Developing a digital solution to container triangulation in China

by

Jieming Feng
Bachelor of Arts in International Trade, Sun Yet-Sen University, 1997

and

Mauricio Moreno Sánchez Briseño
Bachelor of Science in Industrial Engineering, ITESM, 2017

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

June 2021

Signature of Author: ____
Jieming Feng
Department of Supply Chain Management
May 14, 2021

Signature of Author: ____
Mauricio Moreno Sánchez Briseño
Department of Supply Chain Management
May 14, 2021

Certified by: ____
Ozden Tozanli Yilmaz
Postdoctoral Associate
Capstone Advisor

Certified by: ____
Pascal Wolff
Assistant Professor
Capstone External 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

Developing a digital solution to container triangulation in China

by

Jieming Feng

and

Mauricio Moreno Sánchez Briseño

Submitted to the Program in Supply Chain Management
on May 14, 2021 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

Improving efficiency and sustainability in logistics and transportation has been a strategic priority for companies and countries as they compete in the era of globalization. However, how to optimize the container transportation and improve container turnaround has become an increasing challenge for the industry, especially in the growing trade imbalances and more frequent disruptions. To overcome this challenge, container triangulation offers remarkable opportunities to the carriers to reduce the transportation of the empty containers, and therefore, improve the turnaround. Container triangulation can be identified as the reuse of the import containers for export. Despite a number of potential benefits of container triangulation may offer, it is challenging to scale-up in China due to the fragmented market and the lack of accurate location data. To focus on this challenge, this research investigates the digitalization of container triangulation as an alternative solution, where matching decisions are automated in a digital platform. This research examines the current process and challenges of automating container triangulation in China for Maersk and explores how to optimize and accelerate this solution. With this motivation, we conducted expert meetings, analyzed data, and applied machine learning algorithms and mixed-integer linear programming to enable container triangulation routing optimization on the company's digital platform. The result showed a trucking cost savings from 11% - 14%, a transportation lead time reduction from 8% - 10%, and a reduction in CO_2 emissions from 8% - 10%. However, the savings would be further reduced with more restrictive conditions for execution. To scale up the solution, we recommend the cooperation of different parties of the container transport industry to share the incentives and adopt the digital solution.

Capstone Advisor: Ozden Tozanli Yilmaz
Title: Postdoctoral Associate

Capstone External Co-Advisor: Pascal Wolff
Title: Assistant Professor

ACKNOWLEDGMENTS

We would like to thank our advisor, Ozden Tozanli and Pascal Wolff, for providing guidance and support throughout this project. Their continuous feedback and insights have been invaluable not only to this project but also to our learning experience at MIT.

We are immensely grateful to the sponsoring company Maersk for providing us with this project that is critical to their business and allowing us this opportunity. In special, Warren Huang for generously connecting us to all the subject matter experts and committing time every week to provide feedback, suggestions, and guidance.

We would like to thank the CTL at MIT for providing us with an opportunity to apply everything we learn through the Supply Chain Management program. We would also like to thank Toby Gooley for reviewing our reports numerous times and providing detailed feedback over areas of improvement.

We are grateful to all our colleagues and friends here at MIT who supported us along the way and allowed us to have such a good time and wonderful memories in Cambridge.

Finally, we would like to thank our family for their love, support, and encouragement throughout the completion of our master's program. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Introduction and Motivation

Along with China's strong economic growth in the past four decades, the inefficiency in its logistics industry has become a severe problem. China's Gross Domestic Product (GDP) growth is one of the most studied economic cases given its constant and rapid growth since the late 1970s, and it is becoming the world's second largest economy. Its rapid economic growth, called the "China Miracle," is the consequence of several factors, such as globalization, urbanization, and international trade, with logistics playing a fundamental role in all of them (Zhang and Cheng, 2009). However, China's logistics expenses to GDP represented 14.6% in 2017 (Statista, 2020). Although China has cut logistics spending by 18%, it remains below developed countries.

Remarkable promotion of logistics efficiency improvement and cost reduction has become an ongoing effort for the Chinese government (PR Newswire, 2019). Duzbaievna Sharapiyeva et al. (2019) stated that creating port infrastructure and logistics efficiency could lead to a positive impact on China's GDP. China and its neighboring regions could increase their per capita GDP growth by 7-8% by increasing port's performance by 1% (Duzbaievna Sharapiyeva et al., 2019). Based on Seabury's customs database in 2019, total import into China was 14.8 million TEUs and export from China was 47.5 million TEUs. Additionally, Seabury Consulting forecasts that total import and export will grow to 16.8 million TEUs and 54.0 million TEUs up to 2023, respectively (*Seabury Consulting*, 2020).

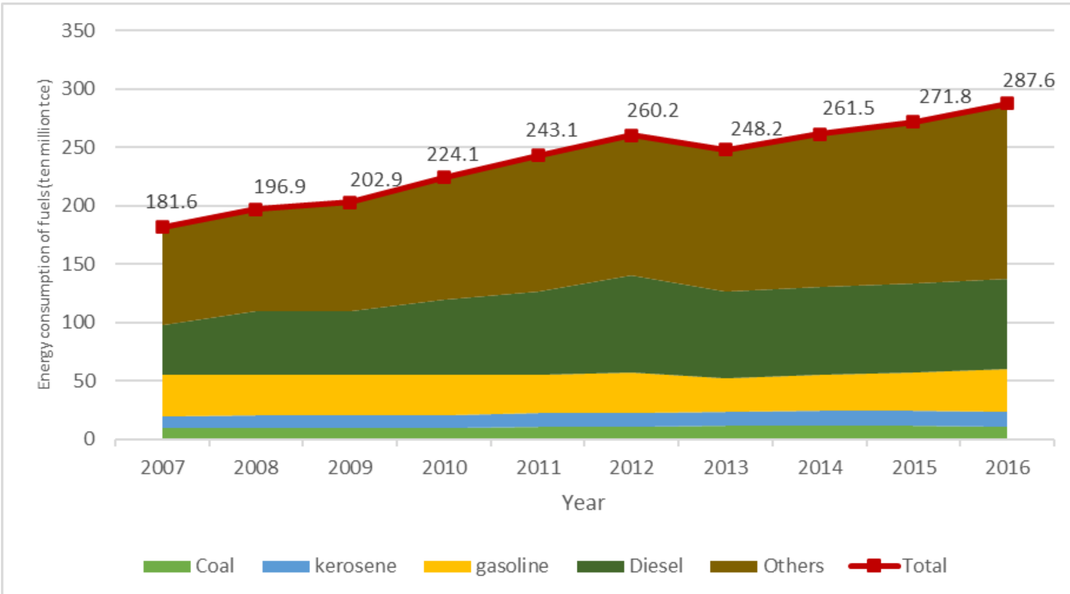
Another critical issue for China is that the economic growth has been accomplished at the expense of sustainability and energy consumption. China is the country that emits the most CO_2

in the world with 10.06 metric gigatons, representing 28% of global greenhouse gas emissions. This is 185% and 380% of the emissions from the United States and India, which are in the second and third positions, respectively (Union of Concerned Scientists, 2008). In its ‘Made in China 2025 blueprint’, China has set an ambitious target to decrease CO_2 emission intensity by 22% in 2020 and by 40% in 2025 compared to 2015 (*MADE IN CHINA 2025*, 2016., p. 19). On September 22, 2020, in the UN virtual General Assembly, China committed to limit greenhouse gas emission by 2030 and become carbon neutral by 2060 (China Pledges to Be Carbon Neutral by 2060, n.d.).

Along the same line, energy consumption in the China’s logistics industry has been exponentially increasing in the last decade, having an annual rate of increase of 5.2%, as depicted in Figure 1 (Liu et al., 2020). Recently, the transportation sector stands out in terms of CO_2 emissions, representing 20-25% of the global emissions, with road transport accounting for 74% of transport total emission (Berg & Langen, 2017).

Figure 1

The logistics-related energy consumption of China's three regions during 2007-2016



Adapted from: The logistics-related energy consumption of China's three regions during 2007-2016. (Liu et al., 2020)

A new challenge is the unexpected disruptions such as the outbreak of the COVID-19 pandemic. The global shipping industry has been experiencing unexpected crisis of container shortage. It started in the first quarter of 2020 and is still expected to last until the end of 2021. One of the reasons of the container shortage is that containers are being shipped from Asia to USA or European countries, but they are not returning to Asia due to the COVID-19 restrictions and port congestion in those countries. Another reason is that the severe gap in the container supply still exists during COVID-19, even with the container manufacturers in China increasing the production capacity. Hence, how to shorten the turnaround time for containers and alleviate the shortage has become a burning question for the shipping industry.

In recent years, companies in almost all industries, including transportation, have conducted several initiatives to explore new digital technologies and data-driven solutions to diminish inefficiencies. Potential benefits are manifold and include increases in sales and productivity, with the usage of data mining, automatization of processes, prescriptive algorithms, and many others with the purpose of acceleration decision-making (Matt et al., 2015). For inland container routes, digital solutions can help accelerate the allocation of import and export demand, potentially increase utilization by reducing container downtime and distance traveled compared to current processes, explained in Section 1.2.

Maersk is an integrated container logistics company, a member of the A.P. Moller Group with the headquarter based in Copenhagen. The company has three strong businesses – Ocean, Logistics & Services, and Terminals. With presence in 130 countries over 80,000 employees, Maersk is representing 20% of the global container shipping market share. The company has started the digital transformation journey since 2016 to become the global integrator of container logistics. With its digitalization strategy to drive competitive advantage, the company already launched several digital products in the market. As one of these digital products, the company is now developing a digital platform to improve service and communication between all parties. This platform allows the company to collect relevant information within the container transfer process such as the customer bookings through the platform and the detection of damage through photographs to determine if a container needs repair or not. Maersk is also committed to environmental sustainability and aims to generate net zero CO_2 emissions from its operations by 2050 (Maersk Annual Report 2019, 2019).

Digital solutions can be developed to reach a totally new level of creating systems with high reliability and efficient operations without any physical effort such as the actual process for the allocation of containers for importers' and exporters' orders (Efimova et al., 2020). How can a digital platform solution match millions of import containers and reuse them for export in China? This research project investigated the current process and challenges of automating container triangulation in a digital platform for Maersk in China. The transportation cost, lead time and CO_2 emissions of executing container triangulation were evaluated, using technological capabilities, such as Machine Learning (ML) or Mixed Integer Linear Programming (MILP).

1.2 Problem Statement and Objectives

Maersk's process of handling import and export containers consists of six steps as illustrated in Figure 2: 1) a trucking supplier collects import laden container from the port and delivers it to the importer's facilities; 2) the trucking supplier returns the empty container to the port or depot after cargo unloading at importer's premise; 3) the container is inspected and cleaned or repaired if it is dirty or damaged; 4) more empty containers are positioned by shipping company to the port or depot due to the imbalance of imports and exports in China; 5) a trucking supplier picks up an empty container from the port or depot and delivers to the exporter's facilities; and 6) the trucking supplier returns the laden container to the port after cargo loading at exporter's premise. In above steps 2 to 5, a depot for empty containers can be located outside the port area, ideally in locations minimizing the cost of storage, handling, and transportation from importers to exporters.

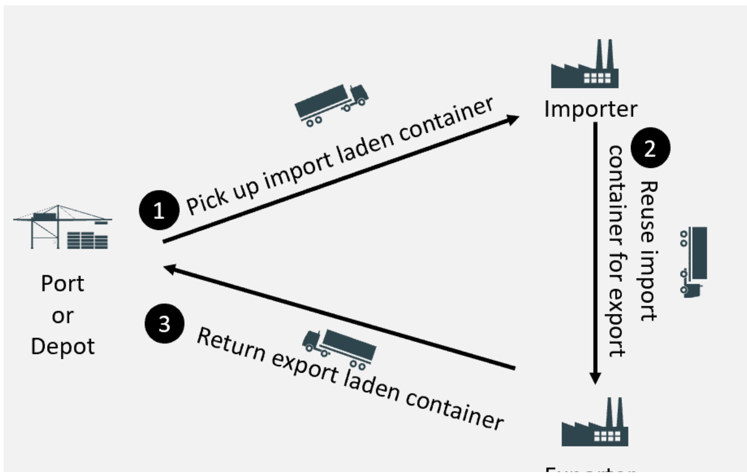
Figure 2

Current process of import & export

In order to improve the container turnaround and optimize the container transportation routing, a solution can be designed to reuse the import containers for export, matching the needs for importers and exporters. This process is called "container triangulation" as illustrated in Figure 3, the process includes: 1) a trucking supplier picks up import laden container from the port to the importer's premise for cargo unloading; 2) the trucking supplier sends the empty container to the exporter's premise after following a digital process for container inspection and 3) the exporter loads the cargo, the laden container is returned to the terminal for export. This solution would reduce total truck trips and distance as well as lead times. As a result, different parties in the supply chain get benefit from this solution. For the importer and exporter, the total trucking cost is reduced, the container turnaround time is shortened. For Maersk, the asset utilization is improved. For all parties involved, the greenhouse gas emissions decrease as the trucking distance is reduced.

Figure 3

Future process with container triangulation



Maersk tested the container triangulation with a pilot project for 50 containers in China. However, it is hard to scale up the volume for container triangulation. These are the typical challenges: 1) the importer and exporter normally contract their own trucking suppliers, while Maersk's own trucking suppliers should be used in container triangulation to ensure the safety of the container asset; 2) it is difficult to reach customers (importers and exporters) as there are several intermediaries between the ocean carrier and the customer; 3) reusing the import container for the export depends on the inspection result, and (4) matching of time and location for both parties. To resolve the inspection issue, Maersk has developed a digital platform applying artificial intelligence (AI) and machine learning (ML) to provide feedback to the trucking suppliers.

Based on nine pictures of the container taken by the trucking suppliers, the digital platform tells whether the container is sound and seaworthy for export with accuracy level at 94%.

As for the location and time matching, it is not practical to just have location data at city levels. Accurate location data down to the street or county level is needed. Currently, this level of location data is input into the digital platform by customers case by case; hence, Maersk is not able to scale up with more customers to use this digital platform for container triangulation, and predictive analysis cannot be performed to plan the match in advance due to lack of accurate location data.

With this motivation, the objective of this research project is to investigate this real-life problem for Maersk and develop a proposal for investigating the scale-up of a container triangulation digital platform in China and evaluation of the tangible value of saving in cost, time, and CO_2 for Maersk and its customers and suppliers. The geographical scope is based on China with a container flow outlined in Figure 3 with two years of data. The main deliverables of this project cover an exploration of the current global state in relation to the use of container triangulation. We studied how an algorithm can be modeled to suit the current situation in China. The approach for automating the container triangulation digital platform was examined as: 1) matching the demand between importers and exporters; 2) clustering customers' locations through machine learning algorithm and; 3) a route optimization model to obtain and quantify the savings of container triangulation. Finally, the study includes an in-depth discussion on the challenges to strengthen the model and the next steps required to improve the results, and our recommendations and insights for the implementation of the container triangulation on the digital platform.

2 LITERATURE REVIEW

This chapter explores the challenges and solutions for container triangulation in the logistics and transportation industry outside of China, as a digital solution that can be rapidly scalable in China has yet to be found. It also explores previously published work in the literature on topics relevant to developing a container triangulation solution and digital transformation in the transportation industry. It outlines the types of digital platforms developed to solve this problem and methodologies and algorithms applied to improve the matching rate of container triangulation. It describes the scope of similar problems, how they are solved, and their results and impacts. Finally, it also discusses the gaps in previous research and current solutions available in the market and what this research project contributes to the container triangulation solution proposed.

2.1 Digital transformation in the logistics industry

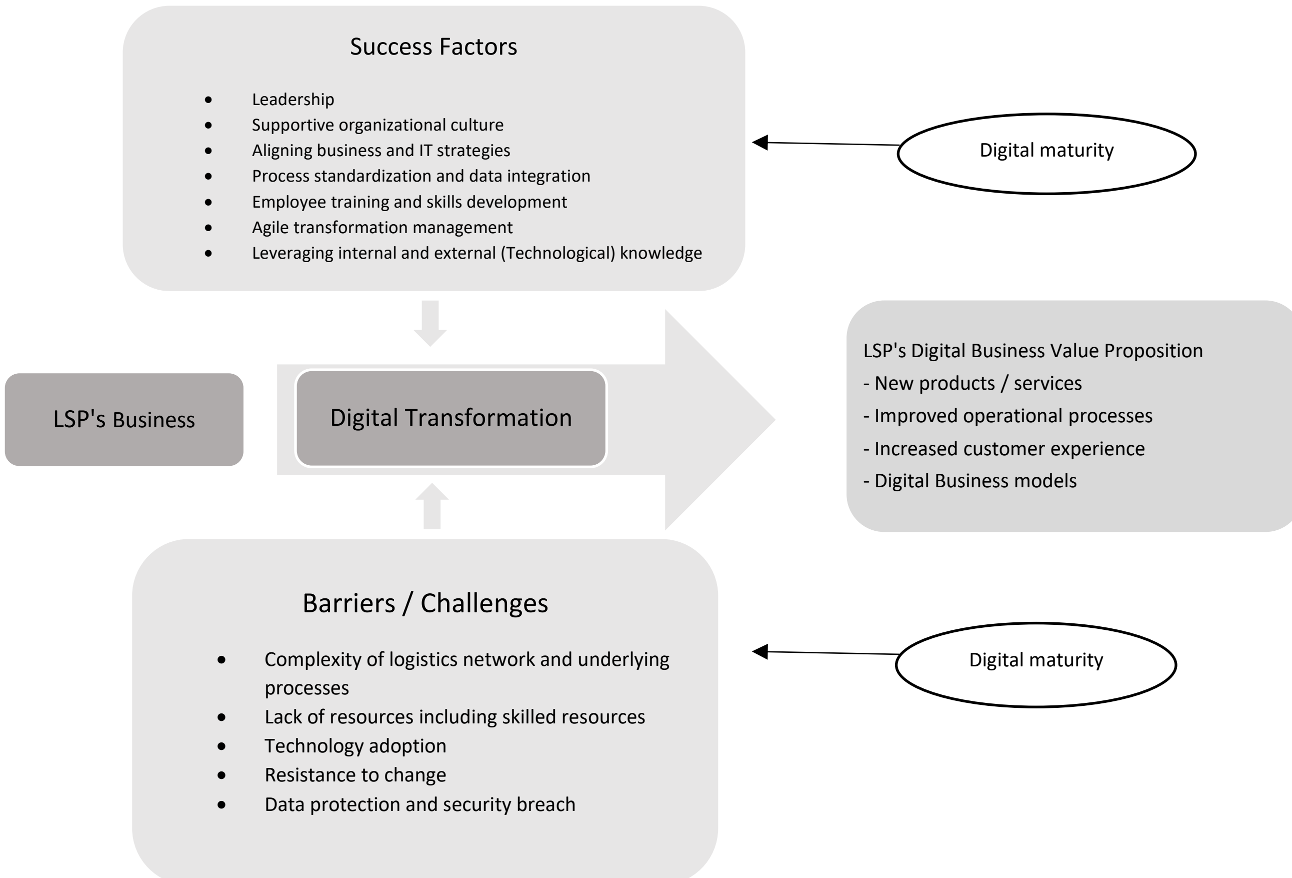
The digital age has fundamentally changed the competitive dynamics of industries. With the emergence of innovative newcomers such as Amazon and Alibaba (e-tailers), more technology-supported warehouses and transport have been invested (Cichosz et al., 2018). Crowd logistics platforms have also become very popular in the logistics market and had challenged current business practices and caught the attention of incumbent logistics service providers (LSPs) (Castillo et al., 2018).

Digital technologies have changed the competitive dynamics of the logistics service industry. Companies like Maersk have started upgrading their traditional services towards technology-supported transportation solutions. To offer a better customer experience with smarter, faster, and more sustainable logistics, companies must increase operational efficiency by addressing industry problems such as highly fragmented market, low transparency, underutilized assets, costly manual processes, and, in many instances, outdated customer interfaces (Cichosz et al., 2018). This new market environment requires the logistics companies to set strategic priority on investing in technologies for efficient performance, communication, and data visibility. As companies adopt technologies for their internal and external processes, digital transformation today becomes a must for better control and management of their new capabilities and resources.

Digital transformation is defined as the adoption of digital technologies towards major business improvements, increase productivity and value creation (Fitzgerald et al., 2014). A digital transformation strategy facilitates the collaboration of cross functional areas and accelerates the pace of business with the connection of systems to get valuable data. It is estimated that by 2025 digital transformation of logistics services could grow to $1.5 trillion of value at stake for logistics players and have a social benefit equivalent to $2.4 trillion (World Economic Forum White Paper Digital Transformation of Industries, 2016). Digitalization strategy, as part of a digital transformation, is a vision that the leading companies in the logistics industry have adopted and, in the same way, small businesses have found an opportunity by capitalizing on the advantages of information technologies (IT). Cichosz et al. (2020) modeled the barriers and success factors that logistics service providers (LSPs) have during a digital transformation (Figure 4).

Figure 4

Barriers and success factors to digital transformation for LSPs



Adopted from Cichosz et al., 2020

2.2 Methodology review in developing a digital platform

Digital platforms can be understood as a digital interface or ecosystem for the exchange of information, products or services to occur between suppliers and consumers (Hein et al., 2020). Digital platforms combine and deploy new ways to integrate and coordinate an ecosystem of supply and demand for a given type of product or service. Another advantage of a digital platform is that several stakeholders can interact and share information without the need for a direct relationship within the business. For example, Maersk, as the container provider, can be the link for coordination between the shipping company, the importer and the exporter. The model developed through this project can be integrated into the digital platform created by Maersk. By doing so, Maersk can include container triangulation as a new feature inside the platform for a better and faster execution of import and export container matching. Additionally, the digital platform can facilitate relevant data to company, such as the exact loading and unloading location of importers and exporters, respectively.

For this project, the integration, alignment, and visibility of the main stakeholders is critical. Suppliers (trucking companies), products (containers), and customers' (importers and exporters) demand and their characteristics is essential for the efficient execution of containers triangulation. In this section, the algorithms necessary to create the matching of all the participants through the use of Maersk's digital platform and its expected result are analyzed. With the use of large data and technology, the process of allocation for container demand is accelerated and scaled up.

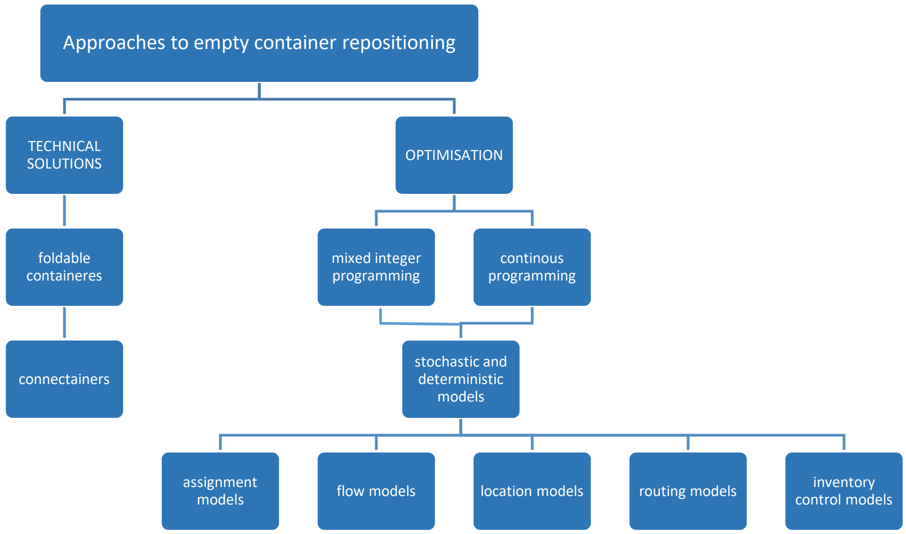
Technological advances have made it possible to incorporate tools in businesses that were previously impossible to manage digitally. Specifically, for the logistics industry, artificial intelligence (AI) has become a value-added to solve different business problems. AI is typically defined as the ability of a machine to perform cognitive functions that are associated with the human mind. Machine learning is one example of technologies that enable AI to solve business problems (*Artificial Intelligence*, 2020). The most recent advances in AI have been achieved by applying machine learning algorithms to very large data sets. These algorithms can detect patterns and predict or recommend a possible outcome, processing data and improving efficiency results over time and becoming a very powerful tool (*An Executive's Guide to AI | McKinsey*, 2020).

For the transportation sector, machine learning as one of AI techniques has been an ally to improve efficiency. ML can be used to improve performance over 89% beyond that provided by other analytic techniques for Transportation and Logistics (Chui, 2018). As another instance, transportation management systems (TMS) can provide predictive analytics for various business problems. These systems work with algorithms to identify patterns and guide business operations to improve performance and service and predict future trends (Bhavsar et al., 2017). An example of the use of predictive analytics techniques is the prediction of On-Time Delivery for Coyote Logistics by using logistics regression and resulting in cost reduction (Alcoba, 2017).

Several studies have investigated the best solution to manage inland loaded and empty container movements. One attractive solution for most of the participants in this process is the container triangulation, moving containers from importers directly to exporters. As Kuzmicz and Pesch (2019) stated, there have been different approaches to find optimization of empty container movements using container triangulations (or street-turns), mainly deploying MILP and continuous programming (Figure 5).

Figure 5

*Approaches in empty container repositioning**



*Note: Connectainer is a new 20-foot container that can be joined to another connectainer to form a 40-foot container.

Adopted from Kuzmicz & Pesch, 2019

Zhang et al. (2010) proposed a mathematical model using an extension of the Multiple Traveling Salesman Problem (MTSP) with time window for expected delivery date. Zhang's approach demonstrates high-quality results for the equivalent truck scheduling and inland container movement problem in container drayage operations. Deidda et al. (2008) studied both the truck routing problem and the optimization model determining the allocation of empty containers between importer and exporter. They compared their results with a real shipping company's practice and showed that their optimization model could produce a significantly better solution for the truck routings. Sáinz Bernat et al. (2016) used new stochastic review policies incorporating a realistic allocation scheme for empty container emissions, realistic maintenance, and repair options as well as street-turns. Evaluating with a simulation model, the results of their analysis shows a reduction of transportation costs in the repositioning of empty containers. Furió et al. (2013) proposed two different integer programming models to make inland empty container assignment under the empty container repositioning with and without container triangulation. Furió et al. (2013) showed that by applying the container triangulation strategy, savings reached up to 2% of total costs compared to no triangulation. Braekers et al. (2011) attributed this low level of improvement to the procedure's due to the complexity, the participants involved, and discrepancies in time, location, container ownership, and container type that the environment has. However, in recent years, attention to this problem has grown together with advancing technological developments. Some of those solutions complement the approach presented in this project, considering the Maersk's business model.

In this project, machine learning is applied with the usage of the k-Means algorithm. The algorithm defines a cluster (group) for each importers and exporters taking their location (longitude and latitude) as features (Section 3.2.2). In the second stage of our approach, we formulated and solved a Mixed-integer linear programming (MILP) to find an optimal vehicle routing model for the planning and allocation of containers. A MILP model consists of a mathematical formulation, often used for optimizing the integration of complex industrial systems (Vielma, 2015). For this project, we used MILP model, considering the location of each cluster and suggest the optimal route for each order/container. This model aims to minimize the distance traveled for each container (Section 3.2.3). By using machine learning and MILP algorithms, Maersk can increase the matching rate of container triangulation in China.

2.3 Current solutions to container triangulation outside China market

Container triangulation has been in operation for more than 30 years, supported with manual processes, which are not easily scalable. Literature provides an extensive body of studies on empty container management and repositioning of empty containers between ports; however, on many occasions, the environmental impact is not considered in the literature (Sáinz Bernat et al., 2016). In the past decade, several tech startups have been developing a digital solution for container triangulation for markets outside China. Today, three large digital platforms are being used by multiple shipping lines and transporters globally: Avantida, MatchBox, and MatchLog. Referring to Table 1 below, the key characteristics (market, service scope, volume, business partners, among others) are summarized based on the expert meetings with Avantida, MatchBox, and Matchlog (personal communication, October – November 2020). According to Table 1, MaerskMatch - a digital platform developed by Maersk, was launched in 2020 as a mobile application. In its initial phase, it is being used by some of the company's business partners for container triangulation.

Table 1

Comparison of different solutions in the markets outside China based on expert meetings

	Avantida by E2Open	MatchBox Exchange	MatchLog	MaerskMatch
Year started	2012	2016	2019	2020
Markets	Europe, USA total 26 countries	Australia, South East Asian countries	India only	China only
Service scope	Marketplace for container exchange, container reuse and value-added services			Container reuse
Volume matched	3,200 containers per month	Thousands of containers per month (specific number is not disclosed)	2,400 TEUs per month	Below 50 containers per month
Business partners	Multiple shipping lines, NVOCCs, transporters, depots, Imp and Exp customers			Maersk business partners only, e.g., contracted transporters, depots, Imp & exp customers
Methodology used	Transporters upload requests on the digital platform with map display, shipping lines approve container reuse or exchange with manual input on the platform. No need for AI or machine learning algorithms, online help is provided by country. Container inspection is done by the transporters and the exporters	AI and machine learning algorithms are used, depot network is optimized for container repair or exchange, cost saving calculator is available on the platform. Container inspection is done by the transporters and the exporters	Transporters upload booking details on the platform, OCR is used to capture data required for the match engine. Transporters take video of the container status and upload to the platform, AI and machine learning algorithms are used to confirm if the containers are seaworthy or not. Depot network is optimized for container repair or exchange. RPA is used to automate the process	Customers place booking on the platform for the match engine, Maersk teams manually check commodity and container off-hire status on systems and approve container triangulation, transporters take 9 photos and upload to the platform, AI and machine learning algorithms are used to confirm if the containers are seaworthy
Charge mechanism	Transporters pay the charge for container reuse and or container exchange, or large shipping lines choose to pay for subscription fee			Maersk contracted transporters get 10-20% discount of the trucking tariff
Learning from the rollout	Shipping lines force the transporters to use Avantida, phone calls or emails are not accepted for booking	Replicate the success in Australia to Singapore, then other South East Asian countries	Key account team is set up to ensure customer satisfaction, convince the partners that data is secure, the process is smart and highly automated	Difficult to get customers and transporters onboard the digital platform

Avatida, Matchbox and MatchLog (personal comuncation, October - November, 2020)

Avantida started in 2012, providing a cloud-based, online platform to enable container triangulation with container reuse from import to export and container exchange among shipping lines or transporters (Avantida - Home, n.d.). It now expands into 26 countries in Europe and USA markets, and 3,200 containers are successfully matched monthly. Noticeably AI and machine learning algorithms are not used; instead, Avantida provides online help for each country. The container inspection process is not automated yet, and it relies on the transporters and exporters to do the container survey to ensure that it is seaworthy. The main users are the transporters; hence, the key to a successful rollout is to mandate that the large shipping lines use Avantida.

MatchBox launched its platform in 2016 in Australia and expanded into Singapore and other Southeast Asian countries in 2017. AI and machine learning algorithms are used in the match engine, and a cost-saving calculator is provided on the platform. They partner with 14 shipping lines and 600 truckers (MatchBox Exchange | Better - Faster - Smarter, n.d.). Maersk has been partnering with MatchBox since 2016 in Southeast Asian countries only (MATCHBOX Exchange - Instant Container Reuse and Exchange Online Platform, n.d.). Similar to Avantida, the container inspection process has not been automated yet.

MatchLog started its operations in 2019 with a focus on India only (MatchLog Solutions, n.d.). The platform is designed on RPA (Robotic Process Automation), and OCR (optical character recognition) technology is used to capture the data from the bookings uploaded by transporters for the match engine to process the result. The container inspection process is enhanced with the transporters taking video of the exterior and interior views of the container and uploading it to the platform. AI and machine learning algorithms are used to confirm whether the containers are seaworthy or not within 13 minutes. MatchLog also sets up its own depots and partner depots to optimize the network for container exchange and repair in India. They are driving the automation in the end-to-end process to ensure contactless process enabled by digitalization and reduce handover points, especially during the COVID pandemic.

However, these companies have no presence in China market yet, as there are some entry barriers for foreign tech companies to penetrate the Chinese market. Some international services such as APIs (Application Programming Interface), SKDs (Software Development Kit) from overseas can be blocked by the Great Firewall of China. Hence a successful digital platform works in overseas markets might not work in China unless they find a local partner or switch to a local platform.

2.4 Conclusion

Maersk needs a digital solution that can scale up the container triangulation offering to its customers in China. Compared to Avantida, MatchBox and MatchLog with their digital platforms, Maersk just launched its digital platform MaerskMatch in 2020, and manual processes are still required to check the commodity and container off-hire status in various systems to approve container triangulation. The depot network is not yet used to enable a better match for container re-use. Business partners are constrained to Maersk customers and Maersk-contracted transporters, limiting the scope for the container match.

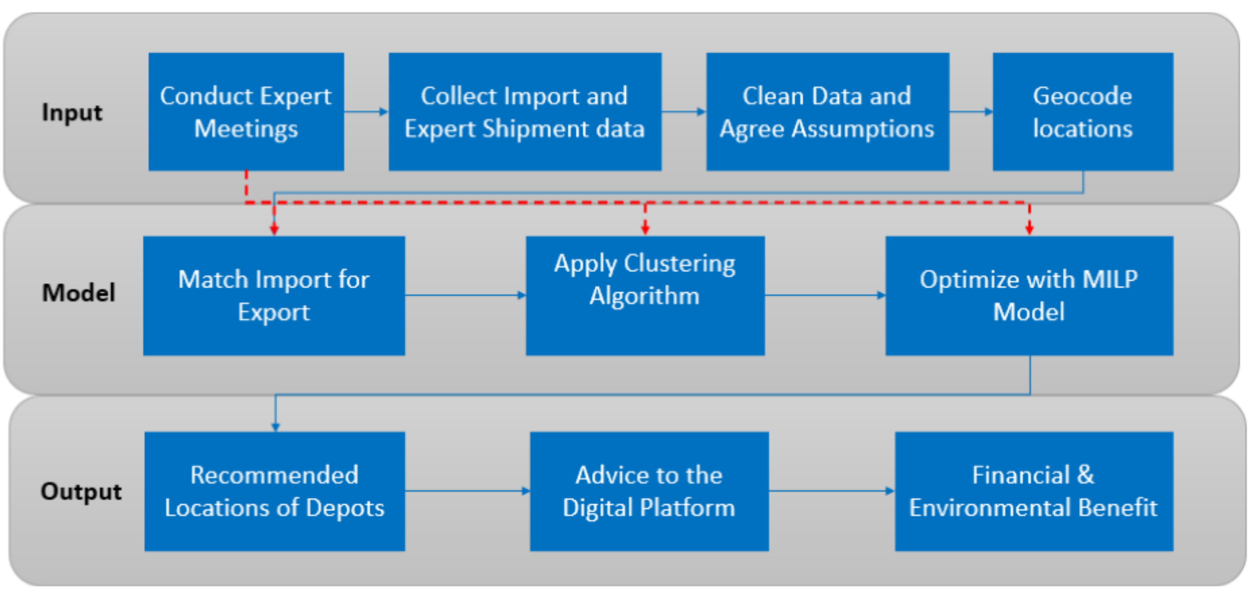
Using the information learned during the literature review, this research project contributes to developing a digital transport flow optimization model of the container triangulation process in China. Using ML algorithms and a mathematical optimization program to increase decision-making speed. Additionally, the model provides insight into critical factors to accelerate the import-to-export container match ratio using machine learning, and finally quantifies the solution's cost, time, and CO_2 emission savings.

3 DATA AND METHODOLOGY

This chapter outlines the key steps to gather the input and build the model to examine how to match import container for export in China. As shown in Figure 6 below, at the input stage, multiple rounds of meetings were conducted with the experts from the sponsor company, and practitioners in the markets covering North America and Europe, India, and Southeast Asia. With a good understanding of the problem and experts' insights, sample data was framed and collected from the sponsor company, followed by data cleaning and analysis and geocoding. Moving to the model stage, match and clustering algorithms were applied, then a MILP model was developed to run the transport flow optimization. Finally, at the output stage, financial

Figure 6

Methodology Overview



savings and environmental benefits were quantified based on the container distance traveled. Recommendations are concluded by combining quantitative analysis and expert advice.

3.1 Input Stage

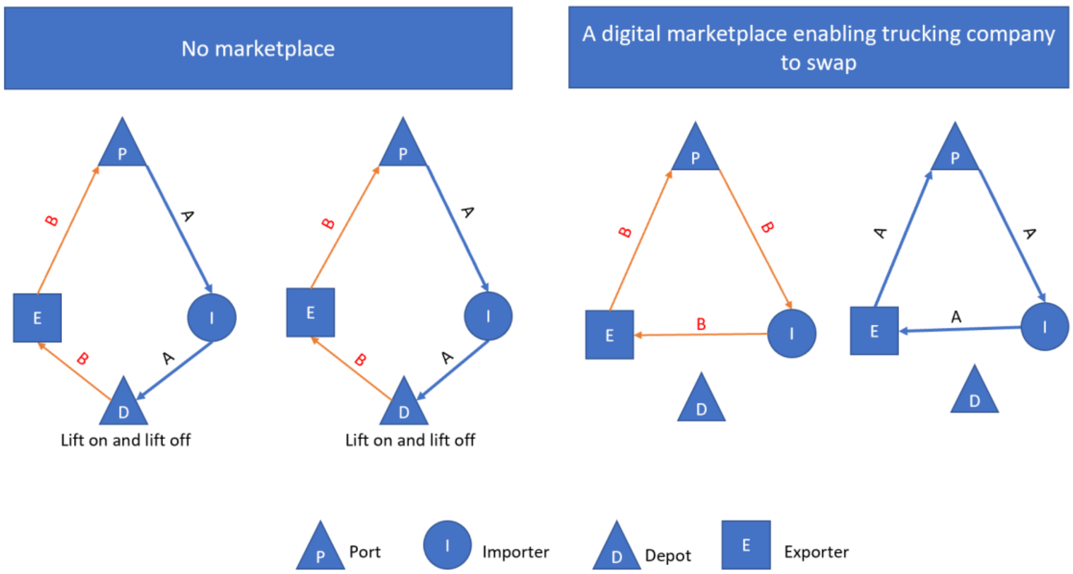
This stage covers expert meetings, data collection and data types, data cleaning, and geocoding.

3.1.1 Expert Meetings

Experts in trucking, equipment management, and intermodal transports from the sponsor company were invited and actively engaged in this research project. The main drivers during the meetings were to explain their process and their relationship with the key participants in the process (container companies, trucking companies, exporters, importers). The key findings are: 1) with the digitalization in ports such as the Shanghai port, the paperless process supported by e-EIR (electronic Equipment Interchange Receipt) creates a rising opportunity for trucking companies to exchange orders; 2) new entrants and incumbents are using a digital platform to match the importer and exporter container demand. A marketplace function is added in the platform, which enables different trucking companies to swap orders in a faster way and scale up the matching. For example, one trucking company can swap with others to handle both the import delivery and the export of the container. This increases the base of matching import containers for export and helps avoid additional moves of lifting off a container from one trucking company to the other in a depot with the required top-loader. As illustrated in Figure 7 below, without a marketplace function trucking companies A and B need to go to a depot with the top loader to lift off the container from trucking company A and lift on to trucking company B. With a marketplace function, trucking companies A and B can swap orders without going to a depot to complete one triangulation trip.

Figure 7

Comparison of scenarios with and without marketplace



3.1.2 Data Collection

Two years of transactional data for import and export orders at the container level were collected from the sponsor company. A sample data spanning three months from July 2020 to September 2020, containing 1.13 million records and 33 features was used for the preliminary analysis and methodology testing. Knowing that the location data for importers and exporters is only at the city or county level in the shipment data, the street-level address data was collected from a separate *Customer Master* database, which contains 2 million records globally including 300,000 records for China.

As for the China market, Seabury data for China import and export was collected (*Seabury Consulting*, 2020), which covers the import and export volume in containers based on customs data. The data period covers over 10 years.

3.1.3 Data Types

The shipment data variables and the structure of import and export containers are outlined in Table 2 below. Each record represents a unique shipment at the container level. Key data elements constitute the name and location of importer and exporter, container type, cargo description and weight, date of import and export, port name, and place of receipt (POR) for export shipment and place of delivery (POD) for import shipment, which are used in data analysis.

Table 2

Import and export shipment data

Shipment Data variables	Type	Remarks
Shipment Number	Text	Shipment unique ID e.g., '1KT000019'
Consignee City Name	Text	Importer city or county e.g., 'Dongguan'
Consignee Cust Code	Text	e.g., '40922179DPC'
Consignee Cust Name	Text	e.g., ABC China
Container Number	Text	Container unique ID e.g., 'MRKU6315056'
Container Size	Categorical	e.g., 20', 40', 45'
Container Height	Categorical	2 types: standard 8'6", high cube 9'6"
Container Type	Categorical	Mainly dry, reefer, flat, open etc.
Shipper Owned Container	Binary	Whether the container is owned by the customer or by shipping line, e.g., Y, N
Cargo Type	Categorical	e.g., 'Dry' or 'Reefer'
BTN Name	Text	Brussels Tariff Nomenclature (BTN) cargo description 1st level e.g., 'Metal'
Item Name	Text	Cargo description 2nd level e.g., 'Aluminum, aluminum articles, metal'
IMCO Detail	Text	IMCO dangerous goods class info e.g. '2.1'
Port Arrival Last Date	Date	Date of last arrival in port for import
DIPLA City Name	Text	Discharge Port Last (DIPLA)- for import shipment e.g., 'Shanghai'
Port Departure First Date	Date	Date of first departure from port for export
LOPFI City Name	Text	Load Port First (LOPFI) for export shipment e.g., 'Shanghai'
POD City Name	Text	Place of delivery (POD) for import shipment, city or county level
POD Country Name	Text	Place of delivery (POD) for import shipment, country level
POR City Name	Text	Place of receipt (POR) for export shipment, city or county level
POR Country Name	Text	Place of delivery (POR) for export shipment, country level
Receipt Service Mode	Text	Service starting from origin for exporters mainly CY, Store Door, CFS etc.
Delivery Service Mode	Text	Service to destination for importers mainly CY, Store Door, CFS etc.
Shipper City Name	Text	Exporter city or county e.g., 'Suzhou'
Shipper Cust Code	Text	e.g., '40600829054'
Shipper Cust Name	Text	e.g., Suzhou XYZ co.

FFE	Numeric	Converted into Forty-Foot-Equivalent (FFE), e.g., 20' = 0.5 FFE, 40' and 45' = 1 FFE
TON	Numeric	Cargo weight in tons

The *Customer Master database* contains customer code and location data down to street level as per Table 3, which is matching the Consignee and Shipper Customer Code in *Shipment Data* in Table 2.

Table 3

Customer Master Database

Customer Master Data	Type	Remarks
Customer Code – SCV	Text	Customer Code e.g., '40600829054' = Consignee and Shipper Cust Code
Street Name	Text	e.g., 'NAN JING RD (W)'
City	Text	e.g., 'SHANGHAI'
Province/State	Text	e.g., 'SHANGHAI'
Country	Text	e.g., 'CHINA'
Postal/Zip code	Text	e.g., '200000'

The external data from *Seabury China Shipper Database* is outlined in Table 4

Table 4

Seabury – China Shipper Database

Seabury China Shipper Data	Type	Remarks
Dates	Date	Aggregated into Year - Month, no breakdown into date
Clearance Points	Text	Country, province, county level
Direction	Categorical	e.g., 'Import', 'Export'
Shipper Name	Text	Importers and Exporters in China e.g., 'Midea'
Shipper Location	Text	Country, province, county level
Partner Countries	Text	Country and city of overseas partners of China importer and exporter
Commodity	Text	Cargo description of the import and export shipments
Volume in TEU	Numeric	Aggregated in Twenty-Foot-Equivalent e.g., 40' = 2 TEUs

3.1.4 Data cleaning and assumptions

Based on the preliminary study of data presented above, the import and export shipment data according to Table 2 has a better data quality to conduct the next level of data analysis.

There are two main reasons: 1) the locations of importers and exporters, the import and export dates, container types, POR and POD locations already exist in this dataset, and 2) more detailed location data down to street level can be obtained from the *Customer Master* database as per Table 2. However, these locations are often the office addresses of the importers and exporters, and they are not necessarily the real cargo delivery addresses and cargo receipt addresses. At this stage, it is agreed that a practical way to continue the study is to use POR and POD locations at the city- and county-level to test the match algorithm. With more accurate locations in the future, the algorithm can be enhanced to generate more realistic results. The challenge of capturing the accurate location data will be further discussed in Sections 5.1 and 5.2.

The Seabury China Shipper Database cannot be used as a reference to compare the output from the analysis mentioned above since the China Shipper is aggregated on a monthly basis, missing container type details and the breakdown of import and export dates.

3.1.5 Geocoding

To display the locations of importers and exporters on the map more accurately and calculate the distance between locations through network flow optimization, geocoding is needed to obtain the location latitude and longitude. Three geocode methods were tested and compared: Amap API, Geopy Nominatim, and Google API.

Amap was acquired by Alibaba in 2014, and it is the largest location-based data intelligence platform in China (Zhenfei, n.d.). It provides web service API for location geocoding, reverse geocoding, and route planning. The geocoding generates the latitude and longitude data at the city-level and customer-level of 130 and 21,000 different locations, respectively. Testing with a file of 130 locations, the geocoding process takes around 9 minutes. We validated the geocoding results by displaying the location on the map. The geocoding of the Chinese addresses in mainland China is highly accurate; however, it failed to parse English addresses in China.

Nominatim can also perform the geocoding based on name and address using OpenStreetMap data (*Welcome to GeoPy's Documentation! — GeoPy 2.0.0 Documentation*, n.d.). Testing with a file of 130 locations, it takes around 10 minutes to complete the geocoding and write to a CSV file as depicted in the Appendix. Overall, Nominatim is slower and less accurate than Google web service.

Google API works fastest in providing the most accurate results among the three methods. It takes around 30 minutes to geocode 10,000 locations and write to a CSV file. With a complimentary Google API key, it is workable to run multiple times to geocode a large data file with each time capped at 10,000 locations. For this project, this method is recommended as a geocoder as 30,000 locations cover more than 95% of the import and export container volume in China. Even with the geocoding from Google API, it is important to verify of the locations displayed on the map and manually overwrite the incorrectly mapped locations.

3.2 Model Stage

This stage encapsulates the match algorithm, clustering algorithm, and MILP. The purpose of this model is to automate the process of matching the demand of importers and exporters. This accelerates the allocation of a container for an exporter once the importer release (unloads) a container and find his next service. The design of the approach will be described in the following sections.

3.2.1 Match algorithm

Once we have the data ready to work, we can seek potential efficient reuse of a container that has been emptied. This linkage between the import container release and the export appointment must be quick to determine the next stop/route of the container. There are two groups of linkage methods: deterministic and probabilistic (Zhu et al., 2015). The first kind is used for data-rich projects with high-quality identifiers such as address or container ID. The second one reflects the probability that data records fall under the same category. In this project, we investigated both types for the container matching by defining the dependent and independent variables and identifying possible container triangulations.

The deterministic approach takes the identifiers necessary to compare records at once and obtain a binary result (true or false). If the pair of records agree on all identifiers, it is considered as a match, otherwise a non-match. The decision of a match can be implemented step-by-step, prioritizing some identifiers, or match them all at once. The probabilistic algorithm uses u-probabilities and m-probabilities to determine if a record has a 'possible match' or not.

The variables or identifiers that help us determine the match and reuse of import and export containers are mainly the attributes related to each customer's needs. We can define them as follows:

- **Cargo type**: container characteristic (dry, open, flat).
- **Cargo height**: container characteristic (2 types: standard 8'6", high cube 9'6 ").
- **Cargo size**: container characteristic (20', 40', 45').
- **Volume:** Balance between imports and exports in the region for a higher reutilization.

- **Industry/commodity**: Type of materials transported to determine the next possible use of the container. Some of them are more harmful to the container than others (e.g., transit chemical products have a high risks)
- **Container status:** Use of AI photo inspection to determine whether repairs are needed.
- **Appointment date**: The scheduled date to arrive to the importer and the exporter location. A 24-hour period is the expected time to arrive to the importer, unload, and transit to the exporter.

With the proper data cleaning and understanding of these critical variables, a matching algorithm can decide if the container is in correct conditions to be reused and apply clustering method. The algorithm runs based in the deterministic linkage methodology, matching first the 'appointment date' of importer and exporter. The following step is to match the specification of each order, such as 'Cargo type', 'Cargo size' and 'Cargo height'. 'Industry/commodity' and 'Container status' are not included for this project because of the data accessibility, but they could be included as relevant matching features in future researches. Once the matching is achieved, the remaining feasible data is considered the input for the clustering algorithm.

3.2.2 Clustering

Clustering is one of the most common approaches in data analysis. Unlike classification methods, in clustering, there are no known labels to train a model (Bhavsar et al., 2017). It is an unsupervised method to classify elements into discrete groups based on their similarities or discovered patterns (Jain et al., 1999). The clustering algorithms can be categorized into four main groups: partitioning algorithm, hierarchical algorithm, density-based algorithm, and gridbased algorithm (Patel & Thakral, 2016). k -means algorithm represents a partitioning algorithm with useful clustering techniques by competitive learning. Clusters start with an initial partition and then use an iterative control strategy to optimize an objective function (Jianliang et al., 2009).

The algorithm steps are provided as follows (Jianliang et al., 2009):

1. Given a data set, define k centroids: $C = c_1, c_2 \ldots \ldots .. c_k$ one for each cluster, $c_i = \frac{1}{n_i} \sum_{x \in w_i} x$, where n_i is the number of the dataset in the cluster.
2. Each data item is allocated to one cluster that has the closest centroid.

   The algorithm minimizes objective function, being the square error function:

$$J_{KM} = \sum_{j=1}^{k} \sum_{i=1}^{n} d_{ij} \left(\left\| x_i - c_j \right\| \right)^2$$

   where,

   k represents the number of clusters.

   n represents the amount of data items in the data set.

$$d_{ij} = \begin{cases} 1 \; if \; c_j \; is \; the \; centroid \; closest \; to \; x_i \\ 0 \; otherwise \end{cases}$$

   $\left\| x_i - c_j \right\|$ represents the Euclidean distance between each data item x_i and the centroid c_j

3. Assign each instance to its closest center: if $d_{ij}\left(x_i, c_j\right) < d_{im}(x_i, c_m)$, where,

   $m = 1 \ldots \ldots .. k, j = 1 \ldots \ldots .. k, j \neq m, i = 1 \ldots \ldots .. n$ and then assign x_i to cluster C_j;

4. Recalculate the new centroids of the clusters.

   $c^*{}_1, c^*{}_2 \ldots \ldots .. c^*{}_k$

5. Repeat steps 2 and 3 until: a) the centroids do not change with the successive iterations, b) all points remain in the same clusters with the successive iterations or c) the maximum number of iterations is reached.

By completing this algorithm, it is possible to get the grouping objects into meaningful k subclasses so that the members from the same cluster are quite similar, and the members from different clusters are quite different from each other.

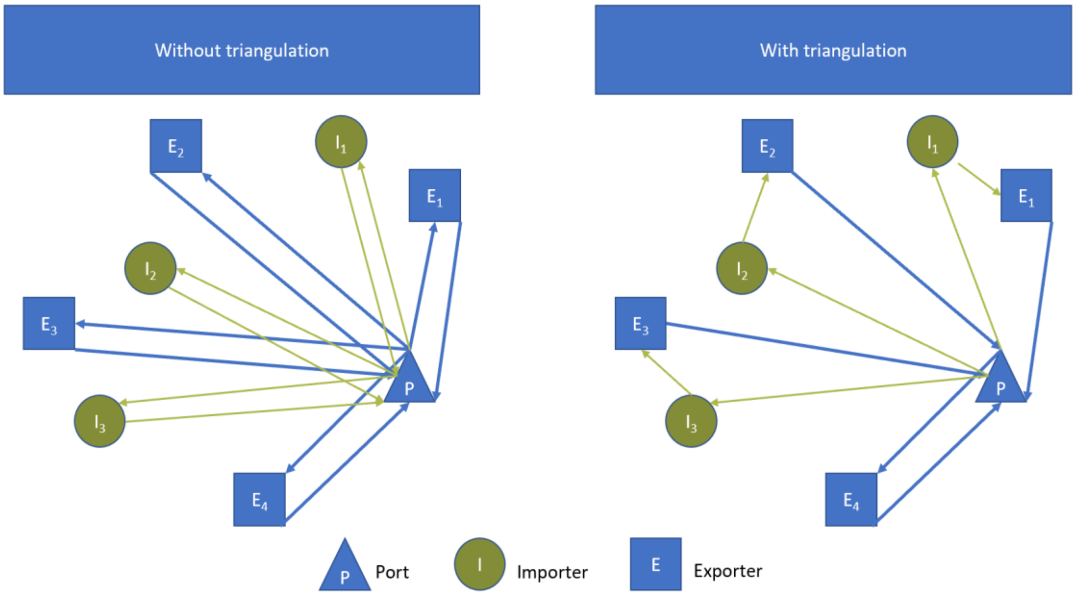
The k-means algorithm is used to define the different geographical zones in China and cluster the customers that are near each other based on distance. Like the matching algorithm, k-means clustering reduces the complexity of the problem that will be taken to run the MILP.

3.2.3 Transport flow optimization – Mixed-integer Linear Programming (MILP)

After the matching and clustering, a subset of the matched data is used as input for transport flow optimization conducted by applying MILP. This subset of data should include the import and export container volume for the same container type and trucking delivery or pick-up window, the geocode data of importer and exporter locations, and port location so that the distance matrix and time matrix among locations are calculated.

Figure 8

Transport flow without triangulation vs. with triangulation



As shown in the illustration of the transport flow in Figure 8, we are given a set of importers, I, and a set of exporters, E. Each exporter j requests to load a number of containers, n_j , and each importer i has a number of containers for unloading, n_i . These containers are of the same equipment type, and only one container is loaded on a truck. In the current scenario without triangulation, the exporters and importers need to arrange trucking transportation from port to their premises with round trips. When the transport flow is optimized with triangulation, the objective function aims to minimize the total container-distance (the number of containers x km) traveled of triangulation - trucking containers from the port to importers, from importers to exporters, and from exporters returning the containers to the port. When the total number of import containers is not equal to the number of export containers, the remaining containers are then transported with round trips between the port and importers, or between the port and exporters. Normally, the container triangulation should be executed within the same day to avoid truck overnight charges. Hence, we added the constraints for the maximum distance travelled per route, L, and maximum service time per route, T, in the optimization model as outlined in Table 5.

Table 5

Nomenclature of formulation

Type	Nomenclature	Description
Index and sets	i	Individual importer, $i \in I$
	j	Individual exporter, $j \in J$
	p	Individual port p used by importer i or exporter j, $p \in P$
Parameters	d_{ij}	Distance of importer i to exporter j
	d_i	Distance of importer i to port
	d_{jp}	Distance of exporter j to port
	n_i	Number of import containers per importer i
	n_j	Number of export containers per exporter j
	M	A large number such as the total number of import and export containers
	L	Maximum distance traveled allowed per route
	t_i	Service time for importer i including the time for waiting and unloading
	t_j	Service time for exporter j including the time for waiting and loading
	v	Average truck speed for all routes
	T	Maximum service time allowed per route
Decision variables	x_{ij}	Total number of containers from importer i to exporter j
	x_{ip}	Total number of containers from importer i to port p
	x_{jp}	Total number of containers from exporter j to port p
	y_{ij}	1 if $x_{ij} > 0, otherwise\ 0$
	y_{ip}	1 if $x_{ip} > 0, otherwise\ 0$
	y_{jp}	1 if $x_{jp} > 0, otherwise\ 0$

The mathematical model of MILP is formulated as below:

Minimize $z = \Sigma_i \Sigma_j ((d_{ij} + d_{ip} + d_{jp})x_{ij} + 2d_{ip}x_{ip} + 2d_{jp}x_{jp})$ (1)

$subject\ to$:

$\Sigma_j x_{ij} + x_{ip} = n_i \quad \forall_i \in I, \forall_p \in P$ (2)

$\Sigma_i x_{ij} + x_{jp} = n_j \quad \forall_j \in E, \forall_p \in P$ (3)

$x_{ij} - y_{ij}M \leq 0 \quad \forall_i \in I, \forall_j \in E$ (4)

$x_{ip} - y_{ip}M \leq 0 \quad \forall_i \in I, \forall_p \in P$ (5)

$x_{jp} - y_{jp}M \leq 0 \quad \forall_j \in E, \forall_p \in P$ (6)

$(d_{ij} + d_{ip} + d_{jp})y_{ij} \leq L \quad \forall_i \in I, \forall_j \in E, \forall_p \in P$ (7)

$2d_{ip}y_{ip} \leq L \quad \forall_i \in I, \forall_p \in P$ (8)

$2d_{jp}y_{jp} \leq L \quad \forall_j \in E, \forall_p \in P$ (9)

$(\frac{d_{ij}+d_{ip}+d_{jp}}{v} + t_i + t_j)y_{ij} \leq T \quad \forall_i \in I, \forall_j \in E, \forall_p \in P$ (10)

$(\frac{2d_{ip}}{v} + t_i)y_{ip} \leq T \quad \forall_i \in I, \forall_p \in P$ (11)

$(\frac{2d_{jp}}{v} + t_j)y_{jp} \leq T \quad \forall_j \in E, \forall_p \in P$ (12)

$x_{ij}, x_{ip}, x_{jp} \geq 0 \quad \forall_i \in I, \forall_j \in E$ (13)

$y_{ij} = \{0,1\}, \quad \forall_i \in I, \forall_j \in E$ (14)

$y_{ip} = \{0,1\}, \quad \forall_i \in I, \forall_p \in P$ (15)

$y_{jp} = \{0,1\}, \forall_j \in E, \forall_p \in P$ (16)

The objective function (1) represents the total container-distance traveled. When the point-to-point cost matrix is available, the distance d_{ij} can be replaced with cost c_{ij} to minimize total cost. Constraint (2) requires that total number of containers allocated in different routes for a specific importer should be equal to the total amount of containers for that importer, Constraint

(3) ensures that total number of containers allocated in different routes for a specific exporter should be equal to the total amount of containers for that exporter. Constraints (4) to (6) are the linking constraints to ensure that when the flow is allocated to a certain route, the binary variable y_{ij} is 1, which means the route is selected, otherwise it is 0. Constraints (7) to (9) ensure that each selected route is below the maximum distance. Constraints (10) to (12) require that total service time from each selected route is below the maximum service time. Constraint (13) ensures that the flow decision variables among importers, exporters, and port are positive values. Constraints (14) to (16) define the auxiliary decision variables for routes among importers, exporters, and port with binary values.

4 RESULTS AND NUMERICAL ANALYSIS

This chapter presents the results and analysis of the modeling approach for the proposed container triangulation. After designing and developing the algorithm described in Section 3, we analyzed the outputs of our model compared to the actual process of handling import and export containers. We focused our analysis on the outputs of two principal methods: k-means and MILP algorithms. The results were obtained from the historical data spanning two years of import and export shipment data at the container level. In order to perform a better analysis, we took different samples from the dataset to run the model and evaluate the outcomes. This samples were selected on specific weeks of year 2019 considered with stable and representative demand. We compared the results and determined the final findings.

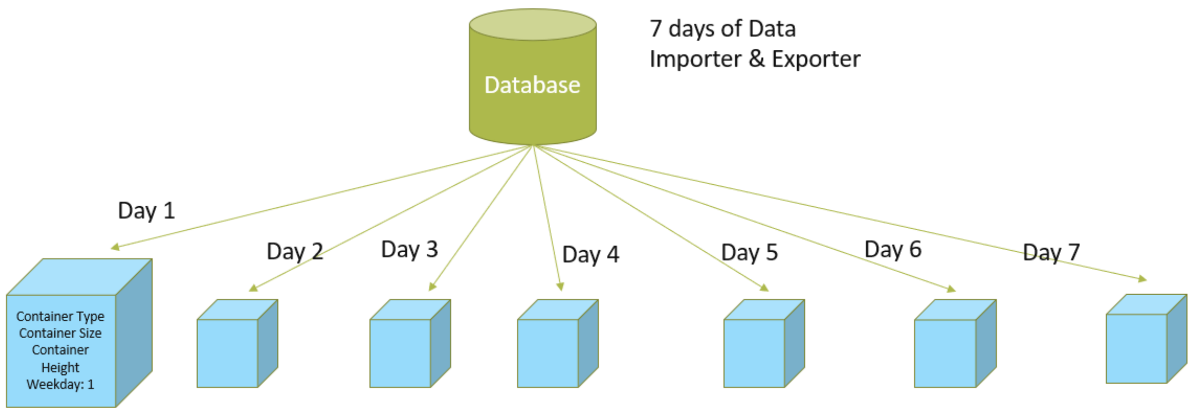
4.1 Matching Analysis

The first step was to create a reliable dataset that had all the relevant variables to determine the best container route (with or without triangulation). That consists of the connection between *Customer Master* database and the geocode of each customer's location. Once we merged the information needed, we matched samples that meet the requirements between exporter and importer. Matching was made to identify feasible opportunities with exporters and importers based on ranked orders, with the purpose to reduce complexity for the MILP algorithm.

The matching algorithm finds all the possible orders with the same Container Type, Container Height, Container Size, and Appointment Date (Figure 9). The results presented in this section are taken from a sample with the following conditions: 1) Year-Week Appointment: 2019-40; 2) Container Size: 40'; 3) Container Type: DRY; 4) Container Height: 9'6"; and 6) Appointment Weekday: Monday. We considered the week 2019-40 as it was a week with a representative volume of demand to measure the business performance, using without disruptions or unusual behavior that might bias the results. We also picked those parameters with the same objective, as they are the most common and relevant for the business. From a dataset of 79,320 records, by applying the matching algorithm, we ended with 652 records matched with the same conditions. This helped us determine our model boundaries, feasibility and reduce complexity to the next stages of the model (*k*-Means and MILP). This dataset will be called *Test Data*.

Figure 9

Representation of the Matching Algorithm by parameters



4.2 k-Means Clustering

The clustering analysis is used to discover similarities between data items and group them into different categories, known as clusters (Jain et al., 1999). In this research, we applied one of the pattern-recognition techniques called k-means clustering. The following section will illustrate the results obtained by applying the k-Means algorithm in our *Test Data*.

4.2.1 Mapping

The clustering was implemented to identify the group of customers based on their locations. This aggregation was done by taking the latitude and longitude of each customer of our sample data. The first step was to map the city-level location of each of the customers, either exporter or importer. This step is illustrated in Figure 10 in a cartesian plane and in Figure 11 on a map representation. By plotting the locations, it was possible to get a perspective on their distribution.

It helped identify outliers, clean up missing data, and get an idea of possible outcomes. Once we determined the location of each customer's order, we calculated the number of clusters needed.

Figure 10

Representation of customers' latitude and longitude within a cartesian

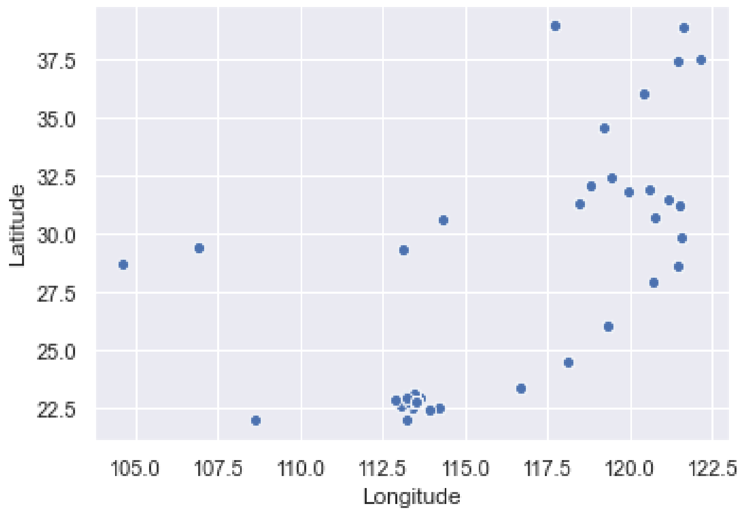
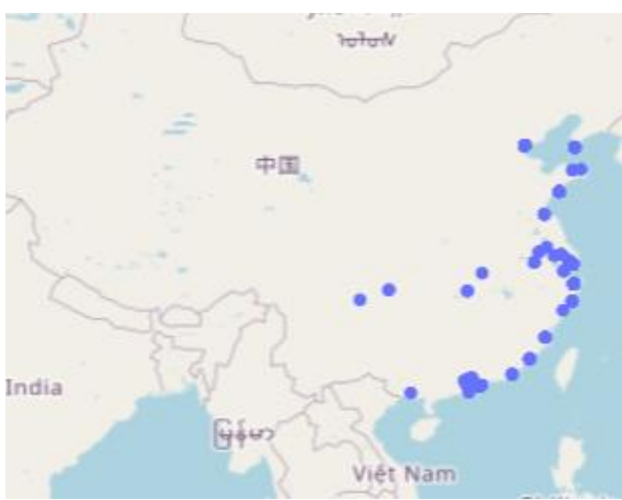


Figure 11

Representation of customers' latitude and longitude on a map

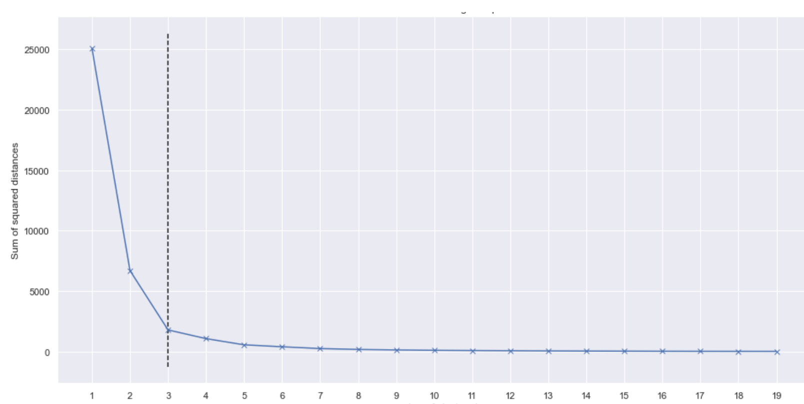


4.2.2 Number of clusters

There are several approaches to specify the number of clusters required for unsupervised records. We used the elbow method, which minimizes the sum of square error (SSE) within clusters and looks for the bend in the curve (Syakur et al., 2018). The best cluster k result will be the basis for clustering. The lower the SSE value and the elbow graph, the better the cluster results. For our *Test Data*, we analyzed the value of the cluster from k = 2 to k = 3, then from k = 3 to k= 4. It shows a drastic decrease to form the elbow at point k = 3 then the ideal cluster k is k = 3 (Figure 12).

Figure 12

Representation of the Elbow Method



4.2.3 k-Means Result

Once we determined the number of clusters (k), k-Means algorithm starts iterating starting with k cluster-centers placed randomly in the data space, and then the following stages are performed repeatedly until convergence:

1) Data points are classified by the center to which they are nearest.

2) The centroid of each cluster is calculated.

3) Centers are updated to the centroid location.

To evaluate if the final clusters present high heterogeneity between the different clusters and high homogeneity within the cluster, graphical representations of the clusters were analyzed. Figures 13 and Figure 14 show the output of the k-means. Figure 13 illustrates the centroids with a bigger red dot and each cluster represented with a shape. Figure 14 represents of the clusters on a map. We found the three clusters of cities (customers) that will be considered for each routing model.

Figure 13

Representation of customers' clusters within a cartesian plane

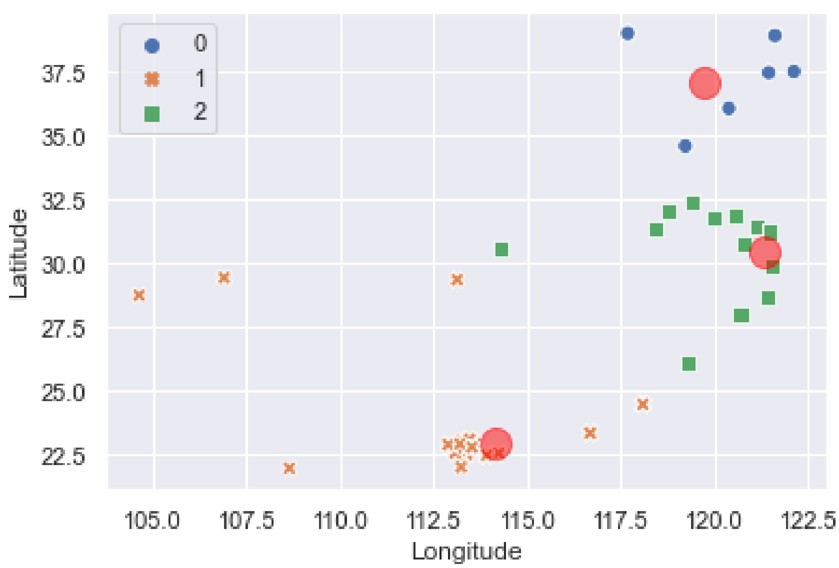
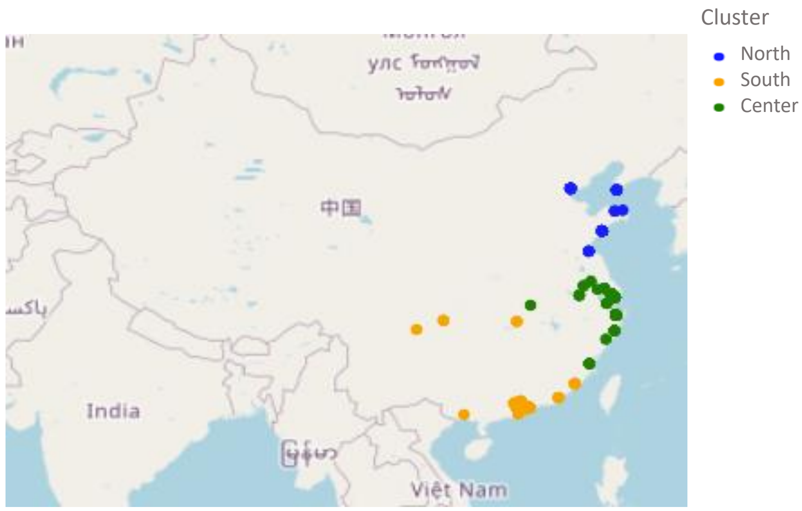


Figure 14

Representation of customers' clusters on a map



4.3 Clustering Results

Once we have assigned each customer's order into a cluster, we analyzed the match created to identify the distribution between exporters and importers. From the 652 records (containers) in our *Test Data,* we obtained three different clusters with a total of 119 importers and 504 exporters (Table 6). Due to the trade imbalance of imports and exports in China, a perfect container triangulation is infeasible. However, there is still an enormous potential of 119 containers to find possible container triangulations between exporters and importers. The clustering analysis allows Maersk to get insights about the demand's behavior and take actions. Once we determined the matches, we took the distribution of the exporters and importers' demand to run the MILP algorithm.

Table 6

Results obtained from the Clustering Analysis.

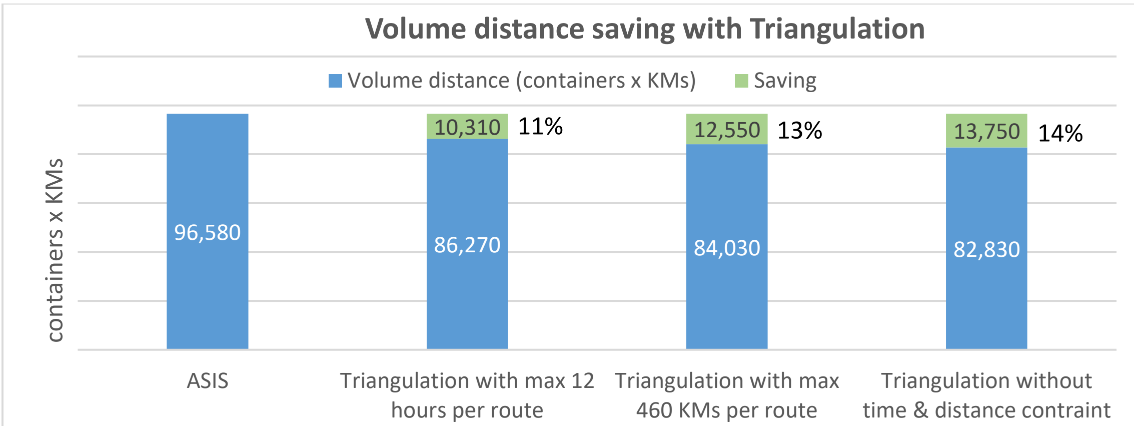
Cluster	Export	Import	Total
North	95	51	146
South	161	23	184
Center	248	45	293
Total	**504**	**119**	**623**

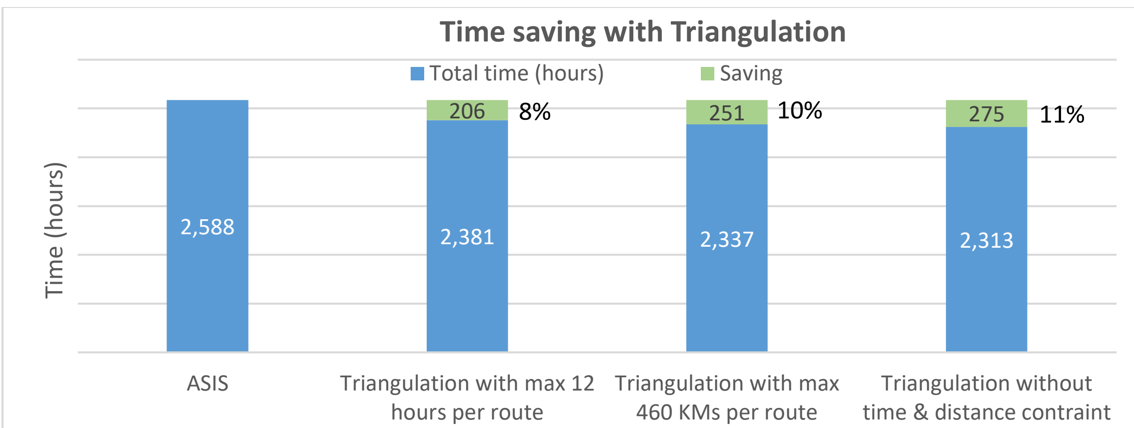
4.4 MILP Results

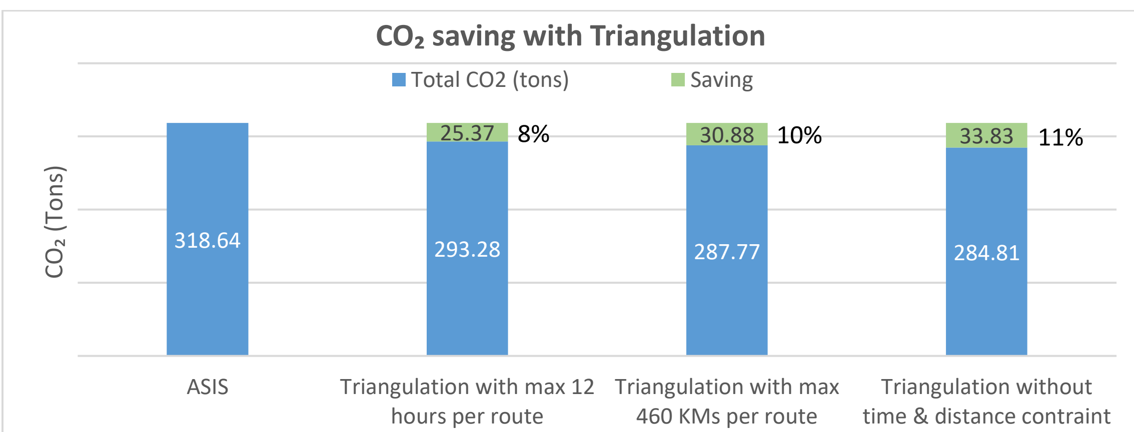
As illustrated in Figure 15 below, the MILP model is tested with a sample dataset from obtained from the clustering analysis – the center cluster consists of 248 export containers, 45 import containers. The total volume distance traveled is 96,580 containers × KMs in the current scenario (ASIS). Our model's optimization results reduce the container-distance traveled by 11–14%; reduce time by 8 –11%; and reduce CO_2 by 8–11%. With the optimized result based on container-distance, the CO2 emission is calculated by weight KM travelled multiplying the CO2 emission factor, which is published based on gram/ton.km. As the travel distance for empty container is reduced, the CO2 emission is decreased. When we add more constraints to the optimization model, such as maximum distance per route 460 KMs, or maximum service time and travel time per route 12 hours, the savings are reduced. If the constraints added are too restrictive, such as setting the maximum service time and travel time per route as 6 hours, it turns out with no feasible solution with this dataset sample.

Figure 15

MILP solving the transport flow optimization for triangulation with time & distance constraints



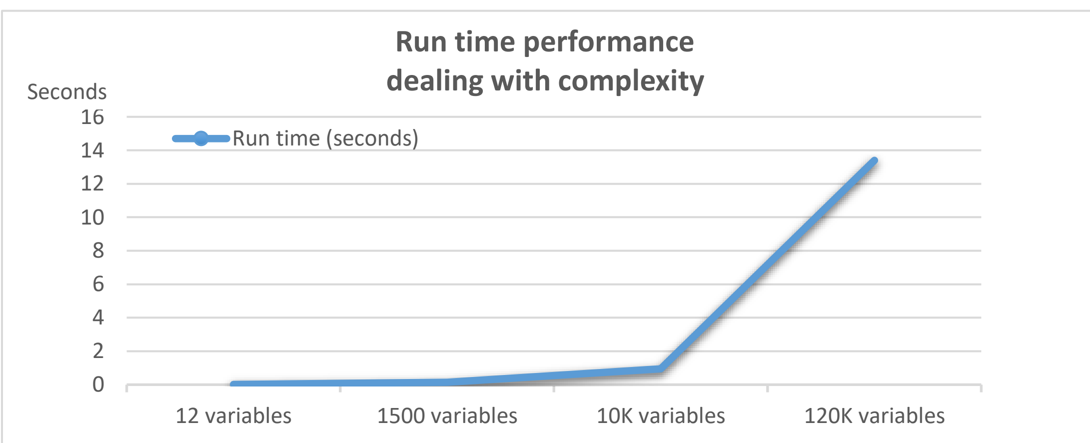




The performance of the MILP model is tested in Python using Gurobi optimizer with increased complexity by adding more variables. We tested four levels of complexity: 1) 12 x_{ij} variables with 3 importers and 4 exporters; 2) 1,500 x_{ij} variables with 30 importers and 50 exporters; 3) 10,000 x_{ij} variables with 100 importers and 100 exporters; 4) 120,000 x_{ij} variables with 300 importers and 400 exporters. The performance is measured by run time of the optimization model handling different levels of complexity. The result in Figure 16 shows that the model can solve the problem within two seconds for variables below 10,000. When the number of variables jumps to 120,000 the runtime increases to around 13 seconds. Based on the validation of the company's dataset and real-life experience, it is estimated that the complexity level for the company's triangulation problem is in the range of several hundred to several thousand variables. Hence our conclusion is that this model can be used in real-time optimization with the current data for the company.

Figure 16

MILP solving the transport flow optimization for triangulation with the increased complexity



5 DISCUSSION

In this chapter, we first discuss the result based on our research project and outline the challenges of container triangulation in China. Secondly, we share the insights developed from this project and initiate our recommendations to the company for review. Finally, we discussed the management implications for the industry and government.

5.1 Principal Findings and Challenges

Our analysis shows that when there is a good match of importers' containers with those for export, container triangulation can provide tangible benefit to the parties involved. The savings in the transportation trucking cost can reach up to 11% - 14%, which can be realized by exporters, importers, or trucking companies who pay for the cost of transportation. The time saving is around 8% - 11%, which comes from the reduced transport distance. The other significant portion of the time, which is spent on loading, unloading, and waiting at the importers' and exporters' premises, remains unchanged. Hence the time saving percentage is lower than the saving percentage of transportation cost. For the carriers, as the transportation time of the container is reduced, the container turnaround is faster than the current state, and asset utilization is improved. Exporters can get the empty containers faster.

With the container shortage crisis in the shipping industry, container triangulation becomes more attractive to those who need the containers urgently. The reduction in CO_2 emissions is 8% - 11%, which supports importers, exporters, and trucking companies' sustainable strategy. This can also be considered as a good result in light of the Chinese government's strategic plan in the logistics sector.

However, the more constraints there are in the planning and execution of container triangulation, the less saving can be realized. For example, each routing of the container triangulation should be completed within the same day to avoid additional truck overnight charges. Nevertheless, importers and exporters might have specific requests on time window which can complicate coordination.

Other challenges also impede the development and scale-up of container triangulation in China. We outline the practical challenges below:

1. Accurate location data for the point of delivery for import containers, and the point of cargo loading for export containers are not captured in the company's system. The current location data is at a city level; some of these locations are the office addresses of the importer, exporter, or middleman. Hence, it is not easy to match and plan the container triangulation perfectly on a larger scale.
2. Delivery date, loading date, and time window information are in different systems of the company. This information is not integrated, hence is missing in the data files we used for the analysis. We therefore used the actual discharge date and load date to estimate the date for the matches on a weekly basis.
3. The trucking market is highly fragmented in China. The importers and exporters have trucking contracts with different trucking companies, for that reason it is difficult to agree on which trucking company can be used to transport the container for both importer and exporter.
4. It is hard to predict the container status before un-stuffing the import cargo from the containers. Import containers are damaged and need to be repaired or cleaned after the

unloading of commodities, such as wood, metal, or waste paper. Based on the operational experience, it was estimated that around 50% of import containers loading those commodities are damaged. Due to the trade imbalance, a large number of damaged import containers reduces the containers available for reuse for export, complicating the possible triangulation of containers.

5. Currently, not many customers are using the company's digital platform, so the volume base for matching container triangulation is small at this stage.

5.2 Insights and Recommendations

To deal with the challenges mentioned above, we share the insights developed from this project and initiate our recommendations to the company for review.

There are several ways to capture accurate location data to improve the match ratio of the triangulation planning. First, current users can provide accurate location data and update it on the Maersk's digital platform. Second, the company can request the transport companies to send the container delivery location via mobile phone to the digital platform when they pick up a container from Maersk. Third, it is practical for the company to make a catalog with the location data of importers and exporters to keep them in the master data. Finally, IoT technology can be considered for application in containers in China, however ROI (return on investment) needs to be assessed.

In case of having sufficient and accurate location data, network optimization can be further explored to decide if a container depot needs to be added to the network. The container depot allows the exchange and triangulation of containers near the facilities of importers and exporters.

The clustering algorithm presented in Section 3.2.2 can be used to explore potential container depot locations.

In addition to the location data, the other three data elements must also be captured: the delivery date, the loading date, and the time windows of the containers for import and export must be captured in the digital platform. If the data is available in other systems, it must be transferred to the digital platform. By having the information integrated, the company can better group importers and exporters on a daily basis and pass the data to MILP to optimize routing for container triangulation. To increase the success rate of matching import containers for export use, the current customer service team can communicate and negotiate more with importers and exporters to adjust the date and time window, as a higher matching ratio creates a larger economic scale and savings benefit for the parties involved. The MILP module can be further enhanced by using a Multiple Traveling Salesman Problem (MTSP) with a time window for the expected delivery date (Zhang et al., 2010).

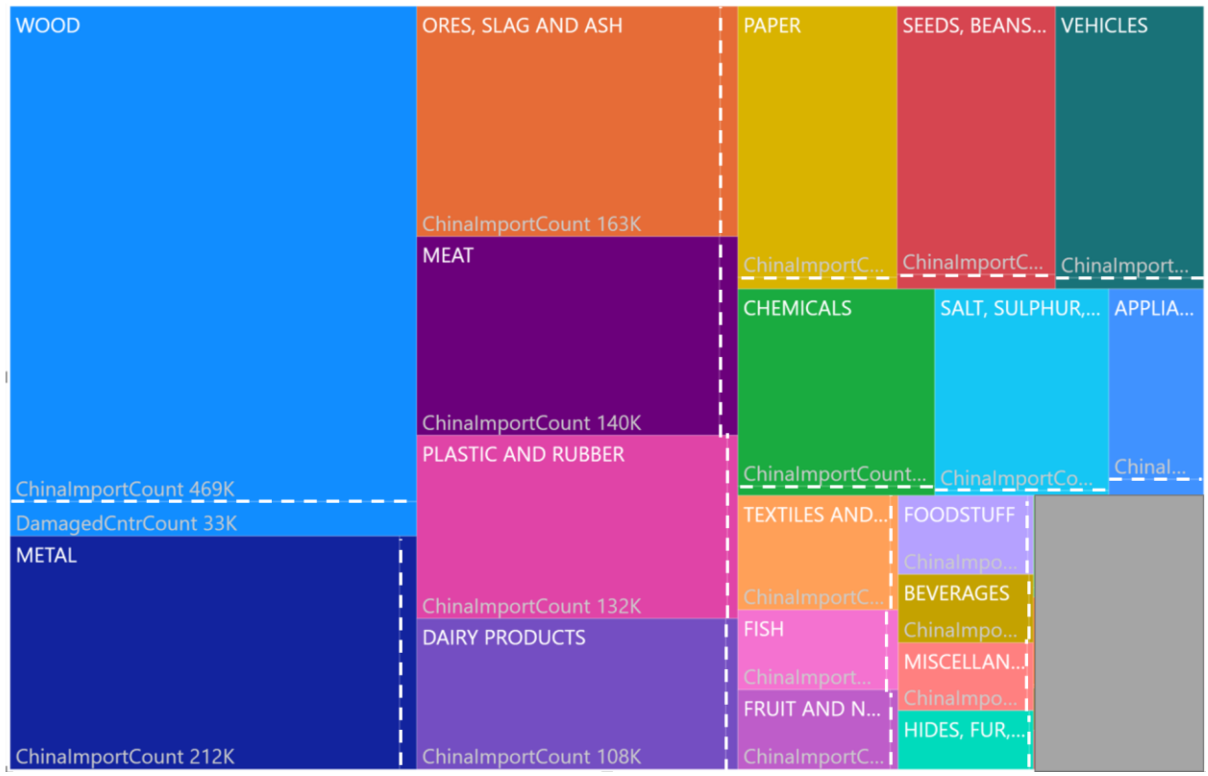
A market module can be added on the company's digital platform, allowing different trucking companies to exchange orders. Therefore, a trucking company can be assigned to run the container triangulation route for both the importer and the exporter. Network optimization with container depot can be explored as mentioned above so that container swapping can be done in the container yard with container handling equipment like top loaders. This approach is adopted by companies outside of China, such as MatchLog and Avantida, as presented in Table 1 in Section 2.3. The company can bundle its current offering of container warehousing and container triangulation products to customers.

Damaged containers can be predicted by applying machine learning models in the digital platform when damaged container data is sufficient to train the model. We explored machine learning models to predict whether the import container was damaged based on features such as stuffed commodity, cargo weight, location, and service. The initial finding was that there was an obvious gap in the information for damaged container reported in the company dataset.

As illustrated in Figure 17 below, we analyzed two years of import container data in China for the company. The split of the sound containers vs. damaged containers by stuffed commodity is visualized in a treemap. For example, the blue box on the top left shows total 469K import containers were stuffed with wood, and 33K containers of them were damaged, which tells us about 7% of import containers loaded with wood were damaged. Similarly, a dotted line inside the dark blue box for metal indicates the split of sound containers vs. damaged containers. When we asked the company to validate these findings, we observed that the results do not match the observations from practice - around 50% of import containers are damaged when the import cargo is wood or metal scrap. With current data, our machine learning model using KneighborsClassifier predicts with 70.65% accuracy. Our recommendation is to validate and further explore with the company whether damaged container information is fully captured and is sufficient for the training of the machine learning model, as this will have a big impact on the triangulation solution.

Figure 17

Split of damaged containers by import cargo



Additionally, the company has deployed another machine learning algorithm to inspect whether the container is damaged or not after cargo unloading. It is based on image recognition technology, and the drivers are required to take photos of the container and upload to the digital platform as outlined in Section 2.3.

5.3 Managerial Implications

The container trucking market is highly fragmented in China, with the trend of digitalization in the logistics industry. The industry can be more connected and integrated with the support of a digital platform. Real-time optimization models become feasible when aided with the increased computing power and improved algorithms. A digital platform can provide greater business visibility and measure the benefits in cost, time and sustainability when using container triangulation. However, to increase the benefit to the industry, different actors in the supply chain are required to accept the necessary changes and improve the level of collaboration and partnership. This is especially important today that the shortage of containers and shipping space have become relevant questions in the industry.

- As the company's digital platform is available for download as a mobile application, a mandatory requirement can be reinforced that the trucking companies need to report the container locations through the mobile application to the carrier.
- Importers and exporters should be open to interchange trucking companies on the digital platform to support container triangulation.
- Carriers can consider a strategic partnership to allow the exchange of containers across different carriers to support the scale-up of container triangulation.

Government and port authorities can develop a policy to promote container triangulation for the industry to drive the reduction of carbon emissions. For example, the port authorities can remove the restrictions that the containers should be imported and exported from the same port, with the increased flexibility the matching of import containers for export can be increased. Furthermore, real container delivery location should be reported in the port's system and the carrier's system.

6 CONCLUSION

This research project investigated the problem of container transportation matching import and export containers in the shipping industry with a focus on China. We conducted a rigorous review of the literature and concluded that this problem can be solved with mathematical models. As leading companies driving the digitalization in the logistics industry, more data is available, the fragmented transport industry becomes more connected.

With the sponsoring company, we built the methodology to conduct expert meetings, collect data, and complete the analysis with clustering and MILP algorithms. The results showed a promising benefit for the parties involved in transport when container triangulation can be planned and executed. Practical solutions were developed, and recommendations were proposed to the sponsoring company. Here we summarize our recommendation and future research required.

Container triangulation is a reality on the market today, as many companies, countries or research are evaluating this new alternative. Maersk, as a leader in the transportation industry, has always been at the forefront of technological solutions to improve its performance and service. As reviewed above, container triangulation leads to reductions in transportation costs, lead time, and CO_2 emissions. By offering this process within its digital platform, Maersk can attract and retain key partners due to the incentives it generates. In addition, this process can be automated as we observed it during the project, planning and executing more quickly.

As we discovered promising benefits, we also detected that there is still much to be done to fully exploit the potential of container triangulation. First, there is a significant need for more detailed and structured data regarding the location of container deliveries and loads for a more accurate model. Such data can be collected using the digital platform and the use of technologies such as GPS. This requirement is vital to measure the impact more precisely. We also recommend the implementation of the container triangulation model within the Maersk digital platform to integrate it into the services that the company presents today. Later, it would be important to publish the benefit of cost savings, time, and CO_2 reduction on the platform to attract more users. Maersk should consider establishing a dedicated team to proactively promote and scale up container triangulation planning. Finally, collaboration and coordination among stakeholders is essential, since their interaction is what makes the solution an attractive alternative. In the following section, we elaborate on the recommendations in further detail.

6.1 Recommendations

Clustering and MILP algorithms can be integrated into the digital platform to enable real-time optimization. As described in Section 3.2.2 and Section 4.1, it is practical to run grouping daily with visibility from 7 days onwards and optimize container triangulation routing for each group. Since the runtime performance for clustering and MILP is in the range of 1 to 20 seconds, real-time optimization is practical on the digital platform.

Here are our suggestions to integrate and scale-up the solution into the digital platform:

- The codes for geocoding, clustering, and MILP are documented in Python programming language and should be handed over to the IT team of the digital platform for integration.
- When the accurate location data is captured in the digital platform, the location of import and export containers can be displayed on the map. The internal users and external users

of the digital platform such as exporters, importers, and the trucking companies can see the container triangulation routing and savings benefit (Cost, time, and CO_2) on a daily basis. The route planning and optimization can be accessible to external users on pay-by-use basis.

- When needed, the company can enable the manual adjustment for the clustering, and the parameters of maximum distance and time per route to simulate different scenarios. Hence the results of saving on cost, time, CO_2 and routings can be easily compared. To optimize the triangulation based on the total cost, the distance matrix can be replaced with the cost matrix in the model. To calculate the time savings, the time matrix using distance divided by trucking speed can be used in the model. As for the savings on CO_2, the CO_2 emission factor based on gram/ton.km and a matrix of ton.km among importers, ports, and exporters can be used in the model. With the real-time optimization, the savings on the above metrics of volume distance, cost, time, and CO_2 can be refreshed in the digital platform very quickly. To use more accurate distance, an API service from Amap for advanced route planning can be further considered.
- With the current product offering of Demurrage and Detention (D&D) management from the company, it provides the tracking of free time storage and proactive overtime cost for the importers for their containers in the port. The D&D product offering can be integrated into the digital platform. Hence the customer service team can get a holistic view of target importers and the potential of container triangulation on daily basis.
- A dedicated team to promote and scale up the container triangulation solution could be useful for a successful implementation of the project. This team should use the digital

platform to select the right import and export customers based on the daily analysis. The cluster algorithm provides a good indication of which clusters to focus on. The time window for import container delivery and export container loading can be adjusted with potential customers, and profit-sharing incentives can be developed to increase the matching ratio.

Savings in time and CO_2 can be reported in the digital platform to add transparency in the system. Customers know the benefits of the container triangulation, attracting new ones to get involved in the project. The same can be published on the company website to demonstrate that the company is acting with committed customers to improve the supply chain efficiency and sustainability in China, which is in line with the strategic plan of Chinese government. The reporting of savings in cost and volume distance can be just shared individually with concerned parties who keep track of their progress on a monthly or yearly basis.

6.2 Future research

Undoubtedly, globalization has brought great challenges in the way and speed of delivering products or services. For future studies, we recommend taking the concepts covered in this research to a global optimization perspective, considering different transport companies or partners. As more participants are involved, more data and more markets are available in the platform, becoming more attractive for new users and creating a feedback loop. This may generate a growth in savings for the participants within the platform.

Applying machine learning to predict container damage to improve container triangulation planning is also a technology that we consider critical for the project continuity. Machine learning helps reduce inspection decision time, enabling digital capabilities such as advanced analytics to develop. This solution is considered very relevant as it could be developed and implemented in a short period of time and produce the positive effect on the company's performance.

Technology has led companies to think that there is no limit to creating something new or solving any problem. Therefore, industries such as transportation must seek to develop and adapt new technological solutions to reduce inefficiencies. Many countries or companies are incorporating technologies like the ones we mentioned above, where the impact is even more positive than what they expected. The larger the scale of container triangulation, the more benefit can be generated for industry, stakeholders and the environment. We hope to see further improvement in the industry with the contribution of more and more researchers.

REFERENCES

Alcoba, R. D. (2017). *Predicting On-time Delivery in the Trucking Industry*. 51.

An executive's guide to AI | McKinsey. (2020). https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai

Artificial intelligence. (2020, August 11). Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence

Avantida—Home. (n.d.). Retrieved November 1, 2020, from https://www.avantida.com/

Berg, R. V. den, & Langen, P. W. D. (2017). Environmental sustainability in container transport: The attitudes of shippers and forwarders. *International Journal of Logistics Research and Applications*, *20*(2), 146–162. https://doi.org/10.1080/13675567.2016.1164838

Bhavsar, P., Safro, I., Bouaynaya, N., Polikar, R., & Dera, D. (2017). Machine Learning in Transportation Data Analytics. In *Data Analytics for Intelligent Transportation Systems* (pp. 283–307). https://doi.org/10.1016/B978-0-12-809715-1.00012-2

Braekers, K., Janssens, G. K., & Caris, A. (2011). Challenges in Managing Empty Container Movements at Multiple Planning Levels. *Transport Reviews*, *31*(6), 681–708. https://doi.org/10.1080/01441647.2011.584979

Castillo, V. E., Bell, J. E., Rose, W. J., & Rodrigues, A. M. (2018). Crowdsourcing Last Mile Delivery: Strategic Implications and Future Research Directions. *Journal of Business Logistics*, *39*(1), 7–25. https://doi.org/10.1111/jbl.12173

China pledges to be carbon neutral by 2060. (n.d.). Retrieved October 10, 2020, from http://country.eiu.com.libproxy.mit.edu/article.aspx?articleid=1350194518

Chui, M. (2018, April 17). *Sizing the potential value of AI and advanced analytics | McKinsey*. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning#

Cichosz, M., Wallenburg, C. M., & Knemeyer, A. M. (2018). Digital transformation at logistics service providers: Barriers, success factors and leading practices. *The International Journal of Logistics Management*, *31*(2), 209–238. https://doi.org/10.1108/IJLM-08-2019-0229

Deidda, L., Di Francesco, M., Olivo, A., & Zuddas, P. (2008). Implementing the street-turn strategy by an optimization model. *Maritime Policy & Management*, *35*(5), 503–516. https://doi.org/10.1080/03088830802352145

Duzbaievna Sharapiyeva, M., Antoni, A., & Yessenzhigitova, R. (2019). The Impact of Port Transport-logistics Infrastructure and LPI for Economic Growth: On the Example of Landlocked Countries. *Pomorstvo*, *33*(1), 63–75. https://doi.org/10.31217/p.33.1.7

Efimova, O. V., Baboshin, E. B., Igolnikov, B. V., & Dmitrieva, E. I. (2020). Promising Digital Solutions for the Efficient Technological and Managerial Processes on Transport. *2020 International Conference Quality Management, Transport and Information Security, Information Technologies (IT QM IS)*, 92–95. https://doi.org/10.1109/ITQMIS51053.2020.9322861

Fitzgerald, M., Kruschwitz, N., Bonnet, D., & Welch, M. (2014). *Embracing Digital Technology*. 16.

Furió, S., Andrés, C., Adenso-Díaz, B., & Lozano, S. (2013). Optimization of empty container movements using street-turn: Application to Valencia hinterland. *Computers & Industrial Engineering*, *66*(4), 909–917. https://doi.org/10.1016/j.cie.2013.09.003

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, *30*(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jianliang, M., Haikun, S., & Ling, B. (2009). The Application on Intrusion Detection Based on K-means Cluster Algorithm. *2009 International Forum on Information Technology and Applications*, *1*, 150–152. https://doi.org/10.1109/IFITA.2009.34

Kuzmicz, K. A., & Pesch, E. (2019). Approaches to empty container repositioning problems in the context of Eurasian intermodal transportation. *Omega*, *85*, 194–213. https://doi.org/10.1016/j.omega.2018.06.004

Liu, B., Su, X., Shi, J., & Hou, R. (2020). Does urbanization drive economic growth decoupled from energy consumption in China's logistics? *Journal of Cleaner Production*, *257*, 120468. https://doi.org/10.1016/j.jclepro.2020.120468

Liu et al. - 2020—Does urbanization drive economic growth decoupled .pdf. (n.d.).

MADE IN CHINA 2025. (n.d.). Retrieved October 10, 2020, from https://kritisches-netzwerk.de/sites/default/files/merics_-_made_in_china_2025_-_the_making_of_a_high-tech_superpower_and_consequences_for_industrial_countries_-_76_seiten_1.pdf

Maersk annual report 2019.pdf. (n.d.). Retrieved October 10, 2020, from https://investor.maersk.com/static-files/984a2b93-0035-40d3-9cae-77161c9a36e0

MatchBox Exchange | Better—Faster—Smarter. (n.d.). Retrieved November 1, 2020, from https://www.matchboxexchange.com/

MATCHBOX Exchange—Instant container reuse and exchange online platform. (n.d.). Retrieved November 1, 2020, from https://www.maersk.com/news/articles/2020/01/21/matchbox-exchange-instant-container-reuse-exchange-online-platform

MatchLog Solutions. (n.d.). Retrieved November 1, 2020, from http://www.matchlog.delivery/

Matt, C., Hess, T., & Benlian, A. (2015). Digital Transformation Strategies. *Business & Information Systems Engineering*, *57*(5), 339–343. https://doi.org/10.1007/s12599-015-0401-5

Patel, K. M. A., & Thakral, P. (2016). The best clustering algorithms in data mining. *2016 International Conference on Communication and Signal Processing (ICCSP)*, 2042–2046. https://doi.org/10.1109/ICCSP.2016.7754534

PR Newswire. (2019, October 16). *Continental releases white paper on intelligent logistics helping to reduce costs and improve efficiency*. https://www-emis-com.libproxy.mit.edu/php/search/doc?pc=CN&doc_dtyp=1&lang=en&dcid=662770928&range=365&indu_local_class=naicsbyid&similar=hide&show_translated_articles=1&coverage=all&title=0&rpp=25&sort_by=relevance&query_entry=quick&abstract=1&display=1&search_selected_tab=all&search_log_id=34d6f990-0d73-4cef-b9f6-

6830a4ec6ba5&get_translated_articles=1&profile_type=0&keyword=Share+of+national+logistics+costs+in+the+GDP+of+China&numresult=8

Sáinz Bernat, N., Schulte, F., Voß, S., & Böse, J. (2016, December 22). *Empty Container Management at Ports Considering Pollution, Repair Options, and Street-Turns* [Research Article]. Mathematical Problems in Engineering; Hindawi. https://doi.org/10.1155/2016/3847163

Seabury Consulting. (2020). https://seaburycargo.com/login

Statista. (2020). *China: National logistics costs as a percentage of GDP 2017*. Statista. https://www.statista.com/statistics/949249/china-national-logistics-costs-as-a-percentage-of-gdp/

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Union of Concerned Scientists. (2008, June 11). *Each Country's Share of CO_2 Emissions | Union of Concerned Scientists*. https://www.ucsusa.org/resources/each-countrys-share-co2-emissions

Welcome to GeoPy's documentation! —GeoPy 2.0.0 documentation. (n.d.). Retrieved November 28, 2020, from https://geopy.readthedocs.io/en/latest/#geopy-is-not-a-service

World Economic Forum White Paper Digital Transformation of Industries. (n.d.). Retrieved November 1, 2020, from http://reports.weforum.org/digital-transformation/wpcontent/blogs.dir/94/mp/files/pages/files/wef-dti-logisticswhitepaper-final-january-2016.pdf

Zhang, R., Yun, W. Y., & Kopfer, H. (2010). Heuristic-based truck scheduling for inland container transportation. *OR Spectrum*, *32*(3), 787–808. https://doi.org/10.1007/s00291-010-0193-4

Zhenfei, L. (n.d.). *Amap-Fundamental-Infrastructure-of-Mobile-Internet*. 32.

Zhu, Y., Matsuyama, Y., Ohashi, Y., & Setoguchi, S. (2015). When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics*, *56*, 80–86. https://doi.org/10.1016/j.jbi.2015.05.012

APPENDIX

Comparison of two geocoding methods: Nominatim vs. Google

	A	B	C	D	E	F	G
1		Nominatim		Google		Difference	
2	MatchCity_geo	Latitude	Longitude	GoogleLat	GoogleLong	Lat	Lon
6	BEIJIAO,China	34.8262356	114.3571132	22.928305	113.207187	152%	101%
7	BEIJING,China	40.190632	116.412144	39.9041999	116.4073963	101%	100%
25	FUZHOU,China	27.9511188	116.3530413	26.074478	119.296482	107%	98%
36	HUANGPU,China	31.2182657	121.4802514	23.106401	113.459749	135%	107%
45	JINZHOU,China	41.494444	121.615278	41.095685	121.1268459	101%	100%
48	JIUJIANG,China	29.38446135	115.680492	29.705077	116.00193	99%	100%
54	LIAN HUA SHAN,China	22.2665052	113.9710013	22.97879	113.504534	97%	100%
70	PUDONG DOWNTOWN,China	34.2644814	117.1760181	31.221517	121.544379	110%	96%
81	SAN SHAN,China	29.1074829	121.4074082	31.219568	118.268101	93%	103%
89	SHEKOU,China	27.211944	119.571389	22.483182	113.916509	121%	105%
91	SHIDAO,China	34.3652	112.879	36.8787218	122.4373306	93%	92%
92	SHUIDONG,China	26.8512901	107.6782749	35.772543	118.795956	75%	91%
97	TAIZHOU,China	32.4585613	119.9181228	28.65638	121.42076	113%	99%
111	XINGANG,China	32.51035395	111.5210891	39.0066626	117.6846985	83%	95%
120	YIBIN,China	28.58689	104.718661	28.751836	104.643159	99%	100%

The color code: from green to red show the order of small to large measured by the percentage to Google's geocoding value of latitude