ON SUSTAINABLE SUPPLY CHAINS: OPTIMAL DESIGN OF A MULTIMODAL LOGISTICS NETWORK WITH SHIPMENT CONSOLIDATION, STOCHASTIC DEMAND, AND MACHINE LEARNING

by

Ibrahim Opeyemi Oguntola

Submitted in partial fulfilment of the requirements

for the degree of Master of Applied Science

at

Dalhousie University

Halifax, Nova Scotia

July, 2022

Table of Contents

List of Tables

List of Figures

Abstract

In this thesis, we consider the logistics network of a multi-echelon multimodal supply chain with multiple products and components taking economic and environmental sustainability, and shipment consolidation into consideration. Procedures for calculating and using both the water and carbon footprints of the network as metrics for its environmental sustainability are also explored. The supply chain logistics network is modelled as a Mixed Integer Linear Program (MILP) and then tested on randomly generated but realistic test instances. The effects of shipment consolidation on the economic and environmental cost of operations are analysed with results showing that consolidation decreases the supply chain (SC) cost especially when the distance between the shipper and receiver is significant.

Considering that in reality, some of the parameters of supply chain network models might be stochastic, experiments are carried out with the designed MILP model having its demand parameter as stochastic. With the continual digitalization of supply chain processes leading to the automatic generation of data, machine learning (ML) has evolved as a methodology with the potential to help optimize stochastic models with its increasingly accurate predictions of future occurrences due to the continuous innovation of new algorithms. ML approaches to predicting stochastic parameters using historical data are evaluated in comparison to the more traditional stochastic programming approaches over multiple prediction periods. The three ML models utilized, Attention CNN-LSTM (AC-LSTM), Attention ConvLSTM (ACV-LSTM) and an ensemble of both models using Support Vector Regression (Ensemble-SVR), performed significantly better than the stochastic programming approaches considered (Simple recourse programming and Chance-constrained programming) in all scenarios. The MILP models using the predictions from the ML algorithms obtained the highest value of stochastic solution (VSS) and had the lowest expected value of perfect information (EVPI). This makes a case for the continued integration of ML prediction methodologies into stochastic optimization modelling.

List of Abbreviations Used

3PL	3rd Party Logistics
AC-LSTM	Attention CNN-LSTM
ACV-LSTM	Attention ConvLSTM
ANN	Artificial Neural Network
AVCU	Average Vehicle Capacity Usage
BoM	Bill of Material
CF	Carbon Footprint
CNN	Convolutional Neural Network
ConvLSTM	Convolutional LSTM
CVaR	Conditional Variance at Risk
DL	Deep Learning
DM	Decision Maker
EVPI	Expected Value of Perfect Information
FTL	Full TruckLoad
LND	Logistics Network Design
LSTM	Long Short-Term Memory
LTL	Less than TruckLoad
MAE	Mean Absolute Error
MFT	Multimodal Freight Transportation
MILP	Mixed Integer Linear Programming
ML	Machine Learning
MOU	Mean Overstocking and Understocking cost
MSE	Mean Squared Error
NN	Neural Network
NVP	NewsVendor Problem
QBL	Quadruple Bottom-Line
RNN	Recurrent Neural Network
SC	Supply Chain

SCL	Shipment Consolidation
SCM	Supply Chain Management
SCN	Supply Chain Network
SCND	Supply Chain Network Design
SSCN/MFT&SCL	Sustainable Supply Chain Network with Multimodal Transportation and Shipment Consolidation
SSSCN/MFT&SCL	Stochastic Sustainable Supply Chain Network with Multimodal Transportation and Shipment Consolidation
SP	Stochastic Programming
SVR	Support Vector Regression
VSS	Value of Stochastic Solution
WF	Water Footprint

Acknowledgements

Al-Hamdu Lillahi Rabbil-'Alamin

I would like to express my gratitude to my supervisor, Prof. Dr. M. Ali Ülkü, for his continuous support throughout my Master’s degree. I am grateful for his motivation, enthusiasm and immense knowledge. I also want to say a big thank you to Assoc. Prof. Dr. Ahmed Saif and Assoc. Prof. Dr. Alexander Engau for being members of my supervisory committee and for offering their time and knowledge in evaluating this thesis.

I am most grateful to my parents for financially and emotionally supporting me throughout my Master’s program. I am incredibly thankful for their love, prayers and sacrifices.

Chapter 1

Introduction

1.1 Problem Definition and Motivation

A supply chain (SC) is a complex network of facilities and organizations that is usually over a wide geographical area with interrelated activities involving the flow of information, products and funds to produce and deliver a product or service to end users (Musmanno et al., 2004). There is a complex web of interactions between these players (facilities and organizations) that stems from the beginning stage of sourcing raw materials to the final stage of delivering the service or product to the end-users. One of the most crucial problems in Supply Chain Management (SCM) is the network design. Inefficiencies in this stage would lead the SC to operate at a higher cost (Kabadurmus and Erdogan, 2020). The goal is therefore to efficiently design the network such that it runs at the lowest cost (economic or as determined by the decision-maker) possible while fulfilling all the criteria or constraints present appropriately.

In the design of the logistical networks of SCs, either local or global, one of the main factors that make the modelling complex is the presence of stochastic parameters. The parameters are uncertain and could affect every other decision variable present. For example, if the demand parameter were stochastic, it becomes difficult to make optimal decisions such as the quantity of raw materials to order, trucks to use, suppliers to contact, and so on. It is thereby necessary for the company/decision-maker to have a method of ensuring efficient logistical planning when stochastic parameters are present. The Logistics Network Design (LND) decisions have to be viable and resilient enough to function well in uncertain environments over long periods (Govindan et al., 2017).

Most companies have to predict what the demand would be in the future to source the raw materials required and make other decisions within the network. This becomes even more critical when the goods involved are perishable and can not be rolled over to the next cycle. However, an over-projection could come with a cost incurred due to storage or damages (especially for perishable goods) while an under-projection leads to a loss in potential sales and possibly, the loss of customer goodwill. It is therefore important to have a methodology for forecasting the stochastic demand with high accuracy or at least to reduce the regret of the decision-maker.

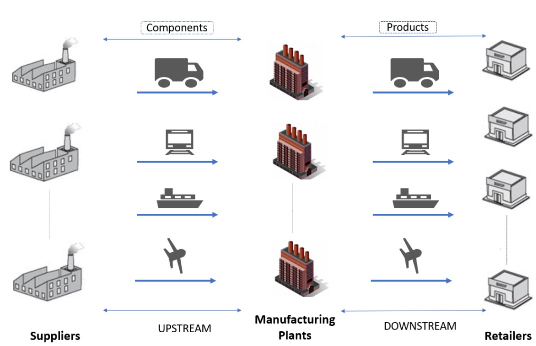
Traditionally, Stochastic Programming (SP) is used for optimization models containing stochastic parameters. The distribution or past scenarios of the parameters are utilized to build the model and the best decision which is expected to lead to the least regret or optimize the goal (cost or profit) is then obtained. However, in today's competitive environment, there is more motivation to have better accuracy in forecasting demand as false predictions could lead to the bullwhip effect (Carbonneau et al., 2008). This refers to the SC phenomenon describing how small fluctuations in demand at the retail level can cause progressively larger fluctuations in demand upstream of the network. With advancements in information technology and the increased availability of a huge volume of data generated in various parts of SCM (Tirkolaee et al., 2021), methods such as Machine Learning (ML) could play a bigger role in ensuring better predictability of demand and other stochastic parameters.

In the design of SC networks, many previous studies mostly considered apparent costs such as order, raw material, transportation and production costs (Kabadurmus and Erdogan, 2020), neglecting other factors that could make the network more robust and resilient while directly or indirectly impacting the cost of the company. Some of these factors include sustainability, multimodal transportation, and shipment consolidation.

One of the major factors under which the logistical operations of an SC network should be analyzed is its impact on the environment it interacts with. A competitive rush to the industrialization has resulted in the rapid depletion of non-renewable resources. The recognition of these impacts has led governments on different levels to sign new treaties [such as the Kyoto Protocol (UNFCCC, 2008), and Paris Agreement (UNFCCC, 2021)] aimed at pressurizing companies into reducing their industrial waste, carbon/water footprints, and so on, thereby leading to global awareness of the demand for sustainability (Ülkü & Engau, 2021). Some of the most important governmental regulations that have been directed at battling the carbon emission rate of companies include the concepts of the carbon tax and carbon cap-and-trade (Kabadurmus and Erdogan, 2020). There have also been recommendations in the literature on how water conservation policies could be developed to monitor and manage water usage levels by factories, especially in drought-prone localities (Farzaneh et al., 2021). This is necessary because a large portion of the world is at risk of water scarcity. With climate change worsening the situation, and the likely increase in human water needs as the years go by, there is a need to proactively set principles in place to manage the situation (Wheeler and Von Braun, 2013; Schewe et al., 2014). Ideally, the impact of SCs on water usage and pollution should be analyzed and considered in an integrated manner in the network design stage. In this thesis, environmental sustainability is considered through the inclusion of water and carbon footprint analysis in the design of the logistics network of SCs.

Multimodal Freight Transportation (MFT) is the utilization of different modes (air, water, rail, road) in the shipment of cargo from one location to another. The presence of MFT widens the choice of the decision-maker (DM) by availing opportunities to use the most appropriate and efficient mode for every scenario encountered. For example, air mode, despite being the most expensive, might be the best option if the delivery deadline is close; while water mode, despite typically being the slowest, might be the best option cost-wise if the delivery deadline is reasonably far. Different modes also have different environmental impacts allowing the DM to set up his transportation options following company policy or governmental regulations. Multimodal transportation could reduce logistical operational costs and carbon emissions (Kabadurmus and Erdogan, 2020). Figure 1 shows a basic three-echelon supply chain network with 4 modes of transportation.

Figure 1. Three-Echelon SC with Multimodal Freight Transportation



Shipment Consolidation is a logistics strategy whereby multiple shipments are merged so that they could be transported on the same vehicle to the same market region potentially saving costs through economies of scale (Ülkü, 2009). It could also provide the SC with an opportunity to achieve its goal of reducing emissions and becoming more sustainable by using fewer vehicles or covering shorter distances. (Ülkü, 2012). Different configurations of SCL are considered in this thesis including shipper-performed and carrier-performed consolidation. Also, an experiment is carried out on the impact of SCL on the network and when it is most profitable.

As competitiveness within industries grows and the complexity of networks increases, it becomes more necessary to factor in strategies such as multimodal transportation and shipment consolidation into the logistical network of any SC to ensure operational excellence and improved network resilience. Also, as more companies become aware of their duties to the environment either due to the cost incurred from government regulations or the recognition of their corporate social responsibility, it becomes more important to factor sustainability directly into the early stages of the SC network design rather than having to accommodate it as a forethought potentially leading to non-optimal decisions. Accurate forecasting of expected demand also plays a role in ensuring that the necessary plans are put into motion to reduce the extra cost incurred due to overstocking or unmet demand.

The problem being tackled in this thesis can therefore be defined as the design of a SCN with the goal of optimizing its economic and environmental (carbon and water footprint) sustainability while simultaneously factoring in MFT and SCL with a focus on the tactical level of SCM decisions (order quantity, suppliers to engage and so on). Afterwards, a case is made for the utilization of ML approaches to stochastic optimization modelling of logistics networks by evaluating its performance in comparison to traditional SP approaches.

1.2 Scope

Considering a holistic approach, in this study, a realistic multi-commodity three-echelon supply chain optimization model is generated which considers factors such as environmental sustainability (water and carbon footprint), multimodal freight transportation (MFT) and SCL. The demand is also considered uncertain. Actual demand data is obtained, which is then used as a baseline to compare several SP methods and ML techniques of modelling with stochastic parameters and their performances over predictions of varying periods into the future. The problem is modelled as a Mixed Integer Linear Programming (MILP) model and solved using a Gurobi optimization solver. Necessary parameters are either obtained from standard resources or realistically generated for each test instance. The goals of this study can be described as the optimization of the logistical network of a stochastic sustainable supply chain with multimodal freight transportation and shipment consolidation (SSSCN/MFT&SCL).

1.3 Contributions

A summary of the comparison between this thesis and other relevant works in literature is provided in Table 1. The main contributions of this research work to literature are outlined as follows:

- Development of a modelling framework for the design and/or operation of a supply chain network that simultaneously considers environmental sustainability factors such as water usage, water pollution and carbon emission, alongside multimodal transportation and shipment consolidation.
- Analysis of the effect of SCL on the economic and environmental sustainability of an SC alongside the factors that enhance its advantageousness.
- Comparative analysis of the efficiency of ML approaches and SP approaches to working with stochastic parameters in designing optimization models.

Table 1: An Overview of the Literature on the Research Work on SCND.

Paper	Economic Sustainability	Environmental sustainability		Stochastic Demand		Multimodal Transportation	Shipment Consolidation
		Carbon Footprint	Water Footprint	Stochastic Programming	Machine Learning		
Le and Lee (2013), Zhang et al. (2017), Liotta et al. (2015), Moradinasab et al. (2018), Kabadurmus and Erdogan, (2020)	✓	✓				✓	
Alizadeh et al. (2019)	✓	✓		✓		✓	
De Boer et al. (2013), Gu et al. (2015)			✓				
Xie et al. (2014)	✓					✓	
Mostert et al. (2018)	✓	✓				✓	
Fan et al. (2019), Muñoz-Villamizar et al. (2022)	✓	✓					✓
Chen et al. (2017)	✓						✓
Ma and Liu (2017)	✓			✓			
Ahmadi and Amin (2019)	✓			✓			✓
Moheb-Alizadeh and Handfield (2018), Golpîra et al. (2017), Sajedi et al. (2020)	✓	✓		✓			
Choudhary et al. (2015), Shaw et al. (2016), Saif and Elhedhli (2016), Kumar et al. (2022), Yu and Hou (2021), Chaabane et al. (2012), Mogale et al. (2022)	✓	✓					
Park et al. (2017), Farazmand et al. (2022), Beresford et al. (2011), Wang et al. (2018)	✓					✓	
Serrano et al. (2017), Glock and Kim (2014), Çapar (2013), Kang et al. (2017), Muriel et al. (2022)	✓						✓
Ridoutt et al. (2010)			✓				
This thesis	✓	✓	✓	✓	✓	✓	✓

1.4 Thesis Organization

The remainder of the thesis is structured as follows: Chapter 2 focuses on related works that have been done in literature exposing the gaps this research wishes to cover. Some background knowledge of each of the general concepts explored is provided in Chapter 3, including logistical network design for a supply chain, sustainability and multimodality, stochastic programming and machine learning. The problem under consideration is described in Chapter 4, including the network model formulation with its objective function and constraints. Chapter 5 explains the procedures utilized for analysing and evaluating the various approaches to handling an SCN with stochastic parameters. The numerical experiments to be performed are then delineated. Chapter 6 summarizes the results and discusses the major findings obtained from this research work. Lastly, Chapter 7 provides a conclusion and suggests potential areas for improvement in future researches.

Chapter 2

Literature Review

2.1 Sustainability in Supply Chains

In the past two decades, sustainability in SCs has become a major concern (Kabadurmus and Erdogan, 2020). Alignment of the SC goals of an organization with the sustainability goals is very important to proactively respond to the possible impact of its operations on other areas such as the environment and the society at large. The "Triple-Bottom-Line" approach (TBL) to sustainability was proposed by Elkington (1994) with the consideration of profit and loss, as well as environmental and social values while analysing the general performance of any organization. The profit and loss performance determine the economic sustainability of an organization. Factors such as the increase in money flow throughput and continual growth despite competition are major goals of any SC. Environmental values are concerned with reducing pollution in the environment by emphasizing biodegradable products and reusing them to avoid further depletion of natural resources. Social values encourage the management of business to ensure increased positive and reduced negative impacts on people. In addition to the TBL approach, there has been a push to consider 'culture' as the fourth pillar of sustainability thereby forming a 'Quadruple Bottom-Line' (QBL) approach to analyzing sustainability (Ülkü & Engau, 2021; Tiller et al., 2022). This is because the consideration of the cultural impact of the design or operations of SCNs enforces policymakers to involve the oppressed local communities in their decision-making. This would further make companies align more with the United Nations Sustainable Development Goals (United Nations, 2021).

Environmental sustainability is a widely considered factor in the majority of studies on sustainable SCs. Water and carbon footprint sustainability are two of the common metrics in the environmental sustainability of an SC. Researchers have shown that carbon emissions from various stages of the supply chain have a significant impact on the environment. The majority of the research works analysed considered carbon emissions as a cost while modelling the SCs. Le and Lee (2013) investigated the economic dimension as cost and the environmental dimension as CO_2 emission of the SC. They considered the CO_2 emissions to only happen during transportation throughout the SC. Choudhary et al. (2015) demonstrated the potential of cutting down on the carbon footprint (CF) of the supply chains through reverse logistics without increasing the total cost. Shaw et al. (2016) investigated carbon emissions and carbon trading issues with the formulation of a chance-constrained model. Results showed that the plants activated as well as material flow throughout the SCN varied with the carbon credit price.

A rational trade-off between environmental issues and total cost has been studied by Zhang et al. (2017). A MILP was formulated and applied to an actual case study of an electric meter company to see the impact. The results showed that the introduction of CF as a parameter within the optimization model had a substantial impact on the SCN, resulting in a decrease in CO_2 emissions per unit shipment. A robust three-stage SP model is proposed for an olefin with the consideration of supply and uncertainty in carbon emission tax rates (Alizadeh et al., 2019). They showed that increasing the carbon tax rate would decrease emissions but increase the total network costs. Kabadurmus and Erdogan (2020) designed a three-echelon sustainable supply chain network considering carbon exploring the effects of cap-and-trade policy on supply chain cost and emission. However, none of these papers considers the analysis of the water footprint (WF) of the supply chain as an environmental sustainability factor.

WF analysis is another important element in sustainability. Most industrial production systems use fresh water as an essential element. The utilization of global freshwater for the agricultural and industrial sectors is 70% and 22%, respectively (United Nations Educational, Scientific and Cultural Organization, 2009). Many of the research works considering WF that were analyzed however focused more on the agricultural sector. This could still be considered as part of the supply chain as raw materials, especially for food supply chains, are sourced from farms. Analysis of WF is critical in all sectors of the SCN for environmental sustainability. Ridoutt et al. (2010) investigated the fresh mango SC network. It was found that the WF during the agricultural stages is 2298 litres per kg, whereas it increased to 5218 litres per kg of mango during the transportation, retailing, and consumption stages. De Boer et al. (2013) studied the impact of freshwater consumption on the environment while considering milk production in the global SC in the Netherlands, including resource depletion, using a Life Cycle Analysis (LCA). They concluded that the amount of water required to produce 1 kg of fat and protein-corrected milk is 66 litres. Gu et al. (2015) inquired about both direct and indirect water consumption in an iron and steel factory in China. Based on the results obtained, the freshwater consumption and water pollution per ton of steel were 5.5m^3 and 146m^3, respectively.

Alternatively, transportation modes, especially shipping serve as a potential source of water pollution through occurrences such as hydrocarbon leakage, sewage discharge and garbage discharge (Shi et al., 2018; Bedaiwi et al., 2019). All of the above indicate that in analyzing the WF of a complete SC network, both the impact of facility (such as farms and factories) operations and transportation on water usage and pollution have to be considered. In this thesis, we model an SCN that simultaneously considers environmental sustainability based on both carbon and water footprint.

2.2 Multimodal Transportation in Supply Chains

Multimodal transportation is another important element in the SCN with the potential of reducing the overall cost of SCs. Multimodal transportation refers to when a supply chain network considers two or more modes of transportation. Liotta et al. (2015) considered a three-echelon supply chain network with the road, rail, and water modes of transportation. It was shown that more multimodal distribution options can help meet environmental and cost-saving goals by leveraging the different costs and emission rates of different modes. A bioethanol SC with multimodal transportation was investigated by Park et al. (2017) considering a combination of truck and rail. Using more than one mode of transportation was more cost-effective than using just one mode of transportation and led to a lower cost for the SC. A multimodal cellulosic biofuel SC was addressed by Xie et al. (2014) to minimize total cost. It was found that railing was more convenient for long-distance transportation while trucking was more effective for short-distance transportation. Beresford et al. (2011) investigated the best multimodal transportation combination for transporting iron ore shipments from Australia to

Northeast China identifying the rail-sea-rail transportation method as the most efficient mode for bulk cargo.

Moradinasab et al. (2018) modelled a petroleum SC while considering rail, road, and pipeline transportation to maximize the total profit and minimize environmental pollution. They formulated the problem as a MILP, and the results showed that profits increased by 11.12% due to the reduced transportation cost throughout the SC. An SC with long-distance freight delivery via multimodal rail-road transportation was worked on by Mostert et al. (2018). They considered roads, intermodal rail, and intermodal inland waterways to find the most economic mode of transportation. Results showed that the combination of two modes of transportation was a cost-effective method, especially when the distance to be covered was more than 300 kilometres. Kabadurmus and Erdogan (2020) considered 5 different modes of transportation (including an environment-friendly truck) and found that the cost and emission level were reduced when compared to unimodal transportation considering only the road mode. Other studies that considered the combination of multiple transportation modes include Zhang et al. (2017), Wang et al. (2018), and Farazmand et al. (2022). Likewise, multiple modes of transportation (air, water, rail and road) are considered simultaneously in this thesis.

2.3 Shipment Consolidation in Supply Chains

Shipment Consolidation (SCL) also plays a huge role in potentially reducing cost and greenhouse gas emissions of SCNs (Ülkü, 2012). Serrano et al. (2017) proposed a global network of cross-docking platforms to connect long-distant assembly factories with first-tier suppliers. The goal is to minimize the overall cost, which includes the cost of inbound and outbound transportation. Fan et al. (2019) examined the flow consolidation of perishable and dry goods. The findings suggest that shipping distance and cargo type affect the performance of flow consolidation in logistics. Çapar (2013) investigated a two-stage combined shipment consolidation with an inventory decision in the distribution system. An exact optimization approach was proposed to handle the issue of optimum replenishment amount at the distribution centre, the order-up-to level at retailers, and a shipment consolidation cycle to measure overall performance. The results indicated that in a shipping consolidation setting, having zero inventory on the retailer's end increased the total cost.

Glock and Kim (2014) established a framework that enables the minimization of the overall cost for shipment consolidation in a context involving multiple vendors and single buyers. According to the numerical findings, deliveries should be arranged such that the buyer gets big shipments at the start of the delivery cycle and modest shipments at the end. They also mentioned that vendors with a high production capacity should have their shipments delivered at the end of the delivery process.

Another type of study on consolidation was the Urban Consolidation Centers (UCCs) idea, which can be implemented considering several cities. A UCC is built on the city's border so that large vehicles from shippers can access it and deliver products. Small trucks are then utilized to deliver products to city recipients. López and Cáceres (2020) discovered that such a system could increase complexity while increasing efficiency. Alves et al. (2019) used delivery lockers as a last-mile alternative to investigate urban freight policy in connection to e-commerce. Carriers may use lockers to decrease the number of trucks needed for delivery. Instead of home delivery, Haider et al. (2020) recommended combining client orders and distributing them to a nearby convenience store. This study designs an SC model that integrates multimodal transportation and shipment consolidation into each other evaluating its impact on especially large networks spanning global scales.

2.4 Stochastic Demand in Supply Chains

A big concern in modelling SCNs is that some of the parameters (such as demand) could be unknown or stochastic. Numerous techniques for building such models with stochastic parameters have been considered in the literature. This includes stochastic programming (Ma and Liu, 2017), robust optimization (Ben-Tal et al., 2009), and fuzzy programming (Zimmermann, 1978). With increasing computing capabilities and available historical data, the use of Machine Learning (ML) to assist DMs in planning is becoming more widely considered. In this thesis, the efficiency of ML approaches is evaluated and compared to some of the more traditional Stochastic Programming (SP) approaches.

2.4.1 Stochastic Programming

SP is a framework that allows the consideration of stochastic parameters in the modelling of optimization problems. Some of the studies utilizing SP in the SCN context design are discussed.

Ma and Liu (2017) investigated a closed-loop supply chain network considering transportation and customer demand as stochastic parameters with known joint distributions. An equivalent deterministic MILP model of the proposed stochastic chance constraint with Value at Risk (VaR) is then formulated and solved using commercial CPLEX software based on the finite discrete distributions of the uncertain parameters. Computational results reveal the significance of the proposed model and solution model. An integrated chance-constrained stochastic programming model for multi-period, multi-product, multi-echelon and multi-customer closed-loop SC mobile network was proposed by Ahmadi and Amin (2019). An equivalent multi-objective linear programming formulation was developed with stochastic product return rate and demand to determine the appropriate location and the optimum number of facilities. An integrated chance-constrained model is formulated for sustainable supplier selection and order allocation with stochastic demand (Moheb-Alizadeh and Handfield, 2018).

2.4.2 Machine Learning

Due to the increasing availability of supply chain generated data, computational power and other advantages of machine learning methods over traditional methods, researchers and practitioners have started incorporating machine learning methods into numerous areas of supply chain management (SCM) including supplier selection, supplier segmentation, risk prediction for the SC, demand forecasting, manufacturing, and inventory management (Tirkolaee et al., 2021). ML provides a good accuracy in demand forecasting which gives a significant advantage to SCM since it helps to reduce the well-known bullwhip effect (Chong et al., 2017). Cao et al. (2017) predicted the customer demand using the least square support vector machine and optimized it by the Particle Swarm Optimization algorithm. The applicability of machine learning methods, namely neural networks, Recurrent Neural Networks (RNNs), and

Support Vector Machines (SVMs) to forecast distorted demand at the end of a supply chain was addressed by Carbonneau et al. (2008). They compared the results with traditional forecasting methods, including naive forecasting, trend, moving average, and linear regression. The RNN and SVMs models outperformed the others. Ampazis (2016) does similar work in a different context utilizing Artificial Neural Networks (ANNs) and SVMs to forecast customer demand at the very first stage of a supply chain.

An intelligent demand forecasting system was developed by Kilimci et al. (2019) using a combined application of time series methods, SVMs, and deep learning methods. A variety of numerical tests show that the suggested demand forecasting system achieves notable outcomes when compared to state-of-the-art studies. A study was carried out by Birim et al. (2022) considering demand forecasting under advertisement expenses using Support Vector Regression (SVR), Random Forest Regression, Decision Tree Regressor and deep learning techniques including ANNs, and Long Short-Term Memory (LSTM). LSTM was proven to be superior to other models in terms of predicting demand based on its significantly higher accuracy. The study we found most interesting though was Okwuchi et al. (2020) which compared traditional ML techniques of predicting time series demand data to deep learning models and observed that three deep learning models performed better than the traditional techniques. These models were the Attention-based CNN-LSTM (AC-LSTM), Attention-based Convolutional LSTM (ACV-LSTM) and an ensemble of both techniques using SVR. Okwuchi et al. (2020) utilized compound deep learning models that performed better than most of the best-performing models in many of the studies stated above (e.g., Carbonneau et al., 2008; Ampazis, 2016; Birim et al., 2022).

The increase in the amount of data generated throughout SCM processes encourages the use of advanced methodologies such as ML to extract insights that can be used to improve such processes. Tirkolaee et al. (2021) recognize that there is a huge gap in the utilization of both ML within the optimization model to design and optimize supply chain networks. This paper thereby integrates ML models into the optimization problem and compares the performance with SP techniques for working with stochastic parameters. To this end, the three best ML models observed by Okwuchi et al. (2020) are considered in this thesis as candidates for ML. This is as it is a fairly recent study and has excellent performing models.

2.5 Summary

This research designs a realistic model that considers a three-echelon SC network, with multiple product types being the flow downstream, and multiple components being the flow upstream serving as raw materials for the products (See Figure 1). Raw material requirements were also a factor considered by Fahimnia et al. (2015), Liotta et al. (2015) and Kabadurmus and Erdogan, (2020). Similar to the majority of the studies (such as Shaw et al. 2016; Zhang et al., 2017; Alizadeh et al., 2019), the SC network in this thesis considers multiple facilities, suppliers and retailers. Kabadurmus and Erdogan (2020) additionally consider carbon footprint and multimodal transportation. To the best of our knowledge, there is no previous work that considers and explores an SC with multimodal transportation, shipment consolidation and environment sustainability (water and carbon footprint) simultaneously under stochastic demand. We also could not find any previous work comparing the effectiveness of utilizing data-driven machine learning approaches in optimizing stochastic models to utilizing basic stochastic programming approaches. These gaps in literature motivates this thesis.

Chapter 3

Background on Supply Chain Design and Related Concepts

3.1 Sustainability

One of the most pressing issues today is climate change, its effects and how much industrial activities have an impact on it. To alleviate the effects of climate change, new regulations and sustainability developmental goals have been put into effect by governing bodies on both regional and international levels. These include the likes of the Paris Agreement, Kyoto Protocol, and Carbon Offsetting and Reduction Scheme for International Aviation (Kabadurmus and Erdogan, 2020). Sustainability in the SC has thereby gained considerable attention in both academia and practice. This is because SC operations play a significant role in climate change on a worldwide scale (Kabadurmus and Erdogan, 2020).

According to Ülkü & Engau (2021), there is a need to assess the performance of an SC network by the quadruple bottom line of sustainable development which includes its economic viability, environmental performance, social responsibility and cultural impact. Based on this, there have been works in academia on the design of supply chain networks while factoring in its effects on these factors (e.g., Kabadurmus and Erdogan, 2020; Najjar et al., 2020; Ülkü & Engau, 2021). This research integrates two dimensions of sustainable development, namely the economic and environmental performance, into the Logistics Network Design (LND). Important factors that contribute to the environmental performance of SCs include the water and carbon footprint of its activities such as transportation and production. Factoring these in the LND stage enables the SC to proactively align its activities to the relevant Sustainable Development Goals as defined by the United Nations. These include access to clean water and sanitation (goal 6), reduced climate action (goal 13), protection of life below water (goal 14) and life on land (goal 15).

3.1.1 Water Footprint Analysis

Nearly 80% of the world's population and 'water affiliated organisms' are at risk of water scarcity (Vörösmarty et al., 2010). This is due to the societal water consumption by humans compounded with large industrial and agricultural usage. Climate change exacerbates this situation (Wheeler & Von Braun, 2013). Considering that human water needs are likely to continue growing (Schewe et al., 2014), it is important to consider the impact of SCs on water usage. This is because poor, unmanaged water distribution systems can lead to agricultural failure, species extinction, food shortages and much more (Farzaneh et al., 2021). The growing world population, climate change and continuing industrialization pose additional stresses to freshwater availability (Manzardo et al., 2014). It is thereby important to factor in the impact of factories and transportation modes on the water when designing supply chain networks. This would motivate the efficient usage of the necessary elements of a supply chain that impact water sustainability.

In considering factory location and its impact on water sustainability, Farzaneh et al. (2021) provide a system for the development of governmental plans for ecological and societal allocation of water within a region considering hydrological factors such as precipitation, evaporation, inflow and seepage after historical patterns on these parameters have been considered. Farzaneh et al. (2021) consider the development of water conservation policies that ensure that during a period, enough water is released to the local reservoir for societal usage, to satisfy downstream ecological needs, and support aquatic ecosystems and ecosystem services. This is to ensure the preservation of the water balance in the reservoir by ensuring that at the end of each period, the total water storage in the reservoir is not less than the dead capacity (which makes drought a risk) and not more than the flood capacity (which makes flooding a risk).

Assuming that such models as proposed by Farzaneh et al. (2021) to determine the industrial allocation of water considering local factors (such as the local drought index) have been built and analysed, factories and manufacturing plants within such locations can then be regulated to limit their water usage on all processes to be within a specified cap. The model provided in this paper assumes that after local water management has run the necessary analysis and decided the amount of water cap is necessary for societal ecological and industrial usage, individual caps or wastewater release caps after which recycling is enforced are then mandated on local industries, factory, manufacturing centres with a charge for any overuse or money spent on wastewater. The water footprint of each product is the average volume of water utilized (excluding the quantity recycled) while producing it (Gerbens-Leenes & Hoekstra, 2011). From the point of view of the company, it (the company) has to ensure all its local processes combined do not consume more water than the cap allocated to it (the company). This might affect its decisions on variables such as the technology to use, the maximum quantity of the item to produce or process, and the amount of water overuse it can afford despite the charge considering the end goal/gain. This is used to calculate the local water footprint of such a company.

The next question is then about how a factory could measure the amount of water it consumes if the quantity can not be gauged automatically. In 2011, the Pacific Northwest National Laboratory released a 33-pages guideline for estimating unmetered industrial water use (Boyd, 2011). This provided a systematic approach to calculating unmetered sources while utilizing engineering estimates. The analysis provided a methodology for estimating water use in systems such as steam boiler systems, and evaporative systems. The approach can however be studied and generalized to other systems that might exist within the industries. The report suggests that the most accurate way to measure water usage would be through the installation of flow meters. However, in the absence or impracticality of installing such devices, the guidelines can be used to obtain an engineering estimate.

Facilities use industrial water generally for purposes such as processing, fabricating, cooling, diluting, washing or through the direct incorporation of water into a product. Systems that utilize water but operate in a closed-loop where there is no significant water loss per cycle would provide negligible impact and can be ignored in estimation (Boyd, 2011). All processes that constitute water usage in Company X have to be analyzed individually and expertly. The quantity of water usage in applications such as within an evaporative cooling system depends on total hours of operation, and the size of the device (i.e chiller tonnage). Within usages similar to the washing processes, the quantity of water depends on the number of units washed, processed or manufactured, the amount of water required for each washing per gallon, and how much of such water is recycled by percentage.

As an analysis of a sample system, Boyd (2011) gives the following formula to calculate annual water usage by an open-recirculating cooling system.

$$Annual\ water\ use\ (gal/year) = n * chiller\ consumption\ * \frac{total\ runtime\ (hours)}{8{,}760\ hours}$$

where *n* is the number of chillers installed within the cooling system, *chiller consumption* is a factor based on the size of each chiller, and total runtime is the number of hours the chillers are run in a year divided by the number of hours in a year. An expert would be required to analyze each of the existing water-consuming systems within a factory and develop the necessary estimation formulae.

In summary, some of the quantity of water usage is dependent on how long the processes are run, while others depend on the total number of units or batches of units produced. The following formula is therefore used as a summary of total water usage within a factory or industry where processes that use water on an hourly basis are indicated as *p* while those that use on a quantity basis are indicated as *q.*

$$Total\ Water\ Usage = \sum_p W_p * H_p + \sum_q W_q * \frac{U_q}{B_q}$$

where W_p is the quantity of water used per hour by each process *p* that uses water on an hourly basis, and H_p is the total number of hours used per process *p*. W_q is the quantity of water used up by each process *q* that uses water on a per-batch basis, U_q is the total number of units worked on during process *q* and B_q is the batch size. W_p and W_q are adjusted to reflect the portion of the water recycled during the process.

Example 3.1

The major sources of water usage in Factory X are an evaporative cooling system, a steam heating system and a washing system. After an expert analysis of all systems, it is determined that the cooling system uses water at a rate of 1,000 gallons per hour. The steam heating system uses water at an average rate of 1,100 gallons per hour (considering factors such as the feed water rate and condensate return). The washing system however uses 200 gallons to wash each batch of the product while reusing 50% of the wash water. 20,000 units are produced annually at a size of 100 per batch, which requires the cooling system to run for 270 hours and the steam system to run for 250 hours. The total water use can thereby be calculated as below with the first, second and third terms representing the water consumption by the cooling, heating and washing systems respectively.

$$Total\ water\ usage$$
$$= 1000\frac{gal}{hrs} * 270hrs\ + 1100\frac{gal}{hrs} * 250hrs + \frac{20{,}000}{100} * 200 * (1 - 0.5)gal$$

= 0.565 million gal

Water is also polluted during transportation. In this thesis, we consider transportation water usage per mile to be negligible. This is as the water footprint directly related to transportation is minimal unless when biofuels (such as ethanol and biodiesel) are used (Aivazidou et al., 2018). An exemption is however made for water transportation. This is because it could directly impact water through means such as oily water discharge, wastewater discharge, garbage waste and ballast water discharge. A general formula to calculate the impact of shipping on the water is provided below. The water footprint is then estimated as the number of containers/vehicles owned by the company multiplied by the time used and the average pollution rate per unit time for transportation mode t.

$$W_T = w_t^{pollution} * \left(\frac{\text{distance}}{\text{average speed}}\right) * Y_t$$

where W_T is the total water pollution, $w_t^{pollution}$ is the average water pollution a vehicle/container is accountable for per hour for mode t, and Y_t is the number of containers/vehicles in use. This gives an estimate of what each vehicle/container contributes to the total SC water footprint.

For water footprint analysis in this thesis, the whole supply chain has to be considered. This includes not just the quantity of direct water consumption but also the sustainable level of water within the different catchment areas impacted considering available resources. To this effect, the following will be considered in the model formulation:

a. The water scarcity level or water cap directives from local water management in facility locations.

b. The direct water consumption (excluding the quantity instantly recycled) while producing batches of each product (from plants).

c. In transportation, all modes are considered to have negligible water impact except water transportation which could pollute water through its direct contact with it. This will be incorporated into the model based on the distance travelled.

3.1.2 Carbon Footprint Analysis

The environmental damage accompanying the activities within SCs calls for action from all parties including the regulatory bodies, business operators and consumers. Greening the supply chain should thereby be an objective while designing an SC network. To this effect, multiple carbon emission regulation schemes have been developed. A carbon cap is set for every company which can be reduced gradually to encourage them to go greener. Regulation schemes include the carbon tax and carbon cap-and-trade. Carbon tax means companies have to pay a tax on every unit of carbon emission above the cap set by the government. This is a widespread initiative used in many countries such as Canada and China (Goulder and Schein, 2013). The unit price paid for excess emission is set by the regulatory bodies. The carbon tax scheme however might cause under-utilization of resources if companies become too careful not to exceed the cap because of the reluctance to pay the tax. This limitation is covered by the carbon cap-and-trade scheme which allows companies to buy and sell the allowance for carbon emissions from other companies if necessary. The unit cost of carbon emission is determined by the market and companies can design the SC network to optimize the overall cost while factoring in emission allowance buying opportunities (Kabadurmus and Erdogan, 2020).

In the SC context, both transportation and facility operations are among the factor that contributes to the overall carbon footprint. In logistics, transportation is considered to be one of the largest contributors to environmental hazards The source of most CO_2 emissions in the United State is the transportation sector (EPA, 2022). In literature, the calculation of the CO_2 emissions from transportation is mostly based on the type of fuel used, the weight of the load, the distance travelled and the type of vehicle used. However, Ülkü (2012) argues that to be more realistic, factors such as vehicle packing efficiency and traffic congestion (especially for road transportation) should be considered. Other factors such as the effective volume and weight utilization of the vehicle should also be considered in the calculation of the CO_2 emissions per trip. In the last few decades, studies on reducing emissions at the SC level, and hence from an SCN perspective have increased (e.g., Elhedhli and Merrick, 2012).

The four major modes of transportation all have to be considered (air, road, rail and sea transportation). The shipping emission factor of all four modes respectively is provided as 1.278, 0.209, 0.021, and 0.0409 kg of Carbon emissions per ton-mile. This is the amount of carbon emitted when shipping a tonne of freight travelling 1 mile (Carbonfund, 2021) . Carbon emission per gallon of fuel is also obtainable from the same source. According to Ulku (2012), carbon emissions from different modes of transportation can be calculated using the following formula:

$$Carb_{transport} = \sum_{t} F_t^{CO2} \frac{D}{M_t} * Y_t + P_t^{CO2} U_t + S_t^{CO2} \frac{U_t D \sum w_p}{(1 - c_f)}, \qquad (3.1)$$

where $cf = 1 - \frac{sp_t}{sp_{max}}$, $U_t = \min\left\{\left[\frac{V_t \gamma_t \theta v_t}{\sum V_p}, \frac{W_t \gamma_t \theta w_t}{\sum W_p}\right]\right\}$, and

i. For each vehicle, θv_t and θw_t are the weight and volume efficiency targets set by the DM.
ii. γ_t is the packing efficiency indicating how well the packages are stacked up.
iii. S_t^{CO2} is the shipping emission factor per ton-mile when vehicle *t* is in use.
iv. D is the total miles travelled.
v. F_t^{CO2} is the carbon emission per gallon of fuel.
vi. M_t is the fuel mileage of the vehicle.
vii. P_t^{CO2} is the average amount of CO_2 emission when packaging each unit of cargo unto vehicle *t*. This includes emissions from loading and unloading processes.
viii. U_t is the maximum number of product *p* that can fit in a single type-*t* vehicle.
ix. sp_t is the average speed of vehicle *t.*
x. sp_{max} is the maximum speed allowed
xi. Y_t is the number of vehicles in use for transportation mode *t*

The first term in Eqn. 3.1 refers to the amount of carbon emitted if the vehicle was travelling empty. The second term refers to emission while packaging the products (loading and unloading), whereas the third term refers to the shipping emission based on the weight of the products being shipped, the distance being travelled and the congestion factor. sp_t and sp_{max} can be assumed equal for some modes of transportation (e.g air, water and rail) thereby having a congestion factor of 0.

Production activities within facilities also add to the total emissions of the SC. It is also assumed that the technology type in different facilities causes varying emission rates even for similar products. If the quantity of carbon emitted per batch production of final product *f* in factory *i* is represented as e_{ip} and the total quantity of product *p* produced is X_{ip}, then the total carbon emission due to production is

$$Carb_{prod} = \sum_{i \in M} \sum_{p \in P} (e_{ip} * X_{ip})$$

3.2 Multimodal Freight Transportation and Shipment Consolidation

Multimodal Freight Transportation (MFT) now plays a central role in SCs majorly due to trends that are fundamentally changing the business strategy of large or global corporations such as the growing demand for speedy product deliveries, the continued economic globalization driving trade and investment, adoption of practices like agile manufacturing and the need to make supply chains increasingly efficient (Rondinelli & Berry, 2000). The presence of multiple transportation modes allows the network to utilize the best combination of modes to ensure products are delivered on time and the SC goals are achieved. With the increasing expansion and integration of transportation systems, the evaluation of the full impact of transportation on the environment (water, air and land resources) becomes more complex (Rondinelli & Berry, 2000). However, researches show that besides potentially reducing the total economic SC cost, multimodality of transportation could also decrease environmental impact (such as the total carbon emitted) by allowing the use of more sustainable transportation alternatives (Kabadurmus and Erdogan, 2020).

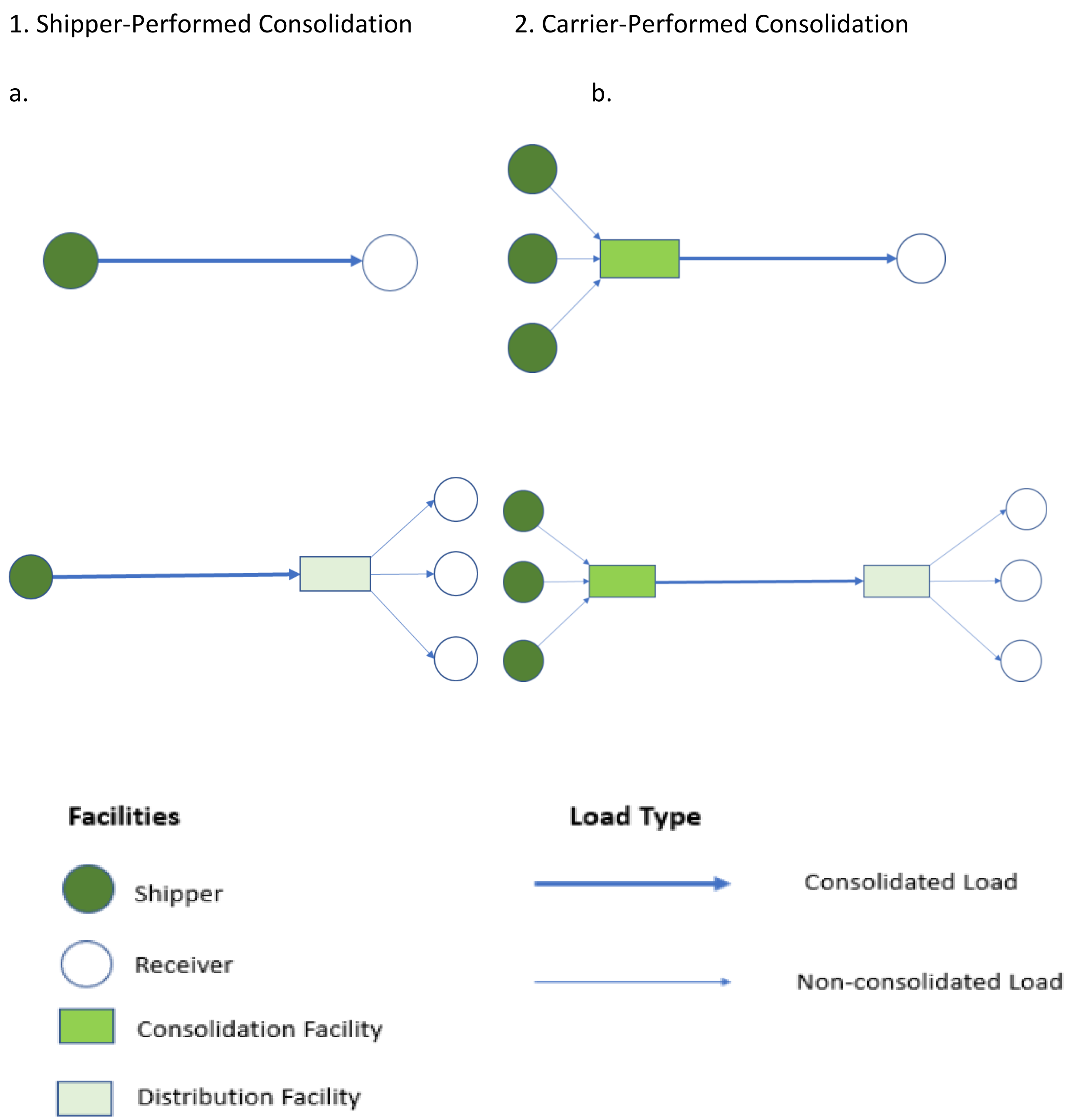
Shipment consolidation is a strategy that combines two or more orders into a larger unit that can be dispatched on the same vehicle usually in the same market region. This strategy is potentially more sustainable than having to dispatch each shipment individually. This is because it can greatly reduce the transportation cost per unit or order (hence economic sustainability) while employing fewer long-haul shipments and reducing total vehicle miles. Ülkü (2012) demonstrates that increased utilization of vehicle capacity (through shipment consolidation) would lead to a reduction in total carbon emission.

The shipment could either be Full TruckLoad (FTL), where the shipper uses most or the entire truck space to make its delivery to its customers, or Less than TruckLoad (LTL), where only a fraction of the truck space is used and paid for. The LTL shipment from multiple shippers is loaded onto a truck and delivered as a whole to a customer or to a distribution center where the shipment is broken down and distributed to respective customers (Ülkü, 2012). FTL shipment is cost-effective if most of the capacity of the truck is utilized by the quantity of the freight. This can be achieved through shipment consolidation. Allowing a mix of FTL and LTL modes of shipments in an SC network can enable cost savings as there is generally more utilization of available trucks (Ülkü, 2012). From the shipper's point of view, FTL mode can be used when a large capacity is required and LTL can be rented when only a fraction of the truck space is needed.

The different shipment consolidation logistics configurations between shippers (e.g, a manufacturer) and receivers (e.g., a retailer) shown in Figure 2 (based on Higginson, 1996) are briefly described.

In configuration 1a, the shipper utilizes its truck, the truck it receives from the receiver or a 3rd party logistics (3PL) truck to directly transport its cargo to the receiver. This might be efficient if the distance is short and/or if the shipper can achieve FTL. Otherwise, there might be great underutilization of vehicle capacities. This could be compounded in multimodal transportation, especially for global SCs where full fixed prices will be paid for underutilized vehicles/containers travelling over long distances.

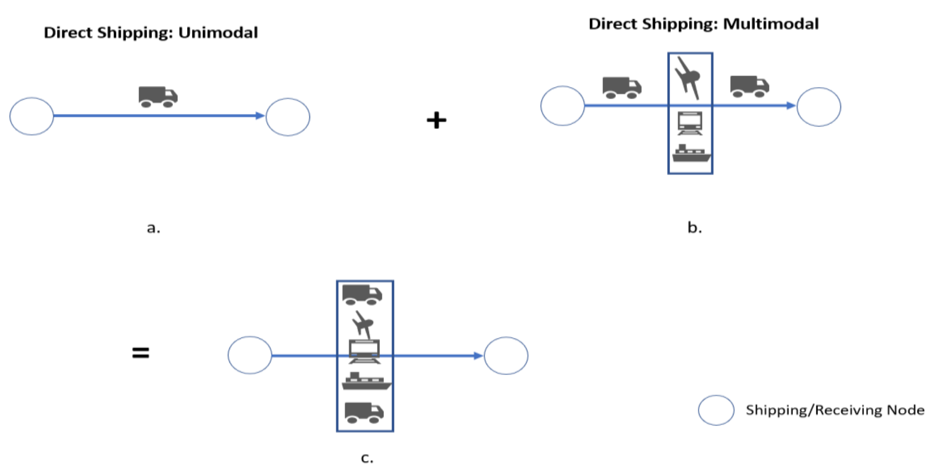
Figure 2. Different Shipment Consolidation Configurations



In configuration 1b, a single shipper intends to supply multiple receivers in the same or relatively close market region. The shipper thereby consolidates all deliveries and sends them over to a distribution facility closer to the region where the receivers are located. The facility breaks down the package and either delivers in individual vehicles to the receivers or utilizes the milk run delivery format. In global SCs, a supplier that has a large customer base in another region can consolidate all its deliveries and thereby save money in the shipping of his consolidated load as he uses fewer vehicles/containers.

In configuration 2a, loads from multiple suppliers are transported to a consolidation facility from where the loads are consolidated and then shipped over to the receiver. Depending on the distance and truckload, this could provide a cheaper alternative to direct delivery as the total distance travelled from the suppliers to the receiver is reduced thereby reducing fuel utilization and carbon emission. Also, in global supply chains where multimodal transportation is required, individual loads are consolidated and shipped out as a singular unit thereby saving costs accrued through less travel distance and increasing the utilization of vehicle capacities. Configuration 2b is similar to 2a. It however recognizes that the shipment might be required to go through a distribution facility if it is to be received by multiple receivers.

Figure 3. Representation of Unimodal and Multimodal Transportation



Each arc on the supply chain network can utilize multimodal transportation. However, in this study, some arcs are indicated as purely unimodal (see Figure 3a). These include trips between facilities in the same region (supplier to local consolidation facility and distribution facility to receiver). Arcs that require multimodal transportation are however considered to require trucking for a portion of the trip to transport the shipment from the shipping node to the intermodal hub (such as the airport), and from the intermodal hub to the delivery node (see Figure 3b). Both modes are sometimes merged and represented as Figure 3c in the coming chapters.

3.3 Supply Chain Stochastic Demand Approaches

Usually, many of the parameters considered in LND such as demand, supply, cost and delivery time are inherently uncertain (Govindan et al., 2017). These uncertainties could be caused by disruptions due to human or natural factors and unforeseen circumstances. The impact of these uncertainties could sometimes be significant, causing decision-making to be more challenging. Methodologies have been generated to solve SCN modelling problems with uncertainty to obtain the best configuration that generally performs well under all or most of the possible occurrences of the uncertain parameter (Govindan et al., 2017). Traditionally, methodologies employed to work with stochastic parameters with known probability distribution is stochastic programming (SP). When no information about the probability distribution is known, robust optimization (RO) models can be employed to optimize the worst-case scenario while fuzzy mathematical programming is capable of handling the modelling when the uncertainty has some level of ambiguity and/or vagueness (Govindan et al., 2017).

Basic SP approaches to uncertain parameters will be considered in this research as the ‘traditional’ methods. The performance obtained will be compared against the performance of other advanced techniques of machine learning.

1. Stochastic programming
    a. Simple Recourse Programming
    b. Chance-Constrained Programming
2. Machine Learning
    a. Attention Convolutional- Long short term memory (AC-LSTM)
    b. Attention Convolutional Neural Network- Long Short Term Memory (ACN-LSTM)
    c. Stacking ensemble of AC-LSTM and ACN-LSTM using Support Vector Regression (SVR)

3.3.1 Stochastic Programming

Stochastic programs are optimization models in which some of the model parameters are considered uncertain. The first forms were by Beale (1955) and Dantzig (1955) which involved a sequence of action-observation-reaction (or recourse). Simple and general recourse assumes risk neutrality of the DM while other models like chance-constraint programming factor in risk-averseness of the DM. SP has attracted a lot of attention in literature because it is a fundamental building block of many supply chain problems including capacity planning, multiperiod inventory, and contract design problems (Chen et al., 2009).

To work with models having uncertainties, several mathematical frameworks have been used including stochastic programming, chance-constrained programming and robust optimization, each with varying degrees of risk-aversion (Li & Grossmann, 2021). SP is a risk-neutral approach that aims to optimize the expected outcome for all scenarios or over the probability distribution. The objective is to find a solution that performs optimally on average while being feasible for all (or most) of the possible scenarios. Chance constrained programming can be seen as a stochastic program wherein some of the constraints have probabilistic parameters and these constraints only have to be satisfied at a given level of probability. Chance-constrained programming has connections with risk management as it allows modelling flexible enough to deal with reliability issues (Li & Grossmann, 2021).

This thesis considers a model with demand as a stochastic parameter with an overstocking cost for stock above demand, and understocking cost for stock below demand. This is quite similar to the popular Newsvendor Problem (NVP) (Scarf, 1958). The vendor must decide on how much stock to order x at a cost c without prior knowledge of what the demand would be. This demand is represented by a random variable ξ. The selling price p could also be considered as the understocking cost u as it is a sales opportunity foregone. The overstocking cost o could be the difference between the sum of the purchasing and holding cost, and the salvage value. However, we can assume perishable goods have no salvage value and would thereby not incur holding costs as they are discarded with no return making the overstocking cost equal to the purchasing cost c. For this thesis, the simple recourse stochastic programming and chance-constrained programming are considered. It is also assumed that the probability distribution of the uncertain parameter (demand) is unknown but characterised by some discrete realizations (scenarios) of this parameter as an approximation of the probability distribution. A finite number of scenarios obtained from historical observations is utilized.

1. Simple Recourse Stochastic Programming: The objective is to find a solution that performs optimally on average while being feasible for all (or most) of the possible scenarios. In two-stage programming, there are two stages of decisions to be made. The first stage decisions are those to be made 'here and now' at the beginning of the period. Second-stage decisions involve uncertainty and are thereby taken as 'wait and see' decisions at the end of the period. These decisions are taken after the true values of the uncertain parameters are disclosed. For simple recourse SP, the recourse action simply involves calculating the penalty based on the deviations of the outcome from the prescribed solution (Dye, 2008). The NVP is an example of a stochastic linear program with recourse (Birge, 1997). The recourse can be seen as the second stage decisions to be made such as the quantity to backorder (if that option is available) or hold as inventory for the next cycle. It can be generally represented as the following model

$$\min_{x \in \Re^m} c^T x \; + \; \psi(\xi, x)$$

$$s.t \;\; Ax = b,$$

$$x \geq 0,$$

where $\psi(\xi, x) = o[\mathrm{x} - \xi]_+ + p[\xi - x]_+$, $[a]_+ = \max\{a, 0\}$ (Birge, 1997).

In a risk-neutral setting, the objective function of the problem can be formulated as the minimization of the expected value of the cost in relation to the probability of each possible demand scenario, or the probability distribution of the demand (assuming a continuous distribution)

$$\min_{x \in \Re^m} \mathbb{E}_F[c^T x + \; \psi(\xi, x)]$$

The minimization of the expected value gives a solution that performs optimally on average in all the possible scenarios (Dye, 2008).

2. Chance-Constrained Programming: This was developed by Charnes and Cooper (1958) and has gained significant attention from researchers in many fields. This approaches the NVP from a service-level perspective to find the optimal order quantity (inventory level) that allows the demand to be met with at least $'1-\alpha'$ probability while minimizing the cost incurred due to inadequate or surplus inventory (van der Laan et al., 2022). This makes chance-constrained programming a risk-averse approach where α is the pre-set risk level of the decision maker.

Supposing that the constraint that links an observation of the demand (scenario) ξ to the target inventory level *x* is linear in *x* (represented as the function $g(x,\xi) \leq b$), then the service-level constraint of the model can be formulated as

$$\mathbb{P}[g(x,\xi) \leq b] \geq 1-\alpha$$

meaning the values of variable *x* the optimal model provides only has to fulfil the constraint often enough to meet the pre-set risk level α. The objective function remains the same while the service-level constraint eliminates some of the scenarios.

In classical stochastic programming, it would be required that such a constraint is satisfied for all possible realisations of the stochastic parameter. However, this might be infeasible or very costly if the model has to adjust its decision variables to account for outliers within the scenario space. The chance constraint programming reformulation allows for the model to ignore as many outlier scenarios as possible that might increase the overall cost if catered for by the program but within the probability limits set by the risk-averse DM.

3.3.2 Machine Learning

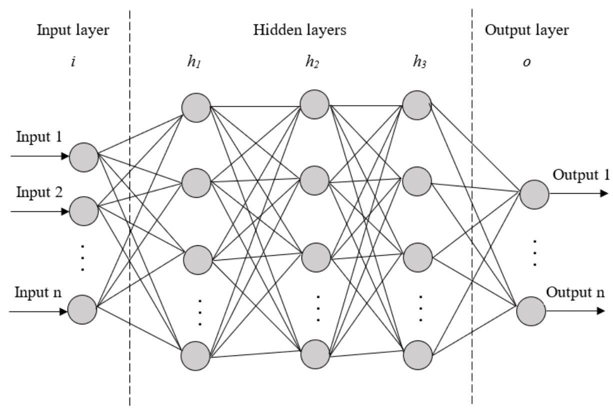
General statistical approaches to demand forecasting employ the time-series method using techniques such as naïve methods, average methods, exponential smoothing, Holt's linear trend method, damped trend methods, moving averages, auto-regressive moving average and auto-regressive integrated moving average (Hyndman & Athanasopoulos 2018). However, machine learning (ML) approaches have been shown to perform significantly better than traditional statistical approaches (Carbonneau et al., 2008). While both statistical and machine learning methods to demand forecasting are based on time-series data, the difference between them is in the computational power of each method. The statistical methods rely on limited historical time-series data while the ML approach can use huge volumes of data alongside several of its related features to forecast future demand. For example, weather data can be used alongside historical data to increase forecasting performance. ML can also do a deeper search to obtain insights into seasonality and correlation within the data to better improve forecasting accuracy (Kilimci et al., 2019).

ML is a subfield of Artificial intelligence (AI) which takes advantage of available computational power and big data to enhance the application of innovative algorithms to improve multiple areas such as supply chain management. ML allows machines to learn without hardcoded programming. Unlike most traditional SP approaches, ML can deal with the huge volumes of data typically generated in supply chain management even when the problem being solved involves non-linearity. This will make the model a data-driven model and can find much use in optimizing supply chain networks, especially nowadays when they generate huge datasets.

ML is usually classified into supervised and unsupervised learning. Supervised ML models are trained using defined labels which are then used to test and improve their (models) performance. These models are mostly used for regression and classification. Unsupervised ML models on the other hand do not require defined labels. The algorithm seeks out hidden patterns and insights from the training data that could then be used to classify or cluster both the training and testing data (Okwuchi et al., 2020). Prediction of a stochastic parameter using historical data would require a supervised learning regression algorithm with the historical data as labels.

In supervised learning, several basic ML algorithms exist for solving regression problems. These include linear regression, support vector regression, random forest regression, and so on. An advanced type of ML is called Deep Learning (DL) which is based on artificial neural networks (ANN) with representation learning. ANNs are computing systems inspired by actual biological neural networks found in the human brain (Chen et al., 2019). An ANN is based on a collection of connected nodes called neurons which model the neurons in a biological brain. Weights are used to connect the neuron. A stack of neurons on the same level is called a layer. An ANN is made up of multiple interconnected layers with each layer performing some sort of transformations on its input yielding a different output. Input data travels from the first layer (input layer), through several hidden layers and comes out of the last layer (output layer). The input to output transition is performed repeatedly. The weights of each connection are updated as the learning proceeds until the best weights that lead to minimal loss (or optimize a cost function e.g accuracy) are obtained. The weights are updated using stochastic gradient descent or other forms of optimizers. ANN serves as the foundation for other types of DL algorithms (Schmidhuber, 2015). Figure 4 (based on Bre et al., 2018) shows the structure of a typical artificial neural network with connecting neurons in multiple layers.

Figure 4. Artificial Neural Network



DL algorithms have been shown to generally perform better than basic ML models (Sarker 2021; Okwuchi et al., 2012). Due to the ability of DL algorithms to learn more complex patterns in data and also improve continuously with increasing data size (Goodfellow and Courville, 2016), the ML techniques used in this thesis are all DL algorithms. They have also been shown to perform excellently on time series data (Okwuchi et al., 2012). For time series data, a spatial relationship can be described as the pattern acquired based on the relative location of data points to each other. Temporal relationships however describe the patterns acquired from the sequential (time-based) arrangement of the data points. Different DL algorithms possess different strengths in observing both types of relationships (spatial and temporal). Examples of DL algorithms include Recurrent neural networks(RNN), convolutional neural networks (CNN), Long Short Term Memory (LSTM), Deep Belief Networks, Deep Neural Networks and so on. For this thesis, only the relevant architectures and concepts are briefly explained.

1. CNN-LSTM

Convolutional Neural Networks (CNNs) are a specialized type of neural network (NN) used for working with two-dimensional grid-like image data. This makes it applicable to fields like image and video recognition, natural language processing, and medical image analysis. It has however also shown good performance in finding spatial relationships within numerical data (Bai et al., 2018). Time series data with samples taken at regular time intervals could be seen as a one-dimensional grid (Goodfellow and Courville, 2016). The CNN is made up of a series of interconnected layers with hidden layers such as convolution, pooling and normalization and fully connected layers. CNNs are difficult to tune, require large datasets and have difficulty in extracting temporal features (Okwuchi et al., 2020).

Meanwhile, Long Short-Term Memory (LSTM) networks are special Recurrent Neural Networks (RNNs) that have an improved remembering capacity. Unlike other NNs, RNNs have a structure in which all inputs and outputs neurons are directly connected to each other. In addition, neuron outputs are applied recursively as inputs recurrently. This acts as feedback which serves as a form of memory that makes previous input data leave a footprint making it good for observing temporal patterns in sequential data. The iterative procedure of the RNN however often leads to the exploding and vanishing gradients problem. LSTM serve as an improvement to RNNs by fixing these problems through the use of special gates (input, input modulation, output, forget) thereby allowing short-term and long-term memories. This makes the LSTM much better than the standard RNNs at capturing long-term temporal dependencies within the data (Yildirim, 2018). LSTMs have high performance in observing temporal relationships in data but do not perform as well in observing spatial relationships.

CNN-LSTM, therefore, involves the stacking of the CNN and LSTM models to take advantage of the strength of each. This creates an architecture that is not only good in observing temporal relationships within data (through the presence of LSTMs) but also good in observing its spatial relationships (through the presence of CNNs). CNN-LSTMs have been applied in time series, text classification and price prediction showing results that outperform either singularly (Okwuchi et al., 2021).

2. Convolutional LSTM (ConvLSTM): The Convolutional LSTM model is an architecture inspired by both CNN and LSTM having a better performance in observing the spatio-temporal relationships in data than either of the models individually (Shi et al., 2015). Compared to standard LSTMs, ConvLSTM is able to model both spatial and temporal relationships simultaneously by encoding the spatial information into tensors (Wang et al., 2017). The ConvLSTM architecture preserves the spatial information within the data by modifying the LSTM through the replacement of its (LSTM) matrix multiplication operations with convolutional operations (Luo et al., 2017). ConvLSTM is typically suitable for a series of time-dependent images or videos (Vrskova et al., 2022). It has however also been applied to time series data (Okwuchi et al., 2021) but requires the reshaping of the input data appropriately.

3. Attention Mechanism: This is a mechanism that can be added to an existing NN model. It mimics cognitive attention and its goal is to help the model focus on important parts of the input data rather than all of the information (Vaswani et al., 2017). For sequence modelling, the attention mechanism allows the modelling of dependencies without regard to their distance in the input or output sequences (Kim et al., 2017). This makes the network devote more attention and focus to small important parts of the data that have a larger influence on the output. Self-attention is an attention mechanism relating different positions of a single sequence in order to compute a representation of the same sequence. A self-attention layer can be added to the architecture of other deep learning models to potentially increase its performance. It is added to the CNN-LSTM and Convolutional LSTM layers previously described to form the Attention-based CNN-LSTM (AC-LSTM) and the Attention-based Convolutional LSTM (ACV-LSTM) respectively. Okwuchi et al. 2021 show that these attention-based models perform better than their counterparts without the attention layer.

4. Ensemble: This is a process where multiple models are combined in the prediction process by either using different ML algorithms or using different training data sets. The predictions are then aggregated using an ensemble model which outputs a single prediction for the unseen

data. This helps reduce the generalization error on predictions. The ensemble model seeks 'the wisdom of the crowd' in making its final predictions (Kotu & Deshpande, 2015). Types of ensembles include stacking and bagging (bootstrap aggregating). Bagging is a technique for reducing the generalization error by combining the predictions from the children models. This is done by having all the children models individually train on the input data, and then vote on the output for the test data. For regression data, the average value of all the predictions would be taken as the final output. On the other hand, Stacking uses a new combiner algorithm to learn how best to combine the predictions from contributing children models taking advantage of where each child model performed better than the other.

For this study, our objective is to forecast future demand based on historical orders. To this end, we investigate the utilisation of multiple machine learning techniques. An increase in forecasting accuracy would lower overall cost as inventory is reduced to what is required while increasing customer satisfaction due to fulfilled requests.

3.4 Summary

This chapter generally introduces the major concept that will be considered in the LND design. This included sustainability, MFT, SCL, SP and ML. The formulas and explanations provided inspire the objective function and constraints used in the formation of the optimization model.

Chapter 4

Problem Description & Model Formulation

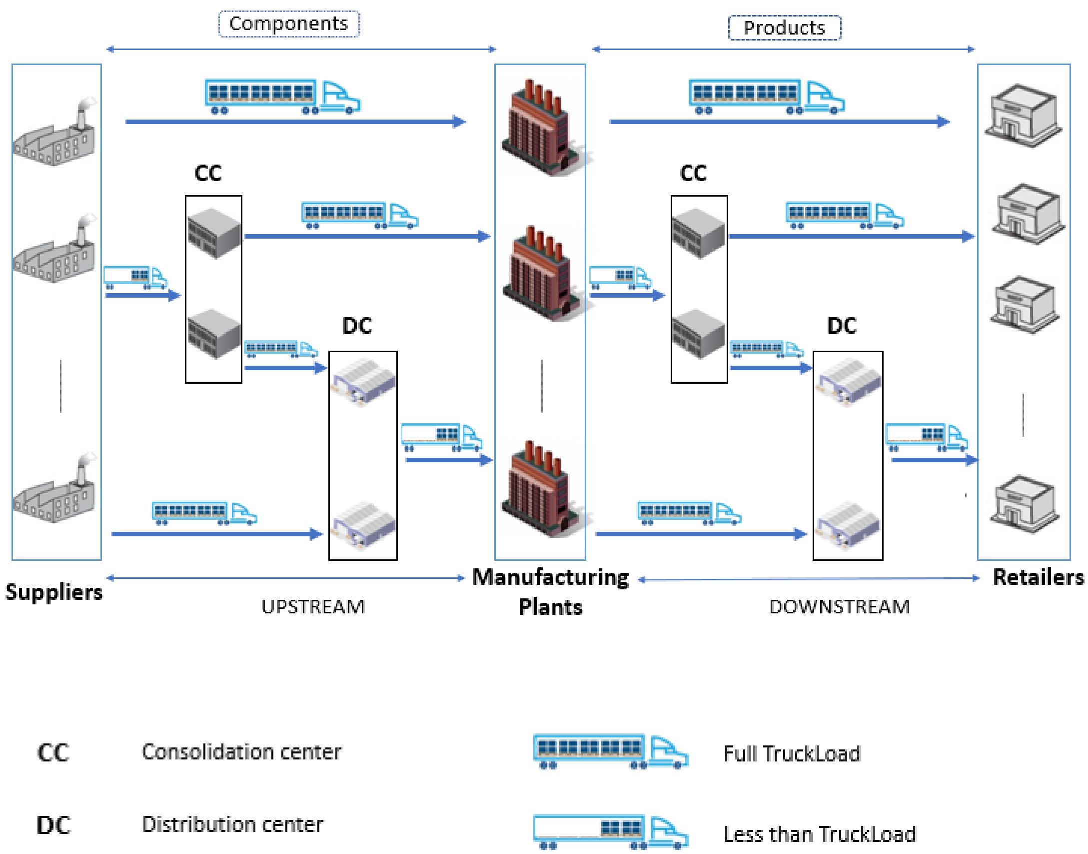
A realistic 3-echelon supply chain network is contrived which considers environmental sustainability with factors like water and carbon footprints. The network also factors in multimodal freight transportation and consolidation of shipment between its agents. This sustainable SCN with multimodal freight transportation and shipment consolidation is termed Model A for convenience. A similar model without shipment consolidation capabilities is also formulated and termed Model B. This allows for a comparative study of both models to analyse the impact of shipment consolidation on economic and environmental sustainability. Both Model A and Model B are considered deterministic.

4.1 Sustainable Supply Chain Network with Multimodal Freight Transportation and Shipment Consolidation (SSCN/MFT&SCL)

The proposed SC model comprises multiple suppliers, plants, retailers, consolidation and distribution facilities/centers, components and final products. Suppliers provide components which are delivered to the manufacturing plants as raw materials. There is a bill of material (BOM) indicating which components are required for the production of each product. The plants then produce the final products and deliver them to the retailers according to their orders. Each plant uses varying technologies which impact the cost, time consumption, water consumption and emission rate of production. The goal is a system-wide optimization of the logistic network of the SCN with a focus on the tactical level logistical decision-making such as order quantity, suppliers to engage and transportation modes to use. Figure 5 shows the three echelon network describing the connections between the supply chain agents. We consider an SCN with 3-echelons because it is the simplest but not trivial case.

The retailers/retail stores base the quantity they order on the local demand from customers. From the time between the initial retailer request to the delivery of the final products, there is a set planning period during which the order-deliver cycle has to be completed. For the deterministic model, it is assumed that the retailers know the exact amount of demand they would receive and plan accordingly.

Figure 5. Supply Chain Network Diagram



The other characteristics of the network and assumptions considered are enlisted below:

- The bill of material (BOM) indicates the components required for the production of each final product.
- All manufacturing plants (suppliers) are capable of producing every final product (component). The capacity for a plant (supplier) to produce a final product (component) can be set to zero to make the plant (supplier) incapable of producing the final product (component).
- The physical network infrastructures are allowed to be in different regions or countries, hence the requirement for multimodal modes of transportation which include road, rail, air and water.

- Different types of components or final products are allowed to be shipped together on the same vehicle/container.
- The consolidation and distribution facilities/centers are set to be geographically located between the suppliers (plants) and the plants (retailers). The location of these consolidation/distribution facilities can however be set as required. The consolidation facilities are similar to urban consolidation centers and the distribution facilities are similar to urban distribution centers.
- The consolidation facilities are assumed to be close to the shippers (suppliers in the upstream and plants in the downstream) thereby requiring road as the only mode of transportation. The model can be adjusted to remove this assumption.
- The distribution facilities are assumed to be close to the receivers (plants in the upstream and retailers in the downstream) thereby requiring road as the only mode of transportation. The model can be adjusted to remove this assumption.
- A short-haul truck with a fixed cost less than that of a normal long-haul truck is used for trucking between the shippers and the consolidation facilities, and between the distribution facilities and the receivers.
- To be more realistic, rail, air and water transportation modes are only available on some of the arcs with long distances. Road transportation is available on all arcs.
- There is an activation cost for each supplier and manufacturing plant. Each supplier (plant) has a limited capacity for producing each component (final product). Consolidation and distribution facilities are however un-capacitated. The capacity limit can be set if required.
- The handling cost (loading, unloading) of items taken through the consolidation or distribution facilities is considered dependent on the number of items being processed.
- Transportation capacity limitations between SC nodes are not considered. It is assumed that as many vehicles/containers as needed can be obtained. There is however a fixed cost for activating each unit of vehicle/containers except for air transportation where cost is solely based on the billable weight (the higher of the volumetric weight and physical weight) of the cargo and the distance travelled.

- When rail, air and water transportation modes are used, the need to utilize short-haul trucking for a portion of the trip is recognized (delivery to and from the intermodal hub. See Figure 3b).
- The portion of each trip through air, water and rail requiring short-haul trucking can be described as a percentage θ. It is assumed that θ on a specific arc for all three modes is the same.
- The cost of utilizing rail, air and water transportation modes on any arc thereby has the added fixed cost of using two short-haul trucks (for delivery from and to the intermodal hub) for each vehicle/container (rail, air or water). The variable transportation cost is also set to be dependent on the percentage of the distance covered by the active mode and the short-haul trucks.
- It is assumed that shippers have to use individual trucks (no milk runs) to directly deliver cargo to the receivers even if the cargo is LTL. This might not be too efficient thereby providing potential benefits to utilizing consolidation facilities between shippers and receivers, especially for long-distance trips. Return trips are not considered.
- The carbon emission from each trip is dependent on both the number of vehicles activated and the weight of cargo on each vehicle.
- Each manufacturing plant has a different emission rate for the production of different products. The total carbon emission due to production is the cumulation of the mathematical product of the quantity of each final product by its production emission rate from all plants.
- For carbon footprint assessment, the carbon-tax scheme is considered. The carbon emission excess of the carbon cap is penalized with a tax.
- Each manufacturing plant has multiple water-consuming processes. Some processes consume water on hourly or daily bases while others consume on basis of the quantity of production.
- The quantity of water impacted by ships through water pollution is also considered.
- The sum of the total quantity of water consumed in each plant above the stipulated regional cap and the quantity of water polluted through shipping is penalized with a tax.

This deterministic model SSCN/MFT&SCL is termed 'Model A' and it considers all possible SCL configurations simultaneously (see Figure 2 and Figure 5). These include:

a. Direct shipping from the supplier (plant) to the plant (retailer).
b. Shipping from the supplier (plant) to a consolidation facility where the package is consolidated with shipments from other suppliers (plants) before delivery to each plant (retailer).
c. Supplier (plant) consolidates shipments destined for multiple plants(retailers) locally and then ships out to a distribution facility where the shipment is broken down and sent to each plant (retailer).
d. Multiple deliveries from different suppliers (plants) to a consolidation facility where the shipments are consolidated and then shipped out to a distribution facility. The consolidated load is broken down and then sent to multiple plants (retailers).

4.2 Mathematical Model: SSCN/MFT&SCL (Model A)

The sets, parameters, decision variables and mixed-integer linear programming (MILP) model of the problem is presented below.

Table 2: Sets

$\boldsymbol{S}$	Suppliers, indexed by *i*
$\boldsymbol{M}$	Manufacturing plants, indexed by *j*
C	Retailers, indexed by *c*
Ce1	Consolidation facilities upstream of the SCN indexed by *e1*
De1	Distribution facilities upstream of the SCN indexed by *d1*
Ce2	Consolidation facilities downstream of the SCN, indexed by *e2*
De2	Distribution facilities downstream of the SCN, indexed by *d2*
$\boldsymbol{F}$	Final products, indexed by *f*
$\boldsymbol{K}$	Components, indexed by *k*
$\boldsymbol{P}$	All products: $\boldsymbol{K} \cup \boldsymbol{F}$, indexed by *p*

$\boldsymbol{B}$	Components *k* required for final products *f* represented as a pair, (*k, f*) for $k \in \boldsymbol{K}, f \in \boldsymbol{F}$
$\boldsymbol{T}$	All transportation modes, indexed by *t* where *t* = $\{t_{road}, t_{rail}, t_{air}, t_{water}\}$
$\overline{\overline{\boldsymbol{T}}}$	Air transportation mode i.e., $\overline{\overline{\boldsymbol{T}}} = \{t_{air}\}$
$\widehat{\boldsymbol{T}}$	Road transportation mode i.e., $\widehat{\boldsymbol{T}} = \{t_{road}\}$
$\boldsymbol{T}'$	All transportation modes except air transportation ; $t' = \{t_{road}, t_{rail}, t_{water}\}$
$\boldsymbol{R}$	Production processes that require daily water use, indexed by *r*
$\boldsymbol{Q}$	Production processes that require water on a per batch basis, indexed by *q*
$\boldsymbol{E1}$	Arcs in the upstream, $(a_u, b_u) \in \boldsymbol{E1} : i, j \in \boldsymbol{S} \cup \boldsymbol{M} \cup \boldsymbol{Ce1} \cup \boldsymbol{De1}$
$\boldsymbol{E2}$	Arcs in the downstream, $(a_d, b_d) \in \boldsymbol{E2} : i, j \in \boldsymbol{M} \cup \boldsymbol{C} \cup \boldsymbol{Ce2} \cup \boldsymbol{De2}$
$\boldsymbol{E}$	All Arcs in the SCN, $(a, b) \in \boldsymbol{E}$; $\boldsymbol{E} \doteq \boldsymbol{E}\,1 \cup \boldsymbol{E}\,2$

Table 3: Parameters

ld_i	Lead time for supplier $i \in S$ [days]
co_t	Variable transportation cost per km per tonne for transportation mode $t \in T$ [$]
f_t	Fixed cost of activating a vehicle/container for transportation mode $t \in \boldsymbol{T}$ where f_{air}=0 [$]
f^{road}	Fixed cost of activating a short haul truck [$]
dem_{cf}	The demand of retailer $c \in C$ for final product $f \in F$
d_{ab}	Distance of arc between node *a* and node *b* for *(a,b)* $\in E$ [km]
θ_{ab}	Percentage of the distance of arc requiring trucking between node *a* and node *b* for *(a,b)* $\in E$
p_{ik}^{k}	Unit purchase cost of component $k \in K$ from supplier $i \in S$ [$]
p_{jf}^{f}	Unit production cost of final product $f \in F$ from plant $j \in M$ [$]
cp	Consolidation/distribution handling cost per unit product [$]
a_i	Activation cost of supplier/plant $i \in S \cup M$ [$]
l_{ip}	Capacity (in units) of supplier/plant $i \in S \cup M$ to produce $p \in P$
α	Carbon tax of one unit (kg) of carbon emission above cap [$]

ϕ	Emission limit [kg]
w_p	Weight of product $p \in P$ [kg]
v_p	Volume of product $p \in P$ [m^3]
w_t^{cap}	Total weight capacity of a vehicle/container/airplane $t \in T$ [kg]
v_p^{cap}	Total volume capacity of a vehicle/container/airplane $t \in T$ [m^3]
w_p^b	Billable weight of product $p \in P$ [kg]
$\bar{e}_t$	Carbon emission per kilometre per weight (metric ton) for transportation mode $t \in T$ [kg/km-Ton]
e_{if}	Carbon emission (kg) of producing one unit of final product $f \in F$ at plant $i \in M$ [kg]
β	Charges for each unit of water usage above the cap (e.g recycling) [$]
b_{iqf}	Batch size of final product $f \in F$ in manufacturing plant $i \in M$ for process $q \in Q$
ω_i	Water limit based on scarcity level of the catchment area in which plant i is located [gal]
wa_{irf}^r	Quantity of water used per day for process $r \in R$ for final product $f \in F$ in manufacturing plant $i \in M$ [gal]
wa_{iqf}^q	Quantity of water used in process $q \in Q$ for each batch of product f in manufacturing plant $i \in M$ [gal]
wa_t	Quantity of water per hour polluted through transportation mode $t \in T$ [gal/hr]
pd_{if}	The average quantity of product $f \in F$ produced per day in the manufacturing plant $i \in M$
f_t^{CO2}	Fuel emissions factor [km/gal]
p_t^{CO2}	Packaging emission rate (while loading and unloading) for transportation mode $t \in T$ [kg/item]
s_t^{CO2}	Shipping emission factor for transportation mode $t \in T$ [kg/Ton-km]
m_t	Average fuel mileage for vehicle type $t \in T$ [km/gal]
θv_t	Target volume efficiency of truck space set by DM
θw_t	Target weight efficiency of truck space set by DM

γ_t	Packing efficiency for vehicle type *t* ∈ *T* [km/gal]
sp_t	Average speed of transportation mode *t* ∈ *T* [km/hr]
cf_t	Congestion factor of transportation mode *t* ∈ *T*
μ_t	0 if road transportation is activated, 1 otherwise
$portD_t$	Dwelling time spent at the intermodal hubs while switching modes *t* ∈ *T* [days].
cons_time	Time taken to consolidate or breakdown cargo at the consolidation/distribution center [days]
plan_period	Maximum time allowed for a full cycle from retailers to suppliers and back to the retailers [days]
M	A very big positive number
Sm	A very small positive number

Table 4: Decision Variables

X_{abtp}	The number of units of product *p* ∈ *P* transported from node *a* to node *b* using transportation mode *t* ∈ *T* for *(a,b)* ∈ *E*
Y_{abt}	The number of vehicles/containers/airplanes *t* ∈ *T* activated from node *a* to node *b* for *(a,b)* ∈ *E*
Z_{ip}	1 if supplier/plant *i* ∈ *S* ∪ *M* is activated for product *p* ∈ *P*, 0 otherwise
$Carb^{excess}$	the amount of excess emission above the emission limit
W_i^{excess}	The quantity of excess water usage above the stipulated limit in the region each plant *i* is located
W^{transp}	The total quantity of water polluted through transportation
K_{abt}	1 if the number of vehicles/containers/airplanes *t* ∈ *T* activated from node *a* to node *b* is greater than zero, 0 otherwise
$Time_{ab}$	The maximum travelling time spent between each pair of SC agents that have direct connection *[(S , M), (S , Ce1), (S , De1), (Ce1 , De1), (Ce1 , M), (De1 , M), (M , C), (M , Ce2), (M , De2), (Ce2 , De2), (Ce2 , C), (De2 , C)]*
$Upstream_{time}$	The maximum time consumed upstream of the SCN

$Downstream_{time}$	The maximum time consumed downstream of the SCN
$MaxProdTime$	The maximum production time used by all plants

Model A

- **Variable transportation cost**

$$EC_1 = \sum_{(a,b)\in E} \sum_{t' \in T'} \left(1-\theta_{ij}\right) d_{ab} co_{t'} \, Y_{abt'} + \sum_{(a,b)\in E} \sum_{t' \in T'} \theta_{ij} d_{ij} \, co_{t_{road}} Y_{abt'} \tag{4.1}$$

For multimodal transportation, the first term is the variable cost dependent on distance and the percentage θ_{ab} of the trip covered by rail, road or water. The second term is the variable cost of using road transportation for θ_{ab} percentage of each trip. Note that θ_{ab}=0 for arcs that use only road transportation.

- **Variable air transportation cost**

$$EC_2 = \sum_{p\in P} \sum_{(a,b)\in E} w_p^b co_{t_{air}} (1-\theta_{ab}) d_{ab} X_{abt_{air}p} + \sum_{(a,b)\in E} \theta_{ab} d_{ab} \, co_{t_{road}} Y_{abt_{air}} \tag{4.2}$$

The first term refers to the variable cost of air cargo dependent on the total billable weight of the cargo and the distance travelled. The second term refers to the variable cost for the 2 short road trips utilized to deliver goods from the start node to the airport and at the other end of the arc, from the airport to the end node.

- **Fixed transportation cost**

$$EC_3 = \sum_{(a,b)\in E} \sum_{t\in T} (f_t \; + 2\mu_t f^{road}) Y_{abt} \tag{4.3}$$

where μ=0 when t=t_{road} and 1 otherwise. This is the cost of activating a vehicle/container/aeroplane. The cost of activating 2 short haul trucks is added if the mode is rail, air or water. f_t of air transportation is equal to 0 as transportation cost is solely based on weight.

- **The purchasing cost of components**

$$EC_4 = \sum_{j \in M \cup Ce1 \cup De1} \sum_{i \in S} \sum_{t \in T} \sum_{k \in K} p_{ik}^{k} X_{ijtk} \tag{4.4}$$

This finds the mathematical product of the outputs from all Suppliers $i \in S$ (to manufacturing plants directly, or through the consolidation and distribution centers) and the cost of purchasing one unit of component *k* from supplier *i*.

- **The production cost of final products**

$$EC_5 = \sum_{i \in C \cup Ce2 \cup De2} \sum_{j \in M} \sum_{t \in T} \sum_{f \in F} p_{jf}^{f} X_{jitf} \tag{4.5}$$

This finds the product of the total output from all Manufacturing plants j$\in M$ (to retailers, consolidation and distribution centers) and the cost of producing one unit of each final product *f.*

- **Make-bulk / break-bulk cost**

$$EC_6 = cp\left(\sum_{e1 \in Ce1} \sum_{i \in S} \sum_{k \in K} X_{ie1t_{road}k} + \sum_{e2 \in Ce2} \sum_{j \in M} \sum_{f \in F} X_{je2t_{road}f} \right) \tag{4.6}$$

This is the sum of the mathematical product of the input to each consolidation centre and the unit cost of consolidation, *cp*

$$EC_7 = cp\left(\sum_{d1 \in De1} \sum_{j \in M} \sum_{k \in K} X_{d1jt_{road}k} + \sum_{d2 \in De2} \sum_{c \in C} \sum_{f \in F} X_{d2ct_{road}f} \right) \tag{4.7}$$

This is the sum of the mathematical product of the output from each distribution center and the unit cost of consolidation cp

- **Facility activation cost**

$$EC_8 = \sum_{i \in S \cup M} \sum_{p \in P} a_i Z_{ip} \tag{4.8}$$

The equation finds the total activation cost of all supplier and manufacturing plants of the supply chain utilized where Z_{ip} is a binary with 1 indicating the activation of the node, and 0 indicating otherwise. It is assumed that there is no activation cost for consolidation and distribution centers.

- **Carbon footprint cost**

$$CF = \propto Carb_{excess} \tag{4.9}$$

This refers to $Carb_{excess}$, the excess carbon emitted above the carbon cap stipulated by the government. An amount of money ∝ is paid per unit extra as tax.

- **Water footprint cost**

$$WF = \beta\left(\sum_{i \in M} W_i^{excess} + W^{transp} \right) \tag{4.10}$$

The first term refers to the accumulated excess water usage in each plant above the stipulated water usage cap in the region where plant *i* is located. The second term refers to the total water pollution/consumption for all transportation on all arcs while β refers to the cost paid per unit to recycle water or as charges.

The cost-minimizing objective function:

$$z = \text{ Economic cost} + \text{ Carbon footprint cost } + \text{Water footprint cost}$$

$$z = \underbrace{EC_1 + EC_2 + \cdots + EC_8}_{Economic\ Cost} + \underbrace{CF}_{Carbon\ Footprint\ Cost} + \underbrace{WF}_{Water\ Footprint\ Cost} \tag{4.11}$$

The objective function is to minimize the sum of the economic, carbon footprint (CF) and water footprint (WF) cost. The economic cost is divided into eight subcategories which are presented as Eqns. (4.1) to (4.8). The CF cost is presented as Eqn. (4.9) while the WF cost is presented as Eqn. (4.11).

Constraints:

- **Flow balance constraints in plants**

$$\sum_{i \in S}\sum_{t \in T} X_{ijtk} + \sum_{e1 \in Ce1}\sum_{t \in T} X_{e1jtk} + \sum_{d1 \in D1} X_{d1jt_{road}k} = \tag{4.12}$$

$$\sum_{(k,f) \in B}\sum_{t \in T}\sum_{c \in C} X_{jctf} + \sum_{(k,f) \in B}\sum_{t \in T}\sum_{d2 \in De2} X_{jd2tf} + \sum_{(k,f) \in B}\sum_{e2 \in Ce2} X_{je2t_{road}f}$$

$$\forall j \in M, k \in K$$

Constraint (4.12) ensures the input into each plant $j \in M$ from all nodes (supplier, consolidation, and distribution centers respectively) equals output from each plant through all nodes (retailer, consolidation and distribution centers respectively). *(k,f)* ∈ *B* ensures the components required for each final product according to the BoM is supplied. B is a set that contains the pair of final product *f* and the required components *k*.

- **Demand balance constraint**

$$dem_{cf} = \sum_{j \in M} \sum_{t \in T} X_{jctf} + \sum_{d2 \in De2} X_{d2ct_{road}f} + \sum_{e2 \in Ce2} \sum_{t \in T} X_{e2ctf} \quad \forall\, c \in C, \in F \tag{4.13}$$

Constraint (4.13) ensures the demand for each product f at retail store c is equal to the input from all relevant locations (consolidation, distribution center, plant) through all modes of transportation

- **Flow balance constraint for upstream consolidation centers**

$$\sum_{i \in S} X_{ie1t_{road}k} = \sum_{d1 \in De1} \sum_{t \in T} X_{e1d1tk} + \sum_{j \in M} \sum_{t \in T} X_{e1jtk} \qquad \forall\, e1 \in Ce1, k \in K \tag{4.14}$$

Constraint (4.14) guarantees that the sum of each component *k* received by each consolidation center *e1* from all suppliers is equal to the sum of each component *k* delivered to all distribution centers and plants.

- **Flow balance constraint for downstream consolidation centers**

$$\sum_{j \in M} X_{je2t_{road}f} = \sum_{d2 \in D2} \sum_{t \in T} X_{e2d2tf} + \sum_{c \in C} \sum_{t \in T} X_{e2ctf} \qquad \forall\, e2 \in Ce2, f \in F \tag{4.15}$$

Constraint (4.15) ensures the sum of each final product *f* received by each consolidation center *e2* from all plants is equal to the sum of each final product *f* delivered to all distribution centers and retailers.

- **Flow balance constraint for upstream distribution centers**

$$\sum_{i\in S}\sum_{t\in T} X_{id1tk} + \sum_{e1\in Ce1}\sum_{t\in T} X_{e1d1tk} = \sum_{j\in M} X_{d1jt_{road}k} \quad \forall\, d1 \in De1, k \in K \qquad (4.16)$$

Constraint (4.16) ensures the sum of each component *k* received by each distribution center *d1* from all suppliers and consolidation center is equal to the sum of components *k* delivered by the deconsolidation centre to all plants.

- **Flow balance constraint for downstream distribution centers**

$$\sum_{j\in M}\sum_{t\in T} X_{jd2tf} + \sum_{e2\in C2}\sum_{t\in T} X_{e2d2tf} = \sum_{c\in C} X_{d2ct_{road}f} \quad \forall\, d2 \in De2, f \in F \qquad (4.17)$$

Constraint (4.17) guarantees that the sum of each final product *f* received by each distribution center *d2* from all plants and consolidation centers is equal to the sum of the final product *f* delivered by the distribution centre to all retailers.

- **Capacity constraints for suppliers and plants**

$$\sum_{t\in T}\sum_{j\in M} X_{ijtk} + \sum_{t\in T}\sum_{d1\in De1} X_{id1tk} + \sum_{e1\in C1} X_{ie1t_{road}k} \leq l_{ik} Z_{ik} \quad \forall k \in K, i \in S \qquad (4.18)$$

$$\sum_{t\in T}\sum_{c\in C} X_{jctf} + \sum_{t\in T}\sum_{d2\in De2} X_{jd2tf} + \sum_{e2\in Ce2} X_{je2t_{road}f} \leq l_{(j+|S|)(f+|K|)} Z_{(j+|S|)(f+|K|)} \quad \forall\, f \in F, j \in M \qquad (4.19)$$

The sum of all the output from each supplier (plant) to the plants (retailers), consolidation center and distribution center should be less than or equal to the capacity of the supplier (plant). *i* and *p* are indexed differently in Constraint (4.19) because in l_{ip} and Z_{ip}, $i \in S \cup M$ and $p \in K \cup F$ respectively.

- **Vehicle/Container volume capacity constraints**

$$\sum_{p \in P} v_p X_{abtp} \leq \gamma_t v_t^{cap} \theta v_t Y_{abt} \qquad \forall (a,b) \in E \; t \in T \quad (4.20)$$

Constraint (4.20) ensures that the total volume of products transported on each arc does not exceed the total volume capacity of that arc due to the number of activated trucks, containers or airplane cargos. The constraint is adjusted to factor in the packing efficiency and target weights efficiency as explained in Eqn. (3.1).

- **Vehicle/Container weight capacity constraints**

$$\sum_{p \in P} w_p X_{abtp} \leq \gamma_t w_t^{cap} \theta w_t Y_{abt} \qquad \forall (a,b) \in E, t \in T \quad (4.21)$$

Constraint (4.21) ensures that the total weight of products transported on an arc does not exceed the total weight capacity of that arc due to the number of activated trucks, containers or airplanes. The constraint is adjusted to factor in the packing efficiency and target volume efficiency as explained in Eqn. (3.1).

- **Air transport billable weight**

$$w_p^b = \max\left(w_p, \frac{v_p}{0.06}\right) \quad \forall p \in P \qquad (4.22)$$

Eqn. (4.22) calculates the maximum of the product's weight and the product's volumetric weight to calculate its billable weight. This is only used for air transportation as the pricing practice for freight transportation through the air is calculated as the billable weight of the consignment multiplied by the distance travelled. The volumetric weight is calculated as the volume (m^3) of cargo divided by 0.06, a value provided by the United States Postal Services (Kabadurmus and Erdogan, 2020).

- **Production water usage constraints based on regional cap**

$$\omega_i + W_i^{excess} = \sum_{f\in F}\sum_{r\in R}\sum_{j\, \in C\cup Ce2\cup De2} wa_{irf}^{r}\frac{X_{ijtf}}{pd_{if}} + \sum_{t\in T}\sum_{f\in F}\sum_{q\in Q}\sum_{j\, \in C\cup Ce2\cup De2}\frac{wa_{iqf}^{q}X_{ijtf}}{b_{iqf}} \quad \forall\, i \in M \tag{4.23}$$

Eqn. (4.23) ensures that the surplus by which water usage in each plant exceeds the regional cap is captured as W_i^{excess} . This excess can then be summed for all plants. The first term indicates the total quantity of water used to perform each activity *r*$\in R$ for each product $f \in F$ in all the plants. Since the number of days (or hours) process *r* will run is required, the total quantity of products produced is divided by the average quantity of product f that plant *i* produces in a day, pd_{if}. It is assumed that all processes *r*$\in R$ have to run simultaneously.

The second term indicates the total quantity of water used to perform all activities *q* for each product f in all plants considering the total quantity of products that were produced and in what batch sizes. It is assumed that the better the technology in plant *i*, the less the water required (W_{iqf}) for each batch and/or the larger the batch size (B_{iqf}) for processes *q*$\in Q$ and/or the less the water required (W_{irf}) per day for processes *r*$\in R$.

- **Transportation-based water pollution**

$$W^{transp} = \sum_{(a,b)\in E} \sum_{t \in T} (wa_t D_{abt}/sp_t)\ Y_{abt} \tag{4.24}$$

Eqn. (4.24) indicates the total water pollution the SC network is accountable for. This is based on average shipping water consumption/pollution per hour for each transport mode (wa_t) for all transportation within the SC and how many containers/vehicles (Y_{abt}) were used in each trip.

- **Total carbon consumption constraints and equations**

$$Carb_{prod} = \sum_{i\in M} \sum_{j\in C\cup Ce2\cup De2} \sum_{t\in T} \sum_{f\in F} e_{if} X_{ijtf} \tag{4.25}$$

$$\begin{aligned} &Shipping\ emission \\ &= \sum_{(a,b)\in E} \sum_{t\in T} \frac{s_t^{CO2}}{1000} * \sum_{p\in P} (w_p\, X_{abtp}(1-\theta_{ab})\ d_{ab})/(1-cf) \\ &+ \sum_{(a,b)\in E} \frac{s_{t_{road}}^{CO2}}{1000} * \sum_{p\in P} (w_p\, X_{abt_{road}p} \theta_{ab}\ \ d_{ab})/(1-cf) \end{aligned} \tag{4.26}$$

$$\begin{aligned} &Empty\ haulage\ emission \\ &= \sum_{(a,b)\in E} \sum_{t\in T} \frac{f_t^{CO2}(1-\theta_{ab}) d_{ab} Y_{abt}}{m_t} + \sum_{(a,b)\in E} \frac{f_t^{CO2} \theta_{ab} d_{ab} Y_{abt_{road}}}{m_t} \end{aligned} \tag{4.27}$$

$$Packing\ emission = \sum_{(a,b)\in E} \sum_{t\in T} \sum_{p\in P} p_t^{CO2} X_{abtp} \tag{4.28}$$

$$Carb_T = \ shipping\ emission + \ empty\ haulage\ emission + packing\ emission \tag{4.29}$$

$$Carb_T + Carb_{prod} - Carb_{excess} = \phi \tag{4.30}$$

These extensive equations for calculating carbon emissions from transportation are adapted from (Ülkü, 2012). The shipping emission is the emission based on the weight of cargo being transported and the distance travelled. The number of vehicles is not factored in as the weight is distributed over the vehicles in use. The empty haulage emission is the amount of CO_2 emissions if the vehicles were travelling empty. It is dependent on the mileage m_t and total distance travelled by all vehicles. The packing emission is the amount of CO_2 emissions while packaging a cargo through activities such as loading and unloading of the vehicle. It is dependent on the number of items handled. The packing emission of modes other than the road is set to be higher. This is to factor in the emissions due to the extra loading and unloading in the transportation hub (e.g transfer of containers from truck to ship in the seaport).

- **Time constraints**

$$TIme_{SM} \geq K_{ijt}\left(ld_i + \frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}} + \frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24} + PortD_t \right) \qquad \forall\, i \in S, j \in M, t \in T \tag{4.31}$$

$$TIme_{SCe1} \geq K_{ijt}\left(ld_i + \frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}} + \frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24} + cons_{time} \right) \qquad \forall\, i \in S, j \in Ce1, t \in T \tag{4.32}$$

$$TIme_{SDe1} \geq K_{ijt}\left(ld_i + \frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}} + \frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24} + cons_{time} + PortD_t \right) \qquad \forall\, i \in S, j \in De1, t \in T \tag{4.33}$$

$$TIme_{Ce1M} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}} + \frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24} + PortD_t \right) \qquad \forall\, i \in Ce1, j \in M, t \in T \tag{4.34}$$

$$TIme_{Ce1De1} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}+PortD_t+cons_{time}\right) \quad \forall\, i \in Ce1, j \in De1, t \in T \tag{4.35}$$

$$TIme_{De1M} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}\right) \qquad \forall\, i \in De1, j \in M, t \in T \tag{4.36}$$

$$Upstream_{time} \geq [TIme_{SM}, TIme_{SCe1}+TIme_{Ce1M}, TIme_{SDe1}+TIme_{De1M}, TIme_{SCe1} + TIme_{Ce1De1}+TIme_{De1M}] \tag{4.37}$$

$$TIme_{MC} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}+PortD_t\right) \qquad \forall\, i \in C, j \in M, t \in T \tag{4.38}$$

$$TIme_{MCe2} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}+cons_{time}\right) \qquad \forall\, i \in M, j \in Ce2, t \in T \tag{4.39}$$

$$TIme_{MDe2} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}+cons_{time}+PortD_t\right) \quad \forall\, i \in M, j \in De2, t \in T \tag{4.40}$$

$$TIme_{Ce2De2} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}+cons_{time}+PortD_t\right) \quad \forall i \in Ce2, j \in De2, t \in T \tag{4.41}$$

$$TIme_{Ce2C} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}+PortD_t\right) \qquad \forall\, i \in Ce2, j \in C, t \in T \tag{4.42}$$

$$TIme_{De2C} \geq K_{ijt}\left(\frac{\frac{\theta_{ij}d_{ij}}{sp_{t_{road}}}+\frac{(1-\theta_{ij})d_{ij}}{sp_t}}{24}\right) \qquad \forall\, i \in De2, j \in C, t \in T \tag{4.43}$$

$$\begin{aligned} &Downstream_{time} \\ &\quad \geq [TIme_{MC}, TIme_{MCe2}+TIme_{Ce2C},\, TIme_{MDe2}+TIme_{De2C}, TIme_{MCe2} \\ &\quad + TIme_{Ce2De2}+TIme_{De2C}] \end{aligned} \tag{4.44}$$

$$MaxProdTime \;\geq \sum_{t\in T}\sum_{f\in F}\sum_{q\in Q}\sum_{j\,\in C\cup Ce2\cup De2}\frac{X_{ijtf}}{pd_{if}} \qquad \forall\, i \in M \tag{4.45}$$

$$K_{abt} \;\leq \frac{Y_{abt}}{Sm} \qquad \forall (a,b) \in E, t \in T \tag{4.46}$$

$$K_{abt} \;\geq \frac{Y_{abt}}{M} \qquad \forall (a,b) \in E, t \in T \tag{4.47}$$

$$Upstream_{time} + Downstream_{time} + MaxProdTime \;\leq plan\_period \tag{4.48}$$

The approximate time spent in the supply chain is equal to the maximum time expended in the downstream, upstream and maximum production time from all plants. To obtain the downstream (upstream) time, the maximum time spent on each of the four possible routes from the supplier (plant) to the plant (retailer) is considered in Eqn. (4.37) and Eqn. (4.44). The maximum production time is also obtained by dividing the number of products to be produced from each plant by the production rate per day. The maximum of all production times is then recorded. The upstream time, downstream time and production time are to all be completed within the allotted planning period (4.48). For routes that include the Supplier, the supplier lead time is added to the time consumed. For the routes that have multimodal transportation capability, dwelling time spent at the port of exchange (*PortD*) is also added. This value is equated to 0 for road transportation. For routes that go through consolidation or distribution centers, an extra time *const_time* is added to reflect the time spent consolidating or breaking down the cargo. Constraints (4.46) and (4.47) ensure that only transportation modes that have vehicles/containers in use on each arc are considered in the calculation of the total time spent in the SCN.

In Eqn. (4.31), $TIme_{SM}$ represents the maximum time spent travelling from the suppliers to manufacturers. K_{ijt} ensures that only the arcs that have at least one vehicle are considered. For each of these arcs, the lead time of the supplier ld_i is added to the travel time (in days) used by the required transportation mode. The time spent in the dwelling ports $(PortD_t)$ is then added. The procedure is repeated for all possible connections in the SCN from Eqn. (4.32) to Eqn. (4.43). Consolidation time *(cons_time)* is added for arcs that end at the consolidation/distribution facilities, dwelling time $(PortD_t)$ is added for arcs that go through intermodal hubs and supplier lead time (ld_i) is added for arcs that start at the suppliers.

- **Boundary and integrality constraints**

$$X_{abtp}, Y_{abt} \geq 0, integer\ ;\ Z_{ip}, K_{ijt} \in \{0,1\} \qquad 4.49$$

$$Carb^{excess}, W_i^{excess}, W^{transp}, upstream_{time}, downstream_{time}, MaxProdTime \geq 0 \qquad 4.50$$

The model is a mixed integer-linear model because of the presence of integer variables such as X_{ijtp} and Y_{ijt}. All the other variables are either continuous or binary. The total number of variables and constraints (exempting the boundary and integrality constraints) in the model can be described in terms of the set sizes.

Number of variables= $(2 + |P|)A^2|T| + |S + M| * |P| + |M| + 17$

Number of constraints = $4A^2|T| + 7 + 2|M| + |P| + |T| (|De2||C| + |Ce2||C| + |Ce2||De2| + |M||De2| + |M||Ce2| + |C||M| + |De1||M| + |Ce1||De1| + |Ce1||M| + |S||De1| + |S||Ce1| + |S||M|) + |M||K| + |C||F| + |Ce1||K| + |Ce2||F| + |De1||K| + |De2||F| + |K||S| + |F||M|$

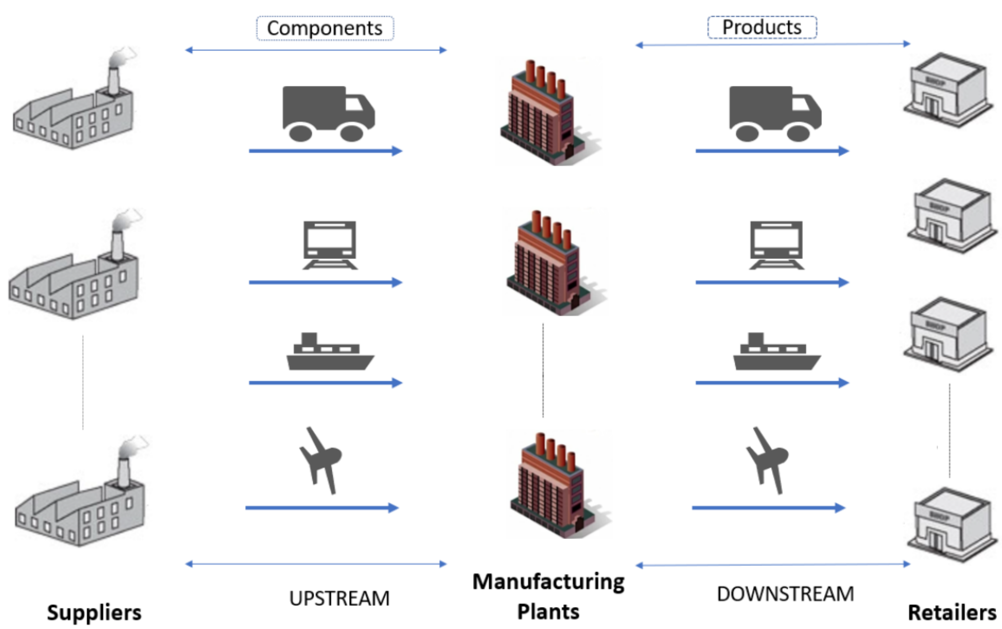
where $A = |S + M + Ce1 + Ce2 + De1 + De2 + C|$

As an example, for an SCN with |S|=|M|=|C|=|4|, |Ce1|=|Ce2|=|De1|=|De2|=1, |T|=4, |K|=6, |F|=4 and |P|=10, the number of variables equals 12,389 and the number of constraints equals 4,485.

4.3 Mathematical Model: SSCN/MFT without SCL (Model B)

A second model is defined with the objective of using it to evaluate the value of incorporating SCL within a supply chain network. This was done by comparing the SSCN/MFT&SC model (termed Model A) and an alternative that excludes the SCL option (termed Model B). Without SCL, only direct delivery from shippers to receivers is allowed (See Figure 6). All other assumptions indicated in the description of Model A are also applicable to Model B.

Figure 6. Integral Backbone of the SC network



Model B

Sets, parameters and decision variables are the same as provided in Table 2, Table 3 and Table 4 respectively.

The cost-minimizing objective function:

$$\mathfrak{z} = \text{Economic cost} + \text{Carbon footprint cost} + \text{Water footprint cost}$$

$$\mathfrak{z} = \underbrace{EC_1 + EC_2 + \cdots + EC_5 + EC_8}_{Economic\ Cost} + \underbrace{CF}_{Carbon\ Footprint\ Cost} + \underbrace{WF}_{Water\ Footprint\ Cost} \tag{4.51}$$

The objective function is to minimize the sum of the economic, carbon footprint (CF) and water footprint (WF) cost. The economic cost is divided into six subcategories which are obtained from Eqns. (4.1) to (4.5) and Eqn. (4.8). This is similar to Model A. However, Make-bulk / break-bulk costs (EC_6 and EC_7) are removed. The CF cost is obtained from Eqn. (4.9) while the WF cost is obtained from Eqn. (4.10).

Constraints:

Constraints (4.12) – (4.50) and Constraint (4.52) provided below.

- **Consolidation/ distribution center elimination constraints**

$$Ce1 = 0, Ce2 = 0, De1 = 0, De2 = 0 \tag{4.52}$$

Constraint (4.52) provides an easy modification of the constraints of Model A to form Model B. It removes all the consolidation/distribution facilities and enforces only the option of direct directly from shippers to receivers.

4.4 Summary

This chapter begins by describing the proposed SCN with multimodal freight transportation and shipment consolidation. This model was termed Model A. Details about the notations were then stated. This included the sets, parameters and decision variables. The full mathematical model was then provided stating the objective function and constraints. The size of the model in terms of variables and constraints was then briefly discussed. Next, Model B, a version of Model A without shipment consolidation was discussed with its notations, objective function, and constraints presented.

Chapter 5

Approaches for the Stochastic Version of Model A

This chapter discusses the various aspects required for consideration when optimizing a supply chain network design problem having stochastic parameters. Models A in the previous section is deterministic. However, we consider its (Model A) stochastic version, where one of its parameters is uncertain. Different approaches to optimizing stochastic models are thereby explored. The approaches considered in this study are generally classified as:

- Stochastic programming approaches (SP)
- Machine Learning approaches (ML)

For the SP approaches, model reformulations (e.g simple recourse) are proposed. Model A is reformulated to factor in either multiple discrete scenarios or the distribution of the stochastic parameter. The objective function and constraints that have the stochastic parameter are also modified. For each of the ML approaches, historical data of the stochastic parameter is fed into models as training data, after which the trained model is used to predict the future values of the parameter. These predictions can then be fitted into Model A and solved deterministically.

In this study, we consider the demand parameter (dem_{cf}) to be the stochastic parameter. To comparatively test the efficiency of each of the approaches mentioned above, real demand data is obtained, and comparative experiments are performed.

In the remainder of this chapter, we go into more detail on the SP (it's model reformulations) and ML approaches considered. We then discuss the experiments to be carried out and the evaluation metrics to utilize. For these experiments, details on obtaining the values of the network model parameters are explained and the sourcing of real data for the stochastic parameter (demand) is discussed.

5.1 Stochastic Programming Approaches

Several Stochastic Programming (SP) approaches have been proposed in the literature including methods such as the simple recourse, chance-constrained programming, Conditional Variance at Risk (CVaR), etc. However, the optimization model has to be reformulated for each method. The demand balance constraint (4.12) of the model is the constraint having a stochastic parameter. The optimal solution is sensitive to the value of the uncertain parameter. Due to uncertainty, we can not find a solution optimal for all scenarios. We seek to balance the optimal solution among all scenarios. We source discrete scenarios from the data points and assume each scenario has equal probabilities. For risk-averse approaches, we seek a solution that minimizes the objective function while considering as many scenarios as possible and fulfilling the risk level set by the DM.

5.1.1 Simple Recourse SP: Model Reformulation (Model A1)

This is a reformulation of Model A which we term Model A1. The demand parameter is represented as a finite set of discrete scenarios to be obtained from historical data. Considering that the model will become infeasible if some of the scenarios are not satisfied, we add a surplus/shortage penalty for each unsatisfied scenario to maintain feasibility.

Model A1

Sets, parameters and decision variables include those provided in Table 2, Table 3 and Table 4 respectively. The parameter dem_{cf} is replaced with dem_{cfs}. To formulate the simple recourse model, it is necessary to consider some extra notations.

Set

Sc	Set of scenarios, indexed by s

Parameters

b_{if}	Understocking cost for final product *f* in plant *i*
h_{if}	Overstocking cost for final product *f* in plant *i*
$prob_s$	Probability of scenario *s*

Variables

z_{cfs}^{+}	Unsatisfied demand for retailer $c \in C$ for product , $f \in F$ in scenario $s \in Sc$
z_{cfs}^{-}	Surplus above demand for retailer r $c \in C$ for product, $f \in F$ in scenario $s \in Sc$

- **Surplus/Shortage cost**

$$SS = \sum_{s \in Sc} prob_s \sum_{c \in C} \sum_{f \in F} (b_f \, z_{cfs}^{+} + h_f \, z_{cfs}^{-}) \tag{5.1}$$

The cost-minimizing objective function:

$$\text{z} = \text{Economic cost} + \text{Carbon footprint cost} + \text{Water footprint cost} + \text{Surplus/Shortage cost}$$

$$\text{z} = \underbrace{EC_1 + EC_2 + \cdots + EC_8}_{Economic\ Cost} + \underbrace{CF}_{Carbon\ Footprint\ Cost} + \underbrace{WF}_{Water\ Footprint\ Cost} + \underbrace{SS}_{Surplus/Shortage\ Cost} \tag{5.2}$$

The objective function is to minimize the sum of the economic, carbon footprint (CF), water footprint (WF) cost and Surplus/Shortage (SS) cost. The economic cost is divided into eight subcategories which are obtained from Eqns. (4.1) to (4.8). The CF cost is obtained from Eqn. (4.9) while the WF cost is obtained from Eqn. (4.10). This is similar to Model A. However, the cost incurred due to the deviation of the total supply to the retailers (manufacturing plant X_{jc}, consolidation X_{e1c} and distribution center X_{d1c}) from the demand scenarios is added as Eqn. (5.1). This is to enforce the model to choose the best values for the supply that would lead to the least cost based on the deviation from all the scenarios under consideration.

***Constraints**:*

Constraints (4.12), (4.14) – (4.50) and Constraints (5.3) – (5.4) described below.

- **Demand balance constraint**

$$dem_{cfs} = \sum_{j \in M} \sum_{t \in T} X_{jctf} + \sum_{d2 \in De2} X_{d2ct_{road}f} + \sum_{e2 \in Ce2} \sum_{t \in T} X_{e2ctf} + z_{cfs}^{+} - z_{cfs}^{-} \ \forall \, c \in C, f \in F, s \in Sc \tag{5.3}$$

$$z_{cfs}^{-}\ ,\ z_{cfs}^{-} \geq 0 \tag{5.4}$$

Constraint (5.3) captures the difference between the input from all sources (manufacturing plant, consolidation and distribution center) to each retailer c for each final product f for each scenario s as either surplus (z_{cfs}^{-}) or shortage (z_{cfs}^{+}). These are then penalized in the objective function. In comparison to Model A, all constraints are used in Model A1 except for Constraint (4.13) which is replaced by Constraints (5.3) and (5.4).

5.1.2 Chance-constrained SP: Model Reformulation (Model A2)

In this reformulation of Model A, the decision maker sets a risk value α^{risk} which indicates how many of the scenarios the model factors in when optimizing. This is useful when there might be outliers within the scenarios. The model ignores the scenarios that are likely to increase the overall cost as much as α^{risk} permits. The demand constraint is an equality constraint so it was expanded into 2 constraints ('≤ ' and ' ≥'). The model is termed Model A2.

Model A2

Sets, parameters and decision variables include those provided in Table 2, Table 3 and Table 4 respectively. The parameter dem_{cf} is replaced with dem_{cfs}. To formulate the chance-constrained model, it is necessary to consider some extra notations.

Set

Sc	Set of scenarios, indexed by s

Parameters

b_{if}	Understocking/Backordering cost for final product f in plant i
h_{if}	Overstocking cost for final product f in plant
$prob_s$	Probability of scenario s
α^{risk}	Probability of violating the chance constraint

Variables

z^{+}_{cfs}	Unsatisfied demand for retailer $c \in C$ for product , $f \in F$ in scenario $s \in Sc$
z^{-}_{cfs}	Surplus above demand for retailer $c \in C$ for product, $f \in F$ in scenario $s \in Sc$
N_s	1 if scenario s is fulfilled, 0 otherwise

The cost-minimizing objective function:

$$\mathbb{z} = \text{Economic cost} + \text{Carbon footprint cost} + \text{Water footprint cost} + \text{Surplus/Shortage cost}$$

$$\mathbb{z} = \underbrace{EC_1 + EC_2 + \cdots + EC_8}_{Economic\ Cost} + \underbrace{CF}_{Carbon\ Footprint\ Cost} + \underbrace{WF}_{Water\ Footprint\ Cost} + \underbrace{SS}_{Surplus/Shortage\ Cost} \tag{5.5}$$

The objective function is the same as that of Model A1 with the goal of finding the best values for the decision variables that minimize the sum of the economic, carbon footprint (CF) and water footprint (WF) costs. This is similar to Model A. However, deviations in the supply from all sources (manufacturing plant X_{jc}, consolidation X_{e2c} and distribution center X_{d2c}) to each retailer *c* for each final product *f* for each scenario *s* is penalised as the Surplus/Shortage (SS) cost [Eqn. (5.1)].

Constraints:

Constraints (4.12), (4.14) – (4.50) and Constraints (5.7) – (5.10) provided below.

- **Demand balance constraint**

The demand balance constraint(4.13) can be changed to a chance constraint

$$\Pr\left(dem_{cf} = \sum_{j \in M} \sum_{t \in T} X_{jctf} + \sum_{d2 \in De2} X_{d2ct_{road}f} + \sum_{e2 \in Ce2} \sum_{t \in T} X_{e2ctf} \right) \geq 1 - \alpha^{risk} \quad \forall\, c \in C, \in F \tag{5.6}$$

Constraint (5.6) can then be reformulated as Constraints (5.7) to (5.10) below.

$$dem_{cfs} + MN_s \geq \sum_{j\in M}\sum_{t\in T} X_{jctf} + \sum_{d2\in De2} X_{d2ct_{road}f} + \sum_{e2\in Ce2}\sum_{t\in T} X_{e2ctf} + z^{+}_{cfs} - z^{-}_{cfs} \quad \forall\, c \in C, f \in F, s \in Sc \tag{5.7}$$

$$dem_{cfs} - MN_s \leq \sum_{j\in M}\sum_{t\in T} X_{jctf} + \sum_{d2\in De2} X_{d2ct_{road}f} + \sum_{e2\in Ce2}\sum_{t\in T} X_{e2ctf} + z^{+}_{cfs} - z^{-}_{cfs} \ \forall\, c \in C, f \in F, s \in Sc \tag{5.8}$$

$$\sum_{s\in Sc} N_s \leq \alpha^{risk} \, |Sc| \tag{5.9}$$

$$z^{+}_{cfs} \, , \, z^{-}_{cfs} \geq 0, \ N_s \in \{0,1\} \tag{5.10}$$

Due to the equality in the chance constraint (5.6), it can be broken down into two constraints with '≥' and '≤'. N_s as a binary variable basically counts the number of times the Constraints (5.7) and (5.8) are violated and Constraint (5.9) ensures the average of the variable N_s for all scenarios is less than the set α^{risk}. Constraint (5.10) ensures the binariness of N_s and sets a boundary for the surplus and shortage variables (z^{+}_{cfs} , z^{-}_{cfs}).

5.2 Machine Learning Approaches

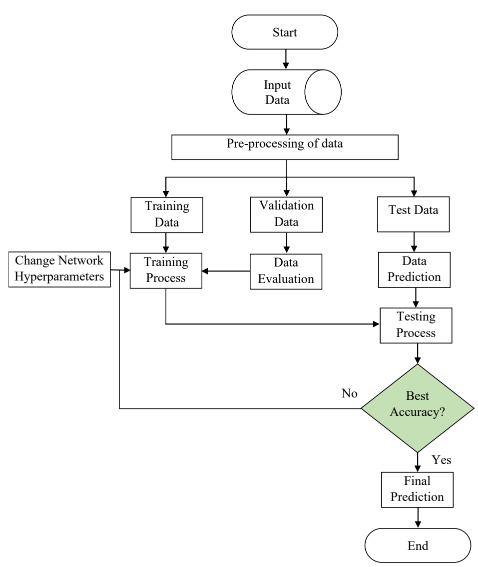
To integrate Machine Learning (ML) approaches into stochastic optimization modelling, the procedure used in this thesis is to utilize historical data in predicting the value of the stochastic parameter in the future. This value can then be used in Model A and solved deterministically. The demand parameter (dem_{cf}) is considered stochastic with univariate time series historical data. A time series is a set of observations recorded at equally spaced intervals, and a univariate time series is a time series with a single observation recorded at each timestamp. The demand data we have fits this description. The ML models explored are the three best-performing models by Okwuchi et al., (2020) who compared traditional machine learning models with deep learning models in predicting the price and yield of perishable goods. These three are the attention-based compound deep learning models named Attention Convolutional

Neural Network-LSTM (AC-LSTM), Attention Convolutional-LSTM (ACV-LSTM) and a stacking ensemble of both using Support Vector Regressor (Ensemble SVR). For each model, the training data was used to build the model as well as evaluate its performance. After adjusting its (the model) hyperparameters to obtain the best performances, the model is used to predict the future demand data points required.

The general procedure of building an ML model (especially for supervised learning) involves the following steps [as shown in Figure 7 (based on Vrskova et al., 2022)]:

1. Choosing the most appropriate model considering the data at hand and the objective.
2. Preprocessing the data and adjusting its shape to that required by the chosen model.
3. Dividing the data into training, validation and testing data.
4. The training data is used as input to train and build the ML model while the performance of the model on the validation data serves as continuous feedback to improve its parameters. Finally, the testing data is the unseen data used at the end of the process to confirm how good the built model is.
5. If the performance on the testing data is poor, the network hyperparameters are adjusted and the training process started again.

Figure 7. Flowchart of ML procedures

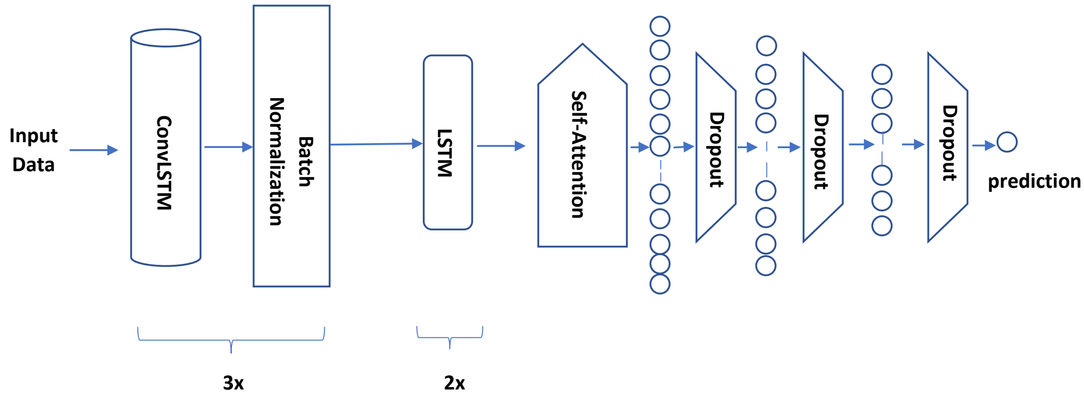


5.2.1 Models

The following are the ML architectures employed in this study.

1. Attention ConvLSTM (ACV-LSTM): After preliminary trials, an architecture comprising three repeated layers of 2D ConvLSTM each with a kernel size of (1,3), and 64 filters is built. Each layer was followed by a batch normalization layer. The output was then mapped into a self-attention layer with sigmoid activation. This is then followed by 4 dense layers with the number of neurons set as 64,32,16,1 times the number of output respectively. Dropout layers with a rate of 0.1 are put between consecutive dense layers. This is described in Figure 8. Rectified linear activation function (Relu activation) is used on each layer (except the self-attention layer), the loss function is set as 'mean squared error' and the optimizer is set as 'Adam'.

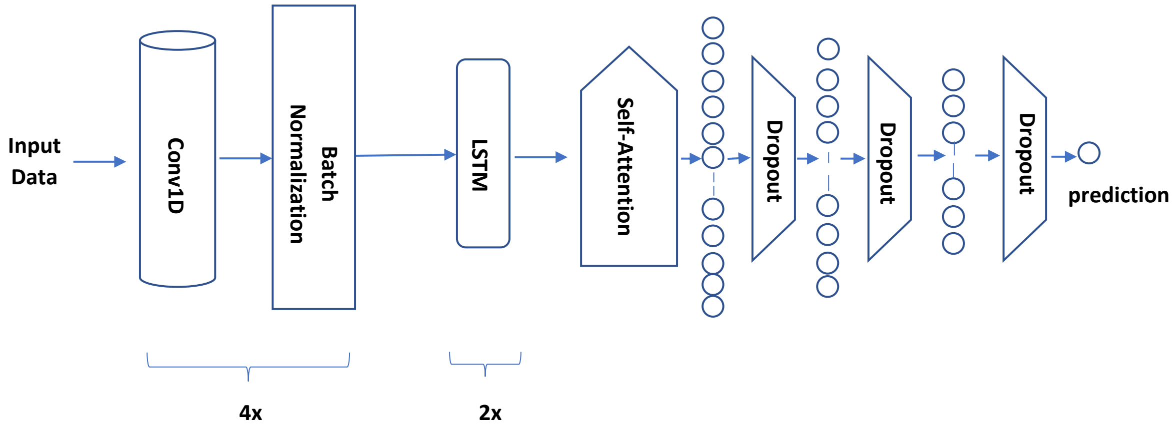
Figure 8. ACV-LSTM Architecture



2. Attention CNN-LSTM (AC-LSTM): After preliminary trials, an architecture comprising of four repeated layers of 1D convolution with 120 filters, kernel size of 3 and stride of 1 is built. Each layer is followed by a batch normalization layer. The output was then mapped into 2 consecutive LSTM layers with 100 units and then into an additive self-attention layer with sigmoid activation. This is then followed by 4 dense layers with the number of neurons set as 64,32,16,1 times the number of output respectively. Dropout layers with a rate of 0.1 are put between consecutive dense layers. This is described in Figure 9. Rectified linear activation

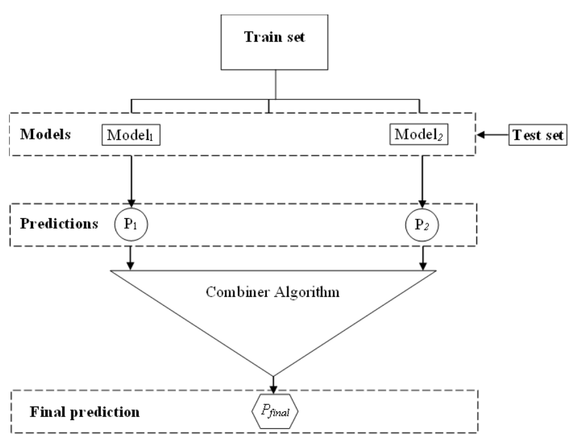
function (Relu activation) is used on each layer (except the self-attention layer), the loss function is set as ‘mean squared error’ and the optimizer is set as ‘Adam’.

Figure 9. AC-LSTM Architecture



3. Stacking Ensemble using SVR: The two models, AC-LSTM and ACV-LSTM are used to create an ensemble. Support vector regression is used as a stacking algorithm. First, both algorithms are individually trained on the data, and then SVR is used as a combiner algorithm to make a final prediction using the prediction from AC-LSTM and ACV-LSTM as additional inputs (See Figure 10). Stacking algorithms typically yield better performances than any of the child models it is built on (Wolpert, 1992).

Figure 10. Ensemble Architecture



5.2.2 ML Performance Metrics

The most frequently used performance measures in literature are the Mean absolute percentage error (MAPE), Mean squared error (MSE), Mean absolute error (MAE) and Root mean square error (RMSE) (Hyndman & Koehler, 2006; Shukla & Jharkharia, 2013).

Mean absolute error is a common metric used for evaluating regression models. It computes the absolute difference between the predicted and actual values. The total of the absolute error for all samples is summed up and the average is calculated. MAE is less impacted by outliers than other measures such as MSE and RMSE.

$$MAE = \frac{\sum_{i=1}^{N} |Pred - Act|}{N} \tag{5.11}$$

where $|a|$ is the absolute value of *a.*

A special metric is developed which takes the ratio of overstocking to understocking cost (OUC) into consideration and uses them as weights to penalize positive and negative deviations. This metric is used as a custom metric within the ML models to ensure the model obtains the best-performing weights and predictions considering our objective (reducing overstocking/understocking cost). This special metric is called the MOU (Mean Overstocking/Understocking cost).

$$MOU = \frac{\sum_{i=1}^{N} (Un\ [Act - Pred]_{+} + Ov\ [Pred - Act]_{+})}{N} \tag{5.12}$$

where $[a]_{+}$is the maximum of a and 0, *Un* is the understocking cost, *Ov* is the overstocking cost and *N* is the number of predictions being compared.

5.2.3 Model Tuning

There are a lot of hyper-parameters in deep learning modelling that can be used to control the behaviour of the algorithm. Hyper-parameters tuning thereby becomes necessary to find the combination that gives the best performance. The model has to avoid underfitting or overfitting the training data. Underfitting refers to when the model is not flexible enough to learn the training data and would normally not generalize well on new testing data. Overfitting refers to when a model is too closely aligned with the training data by learning its details and noise to the extent that it negatively influences its performance on new testing data. How well the model performs on previously unseen inputs is called generalization. The generalization error of a model is measured as its performance on test data is separate from the data it was trained with (Goodfellow and Courville, 2016).

Hyper-parameters such as the learning rate determine the size of the steps the optimization algorithms take toward finding the global minimum. The larger the value, the more the chances of non-convergence. On the other hand, the lower the value, the more the chances of the optimizer getting stuck in a local minimum. To choose the best learning rate, a scheduler is used to continually adjust the learning rate value after every epoch.

Model checkpointing is also used to save the weights of the model at the epoch where the model obtains the least validation error. This helps to avoid inaccuracies caused by overfitting which might be due to having too many epochs.

Other parameters such as the model size, the number of layers, optimizer algorithm, regularization parameter and error functions were determined through manual iterations

5.3 Numerical Experimentations

In analysing all the models already defined, three individual experiments are to be performed. These experiments and the logic behind them are outlined below.

1. Comparative Analysis of Models A and B: The integral backbone of the three-echelon SCN was shown in Figure 1. Model A however includes consolidation and distribution centers which are not integral parts of the network. Model B provides an alternative configuration to Model A by restricting delivery options to only direct delivery from shipper to the receiver. Constraint (4.52) ensures this by excluding the consolidation/distribution facilities forcing the model to take only direct routes between shippers and receivers. This also means that in the network of

Model B, shippers might have to use full trucks even if they want to ship LTL over long distances.

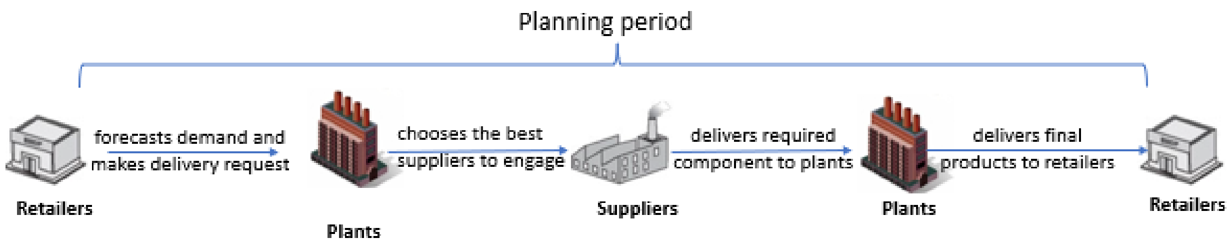
In this experiment, the objective is to observe if the presence of the consolidation/distribution facilities enables economic and/or environmental cost savings or not. Metrics such as the average distance between the integral SCN facilities (suppliers, manufacturing plants and retailers) and the average vehicle capacity usage are varied and the cost of both models are compared.

The average distance is varied to test the relationship between the advantageousness of SCL and the distance to travel between shippers and receivers. The average vehicle capacity usage is varied to evaluate the relationship between the average quantity of outbound cargo from each shipper and the potential cost-saving benefits of using SCL within the SCN. The demand is considered deterministic for this experiment.

The second and third experiments are carried out to examine different approaches to optimizing decision-making when some of the relevant parameters of the SSCN/MFT&SC model are uncertain. For these experiments, the demand is considered stochastic. Data-driven stochastic programming and machine learning approaches are thereby employed to optimize the stochastic model. Real demand data is sourced online, extracted and preprocessed to the format required. For each of the SP approaches, the SC model is reformulated appropriately. This gives new models which we term Models A1 and A2. For each of the ML approaches, an algorithm is built and used to predict the demand which is then used to solve Model A deterministically. Solutions obtained from all the approaches (SP & ML) are then compared with the deterministic solution obtained from using the actual demand assuming perfect knowledge. This helps us evaluate the expected value of perfect information (EVPI) of each approach. The solutions are also compared to that of using the expected value of the demand parameter from historical data. This serves as a lower baseline for comparing how good or bad how approaches are in comparison to simply using the average historical value of the stochastic parameter.

In the sequence of logistical operations in the SCN, the retailers make their order requests and start the cycle. The time spent in the SCN from the time of ordering by the retailers to the final delivery of the requested final products must be less than or equal to the set planning period (See Figure 11). Experiments 2 and 3 are based on these details.

Figure 11. Sequence of Operations



2. Stochastic Model with a planning period of 28-days: This considers a production planning period of 4-weeks (28 days) wherein the demand is forecasted leading to the supply of required components. The sum of the delivery time from suppliers to the plant, supplier lead time, dwelling time at intermodal facilities, consolidation time, production time within the plants and the delivery time of final products from plants to retailers thereby have to be less than 28 days. In real life, this might apply to products that have short expiry dates or quickly run out of fashion. Considering the short period (relative to Experiment 3), the SC might not be able to take advantage of some of the cheaper and more sustainable transportation modes (such as rail or water) (Carbonfund, 2022). For example, sea shipping typically takes 20-45 days or more (Freightos, 2022). Alternatively, the demand quantity might also be considerably large requiring a processing period of more than 28 days. These limitations are considered in the third experiment. The performance of the SP & ML approaches are compared over a planning period of 28-days.

3. Stochastic Model with a planning period of 3-months: The second experiment considers a production planning period of 3 months (84 days) wherein the forecasted demand dictates the number of components demanded from suppliers. Once again, the total sum of delivery time from suppliers to plants, supplier lead time, dwelling time at intermodal facilities, consolidation time, production time within plants and delivery time of final products from plants to retailers

have to be less than 3 months. The longer planning period allows for the optimization model to consider even slower modes of transportation, while also testing the robustness and limit of the approaches under consideration.

5.4 Overall Evaluation Metrics

For Experiment 1, to evaluate the efficiency of Models A and B, their economic and environmental costs are compared while varying the average vehicle capacity usage ($AVCU$) and the *average_distance* parameters. The demand is considered deterministic for this experiment.

$AVCU$ represents the amount of cargo an average supplier has to ship out to meet the final demand from the retailers in terms of the number of vehicles/containers. $AVCU$ is dependent on the total demand for final products from the retailers, the average number of components required to produce each final product, the average weight/volume of these components, the average weight/volume capacity of a vehicle/container, and the number of suppliers required to fulfil the total demand.

$$AVCU = \frac{(total\ demand)(components\ per\ final\ product)K_{w/v}}{|S| vehicle_{cap}} \tag{5.13}$$

The numerator indicates the total weight/volume of components that are required from all suppliers based on the final product demand from all the retailers, average component weight/volume and number of components required for each final product. The denominator indicates the number of suppliers and the average vehicle weight/volume. $AVCU$ thereby represents the average vehicle capacity usage for each supplier. The closer $AVCU$ gets to 1, the more the average percentage of a truck occupied by outgoing cargo. *AVCU* below 1 indicates that an average supplier has to ship out less than FTL cargo while values higher than 1 indicate that the supplier might have to use a mixture of FTL and LTL to meet its requests. The higher the *AVCU,* the more the total demand and number of trucks the average supplier requires.

Eqn. (5.13) can be reformulated to make demand the subject of the formula. Total demand can also be represented as the mean demand of all the retailers *C* for all final products *F*. Eqn. (5.15) allows us to vary the $AVCU$ to obtain different mean demand values for experimentation. This is required because the demand is an actual parameter in the model formulation.

$$total\ demand = mean\ demand * |C||F| \tag{5.14}$$

$$mean\ demand = \frac{vehicle_{cap} * AVCU * |S|}{\#\ of\ components\ per\ final\ product * K_{w/v} * |C| * |F|} \tag{5.15}$$

where *S* is the number of suppliers engaged, $vehicle_{cap}$ is the maximum capacity of the vehicle, and *components per final product* is the average number of components required for the production of one final product according to the BoM. $K_{w/v}$ is the average weight/volume of each component.

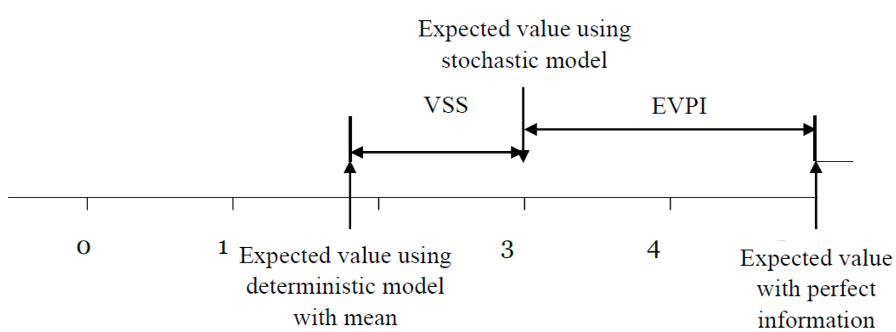
We also evaluate the impact of the average distance between shippers and receivers on the viability of SCL. The locations of the facilities are set to be dependent on the parameter *average_distance* as indicated in Table 4. This allows for the comparison of the total SC cost (economic impact) and the total distance travelled (environmental impact) for Model A and B when the facilities are varying distances apart. Both evaluations metric ($AVCU$ and *average_distance*) serve a function in analysing the settings or conditions under which SCL is advantageous and when it does not provide clear advantages.

For Experiments 2 and 3, the evaluation metrics utilized to compare the SP and ML approaches were the Value of Stochastic Solution (VSS) and Expected Value of Perfect Information (EVPI). For a stochastic optimization model, if we replace the stochastic parameter with its expected value (such as its mean from historical data), and solve the model deterministically, we obtain a simpler problem (Birge, 1997). The solution for the model with mean values could serve as a baseline to compare other approaches to handling stochastic models. VSS is a measure of how good or bad the solutions obtained from these other approaches are in comparison to the model with mean values. Also, the actual future value of the stochastic parameter, assuming perfect knowledge, can be used to solve the model deterministically. This serves as another baseline which can also be used to compare the other approaches to handling stochastic models. The closer the solution from these approaches is to that of the model with perfect knowledge, the lower the EVPI, and vice versa.

In this study, the solution for a specific solution approach (either SP or ML) is considered as the total cost which is calculated as the sum of the model's optimization cost and the overstocking/understocking cost (OUC) obtained from comparing the quantities of the final products provided to the actual demand. As an example, if our approach instructs the SCN to provide 15 final products and the actual demand is 10, the total cost of the approach is the SCN cost incurred due to the transportation, production, and so on, of 15 products plus the overstocking cost of 5 units. On the other hand, if it instructs the SCN to provide 5 final products, the total cost is equal to the cost of providing the 5 units and the understocking cost of 5 units.

We intend to obtain both the VSS and EVPI simultaneously. The lower baseline is set as the solution obtained from solving Model A deterministically with the average of all the past demand scenarios. This is termed the 'Model with mean values'. The higher baseline is set as the solution obtained from solving Model A deterministically with the actual future demand assuming perfect knowledge. This has no overstocking or understocking cost. The solution from each of the approaches (SP from Model A1&A2 and ML from Model A) is then positioned between the 'Model with mean value' and the model with 'Actual demand' [See Figure 12 (based on Birge 1997)].

Figure 12. VSS vs Solution vs EVPI



For the SP approaches, the reformulated models A1 and A2 are solved using the scenarios generated from the data. The final value of the total supply of final products to each retailer [see Constraint (5.3) for Model A1 and Constraint (5.6) for Model A2] is taken as the 'prescriptions' of the model. The base Model A is then solved deterministically with these prescriptions to obtain the SC cost. The OUC is also calculated using the surplus/shortages (difference between actual demand and prescribed). The SC cost is added to the OUC as the final cost of the SP approach.

For the ML approaches, almost similar procedures are taken. After the ML model has been used to predict the future demand, the predicted demand is used to solve the optimization Model A1 deterministically to obtain the SC cost. The surplus/shortages are then used to calculate the OUC. The SC cost and the OUC are then added to get the final cost of each ML approach. The VSS and EVPI are used as evaluation metrics to compare both approaches. The formula below is used to calculate the performance of each model relative to the total cost obtained using the mean values of historical data and actual demand assuming perfect knowledge of the future.

$$Performance = \frac{Model\ with\ mean\ values\ cost - stochastic\ model\ cost}{Model\ with\ mean\ values\ cost - Model\ with\ actual\ demand\ cost} * 100 \quad (5.16)$$

Eqn. (5.16) is set such that the solution from the 'Model with mean values' gives a performance of zero while that of the model with 'actual demand' gives a performance of 100. The closer the solution under consideration is to that of perfect knowledge, the closer the performance factor is to 100. Approaches that give solutions worse than that of the mean demand obtain a negative value for performance. The performance of the solution under consideration relative to the model with mean values is the VSS and its performance relative to the model with actual demand is its EVPI.

The mean absolute error (MAE) is the average of the deviation of the provided units from the actual demand for all final products *F* for all retailer *C* and is calculated using Eqn. (5.17). This is used as an additional evaluation metric. Note that Eqn. (5.17) is an adaptation of Eqn. (5.11).

$$MAE = \frac{\sum_{f \in F} \sum_{c \in C} (\left[Act_{cf} - Pred_{cf} \right]_{+})}{|C||F|} \tag{5.17}$$

The MILP models are coded using Gurobi optimization solver and the ML models are built and trained using Tensorflow. All experiments were executed on an Intel Core™ i7-9750H CPU with 2.60 GHz PC and 6GB dedicated memory.

5.5 SC Network Model Parameter Generation

The size of the model is dependent on the number of elements present (i.e number of suppliers, plants, retailers, products and components). This means the model can be adjusted to fit both large and small SCs. The multimodality of the network could also be adjusted to only include the available modes. These can be used to specify the regionality or globality of the supply chain members.

To analyze the proposed model, realistic data is generated for most of the parameters such as plant and supplier capacity, costs, distances, weight/volume of products or components etc. Other parameters such as emission rates are obtained from publicly-available resources and the literature. Plant capacity is determined such that multiple plants will be needed to satisfy customer needs. Supplier capacity is also determined so that multiple suppliers are required by each plant. Most of the parameters are obtained from Kabadurmus and Erdogan (2020) as the network they analyse contains many of the elements used in our model. The volume, weight and cost of components are designed to be much less than those of the final products because in this case study, we assume that 3 components are required for a single product. Weight and volume capacity for road, rail and water transportation modes are according to the standard for their container types. A 40ft-container is considered for rail and water while a 45ft-container is considered for road transportation. The activation cost of each plant/supplier is considered \$2000 (Kabadurmus and Erdogan, 2020) while there is no fixed activation cost for consolidation/distribution centres. There is however a variable cost dependent on the quantity of cargo processed through the facilities (set as 0.5\$/unit). Water recycling cost is dependent on the region and could range from as low as 0.22$/m3 (0.0008$/gal) to as high as 2.0$/m3 (0.0075$/gal) (Plapally, 2012). The median is used (0.0042$/gal). The quantity (in gal) of water used for each process is guided by the examples provided by Boyd (2011). The quantity of each product f produced per day in each plant is set such that multiple days might be required to fulfil the demand. More details are provided in Table 5.

Table 5: Data Generation for Case Study

Parameter	**Data Generation**
Average demand of a retailer (for deterministic model)	This is set to be dependent on the average vehicle capacity usage $(ACVU)$ for Experiment 1 [See Eqn. (5.15)].
Demand variance of a retailer	Uniform random number between 0.25 and 0.33 of the average demand.
Actual demand of a retailer	Normal distribution according to the average demand and variance.
Plant capacity for a final product	Uniform random number between *c* and 1.5**c*, where c is the total quantity of demand required divided by the number of plants available (for the stochastic model, the maximum demand from historical data is used as *c*).
Supplier capacity for a component	Uniform random number between *s* and 1.5**s*. *s* is calculated as the total quantity of component *k* required divided by the number of suppliers available (quantity of component is based on the BoM).
Production cost of a final product (pf_{jf})	Normal distribution with the mean m and standard deviation 0.05**m,* where *m* is a uniform random number between 10 and

	100.
Purchasing cost of a component (pk_{ik})	Normal distribution with the mean *m* and the standard deviation 0.05**m*, where *m* is a uniform random number between 1 and 10.
Carbon emission (in kg) for producing a final product	Uniform random number between 0.02 and 0.08.
Average distance	This is adjusted to vary the average distance between shippers and receivers for experimentation.
Location of a supplier	Each node is randomly located on a square grid. X and Y coordinates are random numbers between *average_distance**0 to *average_distance**0.5.
Location of a manufacturing plant	Each node is randomly located on a square grid. X and Y coordinates are random numbers between *average_distance**1 to *average_distance**1.5.
Location of a retailer	Each node is randomly located on a square grid. X and Y coordinates are random numbers between *average_distance**2 to *average_distance**2.5.
Location of a consolidation center	Uniform random number betweIn *i* and *i*r*, whIre *i* is the average location of all the shipping (supplier or plant) nodes and *r* is a factor indicating how far the center is from the nodes it serves.
Location of a distribution center	Uniform random number between *j* and *j*r*, where *j* is the average location of all the receiving nodes (plant or retailer) and *r* is a

	factor indicating how close the center is from the nodes it serves.
Distance of arcs	Euclidean distance between nodes.
Percentage of distance of arc for which road transport is required (θ)	Uniform random number between 5 and 10.
Lead time for a supplier (ld_i)	Uniform random number between 1 and 5
Weights (kg) of a final product and a component $\left(w_p\right)$	The weight of a final product is a uniform random number between 1.5 and 3, and the weight of a component is 0.8
Volumes (m^3) of a final product and a component $\left(v_p\right)$	The volume of a final product is a uniform random number between 0.08 and 0.18, and the volume of a component is 0.05
Production batch size for each MC (b_{iqf})	Uniform random number between 40 and 60
Average quantity of product *f* produced daily in each plant $\left(pd_{if}\right)$	Uniform random number between 800 and 1500
Quantity of non-reused water utilized for making a batch of final products in each MC (wa_{iqf}^{q})	Uniform random number between 100 and 200 gal
Quantity of water use per day by each process (with daily use) in each MC (wa_{irf}^{r})	Uniform random number between 15,000 and 20,000 gal
Target volume efficiency (θv_t)	1 for all transportation modes
Target weight efficiency (θw_t)	1 for all transportation modes
Packing efficiency factor (γ_t)	Uniform random number between 0.8 and 1
Consolidation/distribution time (*cons_time*)	1 day
Carbon tax per unit (α)	1.2$/kg
Water recycle cost (β)	0.0042$/gal (Plapally, 2012)

Table 6: Parameter Values of Transportation Modes

	Road	**Rail**	**Water**	**Air**
f_t (per vehicle)	150	170	150	0
f^{road}	50			
c_t (per km per vehicle)	0.075	0.035	0.05	0.0025[a]
wa_t(gal/hr)	0	0	200	0
$\ddot{e}_t$ (per km per kg)	0.075	0.025	0.015	0.6
f_t^{CO2} (kg/gal)	8.78	0	0	0
p_t^{CO2}(kg/item)	1.0 -3.0	3.0-5.0	3.0-5.0	3.0-5.0
s_t^{CO2} (kg/Ton-km)	0.144	0.014	0.028	0.875
m_t (km/gal)	10.5			
v_t^{cap} (m^3)	90	67	67	60
w_t^{cap} (kg)	24000	26000	26000	21000
sp_t (km/hr)	100	32	40	900
cf	0-0.2	0	0	0
$PortD$ (days)	0	1-3	1-5	1-3

[a]The variable cost of air mode is calculated per km per kg

Data for f_t^{CO2} is obtained from EPA (2022). The fuel used in road transportation is taken as motor gasoline. The values for other modes are set to 0 as the cargo only occupies a portion of the vehicle. This is compensated for by slightly increasing the value of s_t^{CO2}. s_t^{CO2} is obtained from Carbonfund (2022). Mileage of truck is extracted from United States Department of Energy (2020). The packing emission rate for each mode is set as a random value within the given range. The total packing emission is dependent on the number of items being loaded onto the truck/container.

Table 7: Probability of Availability of Transportation Modes on Multimodal Arcs

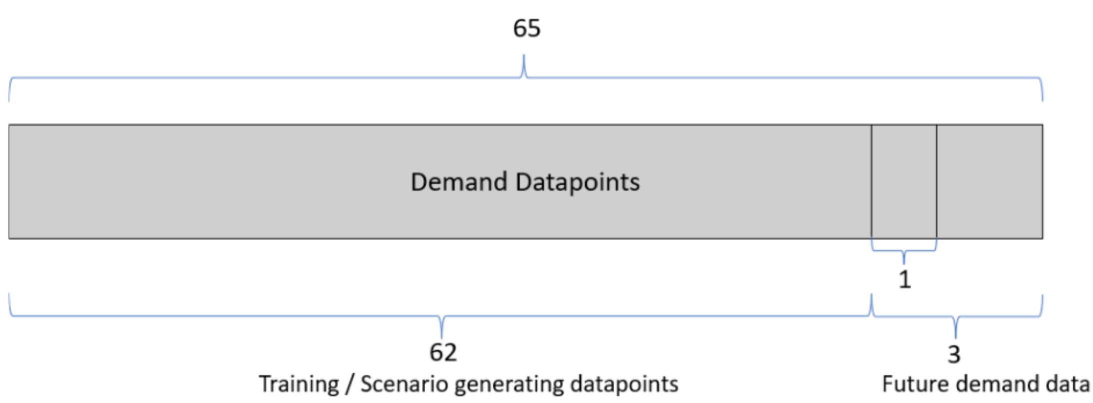
	Trucks	**Rail**	**Water**	**Air**
Minimum distance(>km)	0	300	1000	800
Probability (%)	100	75	25	100

To be more realistic, the probabilities of rail and water transportation modes being available are set to be 75% and 25% respectively. Also, they are only activated if the distance to be covered is more than 300 and 1000 kilometres respectively. Air transportation is available on all arcs with a distance greater than 800. After preliminary tests, values are also chosen for the emission limit and water limit based on the total demand and number of plants.

5.6 Data Sourcing

To examine and compare the effectiveness of all the approaches considered, an actual dataset was extracted from Kaggle.com titled 'Store item demand forecasting challenge' (Kaggle, 2018). The dataset contains 5 years of daily sales data of 50 items in 10 different retail stores. In preprocessing, we divide the data into 10 separate groups with each having 5 different items for the 10 stores. This allows for 10 repetitions of the same experiments with exclusive data. Each supply chain network will therefore have 10 retailers with 5 final products. The obtained data is clean and has no missing values. Some of the data is cut out to represent the unseen actual future demand data while the rest is used to build the models (See Figure 13). The data is preprocessed differently for each approach (SP or ML). For the SP approaches, the scenarios are generated from the data which are then used as parameters while building each model, while for the ML approaches, the data is reshaped to the required input dimension for each model.

Figure 13. Splitting Data into Historical/Training and Future Demand Data



5.7 Summary

This chapter starts with a description of the multiple approaches to handling stochastic optimization modelling explored in this study. The SP approaches were presented and the reformulations to the base Model A where provided. These were termed Model A1 and Model A2. Details on the procedure of applying ML approaches were then presented describing the models/algorithms, training performance metrics and model tuning. We then go ahead to discuss the details of the experiments to be carried out. Details on evaluation metrics, network parameter generation and data sourcing are then provided.

Chapter 6

Experiment Results & Discussion

6.1 Experiment 1

To analyse the influence of shipment consolidation on the supply chain network, six sample networks with 4 suppliers, 4 manufacturing plants, 4 retailers, 5 final products, 6 components and 4 modes of transportation were generated (4S-4M-4C, 5F-6K, 4T) using realistic random test instances according to Table 5 . The mean of the obtained results was calculated and graphs were plotted. The *average_distance* between shippers and receivers is set as 2000km to represent a wide global SC. The goal of this experiment is to evaluate the advantageousness of SCL and the conditions that make it profitable. For the Model A, the number of upstream consolidation and distribution centres were set to 1. The same was set for the number of downstream consolidation and distribution centres. For Model B, the values of the facilities were set to zero. This enforced the model to take only the direct routes from shippers to receivers while Model A could take multiple routes.

The average vehicle capacity usage *AVCU* is set as an evaluating factor to compare both models. The more the *AVCU*, the more the quantity of the items being shipped out on average by each supplier or plant. The cost would also generally increase. The results obtained are also plotted.

According to Figure 14, the total cost accrued by Model A was always less than or equal to that of Model B. This is expected as the only transportation route option available to Model B is the direct route while Model A has more transportation options and could seek cheaper alternatives. This makes it such that the lowest cost possible for Model A is the cost of Model B. The total cost is the sum of the transportation, facility activation, production, carbon and water cost.

Figure 14. Total Cost vs AVCU

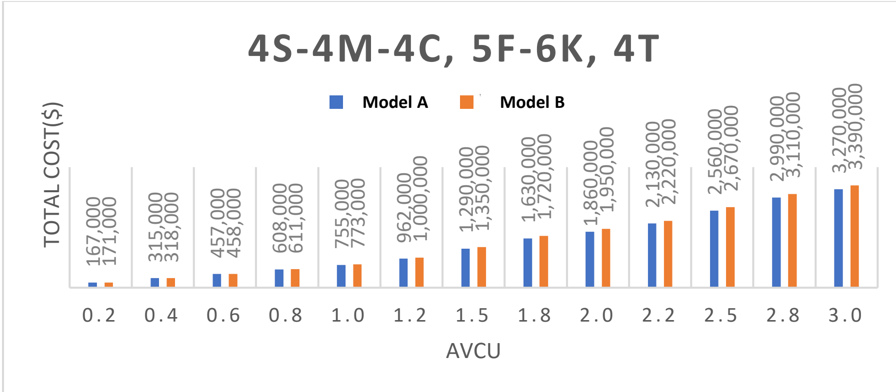


Figure 15. Number of Vehicles vs AVCU

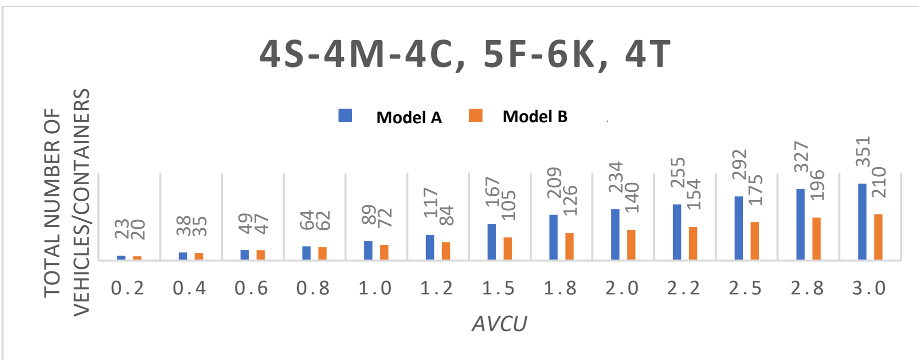
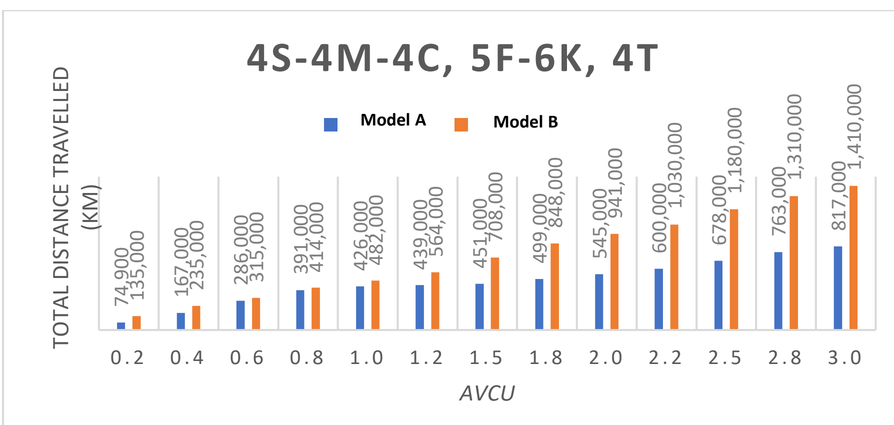
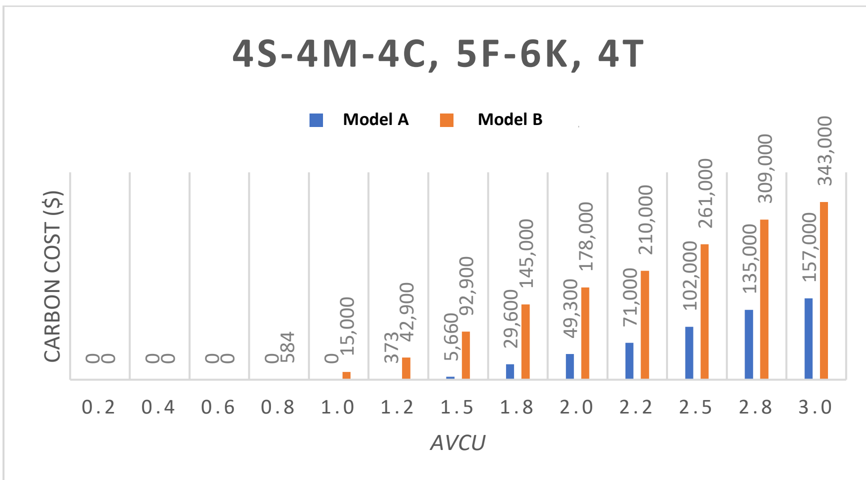


Figure 16. Total Distance Travelled vs AVCU

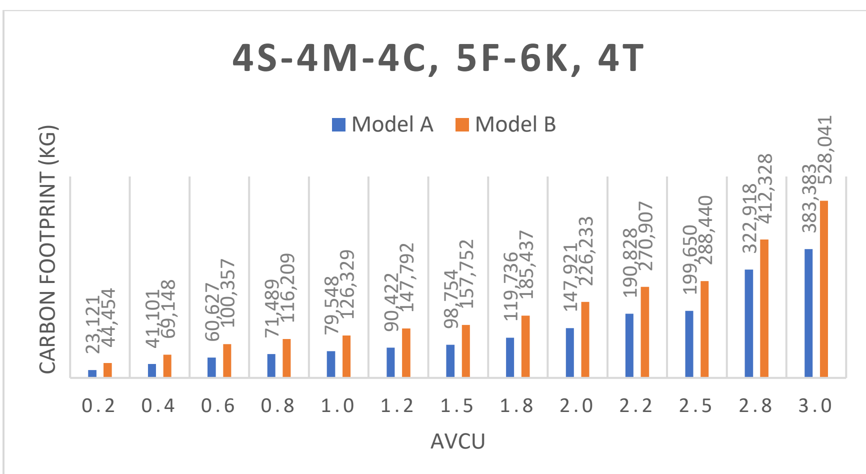


There were more vehicles in total used in Model A (Figure 15). There were a lot of short-haul trips between the shippers and the consolidation/distribution facilities thereby requiring more vehicles. This could have been potentially reduced if there were milk runs between the shippers. From Figure 16, it can be seen that the total distance travelled by all the vehicles/containers in Model A is significantly lower than that of Model B. This is most likely because the shipments were consolidated early in the SC and fewer vehicles/trucks (most likely with FTL) were used for bigger parts of the trips. This option was not available in Model B as shippers had to use as many trucks as required even if they were shipping LTLs, thereby causing Model B to have fewer trucks, but with each needing to traverse the full distance to the receiver.

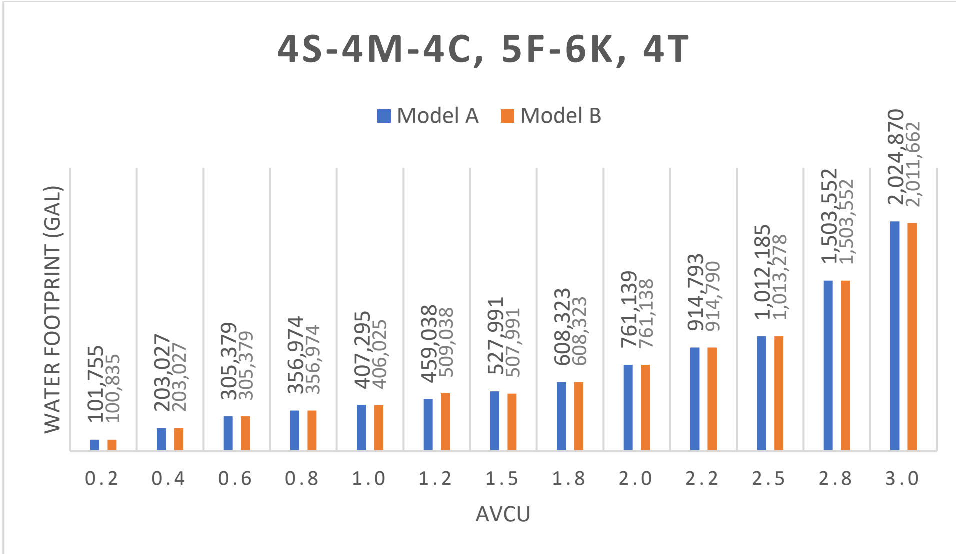
Figure 17. Carbon Cost/CF/WF vs AVCU



17a



17b



17c

The cost accrued due to the carbon tax was higher for the Model B (Figure 17a). The values in the figure however include the carbon emissions during production within the plants. The reduced carbon cost for the Model A is majorly due to the decrease in the total distance travelled (Figure 16). Figure 17b shows that the carbon footprint of Model A is consistently less than that of Model B. This means that Model B is less environmentally sustainable. In Figure 17c, the WF of both models are the same in most of the experiments. There are however samples where the WF of either model is slightly higher than the other. This is because production is the major contributor to the water footprint in our case studies and production WF is the same for both models in most cases. The only other source of WF in our study is the water mode of transportation. The slight difference between the WF of Models A and B seem to occur when either of them chooses to use the water mode of transportation within its SCN.

The experiment was repeated with the value of *AVCU* fixed as 1.2 while varying the average distance between shippers and receivers.

Figure 18. Total Cost vs Average Distance

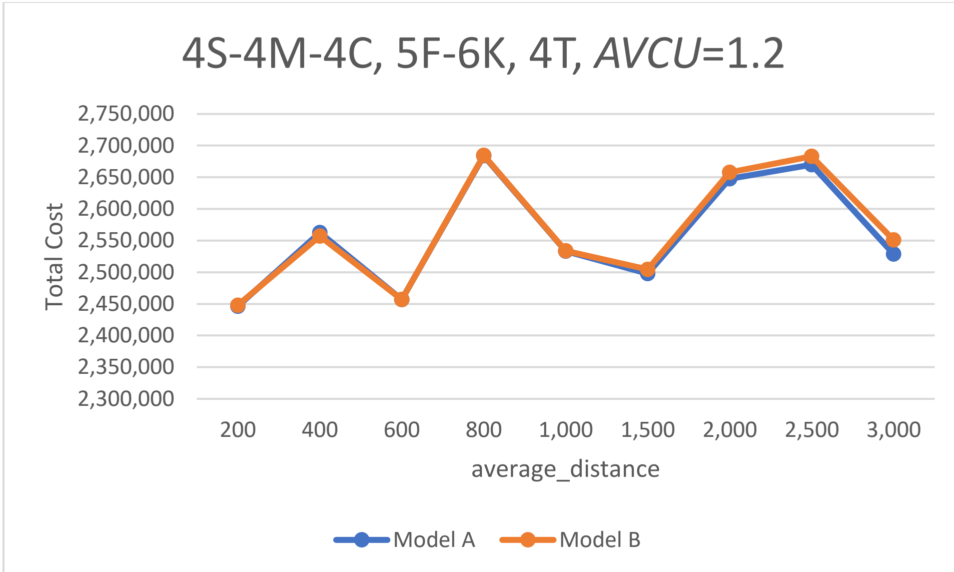
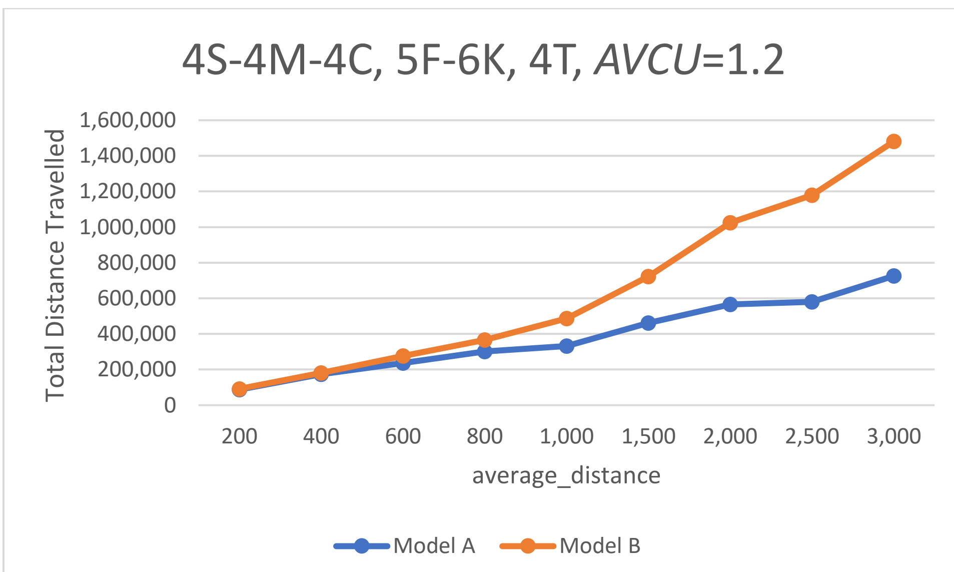


Figure 19. Total Distance Travelled vs Average distance



The total cost was fairly similar until the *average_distance* got up to around 1500km (see Figure 18). The total distance travelled was also similar until the *average_distance* (between shippers and receivers) got to around 600 (see Figure 19). This means SCL might not lead to economical and environmental savings in every scenario. The distance between the SC agents has to be above a certain threshold for SCL to be beneficial. The high spike in Figure 19 between *average_distance* 2000 and 3000 shows that if all the necessary factors were in place, a huge environmental savings opportunity could arise.

In summary, the inclusion of shipment consolidation into the supply chain network generally reduced the total transportation cost, carbon cost due to emissions above the cap, and reduced total distance travelled. However, the number of vehicles used increased due to the absence of milk runs between shippers. The average distance between the shippers and receivers also has to be above a certain threshold before SCL can be considered advantageous.

6.2 Experiment 2

For the ML approaches, monthly demand for the retailers is predicted using historical demand as input. The daily data is aggregated into monthly (28-days) values. Demand for 24 months before the demand prediction date is taken as input with a corresponding output of the demand prediction for 3 months after the prediction date. 3-months timeline was selected so as to be able to use the output of each ML model/algorithm for both Experiments 1 and 2 simultaneously. After training the models and predicting the demand for the final 3 months, the future prediction for the first month is taken as the value used in this experiment. The data has 10 retail stores with 50 items/final products. This was divided into 10 groups with each group having the same 10 retail stores but with 5 unique items/final products each. For each retail store, each product has historical demand data. The goal is therefore to predict the demand for the next month (28-days) for all retail stores and items.

The SCN model was set with 4 suppliers, 4 manufacturing plants, 10 retailers, 5 final products, 7 components and 4 modes of transportation (4S-4M-10C, 5F-7K, 4T). The planning period is set as 28 days. Parameters are generated, and the model is preliminarily solved deterministically with the actual demand values (no overstock and understock) to obtain the first baseline. This gave the actual cost of the supply chain if we had perfect knowledge. The total cost was divided by the total quantity of final products to obtain an idea of the unit cost. The understocking cost is considered as sales opportunity foregone and made equivalent to the selling price which is set as 2 times the per unit cost of the final products (total SC cost/total quantity of final products). Overstocking cost is also considered as half the unit cost (salvage value is set as half the unit cost). Excess stock is however not carried over to the next planning period. The historical data aggregated into monthly data is used as past scenarios for the SP models, and as training data for the ML models. The ML models were trained with MOU (Eqn. 5.12) as the custom metric so the model learns to optimize the output while factoring in the penalties for overstocking and understocking proportionately.

Figure 20 shows the graphical comparison between the prediction demand and actual demand for the AC-LSTM model for Group 1 and 10. The x-axis shows the Store/Item(final product) number while the y-axis shows the demand value. The tight closing between the actual and predicted values in the results shows that the model was able to learn the relationships within the data. The model tends to often overshoot in its predictions. This is because the penalty for over-prediction is less than that of under-prediction (understocking cost>overstocking cost).

Figure 20. AC-LSTM Prediction Graph

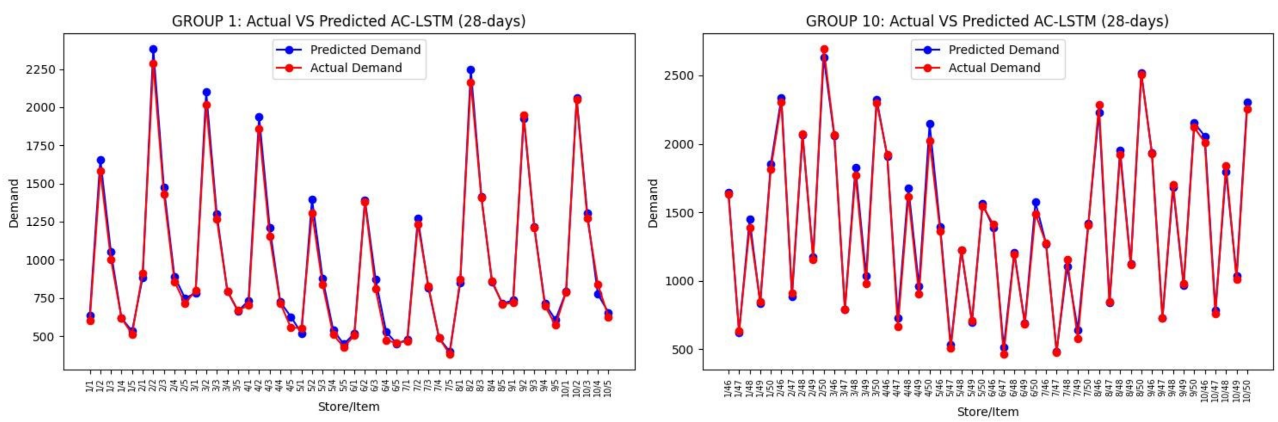


Figure 21 and Figure 22 show that the ACV-LSTM and Ensemble SVR were also able to learn the pattern within the data and give very close predictions. The Ensemble SVR however seems to generally have higher predictions than the ACV-LSTM. This could be advantageous if it leads to a smaller overall SC cost. The overall performance of all model types is evaluated next.

Figure 21. ACV-LSTM Prediction Graph (28-days)

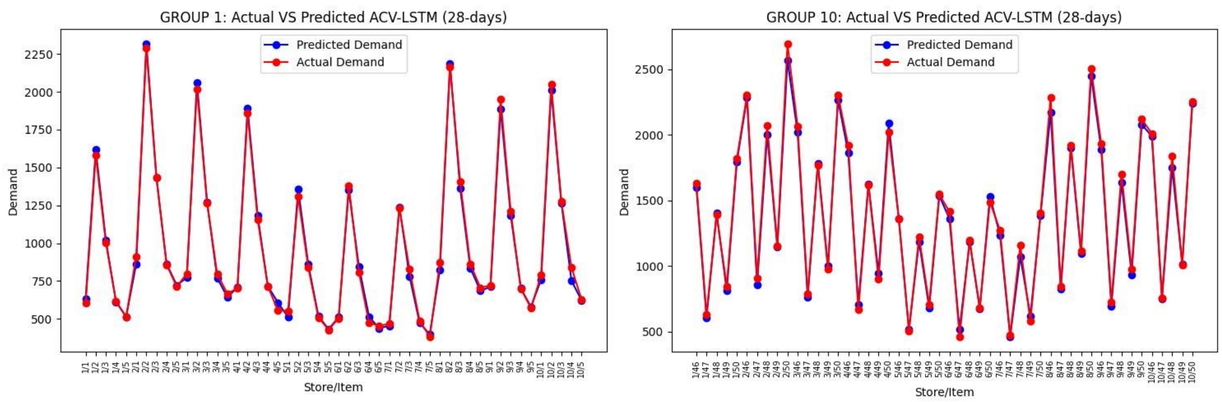


Figure 22. SVR-Ensemble Prediction Graph (28-days)

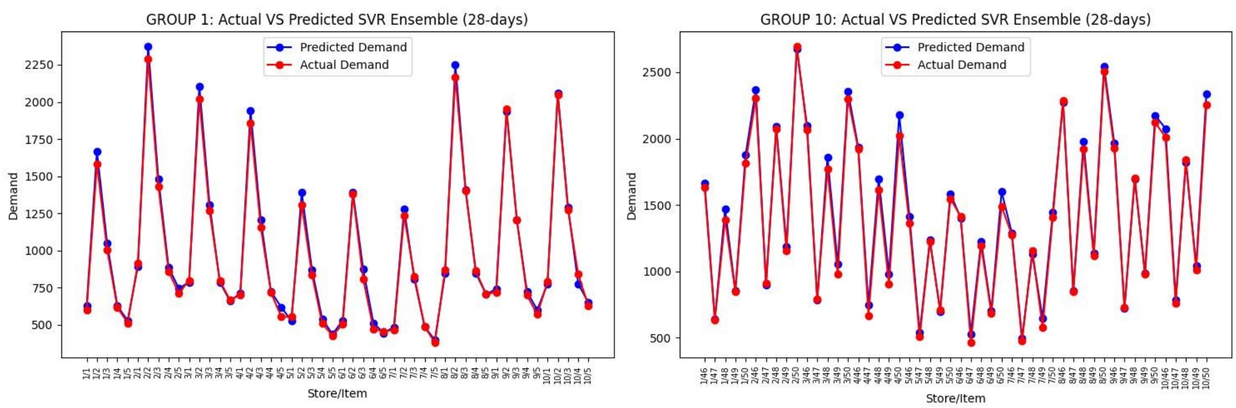


Table 8 shows a summary of all the models considered for Group 1. The first column describes the mean absolute error (MAE) between the predicted and actual demand values. Considering that the goal of the prediction is its usage within a newsvendor MILP model, this evaluation compares all the ML and SP models through the total cost incurred after using the predictions to optimize the MILP model while including the overstocking and understocking costs (OUC) due to inaccurate predictions. The mean value of historical demand is also used as a ‘prediction’ to enable comparison with the ML and SP models and to evaluate the Value of the Stochastic Solution. The solution of the model with ‘Actual Demand’ is added to the table to compare how good the predictions of the ML and SP are in comparison to the ideal scenario where the actual demand is known perfectly.

From Table 8 , the results from the experiment show that generally, the ML models outperform the SP models. MAE is calculated as the average of the absolute difference between the prediction and actual demand [see Eqn. (5.17)]. For each SP approach, the model's 'prediction' is considered as the summation of the values of the variables X_{jc}, X_{d2c}, X_{e2c} [see Constraint (5.3) for Model A1 and Constraint (5.6) for Model A2] after optimization. The Total OUC is also calculated as the cost of understocking and overstocking. Optimization model cost (also Total cost) is the addition of the cost of solving the Model A deterministically with the 'prediction' of each method and the Total OUC. ACV-LSTM gives the smallest MAE. The Ensemble-SVR however performs better, leading to the lowest OUC and total cost.

Table 8: Group 1 Evaluation (28-days)

Model (Group 1)		**Performance Metrics (28-days)**		
		MAE	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	69.2	608,979	4,609,866
	Chance constraint (α=0.2)	63.4	527,516	4,576,183
	Chance constraint (α=0.2)	59.8	469,474	4,547,130
ML	AC-LSTM	31.4	98,839	4,365,955
	ACV-LSTM	23.9	144,385	4,385,431
	Stacking Ensemble	30.8	97,108	4,360,621
Baselines	Model with mean values	70.4	625,594	4,622,319
	Actual Demand	0	0	4,339,078

A very similar trend is seen with Group 10 in Table 9 with the ML approaches having very good performances and obtaining results closest to that of having perfect knowledge (EVPI). The Ensemble SVR once again performs best with the smallest OUC and Total cost while the AC-LSTM has the smallest MAE. The Simple recourse approach (Model A1) obtains a solution very close to the model with mean values while the Chance constraint approaches (Model A2) obtain better results with increasing values of α.

Table 9: Group 10 Evaluation (28-days)

Model (Group 10)		**Performance Metrics (28-days)**		
		MAE	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	100.2	700,864	5,144,707
	Chance constraint (α=0.2)	89.5	620,082	5,117,958
	Chance constraint (α=0.2)	82.4	547,745	5,092,752
ML	AC-LSTM (MAE)	30.4	92,506	4,957,362
	ACV-LSTM (MAE)	35.9	209,997	5,005,574
	Stacking Ensemble (MAE)	40.4	84,515	4,951,500
Baselines	Model with mean values	104.7	725,360	5,145,069
	Actual Demand	0	0	4,884,868

Only evaluation data for Group 1 and 10 are shown in this chapter with the remaining 8 groups illustrated in the Appendix. The average of all ten groups is however discussed for overall comparison. Table 10 shows the summary of all 10 groups with the average of all solutions generally in line with the analysis of Groups 1 and 10.

Table 10: Average Evaluation for all 10 Groups (28-days)

Average of 10 Groups		**Performance Metrics (28-days)**		
		MAE	**OUC**	**Optimization Model Cost**
SP	Simple Recourse	106.4	805,370	5,660,968
	Chance constraint (α=0.2)	98.5	739,430	5,635,534
	Chance constraint (α=0.2)	92.9	654,018	5,603,868
ML	AC-LSTM (MAE)	45.4	119,331	5,466,616
	ACV-LSTM (MAE)	40.1	154,057	5,439,291
	Stacking Ensemble (MAE)	42.2	113,660	5,435,597
Baselines	Model with mean values	125.7	843,018	5,669,537
	Actual Demand	0	0	5,364,076

Figure 23 and Figure 24 plot the three solutions (Groups 1, 2 and the average of all groups) on a two-dimensional graph to show how well placed each model is in comparison to the baselines. The position of each approach relative to the model with mean values is its VSS while its position relative to the model with actual demand is its EVPI.

The performance measures in Figure 24 are calculated using Eqn. (5.16). The closer the value is to 100, the more the cost saved due to the accuracy of the prediction. The Ensemble-SVR has the lowest EVPI showing how close to perfect knowledge the model's predictions were. The ACV-LSTM model is however the closest to the mean values among the ML models. It (ACV-LSTM) still performs significantly better than the SP models. Among the SP models, the Simple recourse (Model A1) performs the worst with a total cost very close to that of the mean values (overall least VSS) showing how close the solution obtained from the method was to barely using the average of the historical data. The Chance constraint models (Model A2) perform slightly better than the Simple recourse (Model A1) but perform poorly in comparison to all the ML approaches.

Figure 23. Mean vs Solution vs Perfect Knowledge (28-days)

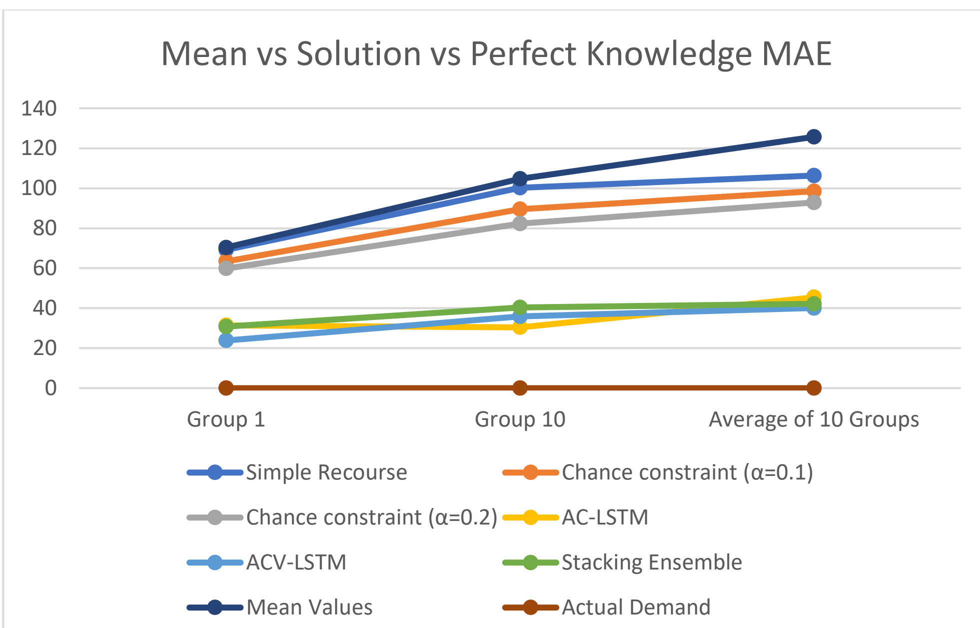
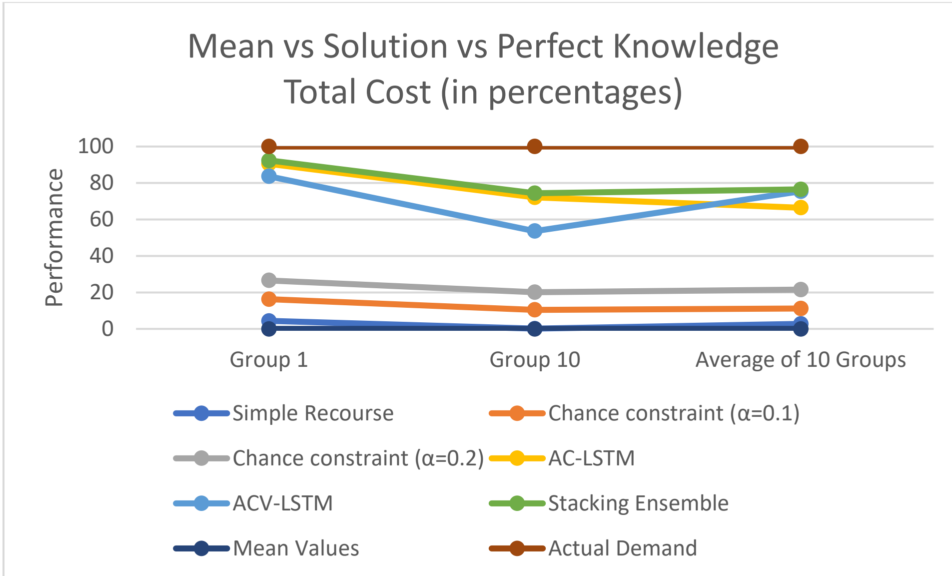


Figure 24. Mean vs Solution vs Perfect Knowledge Total Cost (28-days)



6.3 Experiment 3

The previous experiment considered the planning period to be 28 days. This experiment however considers a larger planning period of 84 days (Note: A month is considered 28 days). The reason for enlarging the planning period was to get an idea of how the SP and ML models would perform if we had to forecast further into the future or expand our planning horizon. For the ML models, historical demand for 24 months before the beginning of the planning period was taken as input to predict the demand for the following 3 months. The available data is for 65 months with the first 62 months used as training data and the last 3 months used as the unseen actual future demand. The training data was split into a sequence such that Month 1 to Month 24 data is used as input for Month 25 to Month 27, Month 2 to Month 25 is used as input for Month 26 to Month 28, up till, Month 36 to Month 59 used as input for Month 60 to Month 62. After training the model, Month 37 to Month 62 is then used as testing data to make the final prediction for Months 63 to 65. This procedure was already carried out in Experiment 2. The 3-month prediction was then summed up into a single value.

For the SP models, the first 62 data points are also used to generate the required scenarios. A rolling aggregate with a 3-months span is used for each scenario. This means, that the sum of demand from Month 1 to Month 3 is the first scenario demand, the sum of demand from Month 2 to Month 4 is the second scenario demand, up till, the sum of demand from Month 60 to Month 62 as the last scenario. This procedure was taken for two reasons, 1. So the generated scenarios can capture some long-term trends and 2. To maintain a relatively large number of scenarios. The model is then optimized with the demand scenarios generated as the discrete samples of the stochastic demand parameter. For each SP model, after optimization, the values of the variables X_{jc}, X_{d2c}, X_{e2c} (see Demand balance constraint 4.13) are extracted and summed up as the quantity of final products the models A1 & A2 prescribe the SCN to supply for each retail store for each product. These values are then used as the demand value to solve Model A deterministically to obtain the model cost. The difference between the quantity of final products supplied and the actual demand is used to obtain the OUC cost.

For the ML approaches, Figure 29 shows the relationship between the actual demand and the predicted demand using the AC-LSTM model. The model's predictions are again very close to the actual demand despite the larger planning period potentially leading to a small OUC.

Figure 26 and Figure 27 also show that the ACV-LSTM and Ensemble SVR have good performances. Graphs for groups 2 to 9 are shown in the Appendix.

Figure 25. AC-LSTM Prediction Graph (3-months)

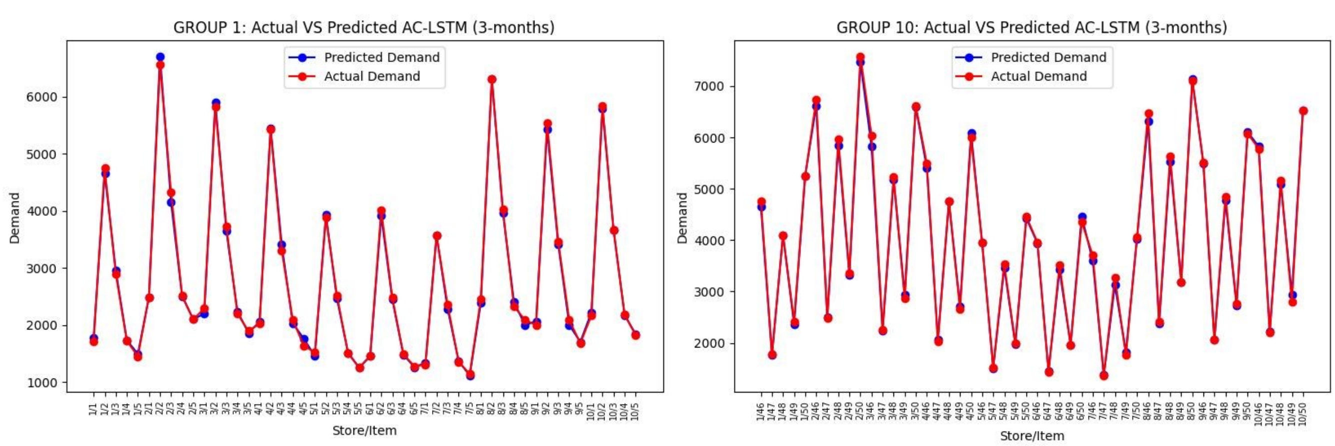


Figure 26. ACV-LSTM Prediction Graph (3-months)

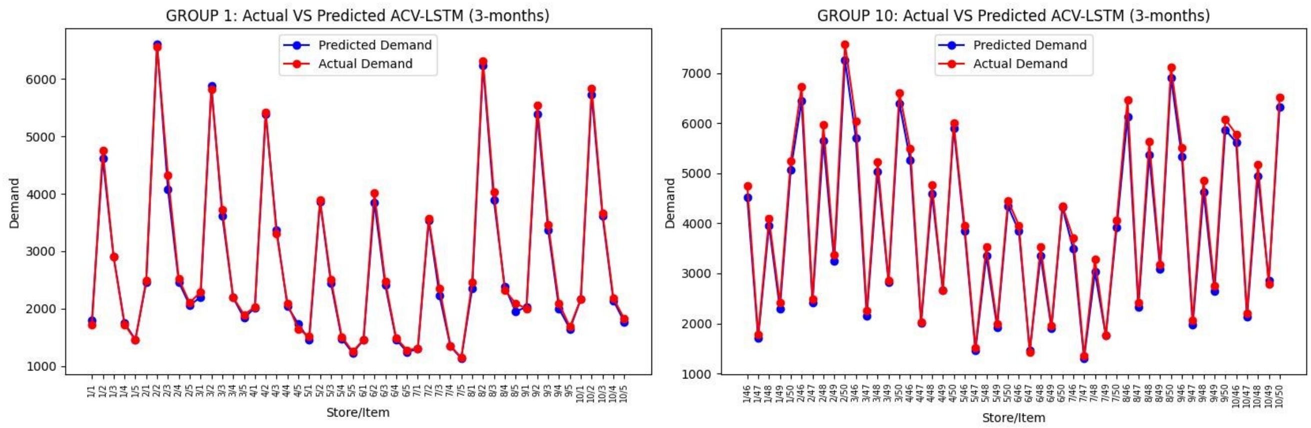


Figure 27. SVR Ensemble Prediction Graph (3-months)

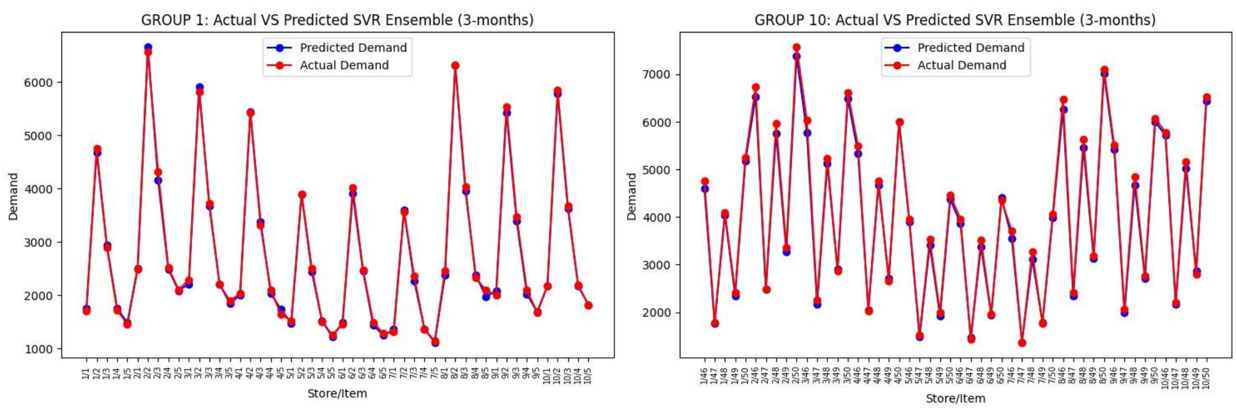


Table 11 presents a comparative analysis of all the methods. All ML models have good performances with values very close to the model with actual demand. The SVR Ensemble gives the lowest total cost. AC-LSTM however comes close with a total cost almost equal to that of the SVR Ensemble. AC-LSTM has the least OUC and MAE. The SP approaches do not perform well as all have MAE, OUC and total costs higher than the model with mean values. Similar results are obtained in Group 10 (Table 12).

Table 11: Group 1 Evaluation (3-months)

Model (Group 1)		**Performance Metrics (3-months)**		
		MAE	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	121.2	1,078,558	11,109,093
	Chance constraint (α=0.1)	135.2	1,224,624	11,152,059
	Chance constraint (α=0.2)	136.8	1,220,792	11,222,144
ML	AC-LSTM	49.1	250,940	10,800,178
	ACV-LSTM	62.8	426,127	10,879,521
	Stacking Ensemble	49.9	287,533	10,793,808
Baselines	Model with mean values	106.1	805,470	11,018,492
	Actual Demand	0	0	10,712,399

Table 12: Group 10 Evaluation (3-months)

Model (Group 10)		**Performance Metrics (3-months)**		
		MAE	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	160.4	1,088,517	11,115,358
	Chance constraint (α=0.1)	165.2	1,222,940	11,190,968
	Chance constraint (α=0.2)	165.1	1,222,406	11,204,924
ML	AC-LSTM (MAE)	54.2	226,672	10,895,784
	ACV-LSTM (MAE)	147.3	775,414	11,034,724
	Stacking Ensemble (MAE)	90.7	455,559	10,942,753
Baselines	Model with mean values	153.4	813,211	11,035,848
	Actual Demand	0	0	10,763,345

Averaging all 10 Groups (Table 13) and comparing the results of all models, the ML models generally outperform the SP models. Among the ML approaches, the SVR Ensemble performs the best overall. This is not surprising as the Ensemble-SVR uses the outputs of the two other ML models as extra input (children models) when training itself. The AC-LSTM and ACV-LSTM also perform close to the model with 'actual demand' and obtain similar results. Analysing the SP approaches, unlike when the planning period was 28-days (Experiment 2), all the models perform poorly (below Model with mean values). The Chance constraint approach (Model A2) has the highest average total cost, OUC and MAE. This might be a result of the model being too conservative thereby impacting the viability of the solution obtained over a longer period. In comparison to Experiment 2 where the planning period was significantly smaller, both the ML and SP models do not perform in this experiment as well as they did. The AC-LSTM and SVR Ensemble with MAE of 69.2 and 65.9 respectively however still perform relatively well in comparison to the actual demand values. This is because Figure 25 shows that actual demand could be as high as 7000 (that is within 1% accuracy).

Table 13: Average of 10 Groups Evaluation (3-months)

Average of 10 Groups		**Performance Metrics (3-months)**		
		MAE	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	181.3	1,539,783	15,731,840
	Chance constraint (α=0.1)	185.4	1,671,597	15,787,402
	Chance constraint (α=0.2)	186.2	1,720,751	15,804,447
ML	AC-LSTM (MAE)	69.2	338,656	15,355,731
	ACV-LSTM (MAE)	103.4	637,870	15,359,121
	Stacking Ensemble (MAE)	65.9	319,260	15,251,771
Baselines	Model with mean values	176.2	1,171,575	15,587,240
	Actual Demand	0	0	15,156,420

Figure 28 and Figure 29 graphically shows the comparison of the three tables (Group 1, 10 and average) and how the models perform in comparison to the baselines.

Figure 28. Mean vs Solution vs Perfect Knowledge (3-months)

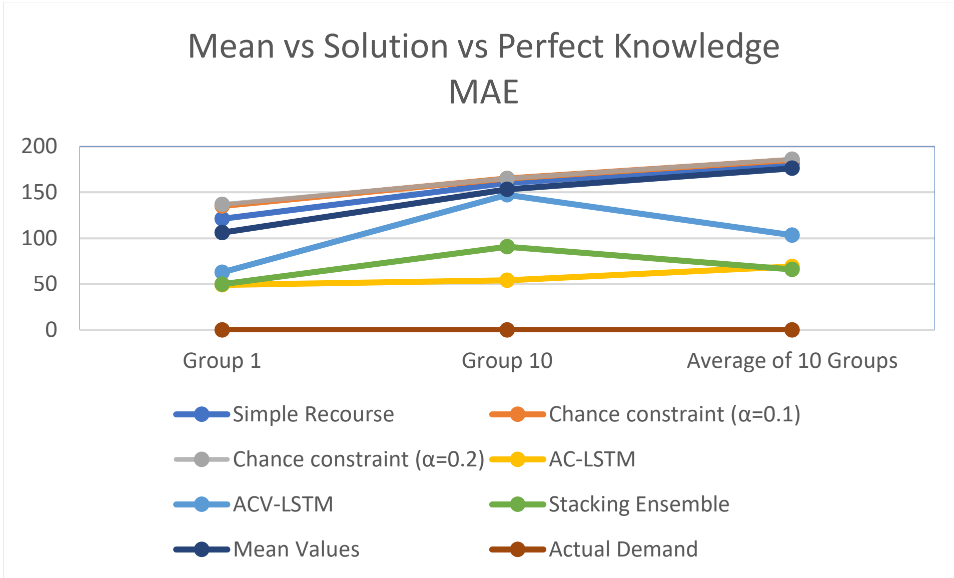
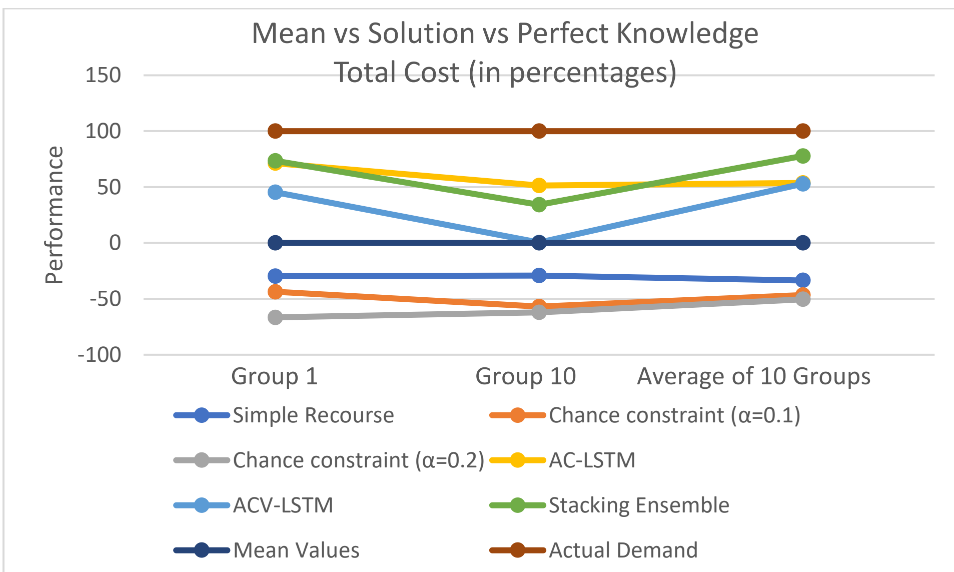


Figure 29. Mean vs Solution vs Perfect Knowledge Performance (3-months)



There is a consistent pattern of the ML approaches performing better than the SP approaches, with the Ensemble-SVR performing best. This trend can be attributed to the ability of ML models to efficiently learn patterns and seasonality within the historical data using advanced algorithms and computational powers, thereby able to predict the future demand fairly closely leading to less penalty incurred due to overstocking or understocking.

In terms of CPU runtime, for Model A1 with simple recourse, it took an average of two minutes to solve the models to optimality. For Model A2 with chance constraint and α = 0.1, it took an average of two hours to solve the models while that of α = 0.2 took about four hours to solve. In comparison, in the ML approaches, it took about 12 hours to run the ML algorithms with an additional two minutes to solve Model A deterministically using the predicted values. This shows that even though the machine learning approaches provided better results and helped save more cost, they were more computationally expensive.

Chapter 7

Conclusion & Future Research

In this thesis, a novel supply chain optimization problem is presented to design sustainable multimodal supply chains which also consider shipment logistics and stochastic demand. One of the objectives of this thesis was to build an SCN model that shows how water footprint and carbon footprint analysis can be factored into the tactical decisions made when planning supply chain logistics. We went through the procedure on how to analyse and incorporate different elements such as water footprint analysis, carbon footprint analysis, multimodal freight transportation and shipment consolidation logistics into the SCN in the design stage or while performing logistical operations for an existing SCN. This involved the building of an adaptable optimization model which could be modified and applied to optimize real-life SCN optimization problems.

We then demonstrated how shipment consolidation logistics could serve as a tool for reducing not just the economic cost of SCs, but also the amount of environmental impact of SC operations due to a reduction in the distance travelled to attain SC goals. We evaluated the conditions that make the use of SCL within the SC most profitable. Model A (with SCL) was compared with Model B (without SCL). For the base case data used, Model A performed equally or outperformed Model B in every single scenario. This is because Model A has the infrastructure of the Model B within it and generally decides to use consolidation/distribution facilities when there are cost-saving opportunities, or ignore those facilities and make only direct deliveries from shippers to receivers when that is better.

While varying the average vehicle capacity usage (*AVCU*) factor, Model A seemed to find significant cost-saving opportunities when the vehicle utilization factor indicated a combination of FTL and LTL. It seemed the LTLs were mostly transported through consolidation/facilities. The number of vehicles utilized within Model A was generally higher than that of Model B. This is mostly because the shippers had to use individual trucks to deliver their LTL cargo to the consolidation/distribution centers before shipping it to the receivers. This led to extra costs due to the fixed cost of activating each vehicle. The overall cost of Model A was however still less than Model B. This is because the optimizer would only choose to use SCL whenever there were significant cost savings to offset the vehicles' fixed costs.

After the *AVCU* factor was fixed as 1.2 and the average distance between the shipping and receiving facilities varied, Model A initially had the same cost as Model B for short distances but found higher cost-saving opportunities as the average distance travelled increased. This demonstrates that utilizing SCL might offer no cost savings when the distance of delivery is low but becomes more profitable when the SCN is large and spans over a wide area (e.g regional, global). The cumulated distance travelled by all vehicles/containers/airplanes within Model A was equal to or significantly lower than that of Model B in all the scenarios considered. This means Model B has more environmental impact as transportation-based CF is roughly proportional to the distance travelled. This highlights the environmental advantage of incorporating shipment consolidation into SCs. The managerial implication of this is that the SSCN/MFT&SCL model can be used by DMs to choose when it is appropriate to use consolidation facilities within the SCN to obtain a feasible solution under budget considerations. Using the results obtained from modifying the model generated in this thesis to specific scenarios, the DM can make logistical decisions and choose the appropriate suppliers, manufacturing plants, transportation modes and order quantity for their operations to minimize economic and environmental costs while concluding all operations within the set planning period.

In tackling the inherent stochasticity of some of the parameters under consideration when designing an SCN or making logistics decisions, different approaches were utilized and compared. These included multiple SP and ML approaches from which all the ML models performed better than the traditional SP models. The ensemble model performed best and obtained results closest to the model with perfect predictions thereby giving the lowest EVPI. Even after extending the forecasting period, the ML approaches still had performances significantly better than the SP approaches with their (ML models) predictions within 1% of the actual demand values. This makes the case for the consideration of ML-based approaches when making logistical decisions on how to run an SC. With the proliferation in the generation of data through the continual digitalization of SC processes, and the innovation of new efficient algorithms, ML-based approaches to uncertainty within the SC promise to be even more accurate thereby enabling more precise planning. The flexibility and effectiveness of ML models in predicting varying forecast periods allow the DM to freely change the planning period in response to the trends in the market or technical changes in network parameters (such as supplier lead time).

Areas of further extension to this work could include:

1. The consideration of the impact of SCs on other pillars of sustainability (social and culture) and how to factor this into the supply chain network design and logistics planning.
2. The expansion of the model to consider milk runs between the shippers and between the receivers. This will potentially reduce the economic and environmental costs of the model.
3. The application of this work to real-life case studies to discover new findings or verify the current findings.
4. The use of other demand-related data (e.g weather conditions, store location, month of the year) alongside the demand data to further improve the forecasting accuracy of ML models.
5. The consideration of other SP methodologies such as sample average approximation using simulated data, Conditional Value at Risk.
6. The utilization of decomposition methods or metaheuristics to handle large-scale cases that can not be easily solved using optimization solvers (such as Gurobi).
7. The analysis of the impact of improved prediction precision through the use of machine learning on the bullwhip effect.

References

Ahmadi, S., & Amin, S. H. (2019). An integrated chance-constrained stochastic model for a mobile phone closed-loop supply chain network with supplier selection. *Journal of Cleaner Production*, *226*, 988–1003.

Aivazidou, E., Tsolakis, N., Vlachos, D., & Iakovou, E. (2018). A water footprint management framework for supply chains under green market behaviour. *Journal of Cleaner Production*, *197*, 592–606.

Alizadeh, M., Ma, J., Marufuzzaman, M., & Yu, F. (2019). Sustainable olefin supply chain network design under seasonal feedstock supplies and uncertain carbon tax rate. *Journal of Cleaner Production*, *222*, 280–299.

Alves, R., Lima, R. da S., de Sena, D. C., de Pinho, A. F., & Holguín-Veras, J. (2019). Agent-based simulation model for evaluating urban freight policy to e-commerce. *Sustainability (Switzerland), 11*(15), 1–19.

Ampazis, N. (2016). Forecasting demand in supply chain using machine learning algorithms. *International Journal of Artificial Life Research, 5*(1), 56–73.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Beale, E. M. (1955). On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society: Series B (Methodological)*, *17*(2), 173-184.

Bedaiwi, M. A., Khomsi, S., Bahabri, A., & Salama, R. (2019). Water pollution disasters caused by the leakage of hydrocarbons during shipping operations “Systematic review for hydrocarbons transportation operations across the sea from the ports of the arabian gulf.” [Master’s thesis, King Abdul Aziz University] https://www.eimj.org/uplode/images/photo/Water_Pollution_Disasters_Caused_by_the_Leakage_of_Hydrocarbons_..pdf [Accessed on 14-06-2022].

Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization* (Vol. 28). Princeton University Press.

Beresford, A., Pettit, S., & Liu, Y. (2011). Multimodal supply chains: Iron ore from Australia to China. *Supply Chain Management, 16*(1), 32–42.

Birge, J. R. (1997). Stochastic programming computation and applications. *INFORMS Journal on Computing*, *9*(2), 111–133.

Birim, S., Kazancoglu, I., Mangla, S. K., Kahraman, A., & Kazancoglu, Y. (2022). The derived demand for advertising expenses and implications on sustainability: a comparative study using deep learning and traditional machine learning methods. *Annals of Operations Research*, *(2022),* 1-31.

Boyd, B. (2011). Guidelines for estimating unmetered industrial water use. *US Department of Energy, Washington, D.C.*

Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings, 158,* 1429-1441.

Cao, J., Jiang, Z., & Wang, K. (2017). Customer demand prediction of service-oriented manufacturing using the least square support vector machine optimized by particle swarm optimization algorithm. *Engineering Optimization*, *49*(7), 1197–1210.

Çapar, I. (2013). Joint shipment consolidation and inventory decisions in a two-stage distribution system. *Computers and Industrial Engineering*, *66*(4), 1025–1035.

Carbonfund (2022). Carbon and usage calculation methods. https://carbonfund.org/calculation-methods/ [Accessed on 14-06-2022].

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154.

Chaabane, A., Ramudhin, A., & Paquet, M. (2012). Design of sustainable supply chains under the emission trading scheme. *International Journal of Production Economics, 135*(1), 37–49.

Charnes, A., Cooper, W. W., & Symonds, G. H. (1958). Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management science*, *4*(3), 235-263.

Chen, Y. Y., Lin, Y. H., Kung, C. C., Chung, M. H., & Yen, I. H. (2019). Design and implementation of cloud analytics-assisted smart power meters considering advanced artificial intelligence as edge analytics in demand-side management for smart homes. *Sensors, 19*(9), 2047.

Chen, Y., Xu, M., & Zhang, Z. G. (2009). A risk-averse newsvendor model under the CVaR criterion*. Operations Research*, *57*(4), 1040–1044.

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, *83*, 187–205.

Choudhary, A., Sarkar, S., Settur, S., & Tiwari, M. K. (2015). A carbon market sensitive optimization model for integrated forward-reverse logistics. *International Journal of Production Economics*, *164*, 433–444.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, *1*(3-4), 197-206.

De Boer, I. J. M., Hoving, I. E., Vellinga, T. V., Van De Ven, G. W. J., Leffelaar, P. A., & Gerber, P. J. (2013). Assessing environmental impacts associated with freshwater consumption along the life cycle of animal products: The case of Dutch milk production in Noord-Brabant. *International Journal of Life Cycle Assessment*, *18*(1), 193–203.

Dye, S. (2008). Simple Recourse Problem . In *Floudas, C., Pardalos, P. (eds) Encyclopedia of Optimization. Springer, Boston, MA*.

Elhedhli, S., & Merrick, R. (2012). Green supply chain network design to reduce carbon emissions. *Transportation Research Part D: Transport and Environment*, *17*(5), 370-379.

Elkington, J. (1994). Towards the sustainable corporation: Win-win-win business strategies for sustainable development. *California Management Review*, *36*(2), 90-100.

EPA. (2022). Emission factors for greenhouse gas inventories. https://www.epa.gov/system/files/documents/2022-04/ghg_emission_factors_hub.pdf [Accessed on 30-05-2022].

Fahimnia, B., Sarkis, J., Boland, J., Reisi, M., & Goh, M. (2015). Policy insights from a green supply chain optimisation model. *International Journal of Production Research*, *53*(21), 6522–6533.

Fan, Y., Behdani, B., Bloemhof-Ruwaard, J., & Zuidwijk, R. (2019). Flow consolidation in hinterland container transport: An analysis for perishable and dry cargo. *Transportation Research Part E: Logistics and Transportation Review*, *130*, 128–160.

Farazmand, M., Pishvaee, M. S., Ghannadpour, S. F., & Ghousi, R. (2022). Green dynamic multimodal logistics network design problem considering financing decisions: a case study of cement logistics*. Environmental Science and Pollution Research*, *29(*3), 4232–4245.

Farzaneh, M., Rezapour, S., Fovargue, R., & Neeson, T. (2021). Stochastic mathematical models to balance human and environmental water needs and select the best conservation policy for drought-prone river basins. *Journal of Cleaner Production*, *291*, 125230.

Freightos (2022). Transit time calculator for sea, air & port to port shipping. https://www.freightos.com/freight-resources/transit-time-calculator/ [Accessed on 06-07-2022].

Gerbens-Leenes, W., & Hoekstra, A. Y. (2011). The water footprint of biofuel-based transport. *Energy and Environmental Science*, *4*(8), 2658–2668.

Glock, C. H., & Kim, T. (2014). Shipment consolidation in a multiple-vendor-single-buyer integrated inventory model. *Computers and Industrial Engineering*, *70*(1), 31–42.

Golpîra, H., Zandieh, M., Najafi, E., & Sadi-Nezhad, S. (2017). A multi-objective, multi-echelon green supply chain network design problem with risk-averse retailers in an uncertain environment. *Scientia Iranica*, *24*(1), 413–423.

GoNoman 2022, Freighter travel costs and question. https://www.gonomad.com/1560-freighter-travelfaqs#:~:text=The%20average%20length%20of%20a,to%20100%20days%20or%20more [Accessed on 27-05-2022].

Goulder, L. H., & Schein, A. R. (2013). Carbon taxes versus cap and trade: A critical review. *Climate Change Economics, 4*(3), 1350010.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Govindan, K., Fattahi, M., & Keyvanshokooh, E. (2017). Supply chain network design under uncertainty: A comprehensive review and future research directions. *European Journal of Operational Research*, *263*(1), 108–141.

Gu, Y., Xu, J., Keller, A. A., Yuan, D., Li, Y., Zhang, B., Weng, Q., Zhang, X., Deng, P., Wang, H., & Li, F. (2015). Calculation of water footprint of the iron and steel industry: A case study in Eastern China*. Journal of Cleaner Production, 92,* 274–281.

Haider, Z., Hu, Y., Charkhgard, H., Himmelgreen, D., & Kwon, C. (2020). Creating grocery delivery hubs for food deserts at local convenience stores via spatial and temporal consolidation*. Socio-Economic Planning Sciences*, *82*(B), 101301.

Higginson, J. K. (1994). Shipment consolidation in business logistics management. [Doctoral dissertation, University of Waterloo].

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. *OTexts.*

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688.

Kabadurmus, O., & Erdogan, M. S. (2020). Sustainable, multimodal and reliable supply chain design. *Annals of Operations Research, 292*(1), 47–70.

Kaggle (2018). Store Item Demand Forecasting Challenge. https://www.kaggle.com/c/demand-forecasting-kernels-only [Accessed on 12-01-2022].

Kang, K., Hong, K. sung, Kim, K. H., & Lee, C. (2017). Shipment consolidation policy under uncertainty of customer order for sustainable supply chain management. *Sustainability, 9*(9).

Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak Bulbul, B., Ekmis, M. A., & Silva, T. C. (2019). An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain. *Complexity*, 2019.

Kotu, V., & Deshpande, B. (2015). Data mining process. *Predictive Analytics and Data Mining*, *1*, 17-36.

Kumar, V., Das, D., & Bardhan, A. K. (2022). Carbon Footprint Implications of Demand and Supply Uncertainties in Supply Chains: A Simulation Study. *International Journal of Modelling and Simulation, 42*(4), 680–705.

Le, T. P. N., & Lee, T. R. (2013). Model selection with considering the CO2 emission alone the global supply chain. *Journal of Intelligent Manufacturing, 24*(4), 653–672.

Li, C., & Grossmann, I. E. (2021). A review of stochastic programming methods for optimization of process systems under uncertainty. *Frontiers in Chemical Engineering*, *2*, 622241.

Liotta, G., Stecca, G., & Kaihara, T. (2015). Optimisation of freight flows and sourcing in sustainable production and transportation networks. *International Journal of Production Economics*, 164, 351–365.

López, A. R., Cáceres, D S. (2020). Defining the concept of city hub in urban consolidation schemes. A case study in the city of Gothenburg. [Master's thesis, Chalmers University of Technology, Gothenburg, Sweden].

Luo, W., Liu, W., & Gao, S. (2017, July). Remembering history with convolutional LSTM for anomaly detection. *In 2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 439-444). IEEE.

Ma, L., & Liu, Y. (2017). A stochastic chance constrained closed-loop supply chain network design model with VaR criterion. *Journal of Uncertain Systems, 11*(4), 306-320.

Manzardo, A.; Ren, J.; Piantella, A.; Mazzi, A.; Fedele, A.; Scipioni, A. Integration of water footprint accounting and costs for optimal chemical pulp supply mix in paper industry. *Journal of Cleaner Production*, *72*, 167–173.

Mogale, D. G., De, A., Ghadge, A., & Aktas, E. (2022). Multi-objective modelling of sustainable closed-loop supply chain network with price-sensitive demand and consumer's incentives. *Computers and Industrial Engineering, 168,* 108105.

Moheb-Alizadeh, H., & Handfield, R. (2018). An integrated chance-constrained stochastic model for efficient and sustainable supplier selection and order allocation. *International Journal of Production Research, 56*(21), 6890–6916.

Moradinasab, N., Amin-Naseri, M. R., Jafari Behbahani, T., & Jafarzadeh, H. (2018). Competition and cooperation between supply chains in multi-objective petroleum green supply chain: A game theoretic approach. *Journal of Cleaner Production*, *170*, 818–841.

Mostert, M., Caris, A., & Limbourg, S. (2018). Intermodal network design: a three-mode bi-objective model applied to the case of Belgium. *Flexible Services and Manufacturing Journal*, *30*(3), 397–420.

Muñoz-Villamizar, A., Velázquez-Martínez, J. C., Mejía-Argueta, C., & Gámez-Pérez, K. (2022). The impact of shipment consolidation strategies for green home delivery: a case study in a Mexican retail company. *International Journal of Production Research, 60*(8), 2443–2460.

1Muriel, A., Chugh, T., & Prokle, M. (2022). Efficient algorithms for the joint replenishment problem with minimum order quantities. *European Journal of Operational Research, 300*(1), 137–150.

Musmanno, R., Laporte, G., & Ghiani, G. (2004). Introduction to logistics systems planning and control. *OR/MS Today*, *31*(6), 66-67.

Najjar, M., Small, M. H., & Yasin, M. M. (2020). Social sustainability strategy across the supply chain: A conceptual approach from the organisational perspective. *Sustainability (Switzerland)*, *12*(24), 1–16.

Okwuchi, I., Nassar, L., Karray, F., & Ponnambalam, K. (2020, October). Deep learning ensemble based model for time series forecasting across multiple applications. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3077-3083). IEEE.

Park, Y. S., Szmerekovsky, J., Osmani, A., & Aslaam, N. M. (2017). Integrated multimodal transportation model for a switchgrass-based bioethanol supply chain: Case study in North Dakota. *Transportation Research Record*, *2628*(1), 32–41.

Plappally, A. K. (2012). Energy requirements for water production, treatment, end use, reclamation, and disposal. *Renewable and Sustainable Energy Reviews*, *16*(7), 4818-4848.

Ridoutt, B. G., Juliano, P., Sanguansri, P., & Sellahewa, J. (2010). The water footprint of food waste: Case study of fresh mango in Australia. *Journal of Cleaner Production*, *18*(16–17), 1714–1721.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, *2*, 21-42.

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, *26*(7), 1443-1471.

Rondinelli, D., & Berry, M. (2000). Multimodal transportation, logistics, and the environment: managing interactions in a global economy. *European Management Journal, 18*(4), 398-410.

Saif, A., & Elhedhli, S. (2016). Cold supply chain design with environmental considerations: A simulation-optimization approach. *European Journal of Operational Research, 251*(1), 274–287.

Sajedi, S., Sarfaraz, A. H., Bamdad, S., & Damghani, K. K. (2020). Designing a sustainable reverse logistics network considering the conditional value at risk and uncertainty of demand under different quality and market scenarios. *International Journal of Engineering, Transactions B: Applications, 33*(11), 2252–2271.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3).

Scarf, H. (1958). A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production. Stanford University Press.*

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N., Clark, D., . . . Kabat, P. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences - PNAS, 111*(9), 3245-3250.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Serrano, C., Delorme, X., & Dolgui, A. (2017). Scheduling of truck arrivals, truck departures and shop-floor operation in a cross-dock platform, based on trucks loading plans. *International Journal of Production Economics*, *194*, 102-112.

Shaw, K., Irfan, M., Shankar, R., & Yadav, S. S. (2016). Low carbon chance constrained supply chain network design problem: A Benders decomposition-based approach. *Computers and Industrial Engineering*, *98*, 483–497.

Shi, W., Xiao, Y., Chen, Z., McLaughlin, H., & Li, K. X. (2018). Evolution of green shipping research: themes and methods. *Maritime Policy & Management*, *45*(7), 863-876.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *28*(2015), 802–810.

Shukla, M., & Jharkharia, S. (2013). Agri-fresh produce supply chain management: a state-of-the-art literature review. *International Journal of Operations & Production Management. 33* (2013), 114-158.

Tiller, S. J., Rhindress, A. P., Oguntola, I. O., Ülkü, M. A., Williams, K. A., & Sundararajan, B. (2022). Exploring the impact of climate change on arctic shipping through the lenses of quadruple bottom line and sustainable development goals. *Sustainability (Switzerland), 14*(4), 2193.

Tirkolaee, E. B., Sadeghi, S., Mooseloo, F. M., Vandchali, H. R., & Aeini, S. (2021). Application of machine learning in supply chain management: A comprehensive overview of the main areas. *Mathematical Problems in Engineering, 2021*.

Ülkü, M. (2012). Dare to care: Shipment consolidation reduces not only costs, but also environmental damage*. International Journal of Production Economics, 139*(2), 438-446.

Ülkü, M. A. (2009). Analysis of shipment consolidation in the logistics supply chain. [Doctoral dissertation, University of Waterloo].

Ülkü, M. A., & Engau, A. (2021). Sustainable Supply Chain Analytics. *Industry, Innovation and Infrastructure, Encyclopedia of the UN Sustainable Development Goals.* 1123-1134.

UNFCC. (2008). Kyoto protocol reference manual on emissions and assigned amount. https://unfccc.int/kyoto_protocol [Accessed 08-04-2022].

UNFCCC. (2021). Conference of the parties serving as the meeting of the parties to the Paris Agreement. https://unfccc.int/kyoto_protocol [Accessed 08-04-2022].

United Nations Educational, Scientific and Cultural Organization. (2009). The United Nations world water development report 3. Water in a changing world. http://unesdoc.unesco.org/images/0018/001819/181993e.pdf [Accessed on 05-05-2022].

United Nations. (2021). The 17 Goals. Department of Economic and Social Affairs. https://sdgs.un.org/goals [Accessed on 20-05-2022].

United States Department of Energy. (2020). Average fuel economy by major vehicle categories. https://afdc.energy.gov/data/10310 [Accessed on 19-06-2022].

van der Laan, N., Teunter, R. H., Romeijnders, W., & Kilic, O. A. (2022). The data-driven newsvendor problem: Achieving on-target service-levels using distributionally robust chance-constrained optimization. *International Journal of Production Economics, 249,* 108509.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998-6008.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555-561.

Vrskova, R., Hudec, R., Kamencay, P., & Sykora, P. (2022). A new approach for abnormal human activities recognition based on ConvLSTM architecture. *Sensors, 22*(8), 2946.

Wang, R., Yang, K., Yang, L., & Gao, Z. (2018). Modeling and optimization of a road–rail intermodal transport system under uncertain information. *Engineering Applications of Artificial Intelligence, 72*, 423–436.

Wang, Y., Long, M., Wang, J., Gao, Z., & Yu, P. S. (2017). Predrnn: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. *Advances in Neural Information Processing Systems*, 30.

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science*, *341*(6145), 508-513.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241-259.

Xie, F., Huang, Y., & Eksioglu, S. (2014). Integrating multimodal transport into cellulosic biofuel supply chain design under feedstock seasonality with a case study based on California. *Bioresource Technology*, *152*, 15–23.

Yildirim, Ö. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology And Medicine*, *96,* 189-202.

Yu, S., & Hou, Q. (2021). Supply Chain Investment in Carbon Emission-Reducing Technology Based on Stochasticity and Low-Carbon Preferences. *Complexity, 2021*.

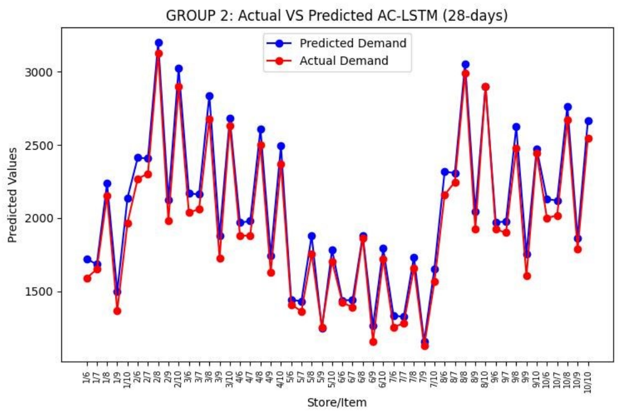
Zhang, D., Zou, F., Li, S., & Zhou, L. (2017). Green supply chain network design with economies of scale and environmental concerns*. Journal of Advanced Transportation 2017*(2017), 1-14.

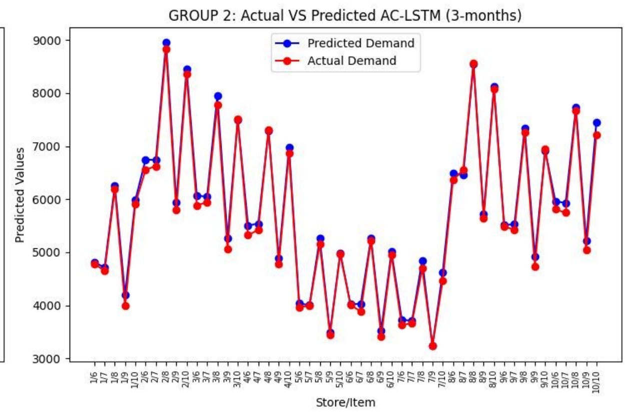
Zimmermann, H. J. (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems*, *1*(1), 45-55.

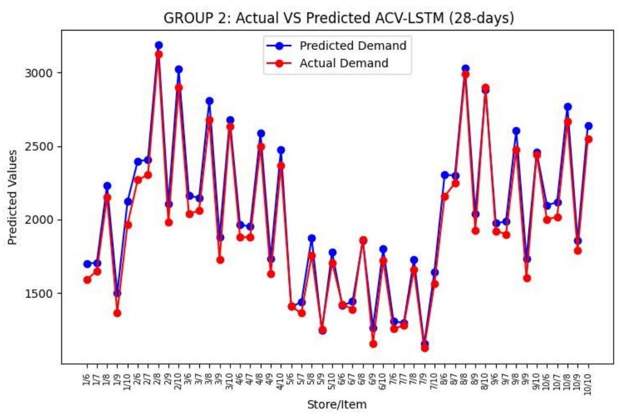
Appendix

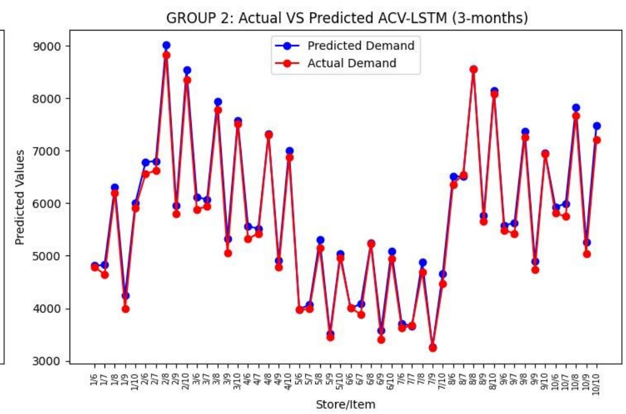
The results and graphs obtained for Groups 2 to 10 are provided in this chapter.

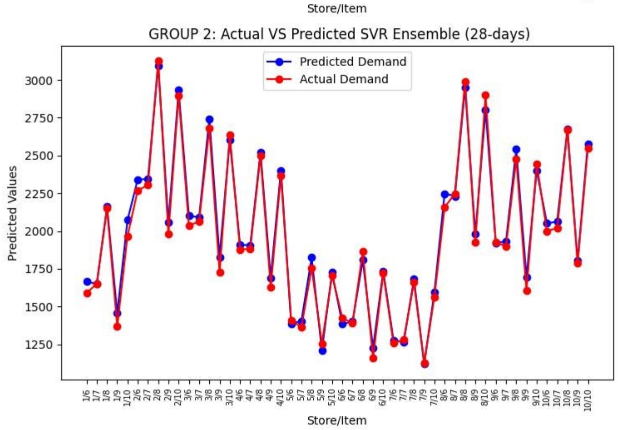
<u>Group 2</u>

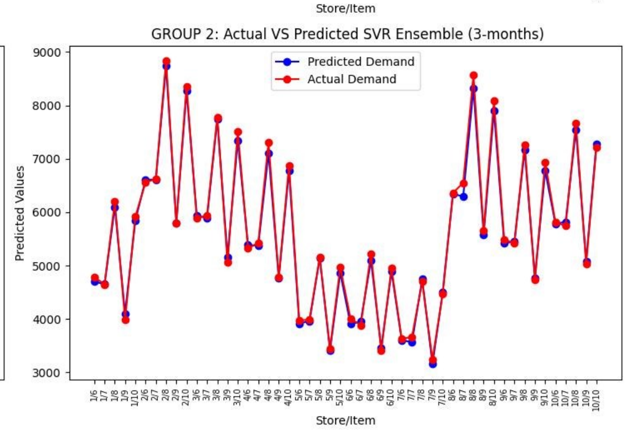






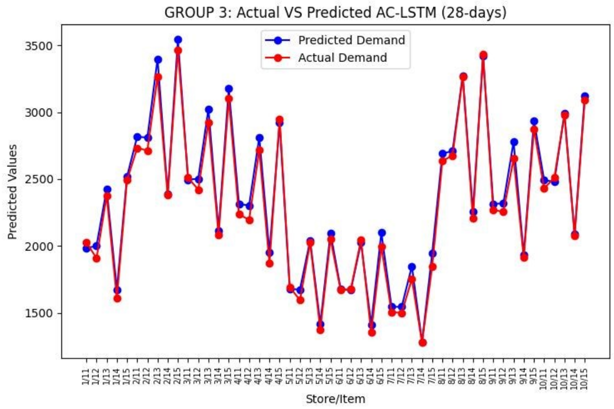


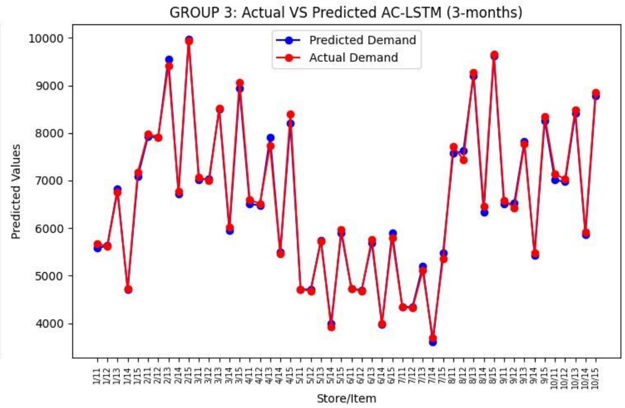


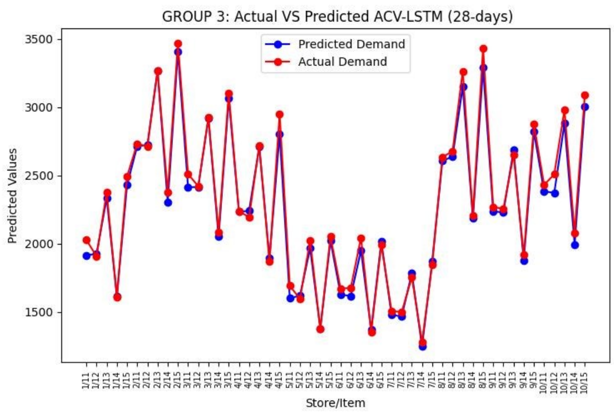


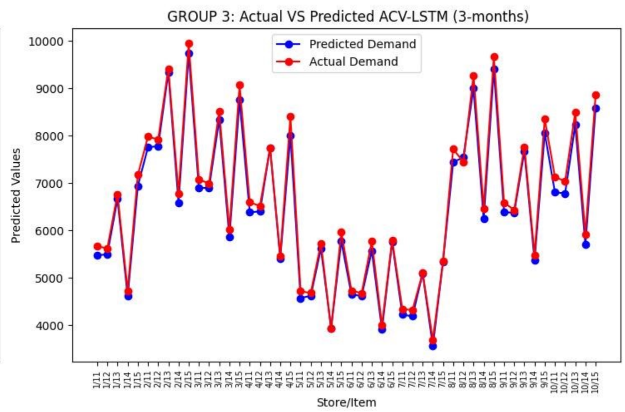
Model (Group 2)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	Total OUC	Optimization Model Cost	MAE	Total OUC	Optimization Model Cost
SP	Simple Recourse	139.6	964,335	7,072,519	215.6	1,984,764	20,231,160
	Chance constraint (α=0.1)	120.6	880,005	7,038,212	220.2	2,262,029	20,227,873
	Chance constraint (α=0.2)	115.9	791,854	7,023,798	219.8	2,259,101	20,222,094
ML	AC-LSTM	90.5	156,037	6,807,077	99.1	190,872	19,948,809
	ACV-LSTM	81.9	144,916	6,790,158	128.9	231,158	20,076,401
	Stacking Ensemble	42.7	117,727	6,740,396	77.0	448,586	19,880,104
Baselines	Model with mean values	147.2	1,001,314	7,085,294	216.0	1,502,304	20,246,691
	Actual Demand	0	0	6,701,768	0	0	19,714,867

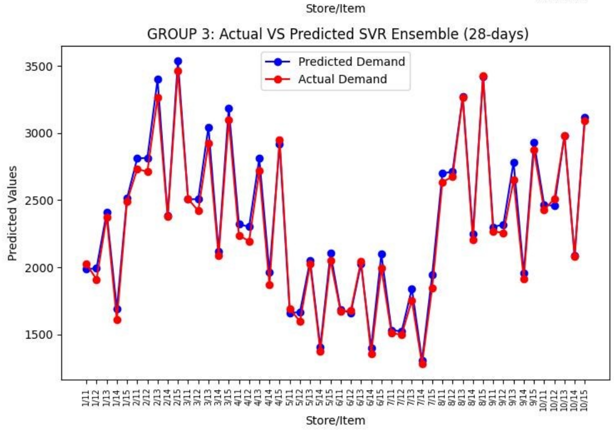
<u>Group 3</u>

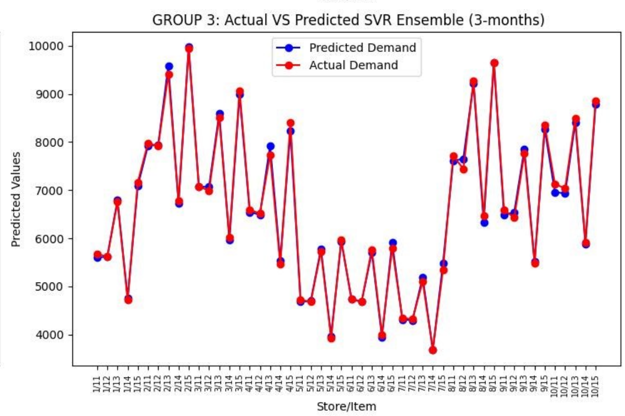

GROUP 3: Actual VS Predicted AC-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 3: Actual VS Predicted AC-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 3: Actual VS Predicted ACV-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

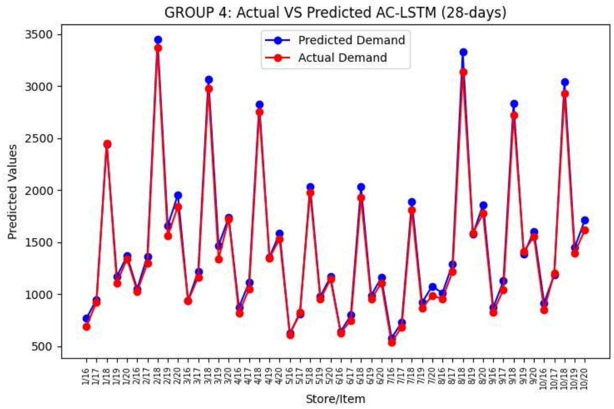

GROUP 3: Actual VS Predicted ACV-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

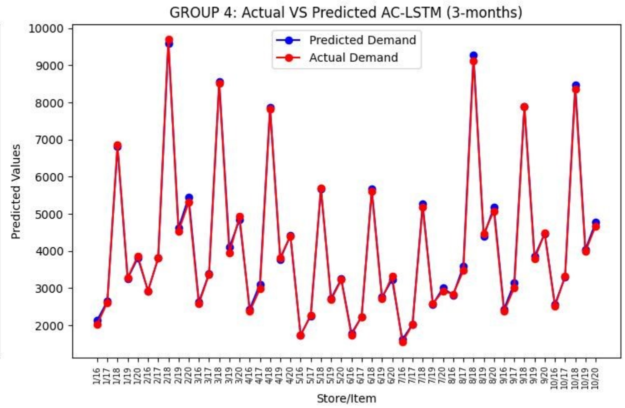

GROUP 3: Actual VS Predicted SVR Ensemble (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

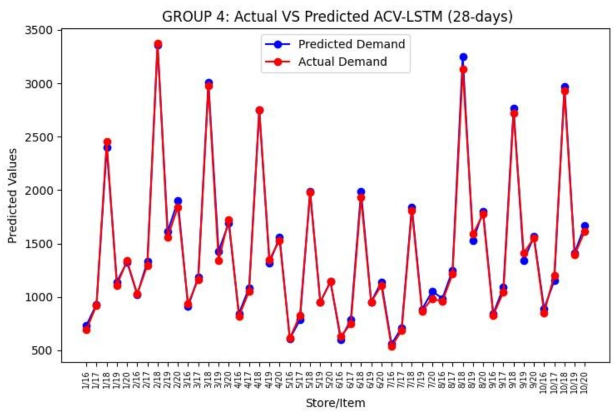

GROUP 3: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

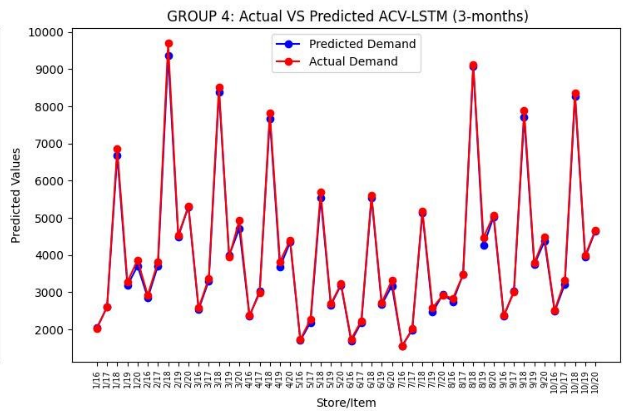
Model (Group 3)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	**Total OUC**	**Optimization Model Cost**	**MAE**	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	70.5	1,140,646	8,132,071	280.4	2,204,239	21,325,571
	Chance constraint (α=0.1)	75.4	1,148,838	8,135,178	282.1	2,459,214	21,403,700
	Chance constraint (α=0.2)	68.3	897,047	8,037,237	279.8	2,456,088	21,402,332
ML	AC-LSTM	53.1	106,390	7,785,767	68.1	302,743	20,741,281
	ACV-LSTM	47.2	290,331	7,831,955	157.7	973,788	20,804,298
	Stacking Ensemble	53.2	109,881	7,736,867	67.4	270,918	20,782,818
Baselines	Model with mean values	180.3	1,207,656	8,155,155	271.6	1,695,665	21,141,360
	Actual Demand	0	0	7,695,000	0	0	20,616,088

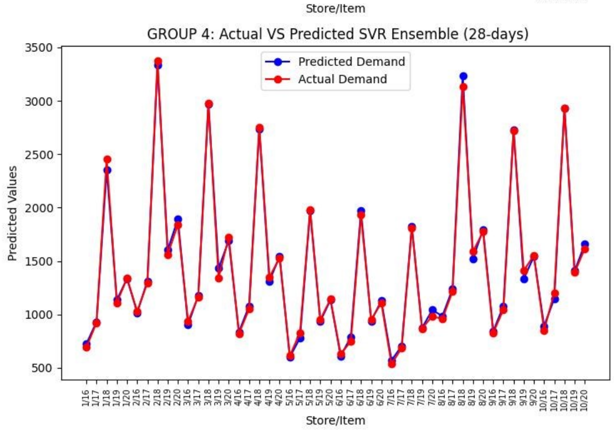
Group 4

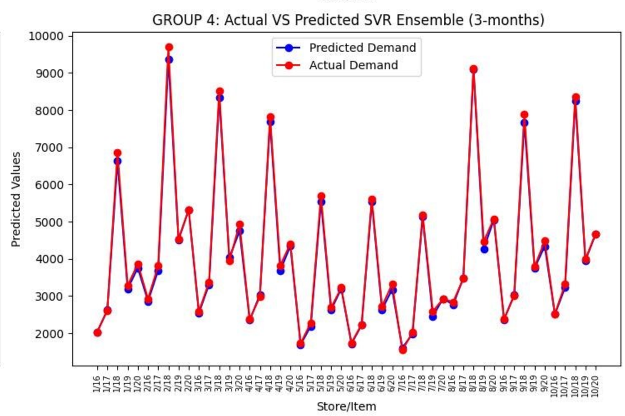

GROUP 4: Actual VS Predicted AC-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 4: Actual VS Predicted AC-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 4: Actual VS Predicted ACV-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

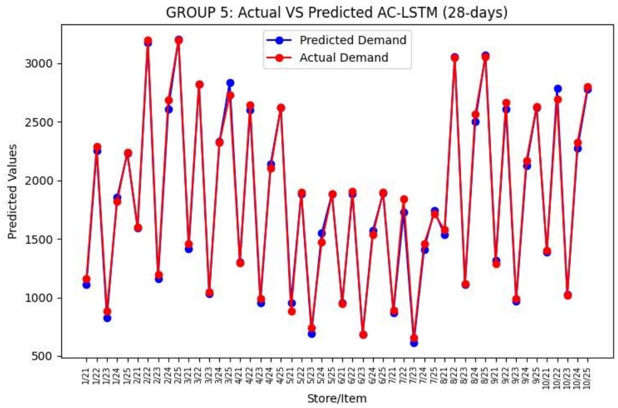

GROUP 4: Actual VS Predicted ACV-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

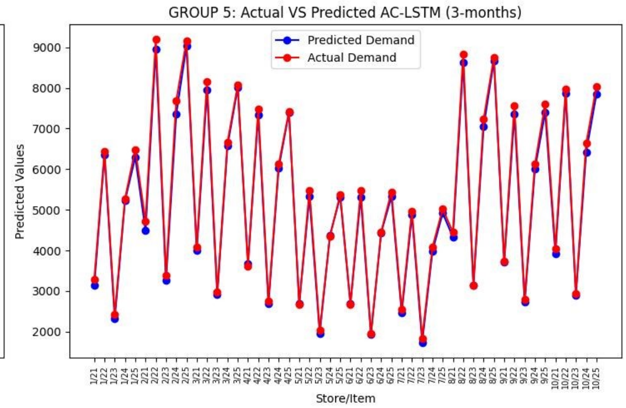

GROUP 4: Actual VS Predicted SVR Ensemble (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

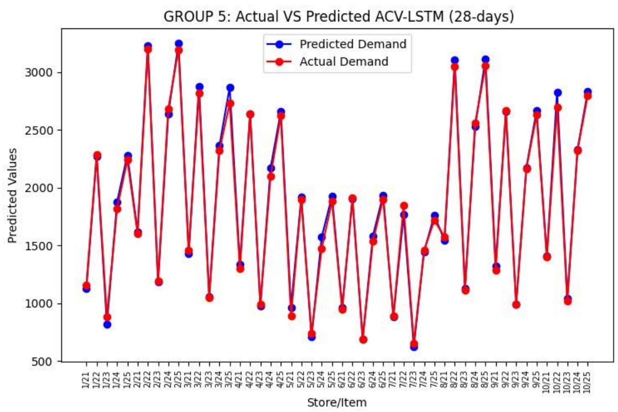

GROUP 4: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

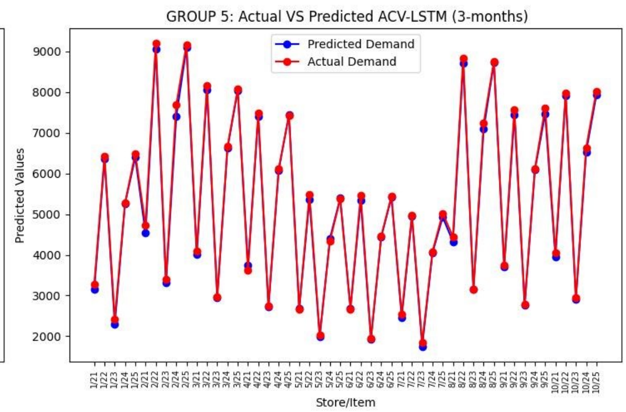
Model (Group 4)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	Total OUC	Optimization Model Cost	MAE	Total OUC	Optimization Model Cost
SP	Simple Recourse	107.2	775,304	5,373,454	150.2	1,238,868	13,428,341
	Chance constraint (α=0.1)	87.7	642,575	5,324,003	152.5	1,400,477	13,491,512
	Chance constraint (α=0.2)	80.3	560,013	5,296,262	151.5	1,270,504	13,447,714
ML	AC-LSTM	58.0	114,095	5,132,994	58.3	153,699	13,052,736
	ACV-LSTM	33.3	105,638	5,112,966	78.1	476,312	13,164,687
	Stacking Ensemble	30.4	119,112	5,166,989	78.6	471,706	13,140,531
Baselines	Model with mean values	108.5	774,731	5,371,192	148.9	942,841	13,303,566
	Actual Demand	0	0	5,106,847	0	0	12,993,957

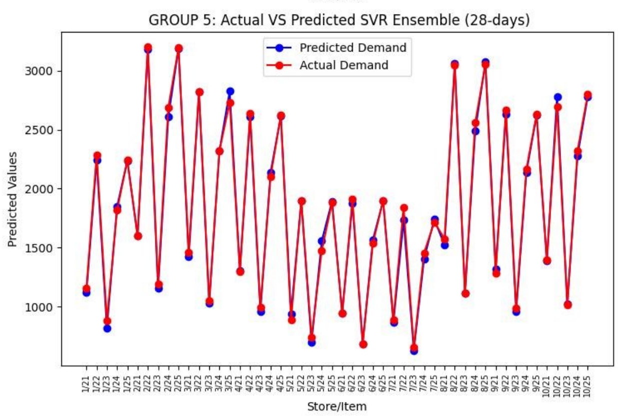
<u>Group 5</u>

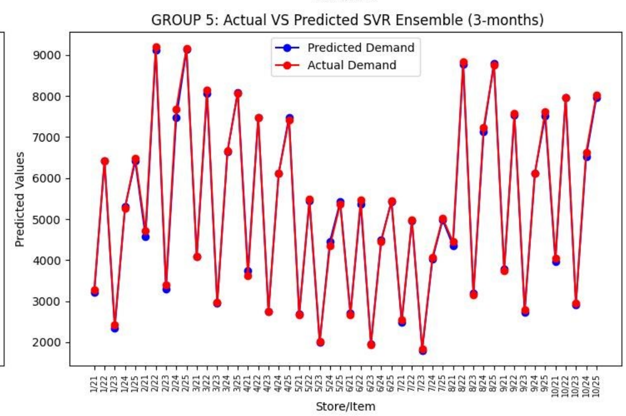

GROUP 5: Actual VS Predicted AC-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 5: Actual VS Predicted AC-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 5: Actual VS Predicted ACV-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

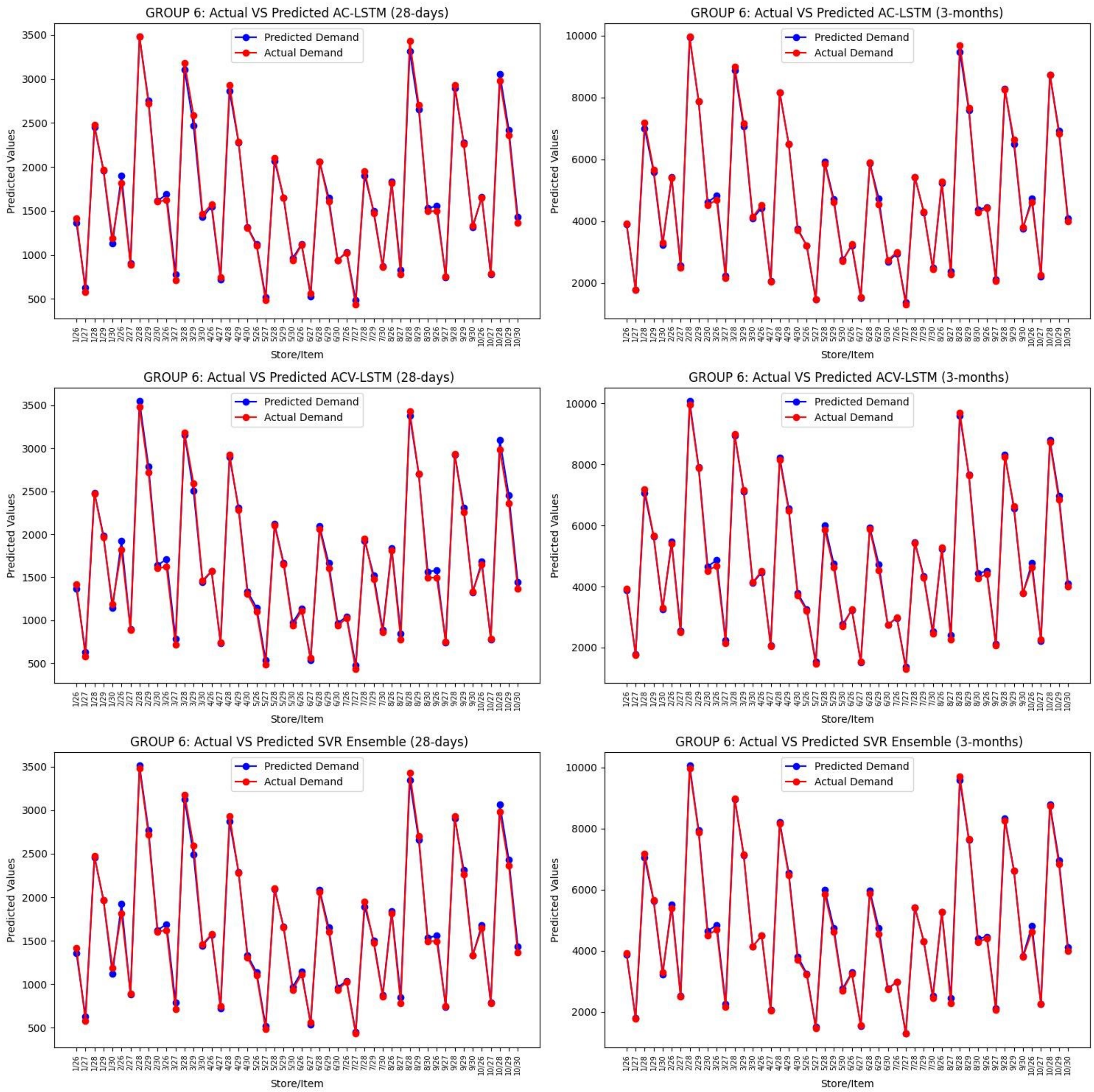

GROUP 5: Actual VS Predicted ACV-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 5: Actual VS Predicted SVR Ensemble (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 5: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

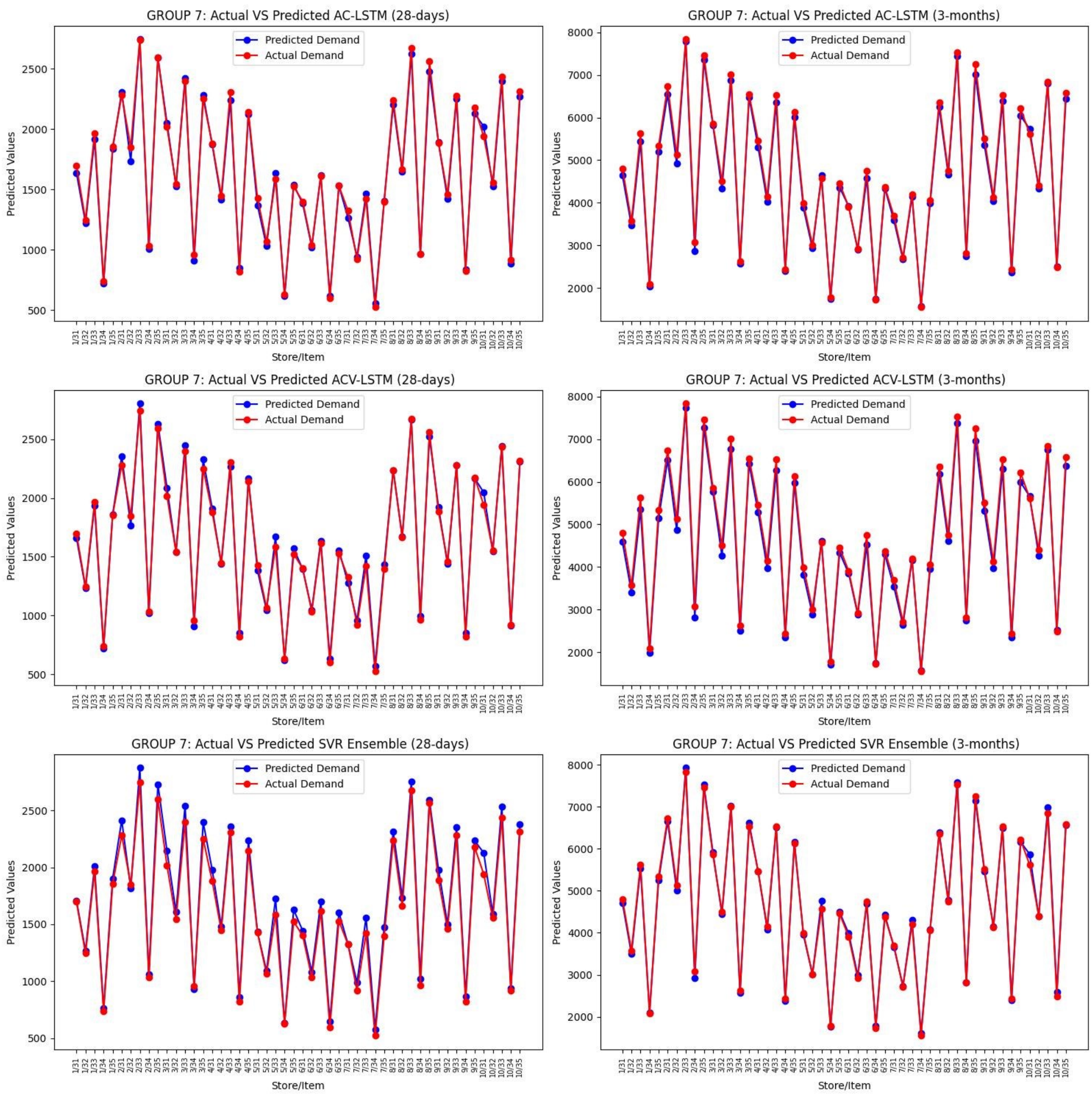
Model (Group 5)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	**Total OUC**	**Optimization Model Cost**	**MAE**	**Total OUC**	**Optimization Model Cost**
SP	Simple Recourse	145.5	920,012	6,368,357	201.9	1,888,855	19,671,473
	Chance constraint (α=0.1)	141.2	908,754	6,364,738	203.5	2,127,231	19,764,097
	Chance constraint (α=0.2)	121.6	734,582	6,290,905	203.5	2,127,231	19,766,110
ML	AC-LSTM	32.9	157,546	6,127,640	109.3	768,058	19,306,171
	ACV-LSTM	37.2	104,005	6,051,385	72.9	492,931	19,180,611
	Stacking Ensemble	30.5	146,124	6,081,689	52.2	304,420	19,053,548
Baselines	Model with mean values	147.89	964,782	6,383,938	202.1	1,452,887	19,507,390
	Actual Demand	0	0	6,024,914	0	0	18,991,453

Group 6


GROUP 6: Actual VS Predicted AC-LSTM (28-days)
GROUP 6: Actual VS Predicted AC-LSTM (3-months)
GROUP 6: Actual VS Predicted ACV-LSTM (28-days)
GROUP 6: Actual VS Predicted ACV-LSTM (3-months)
GROUP 6: Actual VS Predicted SVR Ensemble (28-days)
GROUP 6: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

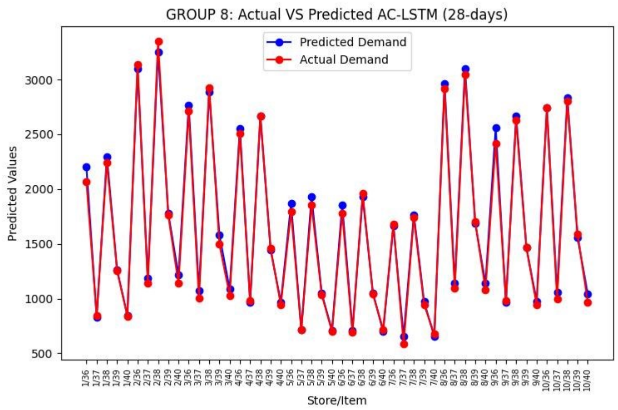
Model (Group 6)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	Total OUC	Optimization Model Cost	MAE	Total OUC	Optimization Model Cost
SP	Simple Recourse	128.5	823,907	5,538,496	180.2	1,408,814	15,023,548
	Chance constraint (α=0.1)	115.3	780,803	5,525,434	170.3	1,206,797	14,967,626
	Chance constraint (α=0.2)	105.6	725,752	5,509,965	175.5	1,610,008	15,064,017
ML	AC-LSTM	35.6	136,884	5,410,440	52.5	233,040	14,606,825
	ACV-LSTM	40.8	102,402	5,309,248	72.1	174,071	14,531,681
	Stacking Ensemble	38.2	122,907	5,334,767	67.8	153,226	14,502,892
Baselines	Model with mean values	133.4	842,543	5,542,981	175.7	1,060,885	14,861,598
	Actual Demand	0	0	5,259,048	0	0	14,451,284

Group 7

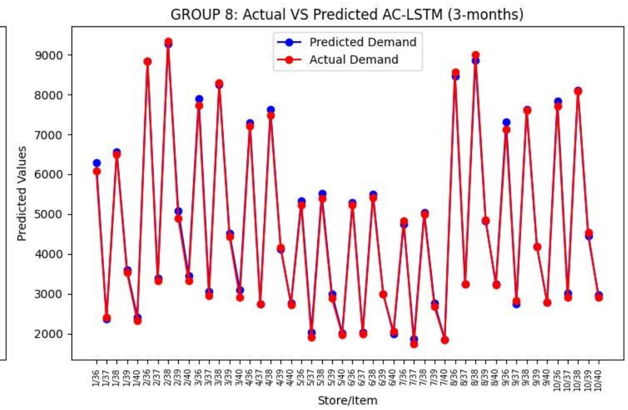

GROUP 7: Actual VS Predicted AC-LSTM (28-days)
GROUP 7: Actual VS Predicted AC-LSTM (3-months)
GROUP 7: Actual VS Predicted ACV-LSTM (28-days)
GROUP 7: Actual VS Predicted ACV-LSTM (3-months)
GROUP 7: Actual VS Predicted SVR Ensemble (28-days)
GROUP 7: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

Model (Group 7)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	Total OUC	Optimization Model Cost	MAE	Total OUC	Optimization Model Cost
SP	Simple Recourse	90.2	743,848	5,255,589	181.2	1,749,540	18,091,028
	Chance constraint (α=0.1)	89.9	671,425	5,212,582	193.5	1,987,195	18,189,266
	Chance constraint (α=0.2)	97.3	641,199	5,214,510	190.7	1,987,492	18,188,732
ML	AC-LSTM	32.1	151,892	5,081,991	99.8	709,691	17,703,366
	ACV-LSTM	32.9	95,844	5,036,467	142.2	1,038,734	17,828,332
	Stacking Ensemble	67.5	109,028	5,297,945	59.9	262,453	17,650,717
Baselines	Model with mean values	128.5	777,689	5,260,351	176.8	1,308,846	17,918,642
	Actual Demand	0	0	4,961,295	0	0	17,402,456

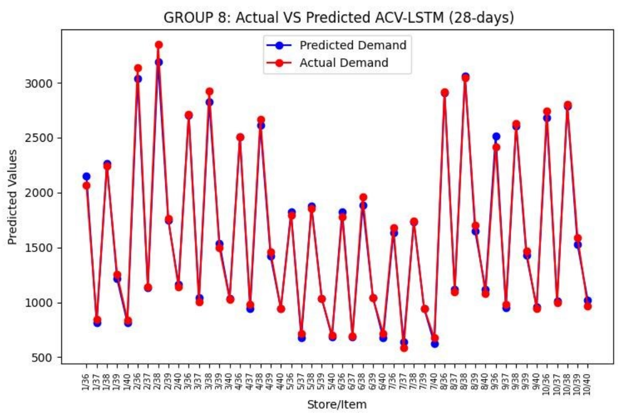
Group 8



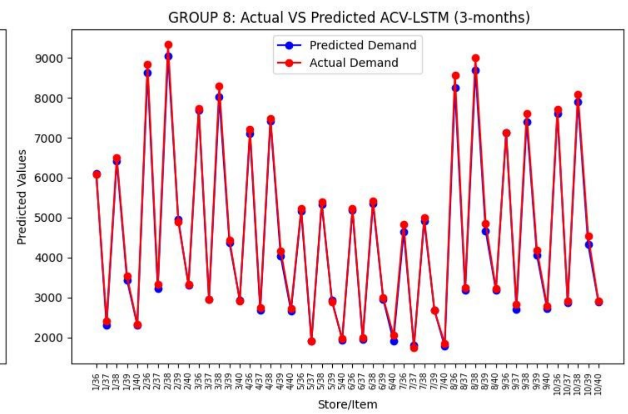
GROUP 8: Actual VS Predicted AC-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item



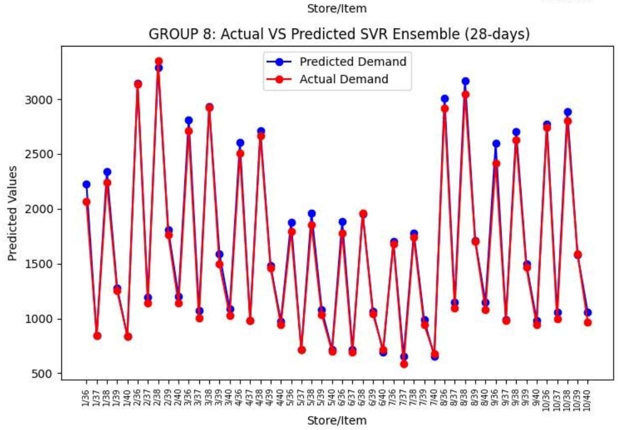
GROUP 8: Actual VS Predicted AC-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item



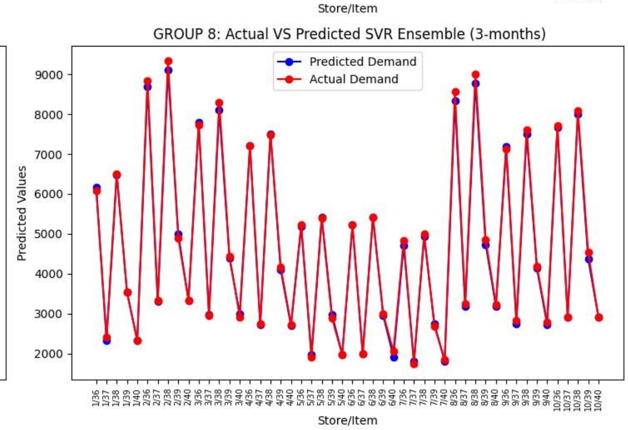
GROUP 8: Actual VS Predicted ACV-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item



GROUP 8: Actual VS Predicted ACV-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item



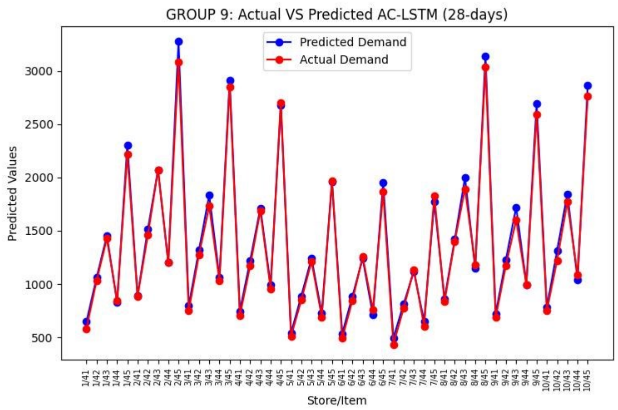
GROUP 8: Actual VS Predicted SVR Ensemble (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

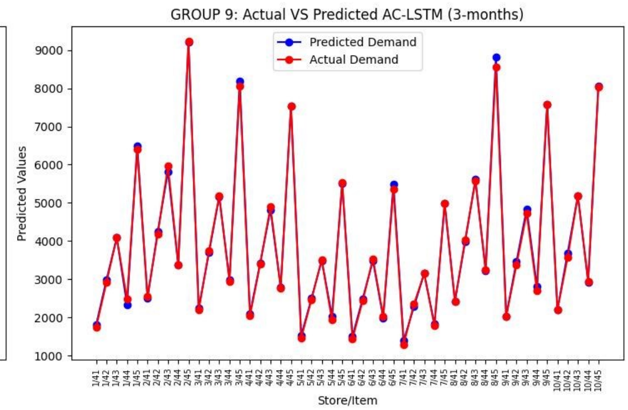


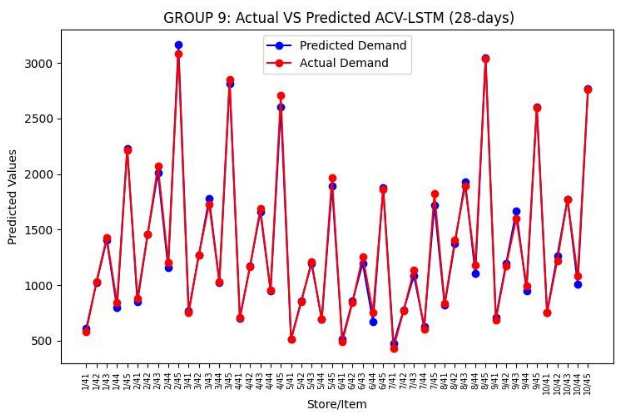
GROUP 8: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

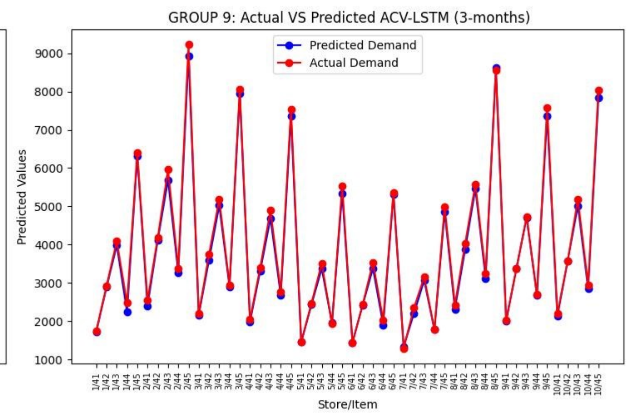
Model (Group 8)		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	Total OUC	Optimization Model Cost	MAE	Total OUC	Optimization Model Cost
SP	Simple Recourse	132.6	946,408	6,197,667	204.1	1,610,210	14,949,048
	Chance constraint (α=0.1)	120.2	831,426	6,161,549	205.6	1,611,193	14,948,670
	Chance constraint (α=0.2)	122.3	844,997	6,166,178	209.3	1,785,206	15,020,518
ML	AC-LSTM	40.9	116,258	6,164,600	77.9	192,710	14,636,752
	ACV-LSTM	35.9	191,112	6,033,082	96.6	568,399	14,283,050
	Stacking Ensemble	52.8	108,524	6,097,639	67.8	341,979	14,046,616
Baselines	Model with mean values	133.2	968,149	6,204,242	203.6	1,247,323	14,803,605
	Actual Demand	0	0	5,893,894	0	0	14,251,628

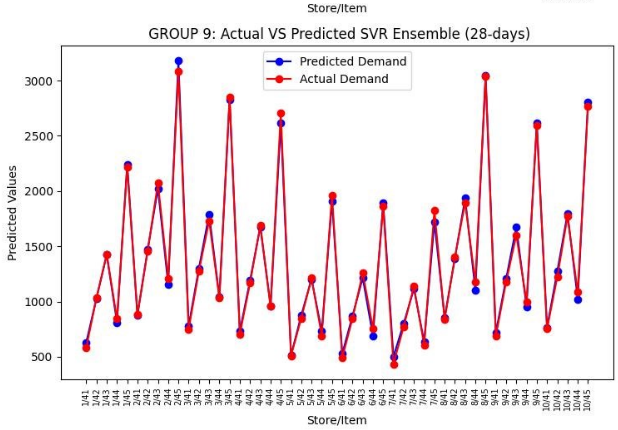
Group 9

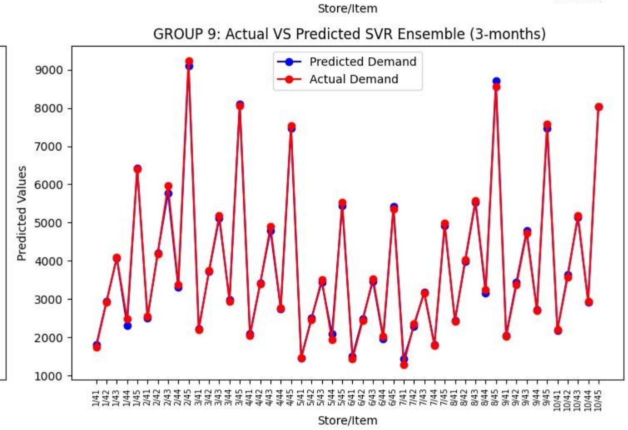

GROUP 9: Actual VS Predicted AC-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 9: Actual VS Predicted AC-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 9: Actual VS Predicted ACV-LSTM (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 9: Actual VS Predicted ACV-LSTM (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 9: Actual VS Predicted SVR Ensemble (28-days)
Predicted Demand
Actual Demand
Predicted Values
Store/Item


GROUP 9: Actual VS Predicted SVR Ensemble (3-months)
Predicted Demand
Actual Demand
Predicted Values
Store/Item

Model (Group 9		Performance Metrics (28-days)			Performance Metrics (3-months)		
		MAE	Total OUC	Optimization Model Cost	MAE	Total OUC	Optimization Model Cost
SP	Simple Recourse	80.2	588,365	4,328,512	152.4	1,590,448	16,866,840
	Chance constraint (α=0.1)	81.9	523,449	4,302,186	161.2	1,804,709	16,939,816
	Chance constraint (α=0.2)	75.6	465,349	4,279,878	163.2	1,807,037	16,940,754
ML	AC-LSTM	49.3	99,570	4,172,795	54.2	210,352	16,362,881
	ACV-LSTM	32.1	142,797	4,187,516	100.5	815,054	16,569,990
	Stacking Ensemble	35.5	123,743	4,162,359	59.1	325,547	16,352,258
Baselines	Model with mean values	102.9	619,657	4,340,593	147.3	1,217,052	16,694,665
	Actual Demand	0	0	4,111,744	0	0	16,225,171