per departure produces a similar performance improvement, especially between 15% and 30% rail adoption. The system absorbs more rail-bound containers without increasing train frequency, which may be more logistically feasible depending on scheduling constraints.

Notably, 25% rail adoption with 750-TEU trains yields the lowest observed dwell time across all scenarios, at just under 86 hours, compared to over 95 in the baseline.

Strategic Takeaways

This scenario illustrates the kind of system-wide insights the simulation is designed to uncover. Even though rail improvements affect only a subset of containers directly, the performance gains—measured through average dwell time across the entire system—highlight how targeted infrastructure changes can deliver broad operational benefits.

- Frequency and capacity are both effective levers. Stakeholders can prioritize based on infrastructure limitations, costs, or scheduling feasibility.
- Rail is currently underleveraged. Raising adoption from 15% to ~25% and increasing train departures from 4 to 6 a day can reduce dwell time by 10+ hours, without adding infrastructure.
- Beyond 20–25%, reductions in container dwell time stall unless the rail system is scaled.

4.4. Commodity Categorization

The implementation of a fine-tuned BERT-based HS code classification model significantly improves the structuring of unstructured shipping manifest data, by mapping free-form product descriptions to standardized two-digit HS codes. Classifying containers by HS code enables analysis of volume patterns and makes it possible to examine dwell times by commodity category. For instance, one could investigate whether certain commodities (e.g., oversized machinery or regulated goods) consistently incur longer dwell times due to specialized handling or customs inspection, versus others (e.g., fast-moving consumer goods) that might clear the terminal more quickly.

As described in the section 3.2, the model was trained using Indian import data labeled with 2-digit HS codes. On this dataset, the model achieved approximately 95% classification accuracy at the HS chapter (two-digit) level, confirming its effectiveness in handling the ambiguity and linguistic variability of raw text. This high accuracy in automated categorization addresses inconsistencies in how commodities are described, yielding reliable and consistent commodity labels from noisy textual data. In practical terms, the BERT-based system reads each item description and assigns it an HS code, effectively learning the domain-specific language (including synonyms and abbreviations) that corresponds to each commodity category.

Commodity Classification Results

After training, the fine-tuned model parameters were saved. Inference was run on a dataset of over 50,000 import records from APM Terminal (Port of NY/NJ) for December 2024. Each of these records includes a known dwell time, extracted from the APM tracking website. The BERT classification output provides a detailed breakdown of the commodities flowing through the terminal.

The model predicted that furniture and related products (HS Chapter 94) would comprise the single largest commodity category in December 2024. This dominance of HS 94 (which covers furniture, bedding, mattresses, lamps, etc.) underscores the prevalence of household goods and furnishings in the port's inbound containerized cargo. It is noteworthy that this finding aligns with broader port and national import trends, where furniture consistently ranks as a top import category (Port Authority of NY/NJ, 2023).

Other highly frequent HS codes in the sample included HS 39 (plastics and articles thereof), HS 84 (machinery and mechanical appliances), HS 85 (electrical machinery and equipment), and HS 95 (toys, games, and sports equipment). Together, these top categories indicate a heavy concentration of consumer goods and manufacturing components in the import mix. The top five HS chapters accounted for 44.6% of

all classified records, suggesting that a relatively small set of commodity groups make up the majority of containerized imports.

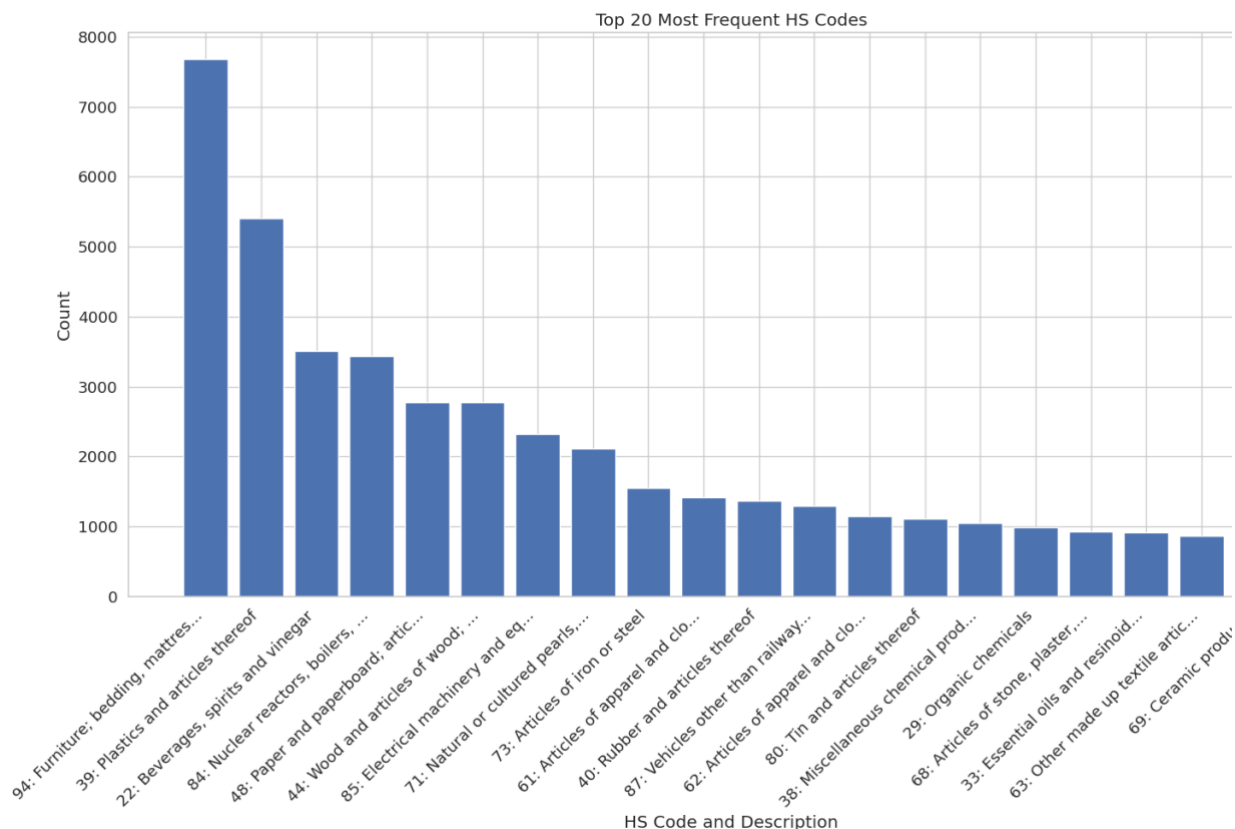


Figure 7: Top 20 HS Code Distribution for December 2024 APM Terminal Imports

Figure 7 shows the breakdown of the most frequent commodity categories, based on the count of containers processed through the APM terminal in December. HS 94 leads in volume. This distribution provides insight into which goods are moving through the port and offers a basis for comparing against official port statistics and validating the representativeness of the sample. According to the Port Authority’s Port at a Glance 2024 (PANYNJ, 2024), eight of the top ten import commodity categories align with those identified by the BERT model for December, reinforcing confidence in the classification results.

Integration with Dwell Time and Temporal Data

Beyond simply categorizing commodities, the enriched dataset allows for quantitative analysis of how these commodities move through the port over time. Each container record classified by the BERT model was linked with terminal tracking timestamps (e.g., discharge time from the vessel, yard availability, and gate-out time) provided by the APM Terminal system. This integration enabled calculation of actual **dwell times** for each container – the duration each container stayed in the terminal from arrival to departure – and facilitated exploration of temporal patterns in the flow of goods. An initial temporal analysis examined container arrival rates by day of week. The data showed distinct weekly patterns, with mid-week days (around Wednesday and Thursday) experiencing the highest container arrival counts, whereas weekends saw significantly lower volumes. This reflects typical operational cycles and vessel scheduling, where large ships often arrive mid-week and fewer unload on weekends.

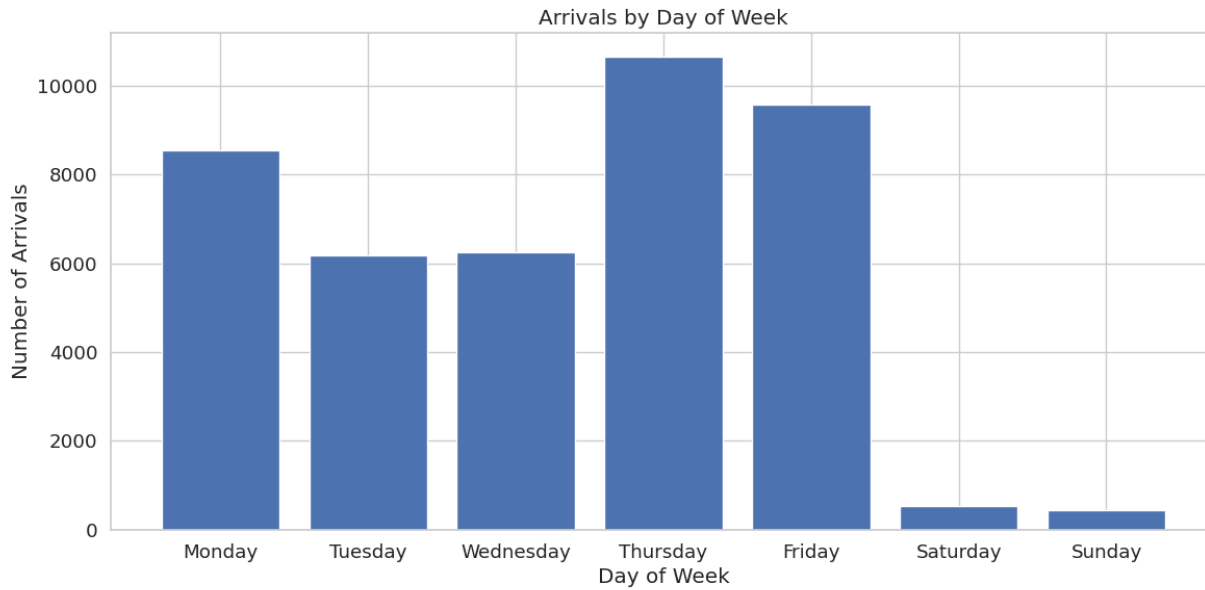


Figure 8: Container Arrival Count by Day of Week

A notable finding is a positive correlation between daily arrival volume and the average dwell time of containers on that day. In other words, days with heavier influx of containers tend to exhibit longer average dwell times per container. The correlation coefficient (Pearson's $r \approx 0.41$) was observed between the number of containers arriving on a given day and the mean dwell time for those containers.

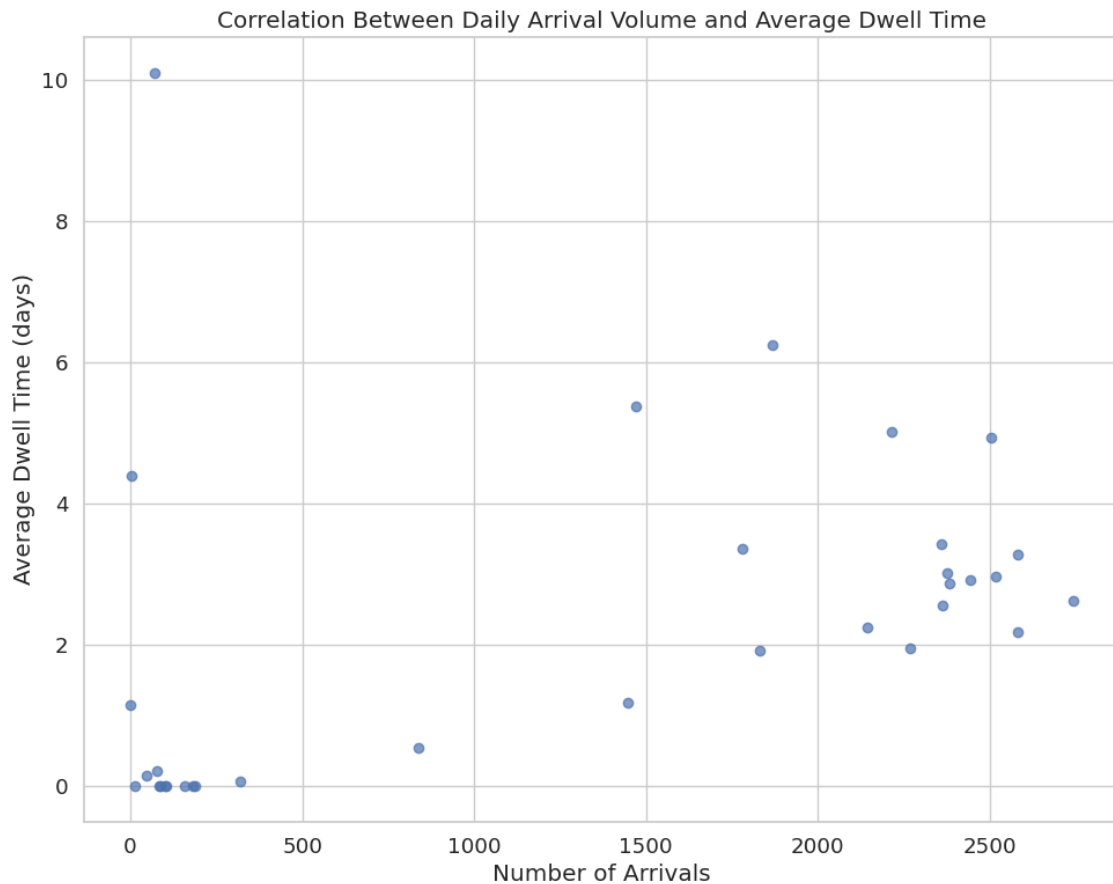


Figure 9: Relationship between Daily Arrival Volume and Average Dwell Time

This moderate correlation suggests that when the terminal is faced with very high volume (for example, when one or more mega-ships berth and discharge thousands of boxes in a short span), the yard and transportation system experience some strain, leading to slower clearance and longer stays for containers. In practical terms, a spike in arrivals can overwhelm available resources—such as yard equipment, labor, or outbound truck and rail capacity—causing containers (regardless of commodity) to experience longer dwell times before they are dispatched. This insight highlights the impact of episodic volume surges on efficiency: even with operations running smoothly, days with very large volumes of arriving containers can push the terminal closer to its capacity limits.

While the December 2024 data did not reveal any extreme outliers in dwell time by commodity in this initial analysis, the framework is in place to monitor such patterns as more data are accumulated. Overall, linking the classification with time stamps adds a critical layer of context, enabling a more nuanced understanding of how commodity mix and arrival timing together influence port performance. By linking each container’s predicted HS code with its APM Terminal timestamps (discharge, availability, gate-out), we created a violin plot showing the distribution of dwell times for the top HS chapters (Figure 10).

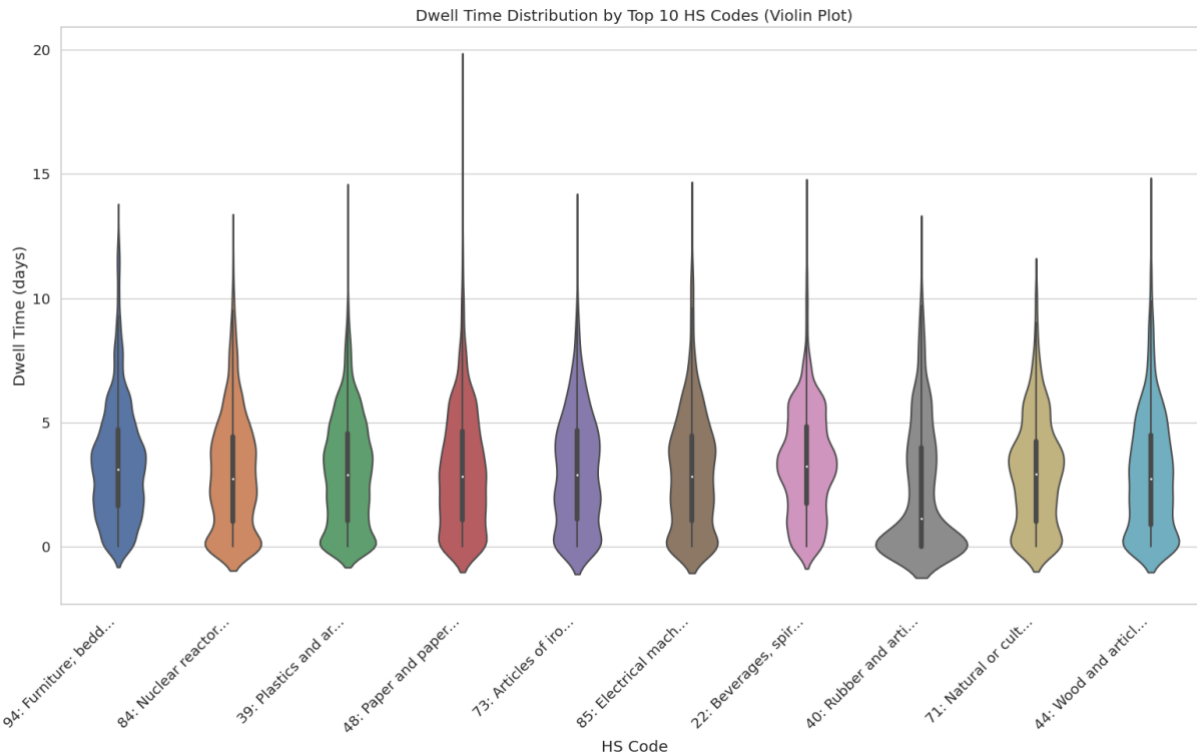


Figure 10: Violin Plot of Dwell Time Distribution by HS Code

This visualization highlights how long containers from different commodity groups remain at the terminal and reveals notable differences in dwell time across the top 10 container volume categories:

- HS 94 – Furniture, bedding, and lighting: Despite being the most common commodity class, this group had one of the tightest dwell time distributions, centered around 3–4 days, with a relatively short tail. This indicates consistent and efficient clearance, likely due to high-volume, well-coordinated supply chains.
- HS 84 – Nuclear reactors, boilers, machinery and HS 85 – Electrical machinery and equipment: These two categories show wide distributions and long upper tails, with some containers dwelling up to 20 days. This suggests operational complexity—perhaps due to inspection delays, specialized handling requirements, or less frequent pick-ups.
- HS 39 – Plastics and articles thereof: This category displays moderate dwell variability, but with a heavy central mass around 3–4 days, similar to HS 94. This suggests that they are processed relatively efficiently, although not as consistently.
- HS 73 – Articles of iron or steel and HS 22 – Beverages and spirits: These also showed long tails, with a small percentage of containers lingering much longer. That could indicate uneven consignee behavior or inspection requirements.
- HS 40 – Rubber, HS 71 – Natural/cultured pearls, HS 44 – Wood products and HS 48 – Paper and paperboard: These four stand out for their bimodal or fat-tailed distributions, indicating operational unpredictability. While some containers clear quickly, others remain for extended periods—pointing to heterogeneity in downstream logistics or regulatory handling.

Operational Implications and Uses of the Classification

The structured HS code classification results have several important implications for port operations and for the discrete-event simulation (DES) model underpinning this capstone project. First, these results can serve as critical inputs to the simulation model. Instead of treating all containers uniformly, the simulation can now

incorporate commodity-specific attributes. Each simulated container can be tagged with an HS code category, allowing the model to differentiate processes or resource needs by commodity type.

This enables more detailed analysis of yard utilization and dwell times by commodity. Key logistics parameters (like how long containers stay in the yard or how they transition to a particular mode of transport) can be modeled with commodity-specific distributions derived from the data. In sum, the BERT-derived classifications have the ability to enhance the fidelity of the simulation, making scenario outcomes more realistic and trustworthy when evaluating changes in the system.

Secondly, the classification insight can allow for more targeted scenario planning, including modal shift strategies. With a clear picture of the commodity breakdown, planners can identify which goods are prime candidates for alternative transportation modes (truck vs. rail). For instance, the data might show that a large portion of high-volume commodities such as furniture (HS 94) and plastics (HS 39) are currently moving out of the port by truck. Using the simulation, we could evaluate a scenario where a percentage of these containers is shifted to rail service – for example, moving more furniture shipments by rail to inland distribution centers instead of by highway. Because we know the volumes and characteristics of these commodities, we can simulate the impact of such a modal shift on port congestion and dwell times.

This approach was one of the motivations for the analysis: the ability to test the feasibility of transitioning specific commodity groups from truck to rail in a data-driven way. If the simulation (fed by our classification data) indicates that shifting, say, 20% of furniture containers to rail is achievable, we could see reductions in gate truck traffic and possibly yard occupancy, thereby improving overall flow. On the other hand, if certain commodities are unsuitable for rail due to time-sensitivity or lack of rail infrastructure to their destinations, the classification helps flag those constraints as well. Thus, the BERT-based categorization directly supports strategic initiatives like increasing short-haul rail utilization for appropriate cargo groups.

Finally, having granular commodity information enables better management of high-volume commodity groups and other operational optimizations. Terminal operators and port authorities can use these insights to allocate resources more efficiently. If certain commodity surges are seasonal (for instance, HS 95—toys and games—might peak in late summer to early fall ahead of the holiday season), the port can anticipate and plan for those surges in advance. In a broader sense, providing structured, high-quality commodity data enhances predictive analytics and strategic decision-making. The port's management could monitor trends in the commodity composition of cargo flows, for example, detecting an increasing trend in electric machinery imports (HS 85) and then responding with appropriate infrastructure or policy adjustments.

In summary, the integration of the BERT-based HS code classification has transformed a once-unstructured aspect of the data into actionable intelligence. This intelligence not only has the potential to feed into the simulation model for scenario testing and policy evaluation, but also supports real-world operational planning, from daily yard management to long-term strategy for accommodating shifts in trade patterns. The ability to classify and quantify commodities at scale thus strengthens both the analysis within this capstone and the practical toolkit available to port stakeholders for managing flows of goods through one of the nation's busiest gateways.

5. Conclusion

Persistent inefficiencies in U.S. seaports present a significant economic challenge, driven primarily by fragmented stakeholder collaboration and a lack of system-level visibility tools. This research directly addressed this critical gap by developing a discrete-event simulation model of import container flows through the Port of NY/NJ. Built in Python and calibrated using publicly available data, including detailed shipping manifests classified by HS code through a fine-tuned BERT model, the simulation serves as a

decision-support framework designed to help stakeholders understand the complex, interdependent consequences of their actions within the port ecosystem.

Through the application of this simulation model, we quantified the impact of key variables such as container volume, resource allocation, and modal mix on essential performance metrics like dwell time and congestion. The baseline scenario analysis revealed that yard waiting time constitutes the largest component of overall container dwell time, highlighting the critical role of inland transportation and gate operations in driving efficiency. Furthermore, the simulation demonstrated the potential for significant performance gains through strategic shifts in modal split. Our scenario analysis on increased rail adoption specifically showed that the port system, as modeled, possesses underutilized rail capacity, and that increasing rail's mode share, particularly when coupled with increased train frequency or capacity, can substantially reduce average container dwell time by alleviating pressure on truck gates and improving yard fluidity, up to a certain threshold where the rail network itself becomes the bottleneck. Complementing this, the successful classification of unstructured commodity descriptions into standardized Harmonized System codes provided granular insights into cargo flow patterns and dwell time distributions by commodity type, enabling a more nuanced understanding of operational complexities and supporting the identification of specific cargo groups suitable for targeted interventions, such as modal shifts.

The simulation model and the data-driven insights derived from it offer a valuable tool for port authorities, policymakers, and private stakeholders, providing a shared platform for structured experimentation and evaluating the system-wide effects of potential operational or policy changes. By making the downstream consequences of localized decisions visible, this framework fosters a more coordinated and informed approach to improving port performance and enhancing supply chain resilience. While the current model incorporates necessary simplifications, such as aggregate yard operations and fixed gate hours, to balance realism with usability, it provides a stable and credible foundation for exploring strategic interventions. Building upon this foundation, future research can further refine the model's granularity, explore a wider array of policy scenarios, and extend its application to other ports, thereby contributing to the development of more efficient and sustainable freight transportation networks.

6. Future Research

This capstone project establishes a foundational framework for modeling import container flows and evaluating operational interventions at the Port of New York and New Jersey. Building on this foundation, several promising research directions can enhance the model's fidelity, broaden its applicability, and increase its strategic value for both public and private stakeholders.

Incorporating Export Flows and Shared Resources

One immediate extension involves modeling export container flows and their interaction with imports. In real-world operations, yard space, cranes, and labor are shared across inbound and outbound processes. Capturing these shared resource dynamics would allow the model to represent total terminal throughput more accurately and analyze interdependencies between export surges and import congestion.

Refining Terminal-Level Operational Logic

Improving the granularity of internal terminal operations would significantly enhance simulation fidelity. Specific areas of refinement include:

- Container stacking and reshuffling strategies within the yard
- Crane allocation and movement behavior
- Truck flow logic inside and around terminal gates
- Container positioning rules and spatial constraints

Such enhancements would move the simulation closer to a digital twin and allow for micro-level performance testing under varying congestion and resource conditions.

Scaling to Multi-Terminal and Multi-Port Networks

Expanding the model to simulate operations across multiple terminals within a single port complex would enable analysis of terminal interdependencies—particularly where rail corridors, roadways, or labor resources are shared. Over time, the framework could be extended to simulate inter-port dynamics at a regional or national scale, especially where competitive or cooperative relationships influence system-wide efficiency.

Integrating Exogenous Disruption Scenarios

Future iterations of the model could incorporate external disruptions—such as weather events, labor actions, tariffs, or geopolitical shifts—as stochastic or scenario-based inputs. These additions would enable stakeholders to test the resilience of various operating strategies and prepare contingency plans that reduce the risk of cascading supply chain failures.

Linking to Economic and Environmental Metrics

To enhance decision relevance, simulation outputs could be linked with cost and emissions models. This would allow users to evaluate:

- The financial impact of investments in equipment, infrastructure, or extended operating hours
- The potential emissions reductions from shifting container flows to rail
- The value of specific short-haul rail corridors (e.g., to the Lehigh Valley) in relieving truck congestion and lowering carbon footprints

In doing so, the model could support economically and environmentally optimized decisions, beyond operational efficiency alone.

Generalizing the Framework for Broader Adoption

For broader adoption across U.S. ports, future work should focus on automating the ingestion and structuring of port-specific data to generate model-ready configurations. Developing a scalable onboarding pipeline would make the tool deployable in new contexts with minimal customization, facilitating comparative analysis and coordinated planning across the national port network.

References

- Abu-Aisha, T., Audy, J.-F., & Ouhimmou, M. (2024). Toward an efficient sea-rail intermodal transportation system: a systematic literature review. *Journal of Shipping and Trade*.
- Acero, A. (2021). Synchromodality: Enhancing supply chain competitiveness, resilience, and sustainability. *International Journal of Logistics Research and Applications*, 345-262.
- Ambrosino, D., & Peirano, L. (2016). Truck Arrival Management at Marine Container Terminals. *European Council for Modelling and Simulation*.
- Ambrosino, D., & Xie, H. (2023). Optimization approaches for defining storage strategies in maritime container terminals. *Soft Computing*.
- Ambrosino, D., & Xie, H. (2024). Machine Learning-Based Optimization Models for Defining Storage Rules in Maritime Container Yards. *MDPI Modelling*.
- Aristov, N., & Li, Z. (2024, September). Elucidating US Import Supply Chain Dynamics: A Spatial-Temporal Graph Neural Network Approach. *IEEE-HPEC*. Retrieved from IEEE-HPEC: <https://ieeehpec.org/wp-content/uploads/2024/09/139.pdf>
- Aristov, N., Li, Z., Koch, T., & Dugundji, E. (2024). Elucidating US Import Supply Chain Dynamics. *Procedia Computer Science*, 238, 216-223. doi:<https://doi.org/10.1016/j.procs.2024.06.018>
- Bland, W., Escobar, L., Hong, A., Kenneally, G., Liberatore, A., & Rosen, S. (2023). A Hybrid System Dynamics/Input-Output Model for Studying the Impact of Transportation Delays on the Resiliency of National Supply Chains. *IEEE*.
- Carboni, A., Deflorio, F., Caballini, C., & Cangelosi, S. (2024). Advances in terminal management: simulation of vehicle traffic in container terminals. *Maritime Economics & Logistics*.
- Caris, A., Macharis, C., & Janssens, G. K. (2008). Planning problems in intermodal freight transport: Accomplishments and prospects. *Transportation Planning and Technology*, 31(3), 277-302.
- Caris, A., Macharis, C., & Janssens, G. K. (2013). Decision support in intermodal transport: A new research agenda. *Computers in Industry*, 64(2), 105-112.
- Cavada, J. P., Cortés, C. E., & Rey, P. A. (2023). Comparing allocation and relocation policies at a logistics service container terminal: a discrete-event simulation approach. *Central European Journal of Operations Research*.
- Chen, D., Feng, H.-Y. F., Huang, Y., Chen, Q.-Y., & Xiao-Shu, W. (2023). Robust Control of Bullwhip Effect for Supply Chain System With Time-Varying Delay on Basis of Discrete-Time Approach. *IEEE Xplore*.
- Craig, A. J., Blanco, E. E., & Sheffi, Y. (2013). Estimating the CO₂ intensity of intermodal freight transportation. *Transportation Research Part D: Transport and Environment*, 22, 49–53.
- Davies, B., Islam, A., Gheith, M., & Eltawil, A. (2024). Simulation-Based Optimization of Truck Appointment Systems in Container Terminals: A Dual Transactions Approach with Improved Congestion Factor Representation. *MDPI Logistics*.
- De Armas Jacomino, L. M.-P.-R.-P. (2021). Dwell time estimation of import containers as an ordinal regression problem. *Applied Sciences*, 11(20), Article 9380.
- Dinwoodie, J. (2006). Rail freight and sustainable urban distribution: Potential and practice. *Journal of Transport Geography*, 309-320.
- Du, L., Zhang, T., & Zhang, J. (2023). Simulation of road traffic flow in the port. *Fourth International Conference on Signal Processing and Computer Science*. SPCS.
- Giuliano, G., & O'Brien, T. (2007). Reducing port-related truck emissions: The terminal gate appointment system at the Ports of Los Angeles and Long Beach. *Transportation Research Part D: Transport and Environment*, 12(7), 460–473.

- Guimarans, D., Harabor, D., & Van Hentenryck, P. (2015). Simulation and analysis of port rail operations: A case study at Port Botany. *Transportation Research Part E: Logistics and Transportation Review*, 46-61.
- Jackson, I., Saenz, M. J., & Ivanov, D. (2023). From natural language to simulations: applying AI to automate simulation modelling of logistics systems. *International Journal of Production Research*.
- Kishore, L., Pai, Y. P., Ghosh, B. K., & Pakkan, S. (2024). Maritime shipping ports performance: a systematic literature review. *Discover Sustainability*.
- Kordnejad, B. (2016). Intermodal transport cost model and intermodal distribution in urban freight. *Procedia—Social and Behavioral Sciences*, 230, 76-83.
- Lange, A.-K., Schweintek, A., & Jahn, C. (2017). Reducing Truck Congestion at Ports – Classification and Trends. *Hamburg International Conference of Logistics*. Hamburg.
- Macharis, C., & Bontekoning, Y. M. (2004). Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research*, 153(2), 400-416.
- Maguire, T. J., Ivey, S. S., Golias, M. M., & Lipinski, M. E. (2010). Evaluating strategies to improve truck operations at intermodal marine container terminals. *Journal of Transportation Engineering*, 136(10), 926–935.
- Maina, J., & Mwangangi, P. (2020). A Critical Review of Simulation Applications in Supply Chain Management. *Journal of Logistics Management*, 1-6.
- Marufuzzaman, M., Eksioglu, S. D., Li, X., & Wang, J. (2014). Analyzing the impact of intermodal-related risk to the design and management of biofuel supply chain. *Transportation Research Part E: Logistics and Transportation Review*, 69, 122-145.
- Musso, E., & Sciomachen, A. (2020). Impact of megaships on the performance of port container. *Maritime Economics & Logistics*, 432–445.
- Oliver Wyman and Leachman and Associates for the Pacific Harbor Line. (2023). Feasibility and benefits of intermodal service short-haul markets.
- PANYNJ. (2024). *Port at a Glance*. PANYNJ.
- Perrykkad, K., Ernst, A., & Krishnamoorthy, M. (2020). A model for demand shifting in short-haul intermodal container transport. *Transportation Research Part E: Logistics and Transportation Review*.
- Phandis, S., Sheffi, Y., & Caplice, C. (2022). *Strategic Planning for Dynamic Supply Chains: Preparing for Uncertainty Using Scenarios*. Boston.
- Pinho, T. M., Coelho, J. P., & Boaventura-Cunha, J. (2016). Forest-based supply chain modelling using the SimPy simulation framework. *IFAC-PapersOnLine*, 90-95.
- Port Authority NY/NJ. (Retrieved: 2024). *Container Terminals*. Retrieved from Port Authority NY/NJ: <https://www.panynj.gov/port/en/our-port/container-terminals.html>
- Port Authority of NY/NJ. (2019). *Iconic Bayonne Bridge Rededication Caps Port Authority's \$1.7 Billion Project*. New York: PANYNJ.
- Port Authority of NY/NJ. (2023). *PONYNJ at a Glance*. New York.
- Statista. (2024, July). *Most used programming languages among developers worldwide as of 2024*. Retrieved from Statista: <https://www.statista.com/statistics/793628/worldwide-developer-survey-most-used-languages/>
- Tongzon, J. (1995). Determinants of port performance and efficiency. *Transportation Research Part A: Policy and Practice*, 29(3), 245–252.
- United States Department of Transportation. (Retrieved: 2024). *Freight Logistics Optimization Works*. Retrieved from United States Department of Transportation: <https://www.bts.gov/flow>
- Vidović, M., K. M., & Zrnić, N. (2011). A combined approach for intermodal terminal location problem in Serbia. *Transport Policy*, 18(5), 654–663.

- Wang, H., & Wang, Z. (2013). Study on the method and procedure of logistics system modeling and simulation. *2nd International Conference on Science and Social Research*. Hunan.
- Wang, M., Zhou, C., & Wang, A. (2022). A cluster-based yard template design integrated with yard crane deployment using a placement heuristic. *Transportation Research Part E: Logistics and Transportation Review*.
- Wang, X., Zhao, N., & Mi, C. (2023). Yard Space Allocation Algorithm for Unloading Containers at Marine Terminals. *J. Mar. Sci. Eng.*
- Yıldırım, M., & Kawasaki, T. (2023). Priority Management for Decreasing Truck Traffic Congestion in Ports. *The Japan Society of Naval Architects and Ocean Engineers*.
- Zhou, C., Wang, W., & Li, H. (2020). Container reshuffling considered space allocation problem in container terminals. *Transportation Research Part E: Logistics and Transportation Review*.
- Zhou, C., Zhao, Q., & Li, H. (2021). Simulation optimization iteration approach on traffic integrated yard allocation problem in transshipment terminals. *Flexible Services and Manufacturing Journal*.
- Zohuri, B., Mossavar-Rahmani, F., & Behgounia, F. (2022). Python programming–driven artificial intelligence. *Academic Press: Knowledge is Power in Four Dimensions: Models to Forecast Future Paradigm*, 827-836.
- Abu Aisha, M., Audy, J.-F., & Ouhimmou, M. (2024). Simulation-based analysis of sea-rail intermodal transportation: A case study of the Port of Trois-Rivières. *Transportation Research Part E: Logistics and Transportation Review*, 172, 102987. https://doi.org/10.21872/2024IISE_7356

Appendix A

Understanding the Port of New York/New Jersey

The Port of New York/New Jersey is the largest on the East Coast and third-largest nationwide. Approximately 16% of all loaded containers entering the United States pass through this port. This accounts for over 34% of import containers bound for the North American East Coast, spanning from Miami to Montreal. In 2023, the port facilitated over 2,200 container vessel calls. The primary imported commodities included furniture, appliances and machinery, plastics, beverages, and electronics. China remains the top country of origin for imports, with the leading destination states being New Jersey, New York and Pennsylvania. In total, the port handled approximately \$238 billion worth of goods in 2023 and completed nearly 630,000 rail lifts (Port Authority of NY/NJ, 2023). The port is managed by the port authority, which leases five terminals to four separate operators:

Table A1: Port Authority of NY/NJ Container Terminals, (Port Authority NY/NJ, Retrieved: 2024)

Terminal Name/Operator	Area (Acres)	Berth Length (Feet)	Depth at Dock (Feet)	Container Cranes	Yard Hustlers	Rubber Tire Gantry Cranes (RTGs)	Straddle Carriers	Reefer Plugs
Port Newark Container Terminal (PNCT)	272	4,400	40-50	13	39	3	100	848
Maher, Port Elizabeth	450	10,128	50	24	50	0	220	1,200
APM, Port Elizabeth	350	6,001	45-50	15	150	32	0	1,964
GCT New York	210	4,012	35-50	6	48	-	-	481
GCT Bayonne	167	6,756	50	12	54	-	-	776

To access four of these terminals (GCT New York, APM, Maher and PNCT), vessels must navigate through the Kill Van Kull strait and under the Bayonne Bridge. To accommodate the increasing size of vessels entering the port, the Bayonne Bridge Navigational Clearance Project, completed in 2019, raised the bridge's roadway to 215 feet. This was a \$1.7 billion dollar project, which is expected to bring more than 400,000 jobs and billions of dollars in regional economic activity over its lifetime (Port Authority of NY/NJ, 2019), highlighting the importance for container ports to accommodate megaships. Figure A! presents a map of the port, with all of the container terminals labelled.

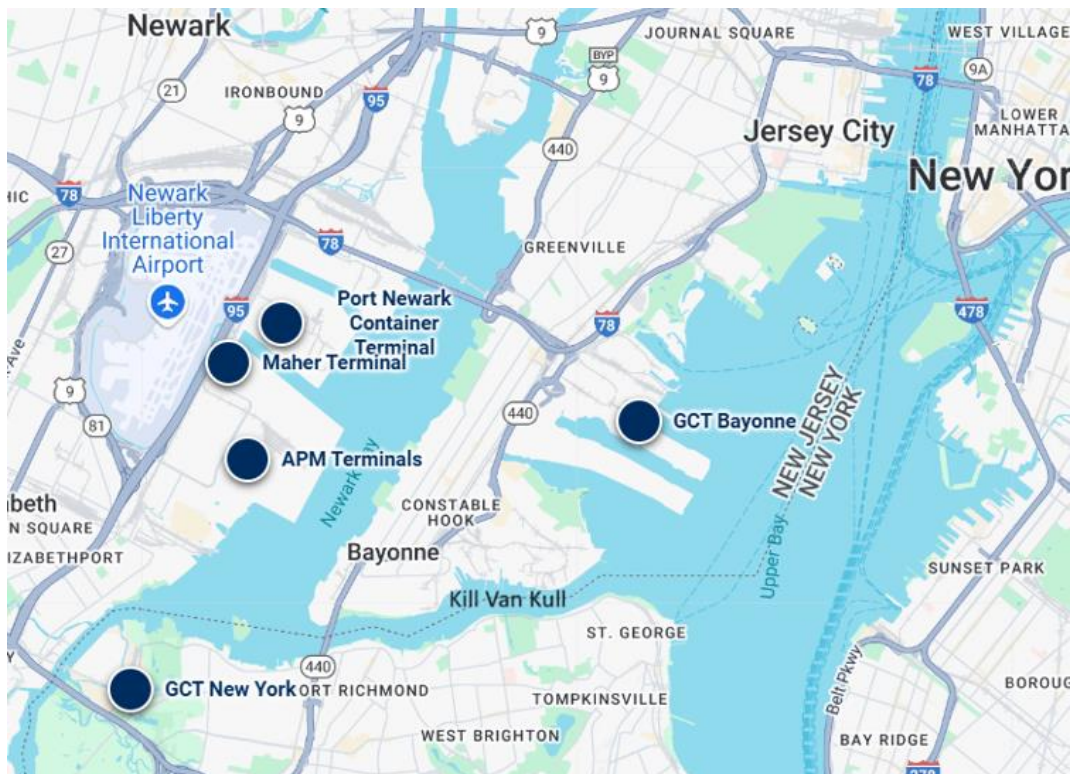


Figure A1: Port Authority of NY/NJ Container Terminal Map, (Port Authority NY/NJ, Retrieved: 2024)

Understanding Inbound Container Processes at the Port of NY/NJ

The following overview summarizes the key steps involved in moving inbound containers through the Port of NY/NJ; from vessel arrival to final pickup. It is based on input from stakeholders and insights gathered during a visit to the port.

Advance Notice of Arrival:

Up to 96 hours before arriving at the port, the vessel provides advanced notice of arrival, which goes to the port authority, terminal operator, U.S. Customs and Border Protection (CBP), and the U.S. Coast Guard. The notice typically includes vessel and crew details, cargo manifests, the number of containers to be offloaded, and the estimated arrival date and time. This information helps each entity prepare for the vessel's arrival and ensure smooth operations.

Vessel Arrival and Anchorage:

The vessel arrives at the port and anchors in the designated staging area, typically the Ambrose Channel. Vessel traffic is managed by the Vessel Traffic Service on Staten Island; it is operated by the Coast Guard, which coordinates vessel movements to maintain safe and efficient flow within the port.

Coast Guard Boarding:

While the vessel is anchored, the Coast Guard will board high-interest vessels for an initial inspection of the vessel and cargo before the vessel is allowed to berth. This inspection ensures compliance with safety, security, and environmental standards. Vessels deemed lower-risk may not require a physical boarding but could undergo remote document checks.

Berthing:

Once a berth becomes available, the waiting vessel is called from the queue. With the help of tugboats and NY/NJ harbor pilots (for vessels over 150 feet in length), the vessel is guided to the designated berthing area. The tugboats and NY/NJ port pilots assist with safe navigation through complex and narrow port channels.

Unloading:

Containers are unloaded from the vessel using quay (gantry) cranes. Each terminal is equipped with [specific types of quay cranes](#). The containers are then placed on chassis pulled by yard hustlers, which transport them from the berth. At selected berths, containers are loaded directly onto trains from the ship to expedite intermodal transfers. Container tracking systems, such as RFID or QR codes, are used to monitor containers' movements and locations within the port.

Moving Containers from Berth:

Yard hustlers transport containers first to the CBP customs zone, where they may undergo initial clearance checks, and then to the designated storage area of the yard. Container destinations within the yard are often determined by cargo type; for example, temperature-sensitive cargo in refrigerated containers will be placed near reefer plugs.

Customs Processing:

Approximately 3% of containers are selected for further customs inspection. These containers are segregated in a different area of the yard and placed in the queue for inspection. Most inspections use Non-Intrusive Inspection (NII), where containers are moved through an x-ray scanner by a straddle carrier. Some containers may also be selected for physical inspection, where they are opened and examined by customs officials.

Placement in the Yard:

The container is offloaded from the yard hustler chassis using rubber-tired gantries, straddle carriers, or reach stackers and placed in the designated area within the yard. Containers are organized based on destination, priority, customer, handling requirements, or shipping line.

Transport Notification:

Once the container has cleared customs, the terminal operator notifies the carrier (if by truck) that the container is ready for pickup. Some terminals use online portals or Electronic Data Interchange (EDI) systems for real-time status updates. If the container is scheduled to depart by train, it is placed in the queue and loaded once space is available on a rail car, typically headed to the US Midwest. Terminals impose storage fees if containers are not picked up within a certain time after clearance.

Truck/Train Arrival:

The designated truck or train arrives at the port. Truck drivers must stop at the gate for the clerk to approve the truck for entry. To access the port, truck drivers need:

- Transportation Worker Identification Credential (TWIC),
- SeaLink Card (truck driver identification card used by the port's automated cargo expediting system),
- A valid Class A Commercial Driver's License (CDL),
- Drayage Truck Registration in PortTruckPass (PTP), and
- A valid RFID tag mounted to the truck.

After arriving, drivers queue up, and a rubber-tired gantry, straddle carrier, or reach stacker retrieves the container from the yard and loads it onto the chassis. Gate hours for terminals can be found [here](#).

Container Exit:

The container exits the terminal. Most trucks transport containers to nearby warehouses before they are routed to their final destinations. Train schedules and departure information can be found online via [Millenium Marine Rail](#) or [Port Newark Container Terminal](#) for planning intermodal connections.