

Shipment Consolidation in Hub Location Modelling

by

Khaled Mostafa Shah

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Management Sciences

Waterloo, Ontario, Canada, 2024

© Khaled Mostafa Shah 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

- External Examiner: Maryam Darvish, Associate Professor
Department of Operations and Decision Systems,
Université Laval
- Supervisor(s): James H. Bookbinder, Professor
Sibel Alumur Alev, Associate Professor
Department of Management Science and Engineering,
University of Waterloo
- Internal Members: Fatma Gzara, Professor
Houra Mahmoudzadeh, Associate Professor
Department of Management Science and Engineering,
University of Waterloo
- Internal-External Member: Christian Bachmann, Associate Professor
Department of Civil and Environmental Engineering,
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 3 of this thesis is based on a submitted paper, which has been co-authored by myself and my supervisors, Dr. James H. Bookbinder and Dr. Sibel Alumur Alev.

I am the sole author of the rest of this thesis.

Abstract

This thesis examines the integration of shipment consolidation and inventory holding decisions with the design of hub networks, where a heterogeneous fleet is used for inter-hub transportation. These problems aim to determine the optimal locations of hubs, allocate demand nodes to them, decide the quantity of inventory to be held at hubs, and assign vehicles from the heterogeneous fleet to operate on hub arcs. This is done while considering inventory storage space at hubs and vehicle capacities. These decisions are made to fulfill a set of demands for flows between origin-destination pairs, distributed across a multi-period planning horizon. Prospective applications of this work include air cargo, less-than-truckload transportation, express shipments, and postal delivery systems.

The first problem studies shipment consolidation and inventory holding in a single allocation hub location problem, which considers a promised delivery time for goods. The goal is to jointly determine the optimal hub locations, allocation of demand nodes to hubs, inventory levels to be held at hubs in each time period, and the type and number of vehicles to be dispatched on inter-hub links in different periods. The objective is to minimize the total cost of infrastructure, transportation, and inventory holding. This problem is formulated as a mixed-integer programming model. To solve larger instances, we propose a variable neighborhood search-based algorithm that incorporates a shipment consolidation sub-problem. Computational experiments with varying parameter values, conducted on a realistically created dataset, demonstrate the comparative effectiveness of the networks generated by both the model and algorithm.

The second problem in this thesis explores the integration of operational decisions re-

garding shipment consolidation and inventory holding within a hub arc location model for a freight network. In contrast to the first problem, the focus is on locating *hub arcs* rather than hub nodes. The allocation of demand nodes to multiple hub arcs is allowed in this setting for greater flexibility in routing and consolidation, and the optimal frequency and types of vehicles operating on those hub arcs are also to be determined. A mixed-integer programming model is proposed to minimize the total cost of hub arcs, transportation, and inventory. It determines the optimal location of hub arcs, allocation of demand nodes to these arcs, inventory levels at the origin nodes of hub arcs in each time period, and the type and frequency of vehicles from a heterogeneous fleet to operate on the hub arcs. To solve the problem efficiently, a Benders decomposition-based methodology is developed and further enhanced with Pareto-optimal cuts and two variable fixing techniques. Computational experiments compare the effectiveness of the proposed solution algorithm against a commercial solver and provide insights into the factors influencing these decisions.

The third problem addresses demand uncertainty in the hub arc location problem, integrated with decisions on shipment consolidation and inventory holding. The aim is to minimize total expected costs of infrastructure, transportation, and inventory, modelling the problem as a two-stage stochastic program. The first-stage decisions involve selecting hub arcs and determining the types and frequencies of vehicles. In the second stage, nodes are assigned to hub arcs, and decisions are made on inventory holding and the optimal routing of flows. To solve this, a Sample Average Approximation methodology is employed. Computational experiments assess the algorithm's convergence, explore the impact of varying costs and frequencies, and estimate the value of considering uncertainty on the decision-making process.

Acknowledgements

I would like to thank my supervisors Dr. James H. Bookbinder and Dr. Sibel Alumur Alev for their advice and kind support. I would also like to acknowledge my relatives in Canada, and friends here - Esma Akgün, in particular. Last, but not least the SHAHs - Farwah and Fakhitah, and Fariha - my family, who played their part in the journey.

Dedication

This is dedicated to

my Parents

Shah Abdul Halim

Khodeja Sultana

Grandmother

Anwara Mannan

and

my Teacher

Ziaul Huq Qutubuddin

Table of Contents

Examining Committee Membership	ii
Author's Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements	vii
Dedication	viii
List of Figures	xii
List of Tables	xiii
1 Introduction	1
2 Literature Review	8
2.1 Hub Location Problems	9
2.2 Shipment Consolidation	12
2.3 Multi-period Hub Location Problems	17
2.4 Economies of Scale and Vehicle Utilization in Hub Location Problems . . .	21

2.5	Shipment Consolidation and Inventory Holding in Hub Location Problems	24
2.6	Hub Arc Location Problems	26
2.7	Benders Decomposition in Hub Location	30
2.8	Sample Average Approximation in Hub Location Problems	33
3	Shipment Consolidation in Hub Location with Multi-period Demand	36
3.1	Mathematical Formulation	38
3.2	Solution Methodology	44
3.2.1	Initial Solution	46
3.2.2	Shipment Consolidation Subproblem	47
3.2.3	Variable Neighborhood Search (VNS) Algorithm	50
3.3	Computational Analysis	57
3.3.1	Data and Parameters	57
3.3.2	Algorithmic Performance	60
3.3.3	Insights from the Results	65
3.3.4	Value of Holding Inventory	73
3.4	Summary	76
4	Shipment Consolidation in the Hub Arc Location Problem	78
4.1	Mathematical Formulation	80
4.2	Solution Methodology	87
4.2.1	Benders Reformulation and Algorithm	87
4.2.2	Improvement Strategies	90
4.3	Computational Analysis	95
4.3.1	Data and Parameters	95
4.3.2	Impact of Frequency Selection on Solutions	97
4.3.3	Algorithmic Performance	109
4.4	Summary	112

5	Shipment Consolidation in Hub Arc Location Problems under Uncertainty	115
5.1	Mathematical Formulation	116
5.2	Solution Methodology for Stochastic Model	119
5.3	Computational Analysis	123
5.3.1	Sample Generation	124
5.3.2	Convergence of SAA-results	124
5.3.3	Quality of Solutions	126
5.3.4	Value of Stochastic Solution	129
5.4	Summary	136
6	Conclusion	138
	References	146
	APPENDICES	169
A	List of Cities in the Data Sets	170

List of Figures

3.1	Examples of resulting hub networks.	70
4.1	Resulting hub networks for <i>low</i> vs <i>high</i> holding cost using 5-frequencies (colors are explained in the text)	104
4.2	Impact of using different frequencies (colors are explained in the text) . . .	108
5.1	Optimality gap and Total CPU Time for the SAA algorithm, across varying sample sizes and number of replications	125
5.2	Resultant networks for <i>med</i> hub arc cost and <i>high</i> holding cost (colors are explained in the text)	134

List of Tables

3.1	Comparison of the SCHL-algorithm with CPLEX on 10-, 15- and 20-city instances.	61
3.2	Results with the SCHL-algorithm for 30-, 45- and 65-city instances.	63
3.3	Numbers of hubs and flights of B-747-F and B-747-100, obtained with CPLEX and SCHL-algorithm on 10-, 15- and 20-city instances.	67
3.4	Results of SCHL-algorithm on 65-city instances with the demands of 2019, 2020, and 2022.	72
3.5	Comparison of the savings against inventory on 10- and 15-city instances. .	74
4.1	Optimal results for 10-city instances using 2, 3 and 5 frequencies	98
4.2	Optimal results for 15-city instances using 2, 3 and 5 frequencies	101
4.3	Algorithmic Performance for 10 city instances	110
4.4	Algorithmic Performance for 15 city instances	113
5.1	Computational results of the stochastic model for 10-node instances	128
5.2	Value of Stochastic Solution for 10-node instances	130
A.1	Cities in different instances	171

Chapter 1

Introduction

The *hub location problem* involves finding the best locations for hub facilities to serve a set of origin-destination (O-D) pairs, between which some flow has to be shipped. In many-to-many distribution networks, hubs serve as facilities, which do loading, unloading, sorting, connecting, switching, consolidation and break-bulk. Demands between O-D pairs in such networks are routed through hubs because direct shipments between origins and destinations are often too high in resource consumption, and hence expensive. Compared to a fully interconnected network, utilizing a subset of these nodes as hubs results in fewer links in the network to serve O-D pairs. Hub networks enable carriers to exploit economies of scale in transportation by consolidating flows through hubs, and thus reduce the costs of establishing a network and its operations. “Economies of scale” in the operation of a hub network means that when the operational activities expand, the cost may be decreased by routing flow through hubs.

Hubs are central transshipment facilities to serve the flow of commodities, people, or information between a large set of O’s and D’s. This is widely used in road, water and air

transportation ([Vidović et al., 2011](#)), postal delivery services ([Ernst and Krishnamoorthy, 1996](#)), and logistics systems ([Tan and Kara, 2007](#)). Particular application areas include, but are not limited to, air passenger transportation, freight transportation, express shipment, rapid transit, truckload and less than truckload transportation. These deal with physical facilities such as airports, depots, stations, warehouses, and physical carriers such as trucks, buses, trains, and aircraft. Additionally, hubs are used in telecommunication networks with particular applications in internet infrastructure, telephone networks, and distributed computing. Here data and information are routed from one endpoint to the other via switches and routers interconnected through cables, and microwave or fiber optic links. Different application areas of hub location have been articulated in the literature by a number of researchers ([Campbell et al., 2002](#); [Alumur and Kara, 2008](#); [Alumur et al., 2021](#); [Contreras, 2021](#)).

Hubs are essential to the functioning of a transportation common carrier, serving key roles in both last-mile operations and inter-hub connectivity. In last-mile operations, hubs facilitate the collection and delivery of shipments from origin points to their respective demand destinations. The efficiency of hub operations in this phase is critical for ensuring the timely fulfillment of customer orders, highlighting the central role hubs play in meeting service level expectations. In inter-hub operations, common carriers employ feeder vehicles—such as large tractor-trailers, ships, or aircraft—to transport freight between hubs. In this context, hubs function as sorting and consolidation centers for partial loads. These hubs store partial loads over one or more time periods, enabling the subsequent dispatch of larger consolidated shipments on inter-hub links. This approach maximizes the utilization of feeder vehicles by ensuring they operate at close to their capacity levels. The

goods stored at hubs, representing “holdover volumes,” are referred to as “inventory” at the consolidation point, in the terminology of this thesis.

Decreasing transportation costs improve profitability and give a competitive advantage to any logistics company. Common carrier companies face the challenge of improving holding cost and customer satisfaction on one hand, and improving transportation costs on the other. A common carrier may dispatch vehicles soon after receiving an order. This would lower holding costs and improve customer satisfaction. However, such a policy would increase the number of vehicles necessary to operate between two nodes of a supply chain, and the utilization of vehicles would be low. As opposed to this policy, the company can consolidate shipment orders and thereby optimally fill a vehicle before dispatching ([Higginson and Bookbinder, 1994](#)). Such consolidation may hurt customer satisfaction, and increase holding cost. Because there is a fixed cost involved in each vehicle dispatch, this will take advantage of economies of scale.

In a freight network, the shipper is connected to a sequence of hubs to route delivery to the destination. Distance from an origin or destination to a hub is typically shorter than the distance between two hubs. Vehicles with large capacities operate between hubs as carriers of deliveries. Deliveries collected at a hub can be consolidated so that the vehicle utilization of inter-hub carriers can be large, thus exploiting economies of scale. To do this, it may be necessary to hold individual shipments at hubs to dispatch a consolidated load at a later time. This stimulates the motivation of why the locations of hubs in a delivery network should be decided in a way that delivery operations can be performed optimally. In other words, consolidation decisions may have an effect on the locations of hubs, and therefore introducing shipment consolidation decisions in the hub location problem can lead

to a more efficient hub network for operations. Hub locations are chosen with a strategic perspective, as their effects influence operations over an extended period. However, these strategic location decisions enable operational-level advantages by facilitating the development of an effective shipment consolidation policy. This alignment between long-term planning and day-to-day operations enhances the overall efficiency and cost-effectiveness of the transportation network.

This thesis makes an effort to bridge the gap between research on tactical issues in shipment consolidation and the strategic considerations involved in hub location. It explores the impact of integrating inventory management within hubs into the overall design of a hub network, to understand the interplay between operational and strategic decisions in logistics optimization. To give a big picture, transportation and logistics play a pivotal role in overall economic activity. In 2023, the U.S. nominal Gross Domestic Product (GDP) was \$27.36 trillion, with business logistics costs—of which transportation accounts for approximately two-thirds—comprising 8.7% of that GDP ([CSCMP, 2024](#)). These aggregate statistics underscore the substantial savings potential within the logistics networks of firms of reasonable size. By optimizing the operation of a hub network, firms can achieve a more efficient balance of key operational costs, leading to significant financial benefits.

Motivated by the above aspects, in this thesis we study how shipment consolidation can improve hub location decisions. We define new problems aimed at developing models and algorithms that determine optimal network design, shipment consolidation and dispatch decisions to improve market competitiveness and enhance customer satisfaction. We contend that achieving optimal performance in freight delivery operations necessitates integrating fleet composition and consolidation decisions when determining hub locations.

This joint consideration ensures that both the selection of vehicle types and the consolidation of shipments are aligned with the strategic placement of hubs, ultimately enhancing operational efficiency and cost-effectiveness. We assume a scenario in which the freight carrier operates multiple types of vehicles between hubs, each associated with a fixed dispatch cost and a variable cost of operation. Local operations, including pickup and delivery to O-Ds, are outsourced to a third-party logistics provider (3PL). Under the terms of the contract, the 3PL handles these services at a specified rate per unit weight per unit distance. In addition to network structure, we are also interested in determining a fleet of heterogeneous vehicles on the hub links to satisfy customer demands.

“Time” is a very important aspect of the supply chain that affects service levels and costs. Most hub location problems considered in the literature assume demand as a rate of flow between each O-D pair that would reoccur in each time period. The demand data on freight *shipments*, however, has a dynamic nature. Those changes with time suggest that data in some hub location problems have time-dependent parameters. The multi-period nature of those problems has been presented in the literature ([Campbell, 1990](#); [Melo et al., 2006](#); [Contreras et al., 2011b](#); [Alumur et al., 2012b, 2016](#)). These papers, amongst other publications, indicate the importance of the time dimension for demand data. In the present work, demands are distributed across multiple planning periods. We employ a representative time horizon, which is divided into discrete time intervals.

In this thesis, we first develop a mixed-integer linear programming formulation for a hub location problem that considers a promised delivery time and a single allocation strategy. With a heterogeneous fleet of vehicles operating on hub links and for multi-period demands, we also account for different cost components: those for establishing hubs, transportation,

and for carrying inventory, to study the impact of inventory holding on network design. We determine the optimal hub locations, network configuration, and extent of shipment consolidation to minimize the total cost of establishment and operations. Additionally, we propose a Variable Neighborhood Search-based solution methodology to solve larger instances of the problem in shorter computational time. We evaluate the performance of the model and algorithm by applying well-defined data from the literature under different parameter settings.

We then adopt a novel modeling approach for a hub arc location problem, which determines the set of hub links and assigns the frequency of vehicles from a heterogeneous fleet to operate on them. The model considers a multiple allocation strategy, multi-period demand, and inventory holding at the origin nodes of the hub links. To efficiently solve the problem, we propose an exact algorithm based on Benders Decomposition. The performance of the algorithm is further enhanced by integrating Pareto-optimal cuts and applying variable fixing techniques. We assess the model and algorithm’s performance under various parameter settings, including different frequency choices, hub arc establishment costs, and inventory holding costs.

Finally, we address demand uncertainty by developing a two-stage stochastic model for the hub arc location problem. We propose a mixed-integer programming formulation to capture the problem’s complexities and utilize a Sample Average Approximation (SAA) scheme to manage stochastic demand. Computational experiments are presented to demonstrate the value of considering uncertainty and the effectiveness of the solution methodology.

The remainder of this thesis is organized as follows. Chapter 2 reviews the literature relevant to this thesis. Chapter 3 presents the mathematical model and the proposed algorithm for the shipment consolidation problem in hub location modeling with promised delivery time. In Chapter 4, we introduce a model and propose a Benders Decomposition-based methodology for determining hub arcs and the frequencies of different vehicle types on these hub links, as well as the amounts of shipments to be held at the origin node of the hub arcs. Chapter 5 addresses demand uncertainty by introducing and solving a stochastic model for the hub arc location problem introduced in Chapter 4. Finally, Chapter 6 summarizes the key points of each chapter, providing concluding remarks and a brief discussion of future research directions.

Chapter 2

Literature Review

In this chapter, we classify and review the literature on relevant hub location problems (HLP). In our discussion of previous relevant publications, the division into the following eight sections enables us to emphasize in turn the distinct sets of references that are strongly related to the respective problem types and methodologies. In particular, Sections [2.1](#) and [2.2](#) give overviews on hub location and shipment consolidation problems respectively. In Section [2.3](#) we present an overview of the hub location problems with multiple periods in the time horizon. Section [2.4](#) focuses on models and methodologies that consider economies of scale and vehicle utilization in hub location. In Section [2.5](#), we take particular interest in literature that attempts to model shipment consolidation and inventory holding in hub location problems. In Section [2.6](#), we consider a particular variant of HLPs – the hub arc location problem. Finally, we indicate publications employing Benders decomposition and Sample Average Approximation in Sections [2.7](#) and [2.8](#).

2.1 Hub Location Problems

Hub location problems involve finding the sites of particular nodes, called hubs, which will make it easier to route flows from origins to destinations. Hubs provide two major functions: switching-sorting-connecting, and consolidation and breakbulk. A hub network consists of O-D nodes, hub nodes, arcs between non-hub nodes and hubs, arcs between two hub nodes, and sometimes direct arcs between two non-hub nodes. To the best of our knowledge, O’Kelly (1986a,b) first introduced the hub location problem in the literature. These papers were inspired by earlier research by Hakimi (1964), who addressed the location of a single facility, the center of a weighted graph, which may be used as a switching center in a communication network.

Hub location problems have several variants that come with a list of attributes. The first is a *demand* that is associated with an O-D pair. There is a benefit (in transportation cost or time) of routing flows through hubs. Thus the objective here depends on locations of hubs and the routing of flows. Some hub location formulations allocate non-hub nodes in the network to only one hub, known as the *single allocation* problem (O’Kelly, 1992). Other models allocate non-hub nodes in the network to multiple hubs; this is known as the *multiple allocation* problem. Campbell (1994a) defined the p -hub median problem as deciding on precise locations for a given number of p hubs. Choices are made from candidate sites while minimizing the total transportation cost, subject to a set of demands for O-D pairs that must be served.

O’Kelly (1987) formulated the single allocation p -median problem as a quadratic integer program and proposed two heuristic algorithms, both of which enumerated all possible

choices of p hub locations. This paper introduced the CAB (Civil Aeronautics Board) data set, which was based on airline passenger flows between 25 US cities. [Ernst and Krishnamoorthy \(1996\)](#) countered with a data set known as Australia Post data, consisting of 200 nodes representing the postal districts of Sydney. Surveys by [Campbell \(1994b\)](#), [O’Kelly and Miller \(1994\)](#), [Campbell et al. \(2002\)](#) and [Alumur and Kara \(2008\)](#) defined other variants of the hub location problem.

[Alumur et al. \(2012a\)](#) proposed a mixed-integer programming model for multimodal hub location problems, where transportation costs and travel times have been jointly considered. [Serper and Alumur \(2016\)](#) considered a hub location problem with multiple transportation modes, with an intent to decide the locations and capacities of hubs, allocation of O-D nodes to hubs, and the number of vehicles to operate on hub links. The objective was to minimize the total cost of opening and material handling at hubs, purchase and operational costs of vehicles. The authors modeled a mixed-integer program to represent the problem and developed a variable neighborhood search algorithm for obtaining a solution.

[Çetiner et al. \(2010\)](#) treated the combined hub location and routing problem in postal delivery systems. [Martins de Sá et al. \(2015\)](#) studied a variant of the hub location problem where located hubs must form a set of interconnecting lines. Subject to a constraint on budget, while minimizing the travel time between O-D pairs, they developed a mathematical formulation, and proposed a Benders branch-and-cut algorithm and several heuristic solution algorithms. [Lopes et al. \(2016\)](#) investigated a hub location-routing problem, where a set of nodes must be partitioned into routes containing exactly one hub each, and then an extra route be determined that interconnects all hubs. They proposed a variable neighborhood descent procedure aided by a genetic algorithm that uses a multi-start local search

technique. [Lopes et al. \(2016\)](#) also formulated an integer programming model and solved it with a branch-and-cut algorithm. To minimize the total cost of transferring flow between hubs and routing flows in the network for the single allocation p -hub median location and routing problem with simultaneous pick-up and delivery, [Kartal et al. \(2017\)](#) suggested mixed-integer programming formulations. They developed heuristic algorithms based on multi-start simulated annealing and ant colony optimization.

[de Camargo et al. \(2017\)](#) studied incomplete hub location problems with and without hop-constraints (i.e. constraints on numbers of legs in an O-D path), offered formulations, and a Benders decomposition technique which can solve large scale test instances. [Hoff et al. \(2017\)](#) considered a single allocation multimodal hub location problem while satisfying constraints on the capacity of edges between hubs. [Meier \(2017\)](#) analyzed the uncapacitated single allocation p -hub median problem with modular capacities, introducing a new mixed-integer program formulation of the problem that uses fewer variables but more constraints. [Mokhtar et al. \(2018\)](#) considered a multiple allocation uncapacitated p -hub median problem and applied a Benders decomposition method. Its performance in generating cuts was enhanced by solving subproblems more efficiently, using minimum cost network flow algorithms. [Mokhtar et al. \(2019\)](#) treated an intermodal hub location problem with two types of hubs and three transfer modes. It was assumed that the underlying network is not fully connected, and that access nodes are not directly connected, in developing a mixed-integer linear program. [Taherkhani and Alumur \(2019\)](#) studied single, multiple and r -allocation hub network design problems with the intent of maximizing *profits*. The special case when the connection between two non-hub nodes is allowed was also considered. The authors formulated mathematical models for the different variants,

testing them with the CAB and AP data sets.

Besides the latter two and USAF (United States Air Flow) (Masaeli et al., 2018), there are two other data sets, commonly known as the Turkish Network (Tan and Kara, 2007) and Iranian Aviation Dataset (Karimi and Bashiri, 2011), that have been used in the literature. However, these two data sets have been primarily utilized by the researchers of the respective geographical domains.

In this section, we have reviewed variants of the capacitated and uncapacitated p -hub median problems because of the strong relevance to our research. Much literature also exists on the p -hub center problems and hub covering problems. Comprehensive reviews on these problems have been presented in Campbell (1994b), Campbell et al. (2002), Alumur and Kara (2008), Campbell and O’Kelly (2012) Farahani et al. (2013), Contreras and O’Kelly (2019), Alumur et al. (2021), Contreras (2021) and Wandelt et al. (2022).

2.2 Shipment Consolidation

Shipment consolidation is a logistics process that, purposefully, combines two or more shipments into a larger load, to be shipped together to a common market or region. It intends to increase vehicle utilization and thereby optimize the total cost of transportation plus inventory. The importance of this process derives from the significance of transportation in the supply chain industry. Saving costs on transportation may bring competitive advantage, leading to profit and sustainability to third-party logistic providers and national economies.

Shipment consolidation may exploit the advantage of economies of scale by reducing transportation costs per unit distance or per unit weight. The challenge here is to decide when to dispatch a vehicle and how much freight to dispatch in the given load. One extreme is to have no inventory cost and dispatch vehicles upon receipt of each order. This would increase transportation costs but would enhance customer satisfaction. The other way is to consolidate as many shipments as would fit in the truck. This would increase vehicle utilization and take advantage of economies of scale on the fixed cost of dispatching a vehicle. That approach, although minimizing the cost of transportation, would greatly increase the inventory cost. An optimal decision here would minimize the sum of transportation plus inventory costs. Research in shipment consolidation is focused on determining optimal policies to make these dispatch decisions, while ensuring reasonable service quality for the customers.

Classical shipment consolidation has three popular policies ([Higginson and Bookbinder, 1994](#)): (a) Time Policy - dispatching vehicles at predetermined intervals, (b) Quantity Policy - holding shipments to a particular destination until a minimum weight threshold is reached, then dispatching the combined load, (c) Time and Quantity Policy - holding shipments until a predetermined shipment time or the combined load reaching a minimum threshold, whichever occurs first. [Higginson and Bookbinder \(1994\)](#) made a quantitative study of these policies, using a simulation model to suggest consolidation policies for different parameter variations of order arrival-rates and maximum holding times.

[Newbourne \(1976\)](#) reviewed some fundamental industrial practices on shipment consolidation. [Higginson and Bookbinder \(1995\)](#) applied a discrete-time Markovian decision process to decide the release time of shipments for both common carriage (a public for-

hire trucking company) and private fleet (the shipper's own vehicles). They considered two objectives - minimizing cost per unit time, or minimizing cost per hundredweight per unit time. Assuming random order size and arrival times, the authors proposed optimal policies for both private and common carriage. [Bookbinder and Higginson \(2002\)](#) applied probabilistic modeling to decide dispatch quantity and maximum holding time to obtain rules for consolidation. Assuming order arrivals in discrete-time epochs and weight of an order measured in discrete units, [Bookbinder et al. \(2011\)](#) considered the case of shipment consolidation by private carrier, and modeled as a Markov chain to study the optimal policies for dispatching the accumulated weight of orders.

[Tyan et al. \(2003\)](#) studied collaborative shipment consolidation between global third-party logistics providers, and proposed a mathematical model to evaluate consolidation policies. A review of the literature on shipment consolidation is presented in [Çetinkaya \(2005\)](#), where the author also reviewed industry practices in the integration of inventory and transportation decisions. Çetinkaya focused on shipment consolidation policies in two scenarios, where inventory and transportation decisions are made separately and then jointly. [Nguyen et al. \(2014\)](#) and [Fan et al. \(2019\)](#) applied shipment consolidation to the dispatch of perishable product freight.

Investigation of variants of shipment consolidation problems with delivery time costs or constraints has been conducted by several researchers. [Ma et al. \(2011\)](#) considered cross-docking distribution networks for shipment consolidation by including an element of time cost in their objective function. They developed an integer programming model and heuristics to solve the problem. [Wei et al. \(2017\)](#) investigated the optimal consolidation policy focusing on the trade-off between economies of scale and expedited shipping costs,

and used dynamic programming to study the optimal policy and its structure. [Hanbazazah et al. \(2019\)](#) treated a case of the multi-period horizon and two-echelon settings for third party logistics, where each shipment has a preset pickup date at the source and a delivery time window at the destination. The authors proposed a mixed integer programming formulation for the problem, which is strengthened by container load relaxation, temporal decomposition, and valid cut generation. [Krishnan and Liu \(2019\)](#) presented a mixed-integer program, and used a branch-and-cut based solution strategy for cargo consolidation in a multiple vehicle pickup and delivery problem with time windows for online market places.

[Li et al. \(2012\)](#) analyzed the shipment planning problem in an airfreight forwarding network. Mixed-integer programming formulations were proposed that use piecewise-linear cargo rates, account for volume and weight constraints, and flight departure/arrival times. The authors' solution methodologies were based on Lagrangian relaxation and a local branching heuristic. [Bookbinder et al. \(2015\)](#) formulated a mixed-integer program and exact solution methodologies based on branch-and-price, a best-fit decreasing loading heuristic, and local branching heuristics for the air-cargo consolidation problem under the *pivot – weight* scheme. The mixed-integer model of [Huang et al. \(2020\)](#) used Lagrangian Relaxation as a solution methodology for shipment consolidation in air-land inter-modal operations.

[Cortes and Suzuki \(2020\)](#) observed that the Vehicle Routing Problem with split deliveries in last-mile delivery has the limitation that it does not take advantage of shipment consolidation. Their model also permits shipment transfers between different vehicles at certain locations. [Zhang et al. \(2019\)](#) studied a similar problem in online retailing, where

multiple trucks may visit the same customer at different times, and suggested a consolidation approach, modeled as an integer program. [Zhang et al. \(2021\)](#) presented an approach for consolidating packages between multiple warehouses, involving items from different orders via inter-warehouse shipments. Their model located transshipment points and flows. [Johansson et al. \(2020\)](#) proposed a time-based consolidation policy in the case of stochastic customer demand, modeled as a Poisson process on a single warehouse and multiple retailer inventory system, to determine shipment intervals and safety stock levels for retailers and warehouse.

[Lin and Wu \(2001\)](#) and [Lin and Lin \(2007\)](#) have explored the practice of holding inventories at hubs to optimize transportation costs. [Lin and Wu \(2001\)](#) focussed on the impact of multiple-frequency-delivery operations on the feeder fleet size and the utilization of hubs within a hub-and-spoke network. This research aimed to optimize the consolidation of partial loads for time-definite freight delivery, using common carriers so as to reduce transportation costs. The authors proposed an integer programming formulation and introduced an implicit enumeration algorithm. In a subsequent work, [Lin and Lin \(2007\)](#) assumed that demand followed a finite discrete distribution; the carriers consisted of a fleet of identical trailers. Lin and Lin developed a heuristic approach that demonstrated the potential for reducing operating costs through the implementation of proactive operational strategies. These strategies encompassed the determination of optimal freight paths, management of trailer fleets, and planning of routes for time-definite common carriers.

Interested readers may refer to references [Guastaroba et al. \(2016\)](#) and [Hanbazazah et al. \(2020\)](#) for further review of freight consolidation problems. Notably, these studies do not address the determination of hub locations. In our research, however, we take a com-

prehensive approach by integrating hub location decisions with transportation planning, providing a more holistic perspective on the problem to optimize the cost of location and transportation.

For a further review of the literature on shipment consolidation policies and models, the reader may refer to [Marasco \(2008\)](#), [Mutlu et al. \(2010\)](#), [Cai et al. \(2014\)](#), [Satır et al. \(2018\)](#), [Kiani Mavi et al. \(2020\)](#) and [Martínez et al. \(2020\)](#).

2.3 Multi-period Hub Location Problems

Demand need not be constant in real supply chain applications, hence the importance of the time parameter. In most of the hub location literature, a common period is assumed for demand between each O-D pair. These demands are assumed to repeat over different time frames, i.e. days, weeks, months, or years. Thus, solutions to such problems are assumed to be applicable to other time frames when demands repeat. However, in real-life, the possibility of a static demand pattern is low. Little attention has been paid to hub location problems with multi-period dynamic demand.

[Campbell \(1990\)](#) studied an uncapacitated, multiple allocation, multi-period hub location problem that assumed cost components for hubs, transportation and relocation. He developed a continuous approximation model to locate hubs for a general freight carrier serving a demand that is increasing (over time periods) between O-D pairs. The model minimized an objective representing the present value of the sum of the costs of hubs, terminals, and relocation. Campbell studied several myopic strategies that provided upper

and lower bounds on the distribution cost. He concluded that the strategy with limited relocation capability is near-optimal, given that relocation costs are small. This may be viewed as the first work on the hub location problem with multi-period demand.

[Gelareh \(2008\)](#) considered a multi-period hub location problem, where the network started with an initial set of hubs. Over the course of the planning horizon, the choice of hubs, along with the network structure, evolved so as to minimize the total cost of transportation, plus the cost of establishing and uninstalling hubs. Only small problem instances could be solved optimally with that mixed-integer formulation. A local-search based heuristic algorithm to obtain a feasible solution to larger instances was presented. [Gelareh et al. \(2015\)](#) proposed a model for uncapacitated multi-period multiple allocation hub location problems with budget constraints. Solution was based upon a metaheuristic algorithm plus Benders decomposition.

[Contreras et al. \(2011b\)](#) also studied an uncapacitated multiple allocation multi-period hub location problem. Their path-based formulation involved an objective composed of cost components on hubs (opening, operation, closing), and transportation. The decisions to be made were establishment, operation and closing of hubs over a finite time horizon. The model allowed a node to be a hub in non-consecutive periods. The authors developed a branch-and-bound algorithm that used Lagrangian relaxation to obtain lower and upper bounds at the nodes of the tree. Instances with up to 100 nodes and 10 time periods could be optimally solved via this algorithm.

[Alumur et al. \(2016\)](#) analyzed single and multiple allocation multi-period hub location problems with modular hub capacities. Here the hub network can be dynamically built,

and the capacities of hubs may be expanded over time. There is an assumed limit on the total flow that goes through a hub, and the allocation between a non-hub and a hub node can change from one period to another. Thus a new hub network is built from scratch, progressively over time. The authors proposed mixed-integer linear programming formulations for each case considered.

[Correia et al. \(2018\)](#) presented a modeling framework for multi-period stochastic capacitated multiple allocation hub location problems. In minimizing the total expected cost, the model decides the capacities and locations of hubs, and the O-D pairs through which transportation occurs. The two-stage stochastic programming model had the hub locations over the planning horizon and their initially-installed capacities as the first stage decisions. In the second stage, the formulation concerned the assignment of non-hubs to hubs, the routing of flows, and the capacity expansion for existing hubs. The model was enhanced by a set of valid inequalities and tested with the CAB data set.

[Khaleghi and Eydi \(2022a\)](#) conducted a systematic review of multi-period hub location problems, identifying various studies that addressed demand distributed over multiple periods within the time horizon. However, it was observed that most of these studies, including but not limited to those by authors such as [Taghipourian et al. \(2012\)](#), [Bashiri et al. \(2018\)](#), [Fotuhi and Huynh \(2018\)](#), [Torkestani et al. \(2018\)](#), [Khaleghi and Eydi \(2021\)](#), and [Neamatian Monemi et al. \(2021\)](#) did not incorporate inventory holding at hubs to improve shipment consolidation, in optimizing transportation costs while addressing hub location problems.

[Pourghader Chobar et al. \(2021\)](#) allowed changes in demand over time in a hub location problem to procure essential commodities. Their mathematical model decides establishment, renovation, or renting the distribution centers, while minimizing the routing cost between production centers and retailers and the pollution from vehicles. Those authors used two Pareto-based solution methodologies to solve the bi-objective problem.

[Khaleghi and Eydi \(2022b\)](#) considered a sustainable hub location problem with time-dependent demand and formulated a multi-objective mixed-integer nonlinear programming model. The change in demand here is a linear function of time, in which the coefficients are fuzzy parameters. The model included hub capacities and modular capacity for the hub links; the objectives involved minimizing costs and emissions, and maximizing job opportunities. [Khaleghi and Eydi \(2024\)](#) treated a multi-period hub network in a continuous-time planning horizon, considering linear time-dependent demand, and developed a non-linear programming model with three objectives on economic, environmental and social aspects. The authors linearized the model, used aggregation functions to make it a parametric single-objective model, introduced valid inequalities, and proposed heuristic algorithms incorporating Benders decomposition in a rolling horizon. Both of these studies considered multi-period demand, but did not model shipment consolidation by holding inventory at hubs.

2.4 Economies of Scale and Vehicle Utilization in Hub Location Problems

Given the demands for flow between a set of O-D pairs, there are different possibilities to minimize cost. If the flow is sent through hubs, in contrast to being sent directly between O-D pairs, large carriers can be used to move consolidated flow between those hubs. That is more cost-efficient than employing smaller carriers. This is due to the sharing of fixed cost between different flow units, and the lower variable cost per unit of larger carriers. Consider a bus or metro-rail network in a city. Big buses and large rail cars, filled with people, running between metro stations, is much more efficient than using automobiles for door-to-door commuting. This reduction in the consumption of fuel per individual commuter exemplifies the economies of scale in the hub network.

Scale economies in designing that network can be achieved in several ways. The first, as above, is to consolidate flows. A second could be the integration of new services to network O-D pairs, in addition to preexisting services, thereby reducing the overall unit cost. Another way is to maximize the network utilization, because an increase in traffic density may lower the unit cost of service ([Jara-Díaz et al., 2013](#)). Hubs are often expensive to establish; they make the routes indirect, so transshipment is thereby required. Nevertheless, hubs can reduce investment in infrastructure and operational cost by concentrating resources at a few nodes to be used by many. Moreover, due to establishment of hubs, the volume of traffic increases on links between hubs. As a result, the frequency of service may increase on the inter-hub links.

Assuming it will be cheaper to use smaller vehicles on allocation links, and larger vehicles on the inter-hub links, classical hub location models (e.g. [O’Kelly, 1987](#)) employ a discount factor α ($0 \leq \alpha \leq 1$) on the inter-hub link traffic. Apart from the costs of hub establishment, the models described in that paper assumed a unit cost only on the route links. But due to the discount factor on inter-hub links, the flow was assumed to concentrate on those links. Papers such as [O’Kelly \(1987\)](#) formed the basis for further research in hub location. However, the discount used in those experiments is on arcs, not on the concentration of flow, and hence does not provide economies of scale in cases of dense flow on an the allocation arcs. On the other hand, the discount factor is still applied on inter-hub links with low traffic. This phenomenon results in using multiple, filled-to-capacity small vehicles on allocation links, instead of using a single large vehicle to reduce cost. Additionally, even when there is lower traffic on inter-hub links, such models would be using large vehicles that are partially filled. Small vehicles would have been more appropriate in such a situation.

The experiments on classical models highlighted the importance of service frequency and scheduling in hub location. To address these problems, researchers including [Horner and O’Kelly \(2001\)](#) have used a non-linear cost function. Other researchers (e.g. [O’Kelly and Bryan, 1998](#)) approximated the non-linearity by a step-wise linear function. The different steps of the latter represent switching to a vehicle with different cost and capacity when flow increases. Still other researchers, including [Alumur et al. \(2009\)](#) and [de Sá et al. \(2013\)](#), employed an incomplete hub network; this tends to lessen the incidences of low traffic on inter-hub links. [Cunha and Silva \(2007\)](#) proposed a step-wise cost function, where the discount α would apply to any link exceeding a threshold, meaning any large or

small vehicles could operate on both allocation and inter-hub links, depending on the traffic volume. In the case when a third party common carrier (e.g. an LTL carrier) is providing transportation and using that cost model, a shipper may over-declare his shipment quantity to enjoy the all-unit discount available by using larger vehicles.

In real scenarios, there is a fixed cost involved in using a vehicle, independent of how long the vehicle runs. That has to be paid every time, as soon as the vehicle is used. [Kimms \(2006\)](#) modeled this fixed cost structure. A larger vehicle has a higher fixed cost than a smaller vehicle. A larger vehicle has a lower variable cost per unit in comparison to that of a smaller vehicle. Such a model results in a combination of large and small vehicles on both the allocation and the inter-hub links.

Researchers including [Yaman and Carello \(2005\)](#) and [Tanash et al. \(2017\)](#) introduced hub networks with modular arc capacities as a contrasting representation of this same problem. [Serper and Alumur \(2016\)](#) applied this idea of fixed and variable costs for different vehicles to choose a vehicle fleet for delivering a given set of O-D demands. [Zhang et al. \(2017\)](#) developed two different mixed integer models for an integrated airplane assignment and hub location problem for an air-cargo delivery service, and developed a two-stage hybrid algorithm for large instances. [Wu et al. \(2023\)](#) studied a service network design problem for an urban same-day delivery system, where the number of vehicles to simultaneously load or unload at a hub is limited, while assuming time constraints for deliveries and capacity constraints on the hubs. The authors formulated an integer program and proposed three heuristics to solve real-life instances. Those papers discussed in this paragraph that incorporate distinct vehicle types, however, do not consider the time dimension of decisions concerning inter-hub service frequency or service schedule.

2.5 Shipment Consolidation and Inventory Holding in Hub Location Problems

To the best of our knowledge, there are only a few instances of considering shipment consolidation decisions in hub location problems. Those publications aim to determine the hub locations, network structure, and the numbers of vehicles and period of dispatching from a hub on an inter-hub link. [Masaeli et al. \(2018\)](#) thus incorporated shipment scheduling decisions into hub location problems. The authors proposed three mixed-integer programming models for different versions of this problem, based on holding costs and multiple vehicle types. Considering the capacities of vehicles, the model also decides on the number of dispatches on each inter-hub link during each time period. The authors used the USAF data set to investigate the impacts of different holding costs and shipment scheduling decisions on the design of a hub network and on routing decisions. This was the first attempt to simultaneously model the decisions of hub location and shipment scheduling to achieve economies of scale by consolidating freight at hubs, where loads might accumulate over one or more periods, due to the scheduling of shipments.

Earlier, [Yaman et al. \(2012\)](#) studied release-time scheduling and hub location problems for next-day delivery to improve service quality at minimum total routing cost. Assuming a predetermined service level and a given closing time at all centers for receiving packages from their customers, they investigated a single allocation p -hub median problem under release-time decisions for trucks from each demand node. They also assumed all trucks must leave a demand center at the same time. The paper presented integer programming models to solve this problem in the special cases of uniform cargo arrival at each demand

center, piece-wise linear arrival patterns, release-time-dependent cargo arrival patterns, and multiple service levels. They proposed several classes of inequalities to strengthen their formulations. The authors presented numerical results on their models using CPLEX for the CAB and Turkish data sets.

[Karimi and Setak \(2018b\)](#) considered shipment scheduling in the incomplete hub location-routing problem to locate hubs, allocate nodes to hubs, and schedule the departure times from nodes. Allowing direct links between O-D nodes but no capacity limits on hubs, the authors proposed an integer programming formulation which ensures that the predetermined flow is delivered at the given service time. Computational results using CPLEX on instances from the CAB, AP, and Iranian Roads Network data sets were presented to prove the efficiency of the formulations. [Karimi and Setak \(2018a\)](#) extended that work in deterministic and stochastic environments when flow arrivals follow a piecewise linear form. They proposed an ϵ -constraint-based method and a normalized weighted sum method to solve bi-objective problems.

Neither [Yaman et al. \(2012\)](#) nor [Karimi and Setak \(2018a,b\)](#) considered holding costs. [Masaeli et al. \(2018\)](#), despite making shipment scheduling decisions, assumed that demands between all O-D pairs are known in the first time period. That work also assumed that those demands should reach the destinations at the end of the planning horizon. Previously, [Miranda et al. \(2009\)](#) modeled inventory holding decisions while addressing a logistic network design problem, which encompassed fleet design and customer clustering decisions. However, their approach did not consider demand between O-D pairs that is distributed over different time periods. [Marufuzzaman and Eksioglu \(2014\)](#) modeled a dynamic, reliable intermodal hub and spoke network problem. Their formulation models inventory

holding, but their demand is not specific to individual O-D pairs.

In a more recent work, [Wu \(2024\)](#) integrated classical multi-plant lot-sizing and hub location problems to jointly decide the location of production facilities, lot-sizing, inventory, hub location, and the distribution of multiple commodities from plants to customers, thereby minimizing the total cost of setup, production, inventory, hubs, and transportation. The author employed an exact solution methodology that integrates dynamic programming and Benders decomposition, aided by multiple cuts in the same iteration, Pareto-optimal cuts, a hub elimination and restoration procedure, and a heuristic. Despite that problem involving multi-period demand, the demand is not between specific pairs of origin-destination nodes, and the hub locations are assumed to hold no inventory.

2.6 Hub Arc Location Problems

Hub Arc Location Problems (HALPs), originally introduced by [Campbell et al. \(2005a,b\)](#), have received little focus in the literature. While other hub location models focus on the sites of hub nodes, these problems emphasize the establishment of hub arcs with discounted costs. Those arcs induce a set of hubs and assign O-D nodes to the hub arcs. Hub arcs thus connect two hubs that are assumed to carry consolidated shipments, and hence have a discounted cost per unit flow to model economies of scale. As opposed to locating hubs and allocating O-D nodes to the hubs, hub arc location models locate hub arcs with lower per-unit transportation costs.

HALPs relax the assumption of general hub location problems that there is a full

interconnection between hubs. HALPs come in both single and multiple allocation models. In HALPs, hub arcs may (or may not) require a particular topological structure of their induced network. Some of these problems assume that O-D paths contain at most one hub arc, limiting O-D paths to have at most three arcs, where the first and last ones are either access or bridge arcs, and the intermediate hub arc is optional. For example, in a passenger transportation network, a passenger should not have to transfer more than twice to get to a destination. Such a bound is common practice in air transportation. In the case of ground transportation, constant limits on the number of arcs may also be applied to restrict time spent in the handling of materials at hubs to enhance the quality of service.

[Campbell et al. \(2005a\)](#) proposed integer programming formulations and optimal solution algorithms for four different types of HALPs. Earlier [Campbell et al. \(2003\)](#) implemented a parallel enumeration-based algorithm to find the optimal solution for large instances of hub arc location problems, and tested them on the AP and CAB data sets. [Sasaki et al. \(2009\)](#) investigated the Stackelberg hub arc location problem and developed an optimal solution methodology. [Campbell \(2009\)](#) studied time-definite motor carriers and modeled as a hub arc location problem, where the author imposed service level constraints by putting a limit to the distance between O-D pairs.

[Contreras and Fernández \(2014\)](#) showed that the minimization of a real-valued supermodular set function can be used to model a class of hub problems that generalize well-known problems in hub network design - uncapacitated hub location, p -hub median, and hub arc location. Here a general class of multiple allocation hub arc location problems containing setup costs and cardinality constraints on hub arcs was represented by the minimization of a supermodular function. The authors presented two mathematical

formulations and compared their performance in computational experiments.

[Rothenbächer et al. \(2016\)](#) focused on hub arc capacities while studying hub arc location problems. Their exact branch-and-price-and-cut algorithm for a service network design problem used a bounding procedure based on column generation to solve a path-based formulation. This was aided by several valid inequalities. Their approach was utilized in hub location planning, and designing the network service frequency in the application to combined road–rail freight transport.

[de Camargo et al. \(2017\)](#) and [Tanash et al. \(2017\)](#) considered hub arc location in the context of flow-dependent costs. [de Camargo et al. \(2017\)](#) treated incomplete hub location problems with and without constraint on the number of legs in an O-D path, proposed formulations, and a Benders decomposition technique that can solve large-scale test instances. [Tanash et al. \(2017\)](#) analyzed a hub location problem that models the flow-dependent transportation costs using modular arc costs. It relaxes the full interconnection assumption between hub nodes and considers link activation decisions as part of the design. The authors proposed a branch-and-bound algorithm based on Lagrangian relaxation to solve the problem.

To address the service-vs.-cost challenges in designing a public transit network for evenings and weekends (times with lower demands for commuting), [Mahéo et al. \(2019\)](#) studied the design of a hub and shuttle model where hubs are connected by high-frequency buses and shuttles connect people from their origin/destination to those hubs. The goal was to optimally decide which hub-links to cover via bus routes. It was proposed that a multi-modal hub network design problem be modeled as a special case of the hub arc

location problem. The authors suggested preprocessing techniques, including trip filtering and link filtering to simplify the problem. Subsequent analyses applied a Benders decomposition technique that uses dedicated solution techniques for solving independent sub-problems, Pareto-optimal cuts, cut bundling, and core point update. The proposed model and algorithm were tested using real-life data from the public transit system of Canberra, Australia.

[Dalmeijer and Van Hentenryck \(2020\)](#) extended the work of [Mahéo et al. \(2019\)](#) by introducing frequencies of vehicles on each link, imposing constraints on the number of transfers on an O-D path, and adding new modes of transport. This was implemented using a special data structure called a “transfer-expanding graph,” so that non-convexity is not introduced to the sub-problem. The authors presented a case study for the city of Atlanta, GA to show the effectiveness of the model.

Recently [Basciftci and Van Hentenryck \(2023\)](#) generalized the work by [Dalmeijer and Van Hentenryck \(2020\)](#) in a context of integrating passenger mode preferences and accommodating demand of new passengers who adapt to the public transit system. This paper showed how to integrate rider mode preferences into the design of On-Demand Multimodal Transit Systems (ODMTS). It is motivated by a common worry in transit agencies that an ODMTS may be poorly designed if the demand of new riders adopting the system is not captured. The authors proposed a bi-level optimization model and an exact solution methodology that uses Benders cuts. The fundamental principles that underpin the emerging ODMTS can be attributed to the multi-modal hub location and hub network design problems ([Alumur et al., 2012a](#)), which were expanded by [Mahéo et al. \(2019\)](#), [Dalmeijer and Van Hentenryck \(2020\)](#), [Auad-Perez and Van Hentenryck \(2022\)](#) and [Basciftci and](#)

[Van Hentenryck \(2023\)](#).

Another recent work by [Auad-Perez and Van Hentenryck \(2022\)](#) extended research done by [Mahéo et al. \(2019\)](#). It considered consolidation in the shuttle legs joining O-D nodes and hubs. The authors relaxed the assumption by [Mahéo et al. \(2019\)](#) that shuttles are always available to serve the allocation links. The newer work proposed ride sharing on legs between O-D nodes and hubs. The shuttle fleet aims to optimize performance requirements, i.e. to minimize the required number of shuttles, important information for the transit companies. The authors developed mixed-integer programming models and a novel fleet sizing algorithm to solve the problem. [Lu et al. \(2024\)](#) studied the impact of congestion and dedicated bus lanes on ODMTS.

2.7 Benders Decomposition in Hub Location

Jacques F. Benders developed the mathematical programming technique ([Benders, 1962](#)) bearing his name. It helps to solve large-scale mixed integer linear programming problems by decomposing the main problem into two smaller and easier-to-manage components: a master problem that involves integer variables, and a subproblem. Thus, it can reduce solution times for large-scale integer programming problems that have a particular relation between the integer and non-integer variables, known as the ‘block structure’. The technique is gaining increasing popularity among hub-location researchers as an exact-solution algorithm because most such problems are computationally hard. Particularly in real-life situations, it is very challenging to find exact solutions in reasonable computational time. Comprehensive reviews of the Benders decomposition algorithm and its applications are

contained in [Şen et al. \(2016\)](#) and [Rahmaniani et al. \(2017\)](#).

[Magnanti and Wong \(1981\)](#) proposed a significant enhancement of Benders decomposition by selectively choosing from the alternate optima of the Benders subproblem to generate ‘strong’ or ‘Pareto-optimal’ cuts, thereby accelerating the convergence of the algorithm.

[de Camargo et al. \(2008\)](#) are the pioneers among researchers who utilized Benders decomposition in solving hub location problems. The authors implemented three variants of Benders decomposition for the uncapacitated multiple allocation hub location problem. In the first, a single cut is generated in every iteration of the algorithm; the second generates multiple cuts in each iteration; and the third terminates with an ϵ -optimal solution obtained. In the same year, [Rodríguez-Martín and Salazar-González \(2008\)](#) developed a mixed integer programming formulation for an incomplete hub network with capacities assumed on hubs and arcs. Those authors also developed two branch-and-cut algorithms based on Benders decomposition techniques, where a two-level nested decomposition scheme outperformed the standard Benders approach.

[Contreras et al. \(2011a\)](#) suggested a Benders decomposition algorithm to solve large-scale instances of the multiple assignment uncapacitated hub location problem. This decomposition, applied to a path-based formulation, was enhanced using multiple cuts per iteration, strong optimality (Pareto-optimal) cuts, and integration of reduction tests plus a heuristic procedure. [Contreras et al. \(2012\)](#) then extended their preceding work to the capacitated case with decisions on hub capacities, utilizing a Benders decomposition algorithm that applies Pareto-optimal cuts and reduction tests to boost speed of convergence.

In addition to the contributions by Camargo, Contreras and their colleagues, we explored several other applications of the Benders decomposition algorithm in solving large instances of various types of hub location problems. Some notable works, not limited to that list, include the following: in flow dependent discounts for inter-hub link (de Camargo et al., 2009), Benders decomposition’s applications by Gelareh and Nickel (2011) in uncapacitated multiple allocation hub location problem for linear shipping networks, and hub location-routing problems that allow routing decisions between non-hub nodes (de Camargo and Miranda, 2012). We note also single allocation hub location with congestion (de Camargo et al., 2013), by Gelareh et al. (2015) in uncapacitated multi-period multiple allocation hub location problems with budget constraints, and incomplete hub networks in de Camargo et al. (2017) and Martins de Sá et al. (2018). Single- and multiple-allocation uncapacitated p-hub median problems have been studied by Mokhtar et al. (2018). Other examples include the design of a hub and shuttle model while designing a public transit network (Mahéo et al., 2019; Dalmeijer and Van Hentenryck, 2020) and humanitarian-aid distribution across refugee camps in war-ridden areas (Neamatian Monemi et al., 2021).

Benders decomposition has been used to find solutions to numerous variants of the hub location problem. For example, Taherkhani et al. (2020) employed the Benders decomposition algorithm to solve the profit-maximizing capacitated hub location problem with multiple demand classes. In their approach, that algorithm was enhanced with a two-phase methodology to decompose the subproblem, separation routines to strengthen optimality cuts, and variable fixing techniques to accelerate convergence of the algorithm. In tackling a profit-maximizing, uncapacitated multiple allocation incomplete hub network design problem that precludes direct connections between non-hub nodes, Oliveira et al.

(2022) employed Benders decomposition integrated with Pareto-optimal cuts, multiple cut strategies, and the Benders branch-and-cut scheme. In another application, Eydi et al. (2022) studied a multiple-allocation hierarchical hub network design problem with a ring-star-star structure, and proposed an accelerated Benders decomposition algorithm to solve the problem.

Very recently, Wu (2024) integrated Benders decomposition with dynamic programming to find an exact solution to the production hub location problem. This problem involves determining the locations of production facilities and hubs, while also deciding lot-sizing, inventory holding, and the distribution of multiple commodities from plants to customers. The objective is to minimize the total costs of production, setup, inventory, hub operation, and transportation. The author proposed enhancements to the Benders approach, including multicut reformulation, Pareto-optimal cuts, elimination and restoration procedures for hubs, and some heuristics.

2.8 Sample Average Approximation in Hub Location Problems

The main challenge in discrete stochastic programming lies in incorporation of the expected value term within the model's objective function. To tackle this, a set of scenarios is generated and the average of the samples is employed to approximate the expected value of random variables. In stochastic optimization problems, including, but not limited to Hub Location Problems, the Sample Average Approximation (SAA) (Kim et al., 2015)

method serves as a scenario generation-based sampling approach.

The SAA is a Monte Carlo simulation-based scheme that generates random samples pertaining to random variables within a stochastic Mixed Integer Programming framework. It employs the sample-average method to approximate the expected value of the random variable. Subsequently, the associated deterministic optimization problem is solved to derive a solution. This process is iterated to generate solutions across replicated samples. Instances documented in the literature have demonstrated that this scheme yields high-quality solutions and may additionally provide a statistical estimation of their optimality gaps. The method was introduced to the literature in a seminal paper by [Kleywegt et al. \(2002\)](#), and has since been widely adopted for addressing various discrete stochastic optimization problems. We remark that [Emelogu et al. \(2016\)](#) proposed clustering techniques to dynamically update the sample size to enhance performance of SAA.

[Contreras et al. \(2011c\)](#) developed a two-stage stochastic Hub Location Problem incorporating uncertainties in both demand and transportation costs. They utilized the SAA method to ascertain the number of scenarios and subsequently employed the Benders decomposition algorithm to solve the model within the framework of the SAA method. [Taherkhani et al. \(2020\)](#) formulated a two-stage stochastic profit-maximizing Hub Location Problem with constrained hub capacities. They employed both the SAA and Benders decomposition to address and solve the proposed model effectively. In an extension of their work, [Taherkhani et al. \(2021\)](#) incorporated revenue uncertainties into the problem that already considered demand uncertainty. They delineated and formulated stochastic and robust stochastic models based on revenue elastic and independent demand.

[Zhang et al. \(2022\)](#) considered an incomplete multimodal multiple-allocation hub location problem with stochastic demands and delivery-time restriction. They developed a path-based formulation and used a two-stage stochastic programming methodology that involved an SAA approach that incorporates Benders-decomposition-based algorithms and a dualization strategy.

In tackling the profit-maximizing hub location problem while considering a cap on regional carbon emissions dictated by population density, [Rahmati et al. \(2023\)](#) devised a two-stage stochastic programming approach. This method accounts for demand uncertainty and incorporates SAA, which operates alongside several variants of the Benders decomposition algorithm. [Rahmati et al. \(2024\)](#) extended their previous work to encompass regulations such as policies on carbon tax, carbon cap-and-trade, carbon cap, and carbon offset policy. For a solution methodology, they used an SAA algorithm enhanced with k-means for clustering and an L-shaped algorithm.

In addition to its successful application in various fields of stochastic optimization, the utilization of SAA by facility location researchers has highlighted its significant potential in the domain of hub location. Several facility location problems, where large-scale instances have been effectively addressed using SAA, include [Romauch and Hartl \(2005\)](#), [Santoso et al. \(2005\)](#), [Schütz et al. \(2009\)](#) and [Kiya and Davoudpour \(2012\)](#). Interested readers may refer to these references for further review of SAA applications.

Chapter 3

Shipment Consolidation in Hub Location with Multi-period Demand

Having reviewed the literature, the present chapter addresses the gap concerning the lack of attention to the time dimension in the hub location literature, specifically the issue of handling multi-period demand in hub locations with inventory holding. We assume a scenario where the freight carrier has several types of vehicles to operate between hubs; for each vehicle type, there is a fixed cost of dispatching and a variable cost of operation. The local operations are outsourced to a third-party logistics provider (3PL) that, under the contract, will do the pickup and delivery to O-Ds for a price per unit weight per unit distance.

We determine how long shipments should be held at hubs, i.e., when to dispatch those goods along with other loads from this hub, and the extent to which shipments should be

consolidated before dispatch, while selecting the optimal locations of hubs and deciding the number of vehicles to operate between those hubs. We develop a mixed-integer linear programming model to formulate this problem, considering various cost components: cost for establishing hubs, transportation costs, and inventory costs. We find the optimal hub locations, network configurations, and degree of shipment consolidation, so that the total cost of network establishment plus operation is kept at a minimum.

In developing a methodology to solve this challenging hub network design and shipment consolidation problem, our solution algorithm decomposes the problem in two. In the first part, we make the location and allocation decisions. These are used to calculate the arrival times of demands to the hubs, and the inter-hub links through which these demands should be routed. For the second part, another mathematical model, referred to as the *shipment consolidation subproblem*, is developed and solved. There we determine the optimal inventory decisions, in addition to the dispatch times, and the types and numbers of vehicles required for transport between each pair of hubs. The hub locations and demand-node allocations are improved using a Variable Neighborhood Search (VNS) algorithm.

The remainder of this chapter is organized as follows: the next section presents a detailed definition of our problem and a mathematical formulation. Section 3.2 presents the shipment consolidation subproblem and describes the solution algorithm. Section 3.3 introduces the data set and presents the results of our proposed solution methodology, as well as a comparison to computations with a commercial solver. Section 3.4 summarizes the chapter.

3.1 Mathematical Formulation

We intend to design a freight delivery network to serve given demands between various O-D pairs. The node set N contains the origins and destinations of demand. The set H denotes the candidate hub locations, which may contain nodes from the set N . We assume all origin and destination nodes may send or receive flow through exactly one hub per node, referred to as *single allocation*, and that the allocation connections are being served by a contracted third-party logistics provider that is responsible for pickup and deliveries to O-Ds.

Decisions have to be made on which locations will be used for establishing hubs and the allocation of demand nodes to those hubs. The planning horizon T is finite and discretized into time windows. Individual time windows are denoted by $t = 1, 2, \dots, |T|$, where the first time window corresponds to interval $[0, 1)$, the second and third windows $[1, 2)$ and $[2, 3)$, and so on. For example, for a planning horizon of one week, each time window might consist of 12 hours, resulting in a total of 14 time windows. All time-related measurements (travel times, arrival times, dispatch times, and promised times for delivery) are discretized and rounded up accordingly. All decisions are taken at the beginning of a time period.

Selecting hub locations is a strategic decision with long-term implications. Demand fluctuations are of course expected over an extended planning horizon, but we assume that the decision-maker specifies a base demand pattern that is both representative and recurring. Throughout the time horizon, the preceding fluctuations will ebb and flow. It is known that operational decisions play a crucial role in influencing the selection of hub locations. For those planning purposes, the base demand pattern is taken to be

representative of the requirements that the chosen hubs must satisfy.

Any demand node may send flow to any other node through the network. At time period $t_0 \in T$ a quantity of freight $w_{ij}^{t_0}$ is originated from node $i \in N$ and destined to reach node $j \in N$. Here, the parameter t_0 , representing the initiation time of demand, can refer to any period within the planning horizon and is not limited to the first period. At the start of period $t_0 \in T$, this demand is shipped to the first hub node along the destination path by the third-party provider. We assume that all demands are divisible for shipment consolidation at first hub along the destination path, and have sufficient time windows within the planning horizon, after their origination, to arrive at their destination nodes.

We also assume that the freight carrier employs its private fleet on the inter-hub network and wants to determine the periods in which different vehicles are dispatched. There is a given set of vehicles $v \in V$ with capacities ϕ_v to operate on the inter-hub links. Vehicles, in this research, may denote different modes of transportation that carry freight between two hubs in the network. Each vehicle, on a link from hub $k \in H$ to hub $l \in H$, has a transit time of τ_{kl} . A flow that enters a (k, l) inter-hub link at time $t \in T$ arrives at l on time $t + \tau_{kl}$. For any demand $w_{ij}^{t_0}$, there is a promised time for delivery π_{ij} . Accordingly, all flows relating to this demand must reach destination node j within $t_0 + \pi_{ij}$.

Freight, upon arrival at hubs, gets unloaded, sorted, and consolidated with other loads. Shipments may be held at hubs for the convenience of consolidation with other freight, given that the respective promised times of deliveries permit such hold-overs. Shipment consolidation decisions, i.e. when and how long to hold inventory at a hub, and when to dispatch a vehicle from a hub, are to be determined for the inter-hub network. In each

period, the quantity of freight to be shipped, and the number of vehicles to be dispatched on different inter-hub links, are to be decided. The inter-hub network may include vehicles operating on all inter-hub links or only a subset of them. Consequently, our model decides whether the designed network will be a complete hub network or an incomplete one.

The model minimizes the total cost of establishing hubs, holding inventory, and transportation, by consolidating shipments at hubs and sending larger nearly-optimally filled vehicles between hubs. We aim to investigate and analyze how changes in hub establishment costs affect network structure and operational decisions such as inventory management and vehicle dispatch. Each vehicle dispatch involves a fixed cost plus a variable transportation cost estimated from the inter-hub distances. Flow on an allocation link, however, incurs a cost proportional to that volume of flow, under the assumption that the 3PL charges a price per unit weight per unit distance. That price encompasses factors such as amortization, driver salaries, and other fleet operation-related costs.

Inventory carrying costs may be incurred if goods are held at a hub for one or more periods. The percentage of inventory value on hand per unit of time may be used as an estimate of the holding costs ([Silver et al., 2017](#)). Additional costs, involving further processing of the freight held, if done at hubs, may also be embedded in carrying costs.

The sets and parameters required for the mathematical formulation are defined as follows:

Sets:

N : Set of O-D nodes.

H : Set of candidate hub locations.

T : Set of time periods.

V : Set of vehicle types.

Parameters:

$w_{ij}^{t_0}$: Flow originated at start of time $t_0 \in T$ from node $i \in N$ destined to node $j \in N$.

Υ_k : The setup cost associated with establishing a hub at location $k \in H$.

η_k : Holding cost per unit of freight per unit time at hub $k \in H$.

θ_k : Inventory holding capacity of hub $k \in H$ for each period.

κ_v : Fixed cost to dispatch a vehicle $v \in V$.

χ_{klv} : Variable transportation cost (including vehicle operational cost) from hub $k \in H$ to hub $l \in H$ for using vehicle type $v \in V$.

c_{ij} : Cost of sending one unit of flow from node $i \in N$ to hub $j \in H$ on an allocation connection.

ϕ_v : Capacity of vehicle $v \in V$.

π_{ij} : The promised time for delivery from node $i \in N$ to node $j \in N$.

τ_{kl} : Transit time from node $k \in H$ to node $l \in H$ in an inter-hub link.

τ'_{ik} : Transit time from node $i \in N$ to node $k \in H$ in an allocation connection.

O_i : The total flow originating at node $i \in N$, $O_i = \sum_{j \in N, t_0 \in T} w_{ij}^{t_0}$.

D_i : The total flow destined to node $i \in N$, $D_i = \sum_{j \in N, t_0 \in T} w_{ji}^{t_0}$.

Decision Variables:

$x_{ijkl}^{t_0 \ t}$: Flow from node k to node l , starting at the beginning of time t with respect to the demand originated at time t_0 , from node i destined to node j ($i, j \in N, k, l \in N \cup H, t_0, t \in T$).

$r_{ijk}^{t_0 \ t}$: Amount of inventory held at hub $k \in H$ at time $t \in T$, with respect to the demand originated at time t_0 from node $i \in N$, destined to node $j \in N$.

f_{klv}^t : Number of vehicles of type $v \in V$ dispatched from hub k to hub l ($k, l \in H$) at time $t \in T$.

$$y_k = \begin{cases} 1, & \text{if a hub is established at location } k \in H, \\ 0, & \text{otherwise.} \end{cases}$$

$$u_{ik} = \begin{cases} 1, & \text{if node } i \in N \text{ is allocated to hub } k \in H, \\ 0, & \text{otherwise.} \end{cases}$$

The single allocation hub location and shipment consolidation model is formulated as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{k \in N} \Upsilon_k y_k + \sum_{i, j \in N} \sum_{k \in H} \sum_{t_0, t \in T} \eta_k r_{ijk}^{t_0 \ t} + \sum_{k, l \in H} \sum_{t \in T} \sum_{v \in V} (\kappa_v + \chi_{klv}) f_{klv}^t \\ & + \sum_{i \in N} \sum_{k \in H} c_{ik} O_i u_{ik} + \sum_{j \in N} \sum_{k \in H} c_{kj} D_j u_{jk} \end{aligned} \quad (3.1)$$

subject to

$$\sum_{k \in H} x_{ijk}^{t_0 \ t} = w_{ij}^{t_0} \quad i, j \in N; t_0, t \in T; t_0 = t \quad (3.2)$$

$$\begin{aligned} r_{ijk}^{t_0 \ t-1} + \sum_{l: t-\tau'_{lk} \in T; l \in N} x_{ijlk}^{t_0 \ t-\tau'_{lk}} + \sum_{l: t-\tau_{lk} \in T; l \in H} x_{ijlk}^{t_0 \ t-\tau_{lk}} = r_{ijk}^{t_0 \ t} + \sum_{l \in N \cup H, l \neq k} x_{ijkl}^{t_0 \ t} \\ i, j \in N; t_0, t \in T; t_0 \neq t; k \in H; t = \{t_0 + 1, \dots, t_0 + \pi_{ij}\} \end{aligned} \quad (3.3)$$

$$\sum_{k \in H} \sum_{t \in T'} x_{ijkj}^{t_0 \ t} = w_{ij}^{t_0} \quad i, j \in N; t_0 \in T; T' = \{t_0 + 1, \dots, t_0 + \pi_{ij}\} \quad (3.4)$$

$$x_{ijkl}^{t_0 t} + x_{ijlk}^{t_0 t} \leq My_k \quad i, j \in N; k, l \in H; t_0, t \in T \quad (3.5)$$

$$\sum_{i,j \in N} \sum_{t_0 \in T} x_{ijkl}^{t_0 t} \leq \sum_{v \in V} \phi_v f_{klv}^t \quad k, l \in H; t \in T \quad (3.6)$$

$$\sum_{i,j \in N} \sum_{t_0 \in T} r_{ijk}^{t_0 t} \leq \theta_k y_k \quad k \in H; t \in T \quad (3.7)$$

$$x_{ijik}^{t_0 t} \leq O_i u_{ik} \quad i, j \in N; k \in H; t_0, t \in T \quad (3.8)$$

$$x_{ijlj}^{t_0 t} \leq D_j u_{jl} \quad i, j \in N; l \in H; t_0, t \in T \quad (3.9)$$

$$\sum_{k \in H} u_{ik} = 1 \quad i \in N \quad (3.10)$$

$$x_{ijkl}^{t_0 t} \geq 0 \quad \forall i, j \in N, \forall k, l \in H \forall t_0, t \in T \quad (3.11)$$

$$r_{ijk}^{t_0 t} \geq 0 \quad \forall i, j \in N, \forall k \in H, \forall t_0, t \in T \quad (3.12)$$

$$f_{klv}^t \in \mathbb{Z}^+ \quad \forall k, l \in H, \forall v \in V, \forall t \in T \quad (3.13)$$

$$y_k \in \{0, 1\} \quad \forall k \in H \quad (3.14)$$

$$u_{ik} \in \{0, 1\} \quad \forall i \in N, \forall k \in H \quad (3.15)$$

The objective function (3.1) minimizes the overall network cost consisting of establishment cost for locating hubs, total inventory cost, and the fixed and variable costs for inter-hub transportation, as represented by the first three summation groups. The last two such groups denote transportation costs on allocation connections. Constraint (3.2) guarantees that the flow concerning a demand is dispatched to potential hub locations at the time of arrival. Equality (3.3) ensures flow divergence, meaning the sum of inflow and on-hand inventory is equal to the sum of outflow and remaining inventory. This is written for every period except the first. Constraint (3.4) confirms that all flows corresponding to a demand arrive at the destination node within the promised time. The relation (3.5) enforces the establishment of hubs. Here, M is a sufficiently large number, we set it to

the sum of all O-D demands in the network Constraint (3.6) calculates the total number and types of vehicle dispatches required on each inter-hub link. Inequality (3.7) enforces the hub capacities for holding inventory in a period. Constraints (3.8) and (3.9) ensure that each node is allocated to a hub facility. Equality (3.10) guarantees single allocation. Relations (3.11) - (3.15) represent non-negative real, binary, and integer variables.

Assuming $|N| = |H| = n$, $|T| = t$, and $|V| = v$, the number of binary variables required for the model follows $\mathcal{O}(n^2)$. The integer and continuous variables amount to $\mathcal{O}(n^2tv)$ and $\mathcal{O}(n^4t^2)$, respectively. There are $\mathcal{O}(n^4t^2)$ constraints. This large number of integer variables and constraints might limit attaining optimality. Accordingly, the next section presents a solution methodology for this problem.

3.2 Solution Methodology

Initial investigations of the proposed model (3.1)-(3.15) indeed showed that it may not be possible to obtain optimal solutions in reasonable computational time using a standard commercial solver (CPLEX in our case). The development of a suitable solution methodology was, therefore, necessary to solve realistically-sized instances.

Our solution method decomposes the problem into two parts. First, we decide on the hub locations, allocate non-hub nodes to hubs, calculate the arrival times of demands from origin nodes to hubs, and the inter-hub links through which those demands should be routed. These initial decisions leave us with a shipment consolidation subproblem to be solved for each inter-hub link in the second part. There we determine the types and

numbers of vehicles to operate on each inter-hub link, in addition to deciding the inventory to be held at the first hub along the link. We iterate through the two parts to improve the quality of a solution.

We first construct an initial feasible solution to the problem. Next, the shipment consolidation subproblem is solved for each hub pair of these initial locations. For the improvement phase, we implement a VNS algorithm. Specifically, we implement five neighborhood structures to the preceding initial solution. These network structures - location and allocations - are then evaluated, and the total cost of neighboring solutions is obtained, again by solving the shipment consolidation subproblem for each hub pair. We then move to the best neighboring solution. These steps are repeated until the stopping criteria are met. Finally, we report the best solution of all these iterations as the solution obtained by the algorithm.

VNS has been employed as a solution methodology in various contexts for hub location problems, as demonstrated by several studies, such as [Ilić et al. \(2010\)](#); [Lopes et al. \(2016\)](#); [Serper and Alumur \(2016\)](#); [Mikić et al. \(2019\)](#). In all these algorithms, the common plan involves generating an initial solution using heuristics and subsequently applying neighborhood search techniques to enhance it. Our approach capitalizes on the inherent problem structure to define a neighboring solution, and then formulate a new sub-problem for optimizing dispatch decisions on inter-hub links. As mentioned, the proposed VNS employs five distinct neighborhood structures to facilitate robust exploration within the solution space. This scheme, in conjunction with the optimal resolution of sub-problems, is anticipated to yield high-quality solutions. Our solution methodology is detailed in [Algorithm 1](#). The following subsections describe each step of this algorithm in detail.

Algorithm 1 Shipment Consolidation in Hub Location (SCHL) Algorithm

Input: All sets and parameters of model (3.1)-(3.15)

Output: S_{Best}

- 1: **Begin**
 - 2: Construct an initial solution, S_{init}
 - 3: Solve the Shipment Consolidation Subproblem for S_{init}
 - 4: Run VNS Algorithm on S_{init} , until meeting stopping criteria, to obtain the best solution S_{Best}
 - 5: **return** S_{Best}
 - 6: **End**
-

3.2.1 Initial Solution

The first step is to construct a greedy initial solution, to be referred to as S_{init} . Although some papers (e.g., [Wandelt et al., 2022](#)) utilize other hub location models to generate an initial solution, such a plan can be time-consuming. Instead, we have chosen to adopt a straightforward heuristic approach to quickly generate an initial solution and spend more time in the improvement phase. In generating that initial solution, the values $(O_k + D_k)$ are calculated for each potential hub location $k \in H$, and sorted in descending order. The first $p\%$ of locations on this descending list are selected as hubs. Next, for each remaining potential hub location $l \in H$, we calculate the total distance from that potential hub node to every other demand node in set N . We then sort those distance values for all potential hubs in ascending order. We select the top $q\%$ of nodes from this ordered list as hubs, so that the size of the initial set of hubs is $(p + q)\%$ of the total number of nodes. In this initial solution, every O-D node is allocated to its nearest hub. To make sure the initial solution S_{init} is always feasible, we assign reasonably high values to parameters p and q and iterate and increase these values, if necessary.

3.2.2 Shipment Consolidation Subproblem

After obtaining the locations and allocations from the initial solution, the goods can directly be sent on to their respective hubs right after their arrival at the origin nodes. This provides us with the arrival times and quantities of freight at different hubs. From that information, we calculate the flow demand on each inter-hub link. The shipment consolidation subproblem decides on the inventory at each hub, and the number and types of vehicles on each inter-hub link, as detailed below.

Let an origin $o \in N$ be allocated to hub $k \in H$, and destination $d \in N$ be allocated to hub $l \in H$. In practice, it is commonly observed that the O-D paths typically involve one or two hubs. This observation has been consistently confirmed through the execution of our model across various data scenarios. Under the assumption of at most two hub stops, let us say that the demand $w_{od}^{t_0}$ flows through the inter-hub link (k, l) . Assuming τ'_{ok} , τ_{kl} , and τ'_{ld} as the transit times on links (o, k) , (k, l) , and (l, d) , the flow arrives at hub k at the beginning of period $t_k = t_0 + \tau'_{ok}$. With π_{od} as the promised time for delivery from node $o \in N$ to node $d \in N$, this demand has a total time of $[\pi_{od} - (\tau'_{ok} + \tau'_{ld})]$ periods for being held at k plus being transported over inter-hub link (k, l) . Hence, this flow may be kept as inventory at hub k for at most $[\pi_{od} - (\tau'_{ok} + \tau'_{ld}) - \tau_{kl}]$ periods, so that it may reach the hub l (before or) at the beginning of time period $t_l = t_0 + \pi_{od} - \tau'_{ok} - \tau_{kl}$.

Let $\hat{D}_{(k,l)}$ denote the set of all demands that need to be transported on hub-link (k, l) . Each $\hat{d} \in \hat{D}_{(k,l)}$ has a total demand of $w_{\hat{d}}$, arrives at hub k at the beginning of time period $t_{\hat{d}k} \in T$, and must reach hub l at (or before) the beginning of time period $t_{\hat{d}l} \in T$. So, for each $\hat{d} \in \hat{D}_{(k,l)}$, the feasible time window for dispatch is $[t_{\hat{d}k}, t_{\hat{d}l} - \tau_{kl}]$. We intend to

make decisions on when and how long to hold inventory at hub k , and when to dispatch a vehicle, of which type v , from k to hub l . That will enable us to minimize the total cost of inventory plus transportation, while ensuring all demands reach their destination nodes within the promised time.

To model the shipment consolidation problem for each inter-hub link, we pre-compute two parameters, based on $t_{\hat{d}k}$ and $t_{\hat{d}l}$, for each demand $\hat{d} \in \hat{D}_{(k,l)}$. The first parameter $\eta'_{\hat{d}t}$ represents how many periods the flow $\hat{d}$ is to be kept in inventory before dispatching, if it is to be dispatched at the beginning of the time period t . For example, if a flow $\hat{d}$ arrived at hub k at the beginning of period 1, and is to be dispatched to l at the beginning of period 3, in that case $\eta'_{\hat{d},2}=1$, $\eta'_{\hat{d},3}=2$ and $\eta'_{\hat{d},t} = 0 \ \forall t \in T \setminus \{2, 3\}$. The second parameter $\eta''_{\hat{d}t}$ represents whether flow $\hat{d}$ was in inventory at period $(t-1)$ or not. The second parameter follows the first, in that $\eta''_{\hat{d}t} = 1$ if $\eta'_{\hat{d}t} \neq 0$, $\forall t \in T$. Those two additional parameters required for the shipment consolidation formulation are described as follows.

$$\eta'_{\hat{d}t} = \begin{cases} t - t_{\hat{d}k}, & \text{if } t \geq t_{\hat{d}k} \text{ and } t < t_{\hat{d}l} - \pi_{kl}, \hat{d} \in \hat{D}_{(k,l)}, \\ 0, & \text{otherwise.} \end{cases}$$

$$\eta''_{\hat{d}t} = \begin{cases} 1, & \text{if } t \geq t_{\hat{d}k} \text{ and } t < t_{\hat{d}l} - \pi_{kl}, \hat{d} \in \hat{D}_{(k,l)}, \\ 0, & \text{otherwise.} \end{cases}$$

We now further define two new decision variables:

$z_{\hat{d}}^t$: Flow on inter-hub link (k, l) , starting at the beginning of time t , with respect to $\hat{d} \in \hat{D}_{(k,l)}$.

g_v^t : Number of dispatches of vehicle type $v \in V$, required on inter-hub link (k, l) at

time $t \in T$.

The shipment consolidation subproblem for inter-hub link (k, l) is then formulated as follows:

$$\text{minimize} \quad \Psi_{(k,l)} = \sum_{\hat{d} \in \hat{D}_{(k,l)}} \sum_{t \in T} \eta_k \eta'_{\hat{d}t} z_{\hat{d}}^t + \sum_{t \in T} \sum_{v \in V} (\kappa_v + \chi_{klv}) g_v^t \quad (3.16)$$

subject to

$$\sum_{t \in T'_{\hat{d}}} z_{\hat{d}}^t = w_{\hat{d}} \quad T'_{\hat{d}} = \{t_{\hat{d}k}, \dots, t_{\hat{d}l} - \pi_{kl}\}, \hat{d} \in \hat{D}_{(k,l)} \quad (3.17)$$

$$\sum_{\hat{d} \in \hat{D}_{(k,l)}} z_{\hat{d}}^t \leq \sum_{v \in V} \phi_v g_v^t \quad t \in T \quad (3.18)$$

$$\sum_{\hat{d} \in \hat{D}_{(k,l)}} \eta''_{\hat{d}} z_{\hat{d}}^t \leq \theta_k \quad t \in T \quad (3.19)$$

$$z_{\hat{d}}^t \geq 0 \quad \forall \hat{d} \in \hat{D}_{(k,l)}, \forall t \in T \quad (3.20)$$

$$g_v^t \in \mathbb{Z}^+ \quad \forall v \in V, \forall t \in T \quad (3.21)$$

The objective function (3.16) represents the overall cost consisting of inventory cost plus the fixed and variable costs for inter-hub transportation. The first summation group of the objective denotes inventory cost, while the second indicates transportation cost. Constraint (3.17) ensures the demanded flow is dispatched from hub $k \in H$ to destination hub $l \in H$ within the feasible time period. Constraint (3.18) calculates the total number and types of vehicle dispatches required in each time period. The relation (3.19) enforces the hub capacity for holding inventory in each period. Constraints (3.20) and (3.21) represent non-negative and integer variables respectively.

Once a set of hub locations and allocations of origin and destination nodes are deter-

mined, the above subproblem may be solved on each inter-hub link to decide on the inventory and vehicle usage. On the other hand, the total costs of that solution are obtained by summing the subproblem objective values over all inter-hub links, plus the cost of establishing hubs and transportation on allocation links. In particular, we calculate the cost of a solution S by $f(S) = \sum_{k \in \mathcal{H}(S)} \Upsilon_k + \sum_{i \in N, k \in H, (i,k) \in \mathcal{A}(S)} c_{ik} O_i + \sum_{j \in N, k \in H, (j,k) \in \mathcal{A}(S)} c_{kj} D_j + \sum_{k,l \in \mathcal{H}(S)} \Psi_{kl}$, where the first term represents location cost, the second and third terms give transportation cost on allocation links, while the last term includes inventory cost plus transportation cost on inter-hub links. Here the sets $\mathcal{H}(S)$ and $\mathcal{A}(S)$ represent hubs and allocations, respectively, in solution S .

We solve the subproblem model using CPLEX to determine the Ψ_{kl} values for each pair of hubs when evaluating a solution. For each solution, we further check the feasibility with respect to inventory holding capacity at each hub. Inventory holding decisions are obtained by solving the subproblem model for every inter-hub link. The total freight quantities to be held over at each hub are compared with the inventory capacity there, and any infeasible solutions are discarded. The next section details how we search the solution space.

3.2.3 Variable Neighborhood Search (VNS) Algorithm

This section presents the VNS algorithm that we developed. Our algorithm iteratively improves the initial solution by performing a sequence of small changes in the structure of the solution to generate different neighbors, and eventually moving to a neighbor with a better cost. We generate various neighborhoods, explore solutions within each of these neighborhoods for potential improvements, and devise a mechanism to transition between

solutions in different distant neighborhoods. This strategy aims to prevent the algorithm from becoming trapped in a locally optimal solution. Interested readers may refer to [Hansen et al. \(2010\)](#) and [Brimberg et al. \(2023\)](#) for extensive reviews of VNS algorithms and applications.

In the sequel, we present the neighborhood structures that we search, the details on implementation of the algorithm’s improvement phase, how we cache the solutions in a tabu list, and the way we enhance the VNS algorithm run-time through memoization.

Neighborhood Structures

Five different neighborhood structures are generated in our algorithm, motivated by the work of [Serper and Alumur \(2016\)](#). Our first way of generating a neighboring solution is to allocate a non-hub node to a different hub. The second way is generated by exchanging allocations of two non-hubs. To obtain the third neighborhood, we increase the number of hubs by one. All non-hub nodes that are closest to this new hub are then allocated to it. The fourth neighborhood is generated by decreasing the number of hubs by one. In this case, all non-hub nodes that had been allocated to the closed hub are allocated to their closest hub. The fifth neighboring solution is based upon exchanging a hub with a non-hub. All non-hub nodes that were allocated to the closed hub are then allocated to the new hub.

To generate random neighbors of a solution, hubs/non-hubs from each of the neighborhoods are selected by using a discrete uniform distribution. All iterations of the algorithm consider each neighborhood structure for improvement. We avoid repeatedly evaluating

the same neighboring solution by the caching technique discussed later in this sub-section.

Improvement Phase

The pseudo-code of Algorithm 2 presents the VNS algorithm in a nutshell. The initial solution S_{init} , generated by greedy construction, is used as the current best solution S_{CB} for the first iteration, and is also assigned to the variable denoting the previous-best solution S_{PB} . Next, we apply the five neighborhood structures to the current best solution and obtain the neighborhood set $\mathbb{S}$.

The elements of the set $\mathbb{S}$, derived from random modifications to the current best solution, may exhibit minimal or significant differences compared to the current best solution. For instance, the neighbors $\mathbb{S}_1$ and $\mathbb{S}_2$ differ from S_{CB} by one and two allocations, respectively. The remaining three neighbors of solution S_{CB} in $\mathbb{S}$, constructed using subsequent neighborhood structures, feature one or two hub differences and additional allocation variations compared to S_{CB} . The selection of which node becomes a hub and which hub is removed from the hub list in each neighbor is random. Therefore, the allocation pattern of non-hubs to hubs may vary significantly among these neighbors. This would depend on which hub is added or removed, and how the non-hubs are allocated to other hubs in the current best solution.

This variability ensures that the VNS algorithm explores both near and distant regions of the solution space in each iteration. The possibility of moving to a close neighbor constitutes the *improvement procedure* as defined by Brimberg et al. (2023), and the possibility that a move may also be made to a distant neighbor ensures the *neighborhood change step*.

Therefore, the neighborhood structures outlined in Subsection 3.2.3 provide a systematic approach to randomly transitioning between close and distant neighbors, constituting the *shaking procedure* as defined in Brimberg et al. (2023). If the current best solution has a higher cost than any of these neighboring solutions, we assign the neighboring solution with minimum cost to S_{CB} .

Algorithm 2 VNS Algorithm

Input: S_{init}

Output: S_{Best}

```

1: Begin
2:  $S_{CB} \leftarrow S_{init}$ 
3:  $S_{PB} \leftarrow S_{init}$ 
4:  $iterCount = 0$ 
5: repeat
6:   for  $i = 1..5$  do
7:     Construct neighboring solution  $\mathbb{S}_i$ 
8:     Solve the shipment consolidation subproblem for  $\mathbb{S}_i$ 
9:     Check feasibility of  $\mathbb{S}_i$ 
10:   end for
11:    $S^* \leftarrow \arg \min_{f(\mathbb{S}_i)} \mathbb{S}_i$ 
12:   if  $f(S_{CB}) \geq f(S^*)$  then
13:      $S_{CB} \leftarrow S^*, count_1 = 0$ 
14:   else
15:      $count_1++$ 
16:   end if
17:   if  $count_1 = limit_1$  then
18:     if  $f(S_{PB}) > f(S_{CB})$  then
19:        $S_{PB} \leftarrow S_{CB}$ 
20:     end if
21:      $S_{CB} \leftarrow S_{init}, count_2++, count_1 = 0$ 
22:   end if
23:    $iterCount++$ 
24: until  $count_2 = limit_2$  or  $iterCount = maxIteration$  or  $runtime \geq allowedTime$ 
25:  $S_{Best} \leftarrow \min(S_{CB}, S_{PB})$ 
26: return  $S_{Best}$ 
27: End

```

We do two things to avoid getting trapped in a local minimum. First, we allow a move to a neighboring solution whose cost equals that of the current best solution. Second, we use a variable $count_1$ that is increased in every iteration when a move to a neighboring solution is not possible, and is set to 0 in the iteration with a possible move. When this variable $count_1$ reaches a threshold $limit_1$, i.e. the current best solution could not be improved for that number of consecutive iterations, S_{CB} is re-initiated to S_{init} . Moreover, S_{CB} is assigned to S_{PB} if the cost of the current best solution is lower than that of the previous best. We employ another variable, $count_2$, to track how many times S_{CB} was re-initiated to the initial solution. This is used as a stopping criterion.

The algorithm ends, based on a set of parameters: a cap on the number of iterations, a cap on the running time, or a cap on the number of times the current best solution was re-initiated to the initial solution. As soon as either of them reaches preset limits, the iterations stop. We report the lower cost, one of S_{CB} and S_{PB} , as the best solution S_{Best} obtained after the iterations.

To ensure the feasibility of solutions to the given problem at all stages of the VNS algorithm, we start with a feasible solution S_{init} , and we never assign an infeasible solution to S_{CB} . This is done in lines 8-9 of the Algorithm 2 pseudo-code, where we solve the shipment consolidation subproblem for all neighboring solutions. Given the location and allocation decisions, for each inter-hub link, the total cost function calculates the quantity of flow demanded, the arrival time of that demand at the first hub, and the travel time allowed on each inter-hub link. The formulation (3.16)-(3.21) introduced in Section 3.2.2 is used to decide inventory and numbers and types of vehicles required. If, under the given location and allocation decisions, any demand cannot be routed to its destination within

the time promised, that is pointed out in this phase. Those are the cases of infeasibility identified. Any neighboring solution that is also infeasible is not considered in subsequent steps of the algorithm. Thus, the feasibility of the best solution is maintained throughout every iteration.

It should be noted that we *additionally* tested our algorithm by randomly generating initial solutions and varying them in each iteration to preserve solution diversity, rather than only employing the greedy approach outlined in Section 3.2.1. We observed that the use of randomly generated initial solutions, both before and during iterations, led to an increase in computational time without a corresponding improvement in solution quality.

The VNS algorithm described above could have been designed to execute the exploration of closer neighboring solutions before transitioning to more distant neighbors, as done in a *Variable Neighborhood Descent* algorithm. Alternatively, another variant of the basic VNS scheme could have been utilized. Such approaches would involve a more exhaustive search of the solution space; hence, the tradeoff between the potential for obtaining improved solutions and increased runtime should be considered.

Caching in a Tabu Solution List

This search algorithm is made efficient by caching solutions in a tabu list. After randomly generating neighboring solutions, the tabu list is consulted to see whether the new solution was previously encountered. This facilitates avoidance of cycling. If an already-generated neighbor is found, a different neighbor is searched using the same structure, until a given maximum number of iterations is reached. If a new neighbor cannot be found within this

limit, then that particular neighborhood structure is not evaluated for the given solution. This avoids repeated evaluation of solutions, and hence, improves the run-time performance of the VNS algorithm.

Memoization

To enhance the run-time of the VNS algorithm, we employed a technique called *memoization*. This helped to prevent solving the shipment consolidation subproblem for any inter-hub link and respective allocations that were previously evaluated. To do that, we created a database containing entries on each inter-hub link, allocations of non-hub nodes to the hubs of that link, and total cost of inventory plus transportation for that inter-hub link. Every time the shipment consolidation subproblem needs to be solved for a pair of hubs and the corresponding allocations, first, a lookup is made to this database for memoized results. If no such entry is found, the subproblem is solved and an entry is made to the database. Otherwise, if one already exists, i.e., the same inputs to the shipment consolidation subproblem occur again, the cost cached at the database is utilized, thus avoiding the call to the subproblem solver. Initially, for a data instance, this database is empty. However, it gets gradually populated as the VNS algorithm explores neighborhoods, and later iterations thus run faster than earlier iterations.

3.3 Computational Analysis

In this section, we furnish some numerical results using the model and algorithm presented in Sections 3.1 and 3.2. For running the computational tests, an IBM laptop PC with an Intel Core-i7 processor and 16 GB RAM was used. The formulation and algorithm presented have been implemented in CPLEX 12.9 and Java (IntelliJ IDEA 21.2.1 using JDK 15.0.2). We first discuss the process of generating data and parameters in Section 3.3.1, followed by assessments of our algorithm’s performance in Section 3.3.2 and some insights from our results in Section 3.3.3.

3.3.1 Data and Parameters

To test our model and algorithm, we generated instances with up to 65 demand nodes, corresponding to the cities of the Civil Aeronautics Board (CAB) data set introduced by O’Kelly (2014). All origin and destination nodes were taken as candidate hub locations. A list of cities is given in Appendix A.

The Freight Analysis Framework (FAF), developed by the Center for Transportation Analysis (CFTA, 2023), provides O-D flow data for 43 commodities and 8 transportation modes. We considered both domestic and international air shipments in this dataset. For each pair from the 65 cities, our data set summed the weights of 43 commodities for the two shipment modes we utilized, in calculating the total demand for each O-D pair. We used the demand quantities for the years 2019, 2020, and 2022 (data for 2021 was not available due to the COVID’19 pandemic). We then randomly distributed those demands

over the different time periods of our planning horizon, using a uniform distribution.

Transportation costs on allocation links have been determined based on average operational costs of trucking per mile, as reported by the American Transportation Research Institute ([Leslie and Murray, 2022](#)). The driving distances between cities were computed using the default settings of Google Maps. Operational costs of trucking are estimated at \$1.70, \$1.65, and \$2.20 per ton per mile for the years 2019, 2020, and 2022, respectively.

We considered two types of cargo planes, B-747-F and B-747-100, for transportation on the inter-hub links, as employed by [Masaeli et al. \(2018\)](#). To estimate the fixed cost per dispatch on those links, the purchasing/leasing costs of aircraft were estimated, and a lifespan of 20 years with four legs of travel per day was assumed ([Masaeli et al., 2018](#)). Respective fixed costs per flight are estimated at \$9247 and \$8082 for B-747-F and B-747-100; the capacities of the two aircraft are 72.58 and 46.38 tons. We assume uniform density for the items carried, thus volume is equivalent to weight.

Variable transportation costs for airplanes on the inter-hub links considered costs such as fuel, maintenance, and flying operations. For approximation purposes, we used the annual average air freight rates published in the 33rd Annual State of Logistics Report by the Council of Supply Chain Management Professionals ([CSCMP, 2024](#)) for realistic costs of B-747-F and B-747-100. With those CSCMP rates and the FAF info, variable costs of \$3.75, \$6.30, and \$6.60 per kg per mile were obtained for B-747-100 for years 2019, 2020, and 2022, respectively. We used the same process followed by [Masaeli et al. \(2018\)](#) to estimate the variable cost for B-747-F from the above information. The air travel distances were calculated using Haversine distances between cities ([Rouberol, 2022](#)).

The Freight Analysis Framework provides different values for 43 item types. Our data set picked four product categories (pharmaceutical, electronics, plastics or rubber, and furniture) covering a wide range of values (from among types provided by the Freight Analysis Framework) for testing different holding cost figures. As in [Silver et al. \(2017\)](#), holding cost is expressed as a percent of the value of the inventory item, where 365 days per year and 25% of the item’s dollar value per year are assumed for calculating holding costs. The resulting costs for pharmaceuticals, electronics, plastics or rubber, and furniture are \$286, \$124, \$19, and \$9 per ton per day, respectively. We categorize these products’ inventory holding costs as *high*, *medium-high*, *medium-low*, and *low*, respectively. We assume a hub capacity of 500 tons per hub for inventory hold-over; the hub establishment cost was varied in { \$5,000, \$10,000, \$25,000 } for all hubs. This cost for location is *annualized*, considering that a hub is in use for a period of 25 years following its establishment. All the parameters are adjusted to a per-unit-of-time scale for the duration of the planning horizon.

We consider a planning horizon of one week with 14 periods, each corresponding to 12 hours, i.e. half a day; a promised time of 2 days (4 periods) is required for all deliveries. Travel times between hub pairs were calculated considering the distances and airplane cruising speeds. We add a total of 5.5 hours to the cruising time to account for loading and unloading operations, sorting, and other hidden times. Travel times on the allocation links were estimated by assuming an average of 55 miles per hour cruising speed for trucks, and that the trucks will be operated by two drivers exchanging hands every 6 hours for an optimal travel time. Two hours have been added to this time for loading, unloading at both ends, and sorting at the hub. The calculated travel times, using this method, were rounded up to the nearest integers, representing the number of time periods required for

the journey.

3.3.2 Algorithmic Performance

Table 3.1 presents results for the model and our SCHL-algorithm based on the 2019 data sets with 10-, 15- and 20-cities. The first column indicates the setup cost for a hub node. For each setup cost, different holding costs ranging from low to high were tested to investigate the impact of these values on location, inventory, and routing decisions. For each instance, the total costs (in \$1000s) from the model's objective value, as well as the cost from the SCHL-algorithm, are reported. The CPU times (in seconds) and optimality gaps are also indicated for runs in CPLEX. For the SCHL-algorithm, we indicate CPU times and the number of iterations taken. We report the median of 10 independent VNS runs for each instance. The table's final column shows the improvement of the SCHL-algorithm relative to the run on CPLEX for each instance. This percent improvement was calculated by $(\text{CPLEX Cost} - \text{SCHL Cost}) / \text{CPLEX Cost}$. For every instance in Table 3.1, CPLEX had a 2-hour run-time limit.

For the 10-city problems, we see that both SCHL and CPLEX reached solutions with equal costs. CPLEX, in this case, obtained solutions with gaps ranging from 4.9% to 14.8% in 2 hours of run time. The SCHL-algorithm, on the other hand, more quickly obtained the same solutions. The longest run time was 31.4 seconds, and the shortest was 9.9 seconds. The algorithm required between 142 to 194 iterations before stopping, while 5000 total iterations were allowed. Next, we also let CPLEX run until optimality for 10-city instances and noticed that all the solutions reported in Table 3.1 under CPLEX columns

Table 3.1: Comparison of the SCHL-algorithm with CPLEX on 10-, 15- and 20-city instances.

	Estab. Cost	Holding Cost	CPLEX Results			SCHL-algorithm			
			Cost (1000\$)	Time (sec.)	Gap (%)	Cost (1000\$)	Time (sec.)	Iter.	Impr. (%)
10-city instances	\$5,000	low	152	7202	7.1	152	14.2	194	0
		med-low	153	7206	13.0	153	10.7	158	0
		med-high	161	7202	11.3	161	10.8	157	0
		high	161	7202	14.8	161	9.9	152	0
	\$10,000	low	162	7202	6.7	162	28.8	181	0
		med-low	163	7208	6.8	163	31.4	156	0
		med-high	171	7218	12.2	171	12.1	169	0
		high	171	7204	12.6	171	11.8	175	0
	\$25,000	low	192	7203	4.9	192	14.5	147	0
		med-low	193	7214	6.0	193	14.8	163	0
		med-high	201	7206	5.1	201	11.8	142	0
		high	201	7202	7.3	201	11.9	173	0
15-city instances	\$5,000	low	350	7202	40.4	334	33.8	322	4.5
		med-low	353	7203	42.2	336	29.0	299	4.8
		med-high	382	7210	41.2	352	31.2	337	8.0
		high	412	7204	43.6	365	33.4	362	11.5
	\$10,000	low	370	7204	36.7	349	37.2	301	5.0
		med-low	373	7208	37.4	351	31.8	249	5.9
		med-high	402	7206	37.5	373	33.0	347	7.4
		high	432	7204	37.8	380	39.1	345	12.1
	\$25,000	low	430	7204	35.4	394	38.5	338	8.3
		med-low	433	7204	28.2	396	41.9	310	8.5
		med-high	462	7203	32.3	418	40.0	297	9.7
		high	492	7204	38.2	425	33.3	326	13.7
20-city instances	\$5,000	low	836	7210	66.9	617	122.9	337	26.2
		med-low	970	7222	70.9	621	116.7	319	35.8
		med-high	1503	7218	79.7	653	117.9	309	55.8
		high	1503	7218	78.8	699	134.4	369	53.5
	\$10,000	low	781	7211	58.7	637	89.3	267	18.5
		med-low	1508	7210	80.2	641	116.2	292	57.4
		med-high	1508	7218	79.4	683	95.4	362	54.4
		high	1508	7215	78.6	719	128.7	375	52.2
	\$25,000	low	787	7208	47.7	697	116.4	297	11.5
		med-low	1064	7208	60.6	701	101.1	298	34.0
		med-high	1523	7218	73.9	743	115.2	315	50.8
		high	1523	7215	100	779	126.6	322	48.8

are actually optimal. Our algorithm are able to obtain the optimal solutions in a shorter amount of time.

For the 15-city problems, the SCHL-algorithm obtained better solutions than that of CPLEX, with a lower run-time in all respective cases. CPLEX reported solutions with optimality gaps ranging from 28.1% to 43.6%. The SCHL-algorithm took between 29.0 to 41.9 seconds and 297-362 iterations. An average 8.3% improvement on CPLEX results was obtained by the SCHL-algorithm in terms of solution cost, in the case of 15-city instances. For the 20-city instances, the improvement made by the SCHL-algorithm was even better. Solution costs resulting from the algorithm were on average 41.6% lower than those of CPLEX. CPLEX yielded gaps ranging between 47.7% and 100% in these instances. The SCHL-algorithm took between about 89 and 134 seconds and 267-375 iterations.

We observe that the mean optimality gap grows with the size of the instances, where the average CPLEX gaps are 9%, 37.6%, and 72.9% for 10-, 15-, and 20-city instances, respectively. For 20-city instances, the lowest optimality gap of CPLEX was 47.7% in a 2-hour run time. In 10-city instances, the SCHL algorithm yielded identical solutions to those produced by CPLEX. However, for 15-city and 20-city instances, the SCHL solutions demonstrated an average cost enhancement of 8.3% and 41.6%, respectively, when compared with the CPLEX solutions.

Table 3.2 presents results obtained by the SCHL-algorithm for 30-, 45- and 65-city instances. Successive columns present the total cost of a solution (in \$1000s), CPU-time (in seconds), and the number of iterations required, along with the number of hubs opened and the number of flights for the aircraft B-747-F and B-747-100. CPLEX could not run

the model for instances of 30 or more cities, due to memory.

Table 3.2: Results with the SCHL-algorithm for 30-, 45- and 65-city instances.

	Estab. Cost	Holding Cost	SCHL-algorithm					
			Cost (1000\$)	Time (sec.)	Iter.	Hubs	B-747-F	B-747-100
30-city instances	\$5,000	low	1828	289.0	380	8	14	97
		med-low	1840	249.8	309	8	14	97
		med-high	1936	293.4	296	8	15	94
		high	2038	241.6	350	7	16	85
	\$10,000	low	1868	236.7	322	8	14	97
		med-low	1873	106.8	357	8	13	94
		med-high	1976	107.4	398	8	15	94
		high	2073	101.4	391	7	16	85
	\$25,000	low	1988	143.6	342	8	14	97
		med-low	2000	166.3	346	7	15	82
		med-high	2083	143.1	286	7	16	83
		high	2178	102.6	392	7	16	85
45-city instances	\$5,000	low	3268	297.2	471	8	24	91
		med-low	3297	342.7	466	8	24	92
		med-high	3467	299.3	412	8	23	96
		high	3642	268.7	481	8	18	107
	\$10,000	low	3308	465.5	462	8	24	92
		med-low	3337	490.9	426	8	24	92
		med-high	3507	484.1	440	8	23	96
		high	3682	491.7	496	8	18	107
	\$25,000	low	3428	537.2	494	8	24	92
		med-low	3446	331.1	506	8	24	92
		med-high	3627	226.2	431	8	23	96
		high	3814	214.3	426	8	18	107
65-city instances	\$5,000	low	4814	854.9	568	9	38	116
		med-low	4830	900.0	536	9	37	118
		med-high	5046	900.0	512	9	33	124
		high	5270	854.9	600	9	22	146
	\$10,000	low	4850	900.0	546	9	37	118
		med-low	4875	900.0	527	9	37	118
		med-high	5091	900.0	573	9	33	124
		high	5315	889.9	605	9	22	146
	\$25,000	low	4994	900.0	516	9	38	116
		med-low	5015	900.0	537	9	38	116
		med-high	5226	845.0	540	9	33	124
		high	5450	610.5	567	9	22	146

For 30-city instances, the run-time and the number of iterations required for the SCHL-algorithm were between 101–293 seconds and 309–392 iterations. This was a 58% increase in terms of average run-time over that of 20-city instances, while an average of 8% increase in terms of the number of iterations. The growth in run time exceeds that of the number of iterations because, in each iteration, more time is necessary for the SCHL-algorithm to evaluate larger-sized neighborhoods. Note that the growth in run-time from the 15-city to 20-city instances was 227%, although the number of iterations only increased by less than 1%. We see a similar phenomenon in cases of even larger instances.

For 45-city instances, the run-time and the number of iterations required for the SCHL-algorithm were between 214–537 seconds and 412–506 iterations. For a typical case, this was a 104% increase in the average run-time over that of 30-city instances, while a mean increase of 32% in terms of the number of iterations. The growth in run-time is again higher as compared to that of the number of iterations. For 65-city instances, the SCHL-algorithm required between 845–900 seconds of runtime and 512–605 iterations. This was a 58% increase in average run-time over the 45-city instances, while a mean 17% increase in the number of iterations.

As shown in Table 3.2, larger networks utilize more hubs and aircraft. In 30-city scenarios, 7-8 hubs were established, whereas for 45- and 65-city instances, this number increased to 8 and 9, respectively. In comparison to the average number of aircraft used in 30-city scenarios, there was a 12.5% upsurge in the case of 45-city scenarios. However, for the 65-city cases, a more significant increase of 33.3% was observed in the total number of aircraft compared to the 45-city instances. When transitioning from 30-city to 45-city scenarios, the average utilization of B-747-F aircraft increased by 50%, while the usage

of B-747-100 aircraft showed a modest 6.4% increment. However, when comparing these numbers for the 65-city scenarios to those of 45-city instances, there was a 46% rise in the use of B-747-F aircraft, and the utilization of B-747-100 aircraft saw a substantial 33.4% increase. More detailed insights on aircraft usage will be provided in Section 3.3.3.

We observed a decreasing trend in the relative growth of average run times as the instance size increased. This trend is attributed to the implementation of memoization, discussed in detail in Section 3.2.3. As the instance size grows, neighboring solutions result in less significant changes to the location and allocation decisions compared to smaller instances. Consequently, neighboring solutions exhibit substantial overlaps in location-allocation decisions. Since the cost of this overlapping part is retrieved from the memo database, there is no need for redundant computation. This advantage in terms of time-saving is more pronounced in larger instances, as smaller instances have fewer overlaps among neighbors, making the time-saving effect less evident in their case.

3.3.3 Insights from the Results

Let us now comment on the resulting hub networks and the numbers of each aircraft type employed. Table 3.3 presents those findings obtained by CPLEX and the SCHL algorithm, corresponding to the results reported in Table 3.1. For 10-city instances, we observe that decisions on hub locations are not affected by variations in the establishment or holding costs. In all cases, hubs are established in Atlanta and Chicago. However, the transportation mode decisions are affected by the value of holding costs. We see that when the holding cost increases from low to medium, four B-747-F and three B-747-100 aircraft

are used. But when the holding cost becomes medium-high, one less B-747-F and one more B-747-100 aircraft are employed. In case of high holding cost, four of each aircraft type are used. The B-747-F has a larger capacity and lower per-ton fixed cost than that of the B-747-100. When the holding cost goes from low and medium-low to medium-high and high, less freight is held over, to compensate for the fixed cost of transportation. The demand data are not symmetric. Therefore, opposite directions may require different numbers of flights between pairs of hubs.

For the 15-city instances, four and three hubs were established by CPLEX and the SCHL-algorithm respectively. Hubs were opened in Atlanta, Chicago, New York, and Austin by CPLEX, while the SCHL algorithm opened hubs only in Atlanta, Chicago, and New York. Those three sites, where hubs were opened in both cases, were the top locations in terms of demand volumes originating from and destined to.

It may not be meaningful to comment on the number and types of aircraft in the solutions obtained by CPLEX, due to high optimality gaps in the results. But in general, the number of aircraft increased as the holding cost grew from low and medium-low to medium-high and high. The aircraft B-747-100 was more frequently used than B-747-F in all instances, for CPLEX results and for those obtained by the SCHL-algorithm. For low and medium-low holding costs, the SCHL solution utilized 3 B-747-F aircraft plus 17 B-747-100 aircraft. On the other hand, for medium-high and high holding costs, the SCHL-algorithm employed 4 B-747-F aircraft plus 16 B-747-100 aircraft. Utilization of one more B-747-F aircraft, in replacement of one B-747-100, indicates the benefit of larger capacity aircraft with greater fixed costs in avoidance of inventory holdover when the holding cost increases.

Table 3.3: Numbers of hubs and flights of B-747-F and B-747-100, obtained with CPLEX and SCHL-algorithm on 10-, 15- and 20-city instances.

	Estab. Cost	Holding Cost	CPLEX Results			SCHL-algorithm		
			Hubs	B-747-F	B-747-100	Hubs	B-747-F	B-747-100
10-city instances	\$5,000	low	2	4	3	2	4	3
		med-low	2	4	3	2	4	3
		med-high	2	3	5	2	3	5
		high	2	4	4	2	4	4
	\$10,000	low	2	4	3	2	4	3
		med-low	2	4	3	2	4	3
		med-high	2	3	5	2	3	5
		high	2	4	4	2	4	4
	\$25,000	low	2	4	3	2	4	3
		med-low	2	4	3	2	4	3
		med-high	2	3	5	2	3	5
		high	2	4	4	2	4	4
15-city instances	\$5,000	low	4	5	16	3	3	17
		med-low	4	4	18	3	3	17
		med-high	4	4	19	3	4	16
		high	4	5	20	3	4	16
	\$10,000	low	4	4	18	3	3	17
		med-low	4	3	19	3	3	17
		med-high	4	4	20	3	4	16
		high	4	5	19	3	4	16
	\$25,000	low	4	5	19	3	3	17
		med-low	4	3	18	3	3	17
		med-high	4	4	19	3	4	16
		high	4	4	23	3	4	16
20-city instances	\$5,000	low	10	13	52	4	3	27
		med-low	11	10	67	4	5	23
		med-high	1	0	0	4	6	22
		high	1	0	0	4	5	28
	\$10,000	low	7	12	44	4	3	27
		med-low	1	0	0	4	5	23
		med-high	1	0	0	4	5	25
		high	1	0	0	4	5	28
	\$25,000	low	6	14	32	4	3	27
		med-low	8	11	59	4	5	23
		med-high	1	0	0	4	5	25
		high	1	0	0	4	5	28

Because the CPLEX solutions had a gap of at least 47%, it is not easy to find a pattern for 20-city runs. For some instances, CPLEX opened only a single hub; for others, it opened 11 hubs. The number of aircraft employed ranges from as low as *zero* to as high as 77. However, in those cases too, hubs were opened at locations that were almost always in the top 50% of cities in terms of demand. The SCHL-algorithm opened hubs in 4 locations for 20-city instances – Miami, Atlanta, Chicago, and New York. We also observe more use of B-747-100 aircraft than B-747-F in the SCHL solutions. When the holding cost increased from low to medium-low and medium-high, more B-747-F were employed in place of B-747-100; since the former has a larger capacity, it allowed greater shipments of freight. When the holding cost increased to high, more B-747-100 were utilized to avoid holding inventory in this case too.

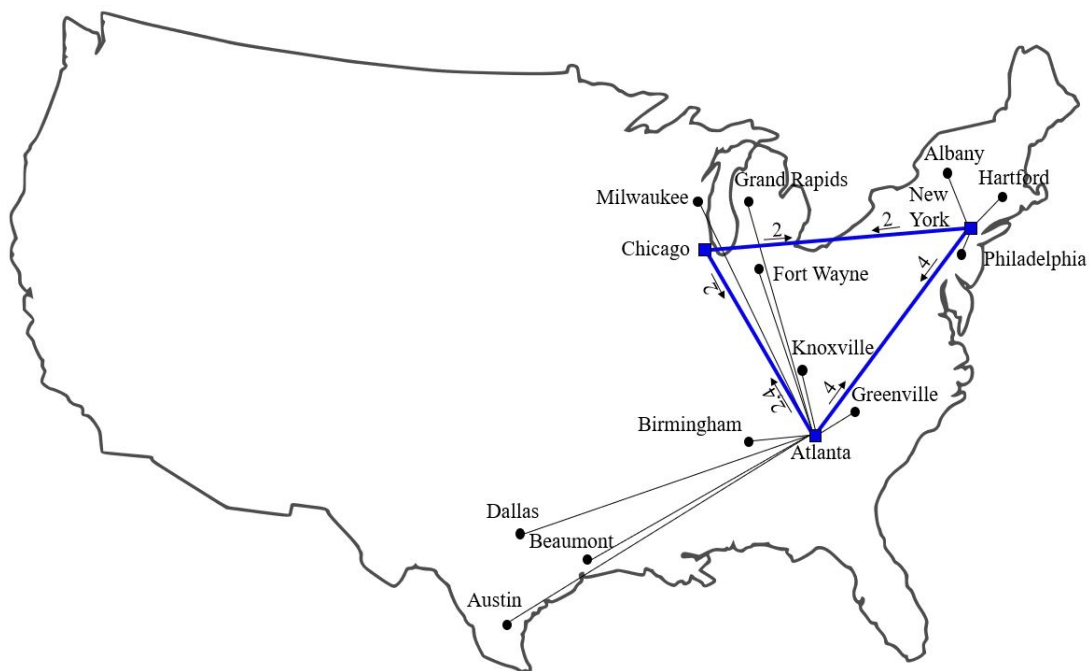
We observe a similar phenomenon in the last three columns of Table 3.2 that correspond to larger-sized instances. For 30-city instances, 7 or 8 hubs were established by the SCHL-algorithm. We note that fewer hubs resulted when there were increases in either the establishment costs or holding costs. B-747-100 was used most of the time for low to medium-high holding costs. For high inventory costs, we see fewer of those aircraft and more B-747-F. The sites chosen as hubs were among the highest-demand-generating cities. For the 45- and 65-city instances, 8 and 9 nodes were chosen as hubs respectively. The number of hubs, in those cases, was unchanged with variations in the cost of hub establishment or holding cost. The use of B-747-F decreased as the holding cost increased, while additional B-747-100 were employed. It is noteworthy to point out that, with its fixed cost per flight and variable costs, the B747-F is about twice as expensive to operate as the B747-100, although its capacity is less than double that of the B747-100. With lower

capacity, the latter's cheaper per-trip cost led to its operation on more flights.

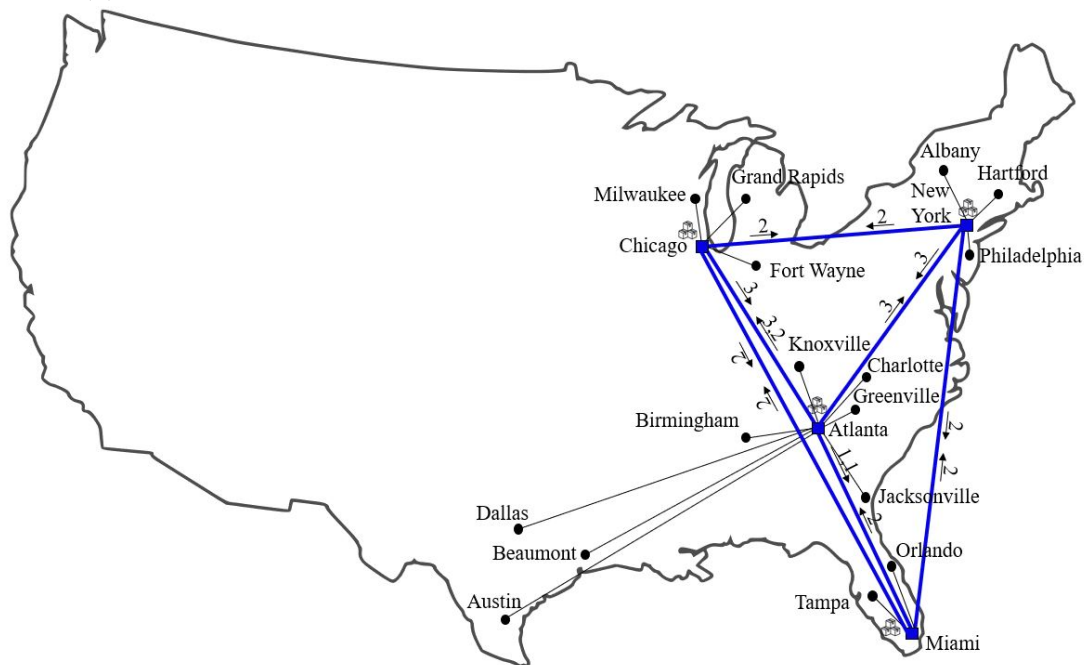
Figure 3.1 shows two network configurations: Figure 3.1(a) for the 15-city instance with high holding costs, and Figure 3.1(b) for the 20-city instance with low holding costs, both having a hub setup cost of \$5,000. In these figures, black circles represent origin-destination (O-D) nodes, which were also considered as potential hub locations. Blue rectangles denote the chosen hub locations. Black arcs represent allocation links connecting O-D nodes to hubs, while blue arcs represent inter-hub links. The types of airplanes used on inter-hub routes are indicated by numbers, with B-747-100 listed first, followed by B-747-F. The arrows on the arcs indicate flight directions. Hubs with inventory holding are marked with a special symbol next to them.

In Figure 3.1(a), three hubs were established in Atlanta, Chicago, and New York. The B747-100 aircraft were predominantly used for inter-hub links, with 4 B747-F and 16 B747-100 airplanes in operation. Despite their proximity to Chicago, the cities of Milwaukee, Grand Rapids, and Fort Wayne, surprisingly, were allocated to Atlanta. That made Chicago solely responsible for its own flows. B-747-100 planes were used on all inter-hub links, with two B-747-F aircraft operating between Atlanta and Chicago. No hubs held inventory, due to high holding costs. For instance, shipments from Chicago to Dallas at the start of $[1,2)$ were consolidated with other flows in Chicago and dispatched to Atlanta on a B-747-100. The consolidated load then moved from Atlanta to Dallas at the start of $[2,3)$.

In Figure 3.1(b), hubs were established in Atlanta, Chicago, New York, and Miami for the 20-city instance with low holding costs and a \$5,000 setup cost. This instance included



(a) 15-city instance with hub setup cost of \$5,000 and high holding cost.



(b) 20-city instance with hub setup cost \$5,000 and low holding cost.

Figure 3.1: Examples of resulting hub networks.

all cities from the 15-city instance in Figure 3.1(a) plus five additional cities. Miami was selected as a hub, while the additional cities were assigned to their nearest hubs. Atlanta served eight cities, Chicago and New York each handled three neighboring cities, and Miami acted as the hub for Tampa and Orlando. All hubs held inventory, with Atlanta holding the most.

In this scenario, 30 flights operated, with 27 using B747-100 aircraft and only 3 using B747-F, on the inter-hub links of Figure 3.1(b). The low holding cost led to freight being held. For instance, consider three sets of flows originating in the first period: Austin to Tampa, Dallas to Orlando, and Greenville to Orlando. These loads were held in Atlanta during [2,3), consolidated with other arriving flows by [3,4), and all were dispatched to Miami on a B-747-100 during [3,4). Each load was then sent to its respective destination at the start of the next period.

Table 3.4 reports results of 65-city instances, with demands of three separate years – 2019, 2020, and 2022. We give the comparative cost of network establishment and operations, the number of hubs established, and the total number of flights required. Recall that 2020 was the first year of the COVID’19 pandemic; 2019 can be considered as a time of “usual” supply chain operations. 2022 observed global inflation caused by the economic after-effects of the pandemic. Table 3.4 thus shows how these changes may have affected supply chain decisions on location and transportation.

Demands in 2020 and 2022 were respectively 8% lower and 5% higher than in 2019. Trucking cost dropped 3% in 2020 but increased by 29% in 2022, relative to that in 2019. The fixed cost of air cargo dispatch, estimated by assuming an aircraft’s lifespan of 20

Table 3.4: Results of SCHL-algorithm on 65-city instances with the demands of 2019, 2020, and 2022.

Estab. Cost	Holding Cost	Cost (in 1000\$)			Hubs			Flights		
		2019	2020	2022	2019	2020	2022	2019	2020	2022
\$5,000	low	4814	4992	6121	9	9	10	154	147	183
	med-low	4830	5014	6146	9	9	10	155	148	183
	med-high	5046	5206	6411	9	8	10	157	123	184
	high	5270	5413	6750	9	8	10	168	131	195
\$10,000	low	4850	5037	6171	9	9	10	155	147	183
	med-low	4875	5059	6196	9	9	10	155	148	183
	med-high	5091	5246	6461	9	8	10	157	123	184
	high	5315	5453	6800	9	8	10	168	131	195
\$25,000	low	4994	5167	6321	9	8	10	154	118	183
	med-low	5015	5187	6346	9	8	10	154	119	183
	med-high	5226	5366	6611	9	8	10	157	123	184
	high	5450	5573	6950	9	8	10	168	131	195

years, was taken as unchanged for the three years. However, the variable transportation cost for airplanes increased significantly, by almost 73% in 2020 and 81% in 2022 compared to the 2019 average per-mile cost. No other cost changes were considered.

Against the variations in demand and variable costs of aircraft and trucking, 2020 saw a 3% increase in overall cost; the corresponding increase above 2019 was 27% in 2022. The latter's large rate was due to increases in all three data parameters above. To cater to the variation in demand, on average 9 hubs were established in 2019, versus around 8 in 2020 and 10 in 2022. This change is consistent with the increase/decrease in demand over the three years. On average, about 159, 132, and 186 flights were used to satisfy demands in the respective three years. The chosen hub locations remained fairly consistent over that time interval, despite notable fluctuations in demand and costs. On the other hand, the number of flights operated exhibited variability during the three years. This observation

underscores that hub location is a strategic decision and is resilient to annual changes, whereas freight dispatch, being a tactical decision, demonstrates considerable flexibility by adapting to changes in demand volume and cost structures.

3.3.4 Value of Holding Inventory

In this section, we aim to highlight the benefits of integrating strategic hub location decisions with tactical decisions concerning shipment consolidation and inventory management. To achieve this goal, we perform a comparative analysis between the networks produced by our model and a scenario where no freight is held over at hub nodes. This comparative analysis provides valuable insights into how the resulting networks differ, in terms of both cost and design, when inventory holding decisions are incorporated. In essence, we seek to evaluate the advantages of our integrated approach by examining how it affects not only the total cost, but also the transportation and inventory management decisions. This analysis will help us better understand the impact of our holistic strategy and guide decision-making in optimizing both the inventory and transportation processes.

Table 3.5 compares scenarios with and without inventory holding. The columns detail cost components, the number of established hubs, vehicle utilization on inter-hub links, and the percentage of improvement in considering our holistic approach. This analysis covers 10- and 15-city instances under different parameter variations. For example, in the 15-city instance with \$5,000 fixed cost and no inventory, five hubs were established in Austin, Atlanta, Chicago, Dallas, and New York. In contrast, when inventory was held at a low cost, no hub was opened in Austin. Without inventory, the total hub establishment

Table 3.5: Comparison of the savings against inventory on 10- and 15-city instances.

	Estab. Cost	Holding Cost	Cost (1000\$)				No. of Hubs	No. of Vehicles	Impr. (%)
			Hub	Transp.	Inv.	Total			
10-city instances	\$5,000	none	10	171	0	181	2	10	-
		low	10	141	1	152	2	7	16
		med-low	10	141	2	153	2	7	15.5
		med-high	10	148	3	161	2	8	11
		high	10	151	0.1	161	2	8	11
	\$10,000	none	20	171	0	191	2	10	-
		low	20	141	1	162	2	7	15.2
		med-low	20	141	2	163	2	7	14.7
		med-high	20	148	3	171	2	8	10.5
		high	20	151	0.1	171	2	8	10.5
	\$25,000	none	50	171	0	221	2	10	-
		low	50	141	1	192	2	7	13.1
		med-low	50	141	2	193	2	7	12.7
		med-high	50	148	3	201	2	8	9
		high	50	151	0.1	201	2	8	9
15-city instances	\$5,000	none	25	440	0	465	5	36	-
		low	20	326	4	350	4	21	24.7
		med-low	20	327	6	353	4	22	24.1
		med-high	20	342	20	382	4	23	17.8
		high	20	358	34	412	4	25	11.4
	\$10,000	none	50	489	0	539	5	39	-
		low	40	327	3	370	4	22	31.4
		med-low	40	325	8	373	4	22	30.8
		med-high	40	346	16	402	4	24	25.4
		high	40	372	20	432	4	24	19.9
	\$25,000	none	100	428	0	528	4	33	-
		low	100	329	1	430	4	24	18.6
		med-low	100	326	7	433	4	21	18
		med-high	100	345	17	462	4	23	12.5
		high	100	376	16	492	4	27	6.8

cost was \$25,000, transportation cost was \$439,727, and holding cost was zero. Opting for freight holdover at a low cost added \$4,074 for inventory, but hub establishment and transportation costs reduced to \$20,000 and \$325,926, respectively, resulting in total cost savings of 24.7%.

Moreover, the number of aircraft employed decreased from 36 to 21 when inventory was held at a low cost. It is expected that transportation expenses incurred without inventory storage surpass those with inventory on hand. This difference can be attributed to the impact of shipment consolidation. When inventory is held, fewer vehicles are needed for transportation, resulting in the avoidance of significant fixed costs associated with dispatching each vehicle. Although this may entail bearing a minimal holding cost, it pales in comparison to the considerable savings achieved through utilization of fewer vehicles and enhanced shipment consolidation.

On average, under different parameter configurations, we achieved 12.3% and 20.1% savings for 10- and 15-city instances, respectively, by allowing inventory holding and optimizing those decisions. The result was an overall mean savings of 16.2%. For 10-city instances, two hubs were consistently opened in all cases, while for 15-city instances, four or five hubs were established with no inventory, compared to four hubs with inventory holding. Without inventory, hubs were in Atlanta, Chicago, New York, Dallas, and sometimes Austin or Beaumont. With inventory holding, hubs were in Atlanta, Chicago, New York, and Beaumont. On average, 33% fewer aircraft were needed with inventory holding, especially for B-747-100 aircraft compared to B-747-F.

The results of this analysis underscore the significance of simultaneously optimizing hub location and shipment consolidation decisions. Such an approach not only leads to substantial cost savings, but also demonstrates that the integration of inventory holding decisions can influence strategic network design choices.

3.4 Summary

In this chapter, we address the integration of shipment consolidation decisions within a hub location model, and explore the role of inventory holding in optimizing hub networks, particularly when O-D demand is spread over multiple periods, and each demand has a promised delivery time. The hub network is designed to enhance operational efficiency by consolidating shipments across different periods and origins, leveraging economies of scale. The problem is formulated as a mixed-integer optimization model, minimizing total costs related to hub establishment, inventory holding, inter-hub transportation, and allocation link transportation. The model accounts for a heterogeneous fleet of vehicles and determines hub locations, node allocations, inventory holding decisions, and vehicle dispatches, ensuring timely delivery and capacity constraints. Numerical experiments showed the model is computationally intensive for large instances. To address this, we developed the SCHL-algorithm, which decomposes the problem into two stages: in the first, the algorithm determines hub locations and allocations, while in the second stage, it addresses shipment consolidation for each pair of hubs. The initial stage's hub locations and allocations are further refined through neighbor generation and a Variable Neighborhood Search (VNS) algorithm, enhancing overall computational efficiency.

We extended a dataset from the literature by enhancing the list of locations and updating the origin-destination (O-D) demands with real data from 2019, 2020, and 2022 to analyze the resulting hub networks and the performance of the proposed mathematical model and algorithm. We employed various parameter settings, including changes to the establishment cost of hubs and inventory holding costs, and analyzed the optimal hub lo-

cations, the structure of the hub network, and the number and types of vehicles used in the solutions. Results provide managerial insights into how the structure of network, and the number and types of vehicles used between hubs, is sensitive to the costs.

Next, in Chapter 4, we study integration of shipment consolidation decisions in hub arc location models with multi-period demand and a heterogeneous fleet of vehicles operating with different frequencies on hub arcs.

Chapter 4

Shipment Consolidation in the Hub Arc Location Problem

Section 2.6 described how the design of transportation networks has been done in the literature by locating *hub arcs*. This idea has not been utilized to help network design for freight transportation. In this chapter, we integrate the operational decisions of shipment consolidation and inventory holding in hub arc location models. We assume a scenario where the freight carrier has a fleet composed of different types of vehicles; it may operate them on the hub arcs with distinct *frequencies*. Similar to what was assumed in Chapter 3, the local operations here are handled by a 3PL who will do pickup and delivery to O-D nodes, charging a price per unit weight per unit distance.

Vehicles of different costs and capacities may be employed on hub arcs. If these vehicles can be loaded to their ideal capacity, transportation costs may be saved by exploiting

economies of scale. Optimal design of such networks may contribute to this cost reduction by scheduling vehicles between hubs in such a way that shipments are consolidated, and large vehicles are used to carry them between inter-hub links.

A major concern for carriers or 3PL firms is that they need to know how many vehicles of which type they should keep in their fleets. This helps their decisions involving vehicle procurement and maintenance planning. In addition, these companies need to know when a vehicle should be dispatched between two nodes. If a ‘frequency’ for each type of vehicle on each link can be decided, keeping in view the cost minimization and service performance, it will help answer the two questions above. In this research, we study the impact of such transportation frequencies under *multiple allocation*, where demand nodes can be allocated to more than one hub arc. Since hub *arcs* are located, rather than hub *nodes* within this framework, the multiple allocation strategy makes more sense because it allows for greater flexibility in routing and consolidation, enabling better distribution of loads across the network and a more balanced use of available vehicles.

The problems discussed in this chapter, as in the previous chapter, consider demands spread over a multi-period planning horizon. This is done with the realization that hub network design is a strategic decision whose impact is felt over a long time interval. We intend to design a hub network, deciding the network structure, type, and frequency of vehicles to operate between nodes.

We propose a mixed-integer programming model to represent the problem. We will determine the types and frequencies of vehicles between different nodes in the hub network, and where and when to hold inventory. This will be done while considering the cost of

establishing hub arcs, so that the total costs of network establishment plus operation in the network are minimized. To solve larger instances of the problem, we develop a Benders-decomposition-based exact algorithm. To enhance convergence, we incorporate Pareto-optimal cuts and two variable-fixing techniques. We also perform extensive computational experiments to evaluate the proposed algorithms and solve large-scale instances.

The remainder of this chapter is organized as follows: Section 4.1 presents a mathematical formulation of the problem, Section 4.2 describes the proposed methodology, Section 4.3 presents computational results using the proposed model and methodology. Finally, Section 4.4 summarizes the chapter.

4.1 Mathematical Formulation

Let the node set N contain the origins and destinations of demands of the freight delivery network we intend to design. Some nodes will serve as transshipment facilities for freight to be moved between origins and destinations. All origin and destination nodes may send or receive flow through more than one location per node; this assumption is referred to as *multiple allocation*. Let $G = (N, A)$ be a complete directed graph, where A is the set of arcs. The elements of set A represent links between nodes in the transshipment network. For each directed arc $a \in A$, $o(a)$ and $d(a)$ represent the first and second nodes on the arc, i.e. $o(a), d(a) \in N$.

Decisions have to be made on which locations to be used for holding inventory, and how many of which vehicle types to be dispatched between these locations. The planning

horizon T is finite and spread into discrete time windows. A time window is denoted by $t = 1, 2, \dots, |T|$; the intervals $[0, 1)$, $[1, 2)$, and $[i - 1, i)$ represent the first, second, and i^{th} time window. All decisions are made at the beginning of a time period. We assume that the above decisions, on inventory and transport type plus frequency, will not change over time, since the planning horizon considered here is relatively short, e.g., a week.

Any demand node may send commodities to any other node through the network. In this study, we assume a demand that repeats over the planning horizon, for example, weekly demand, monthly demand, etc. Such demand is distributed over different periods during that horizon. Let K be the set of demands. For each $k \in K$, the origin and destination nodes $o(k), d(k) \in N$. For each demand $k \in K$, a quantity of freight w_k is originated at time $t^0(k) \in T$ from node $o(k) \in N$, and destined to be routed to node $d(k) \in N$. At the start of period $t^0(k)$, this load w_k may be shipped to the first node along the destination path. Otherwise, that load may be held in inventory at $o(k)$, to be shipped later, if $o(k)$ is the first endpoint of a hub arc. We assume that all demands are divisible for shipment consolidation at the first hub along the destination path, and have sufficient time windows within the planning horizon, after their origination, to arrive at their destination nodes.

Using the notation of [Contreras et al. \(2012\)](#), a *hub arc* $a \in A$ is defined as an ordered set $\{o(a), d(a)\}$ if $|a| = 2$; it is a *hub-loop* if $o(a) = d(a)$, i.e. $|a| = 1$. We assume no direct shipments are allowed between O-D pairs unless they are connected by a hub arc. Exactly one hub arc must be present on a route.

In this research, different modes of transportation are represented by a given set of vehicles $v \in V$ with capacities ϕ_v to carry freight on a subset of arcs joining nodes of the

network. Assume $\hat{F}$ to be the set of frequencies. “Frequency” is measured in the number of vehicles per time period. A vehicle $v \in V$ operates on a link with frequency $f' \in \hat{F}$. For example, three large trucks per day may operate on a link. Let the set F be composed of “vehicle frequencies”. Let δ_f^t be a binary parameter that associates each $f \in F$ with a set of periods $t \in T$. It is 1 if “vehicle frequency” f is operated on an arc during period $t \in T$, and 0 otherwise. Each element $f \in F$ represents a vehicle $v \in V$ operating on a hub arc with frequency $f' \in \hat{F}$ in period $t \in T$. Those elements provide a total transportation capacity $\Gamma_f = \delta_f^t \phi_v f'$ on the arc for each of the periods $t \in T$. For example, a member of the set F can be dispatching 5 small trucks per day on a link on Monday, Wednesday and Friday; or dispatching 2 large trucks on a link every Tuesday and Thursday, etc.

Upon arrival at a node, freight gets unloaded, sorted, and consolidated with other loads. Shipments may be held at nodes for the convenience of consolidation with other freight. Shipment consolidation decisions, i.e. when and how long to hold inventory at a node, and when to dispatch a vehicle from that node, are to be determined for the intended network. The quantity of freight to be held at any node in each period, and the type plus frequency of vehicles to be dispatched on different hub arcs, are to be decided.

The model aims to minimize the total cost of transportation plus inventory, by optimally deciding vehicle types and their frequencies on different arcs, exploiting economies of scale by sending large and optimally-filled vehicles on the hub arcs that will carry consolidated flow from multiple origins along the path to the respective destinations. Inventory will be held at nodes, if necessary, incurring the inventory carrying costs when those goods are held at a node for one or more periods. The percentage of inventory value on hand per unit of time may be used as an estimate of the holding costs (Silver et al., 2017). Additional costs,

involving further processing of the freight held, if done at nodes, may also be embedded in carrying costs. We assume that all nodes have sufficient storage capacity to hold inventory. We will decide the optimal frequencies of vehicles dispatched between any pair of nodes. We assume that the company employs its private fleet to serve the hub arcs.

For each available arc in the network, we decide whether it will be used for transportation. If it is to be used, we decide if it will be a hub arc or an allocation arc. The latter is a link that joins a demand node (origin/destination) with the hub arcs. We also determine the vehicle types that carry freight and their frequencies on hub arcs. For activating each of these vehicle frequencies, there is a cost involved, which contributes to the total cost of establishment and transportation in the network. The cost for a vehicle frequency on a hub arc involves both fixed and variable components. The fixed cost portion accounts for some of the price of the vehicle, labor, and at both ends of the link necessary infrastructure and handling costs. The variable cost (consisting of maintenance and fuel, and other operational costs) may be estimated from the inter-node distances.

Denote by κ_v the fixed cost to dispatch a vehicle $v \in V$ (including a portion of the price of the vehicle, related labor and material handling at both ends of the trip). Let χ_{va} be the variable transportation cost (including vehicle operational cost) on hub arc $a \in A$, and let ζ_f ($f \in F$) be the cost of infrastructure necessary to be built for $f' \in \hat{F}$ times loading/unloading per period on each end of a link operating vehicle $v \in V$. The cost $\mathcal{C}_{af}$ of enabling and operating arc a with vehicle frequency $f \in F$ is defined by $\mathcal{C}_{af} = (\kappa_v + \chi_{va}) f' \sum_{t \in T} \delta_f^t + 2 \zeta_f$. Flow on an allocation arc incurs a cost proportional to the volume of flow on that link. Let $c_{ak} = c_{o(k)o(a)} + c_{d(a)d(k)}$ be this cost for demand $k \in K$ for hub arc $a \in A$. Here, $c_{o(k)o(a)}$ is the per unit cost incurred by sending flow from $o(k)$ to $o(a)$;

$c_{d(a)d(k)}$ is due to sending per unit flow from $d(a)$ to $d(k)$.

The sets and parameters required for the mathematical formulation are defined as follows:

Sets:

N : Set of O-D nodes.

A : Set of arcs between demand nodes.

T : Set of time periods.

K : Set of demands.

F : Set of vehicle frequencies.

Parameters:

w_k : Flow originated at start of time $t^0(k) \in T$ from node $o(k) \in N$ destined to node $d(k) \in N$.

c_{ak} : Cost of sending one unit of flow from node $o(k)$ to node $d(k)$ on an allocation connection, while using hub arc $a \in A$.

τ_{ak} : Transit time from node $o(k) \in N$ to origin of hub arc $a \in A$.

η : Holding cost per unit of freight per unit time.

Γ_f : Capacity obtained by operating vehicle frequency $f \in F$.

$$\delta_f^t = \begin{cases} 1 & , \text{ if vehicle frequency } f \in F \text{ operates at time } t \in T \\ 0 & , \text{ otherwise} \end{cases}$$

$\mathcal{C}_{fa}$: Cost of enabling and operating arc $a \in A$ with vehicle frequency $f \in F$.

Variables:

x_{ak}^t : Fraction of flow on arc $a \in A$, starting at the beginning of time $t \in T$ with respect to the demand $k \in K$.

$$z_{af} = \begin{cases} 1 & , \text{ if arc } a \text{ is enabled with vehicle frequency } f \in F \\ 0 & , \text{ otherwise} \end{cases}$$

The multiple allocation multi-modal hub network design model for demands over a time horizon is formulated as follows:

$$\begin{aligned} \text{minimize} \quad & \sum_{a \in A} \sum_{f \in F} \mathcal{C}_{af} z_{af} + \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} c_{ak} w_k x_{ak}^t \\ & + \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} \eta(t - t^0(k) - \tau_{ak}) w_k x_{ak}^t \end{aligned} \quad (4.1)$$

subject to

$$\sum_{a \in A} \sum_{t \in T: t \geq t^0(k)} x_{ak}^t = 1 \quad \forall k \in K \quad (4.2)$$

$$\begin{aligned} \sum_{k \in K: t^0(k) \geq t - \tau_{ak}} w_k x_{ak}^t &\leq \sum_{f \in F} \delta_f^t \Gamma_f z_{af} \\ &\forall a \in A, o(a) \neq d(a), \forall t \in T \end{aligned} \quad (4.3)$$

$$x_{ak}^t \geq 0 \quad \forall k \in K, \forall a \in A \forall t \in T \quad (4.4)$$

$$z_{af} \in \{0, 1\} \quad f \in F, a \in A \quad (4.5)$$

The objective function (4.1) represents the total cost, which is calculated by adding the cost of operating on hub arcs, costs on allocation arcs, and inventory cost - the three respective summation groups of the objective function. Each demand can be partially satisfied through different hub arcs. However, when satisfied, its flow must be routed via a hub arc by constraint (4.2). Multiple vehicle frequency may be activated on an arc. Constraint (4.3) ensures that flow on an arc in each period is bounded by the total capacity of vehicles operating on that arc. Finally, the domains of the decision variables are defined by constraints (4.4)-(4.5).

The above formulation does not model transit time. The variable x_{ak}^t is defined for all periods $t \geq t^0(k)$ in the time-horizon. There may be data scenarios where the locations are distributed over an expansive geographic area. Assume the case when the transit time from $o(k)$ to $o(a)$ is very large. Suppose that transit time (between $o(k)$ and $o(a)$) plus the time necessary for loading and unloading at $o(a)$, is close to t_T periods. Then x_{ak}^t should be set to 0 for $t = \{t^0(k) \dots t^0(k) + t_T - 1\}$, and the vehicles on hub arc a can carry this demand starting from $t = t^0(k) + t_T$. Such a situation may be handled by pre-computing transit times between all node pairs and fixing the appropriate x_{ak}^t variables to 0, thus strengthening the model.

Assuming $|N| = n$, $|V| = v$, and $|F| = f$, the number of binary and continuous variables required for the model follows $\mathcal{O}(n^2vf)$ and $\mathcal{O}(n^4t^2)$, respectively. There are $\mathcal{O}(n^2t^2)$ constraints. This large number of binary variables and constraints might limit attaining optimality in case of large instances of the problem. Nevertheless, we develop an exact solution algorithm in the following section.

4.2 Solution Methodology

We propose a Benders-decomposition-based exact solution algorithm for addressing our model. Benders decomposition, as discussed in Section 2.7, is particularly suitable for hub and hub arc location problems, especially those with multiple allocation structures. That is because this methodology facilitates the decomposition of the problem into linear sub-problems by fixing the integer variables governing locations. The following section begins by introducing the Benders reformulation of the model and the Benders decomposition algorithm, followed by some strategies we adapted to make the algorithm converge faster.

4.2.1 Benders Reformulation and Algorithm

For our mixed integer model described in formulation (4.1)-(4.5), we manage the hub arc location decisions in the master problem, while the remaining aspects are delegated to the subproblem. We fix the values of binary variables z_{af} to $\bar{z}_{af}$ in Eqs. (4.1)-(4.5), and thereby obtain the following linear primal sub-problem (PS):

$$(PS) \text{ minimize } \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} c_{ak} w_k x_{ak}^t + \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} \eta(t - t^0(k) - \tau_{ak}) w_k x_{ak}^t \quad (4.6)$$

subject to

$$\sum_{a \in A} \sum_{t \in T: t \geq t^0(k)} x_{ak}^t = 1 \quad \forall k \in K \quad (4.7)$$

$$\sum_{k \in K: t^0(k) \geq t - \tau_{ak}} w_k x_{ak}^t \leq \sum_{f \in F} \delta_f^t \Gamma_f \bar{z}_{af} \quad \forall a \in A, o(a) \neq d(a), \forall t \in T \quad (4.8)$$

$$x_{ak}^t \geq 0 \quad \forall k \in K, \forall a \in A \forall t \in T \quad (4.9)$$

We associate dual variables α_k and u_a^t respectively to constraints (4.7) and (4.8). This leads to DS , the dual subproblem of PS :

$$(DS) \text{ maximize } \sum_{k \in K} \alpha_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f \bar{z}_{af} u_a^t \quad (4.10)$$

subject to

$$\alpha_k + w_k u_a^t \leq w_k (c_{ak} + \eta(t - t^0(k) - \tau_{ak}))$$

$$\forall a \in A, k \in K, t \in T, t^0(k) \geq t - \tau_{ak} \quad (4.11)$$

$$\alpha_k \quad \text{urs} \quad \forall k \in K \quad (4.12)$$

$$u_a^t \leq 0 \quad \forall a \in A, t \in T \quad (4.13)$$

Assume P denote the polyhedron defined by the formulation (4.10)-(4.13). Let $\bar{P}$ represent the set of extreme points of P . P is always feasible and bounded. The feasible region of DS does not depend on z . Hence, either PS and DS are both feasible, or PS is infeasible and DS is unbounded. Consequently, the optimal solutions are among the extreme points of P and extreme rays of R . We note that $\bar{P}$ does not depend on $\bar{z}_{af}$. Thus, for any z_{af} , the dual subproblem may be defined as:

$$\text{maximize}_{(\alpha, u) \in \bar{P}} \sum_{k \in K} \alpha_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f z_{af} u_a^t$$

subject to

$$\sum_{k \in K} \alpha_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f z_{af} u_a^t \leq 0, \quad \alpha, u \in R$$

Let θ be the cost obtained by satisfying the demands. The Benders master problem *BMP* is thus formulated as:

$$(BMP) \text{ minimize} \quad \theta + \sum_{a \in A} \sum_{f \in F} \mathcal{C}_{af} z_{af} \quad (4.14)$$

subject to

$$\theta \geq \sum_{k \in K} \bar{\alpha}_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f z_{af} \bar{u}_a^t \quad \alpha, u \in \bar{P} \quad (4.15)$$

$$\sum_{k \in K} \alpha_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f z_{af} u_a^t \leq 0, \quad \alpha, u \in R \quad (4.16)$$

$$z_{af} \in \{0, 1\} \quad f \in F, a \in A \quad (4.17)$$

The aforementioned BMP consists of $|\bar{P}| + |R|$ constraints, corresponding to the number of extreme points and number of extreme rays in the *DS*. To manage this, we utilize a cutting-plane method. Beginning with an empty set $\bar{P}$ and empty set R , we iteratively solve the BMP by incorporating new extreme points and extreme rays obtained from solving the subproblems, until achieving an optimal solution to the *BMP*.

The Benders decomposition framework is delineated in Algorithm 3. Here, LB and UB denote the current lower and upper bounds at each iteration, respectively. In each iteration, we initially solve the *BMP* to acquire $\bar{z}$ and the corresponding objective value, thereby updating the lower bound. Subsequently, utilizing the obtained $\bar{z}$, we solve the *DS*

to obtain $\bar{\alpha}$, $\bar{u}$, and the DS -objective value. We then generate a new extreme point $(\bar{\alpha}, \bar{u})$ and append it to $\bar{P}$. Next, we update the upper bound using the current DS -objective value, as outlined in the algorithm. The iterations terminate when the difference between LB and UB is sufficiently small.

Algorithm 3 Benders Decomposition Algorithm

```

1: Begin
2:  $LB \leftarrow -\infty$ 
3:  $UB \leftarrow +\infty$ 
4:  $\bar{P} \leftarrow \phi$ 
5: repeat
6:   SOLVE MP to obtain  $\bar{z}$ 
7:   Calculate  $LB \leftarrow$  BMP-objective (4.14)
8:   SOLVE DS to obtain  $\bar{x}, \bar{\alpha}, \bar{u}$ 
9:   Calculate  $UB \leftarrow \min(UB, \sum_{a \in A} \sum_{f \in F} \mathcal{C}_{af} \bar{z}_{af} + \text{value of (4.10)})$ 
10:  Add  $(\bar{\alpha}, \bar{u})$  to  $\bar{P}$ 
11: until  $UB - LB < \epsilon$ 
12: End

```

The computational efficiency of the Benders decomposition algorithm is typically assessed based on the convergence rate of LB and UB , i.e. the number of iterations needed for convergence. To enhance efficiency, we have integrated the addition of stronger non-dominated cuts to the BMP, and implemented two variable fixing techniques to reduce the size of the problem. These approaches are detailed in the following subsections.

4.2.2 Improvement Strategies

Here we delineate two strategies aimed at enhancing the computational efficiency of the Benders decomposition algorithm. The first involves incorporating Pareto-optimal cuts,

while the second strategy entails variable fixing techniques. These approaches are elaborated in the next subsections.

Pareto Optimal Cuts

As outlined in Section 2.7, Magnanti and Wong (1981) introduced the concept of constructing stronger, non-dominated cuts, referred to as Pareto-optimal cuts, to enhance the convergence of the Benders decomposition algorithm. This technique has been widely employed in the literature to generate strong optimality cuts. Let $(\bar{\alpha}^1, \bar{u}^1)$ and $(\bar{\alpha}^2, \bar{u}^2)$ denote two solutions of the DS . The cut derived from the former dominates the cut derived from the latter when the following relation holds for all values of z , with strict inequality for at least one point:

$$\sum_{k \in K} \bar{\alpha}_k^1 + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f z_{af} \bar{u}_a^{1t} \geq \sum_{k \in K} \bar{\alpha}_k^2 + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f z_{af} \bar{u}_a^{2t}$$

To identify a cut that is not dominated by any other cuts, we define a polyhedron (Ω) represented by $0 \leq z_{af} \leq 1$ for all $f \in F$ and $a \in A$, with $ri(\Omega)$ denoting the relative interior of Ω . At each iteration of the algorithm, we solve the following Pareto-optimal subproblem (DS_{PO}) to generate a Pareto-optimal cut and incorporate it into the BMP .

$$(DS_{PO}) \text{ maximize } \sum_{k \in K} \alpha_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f \hat{z}_{af} u_a^t \quad (4.18)$$

subject to

(4.11) – (4.13)

$$\sum_{k \in K} \alpha_k + \sum_{a \in A} \sum_{t \in T} \sum_{f \in F} \delta_f^t \Gamma_f \bar{z}_{af} u_a^t = \bar{D}S \quad (4.19)$$

Here, $\hat{z} \in ri(\Omega)$ and $\bar{D}S$ represents the optimal solution value of subproblem DS at that iteration. Constraint (4.19) ensures that the optimal solution of DS_{PO} is chosen from the solutions of DS at that iteration. The inclusion of Pareto-optimal cuts in the BMP at each iteration decreases the number of iterations needed to converge. However, this improvement comes with the cost of executing the aforementioned linear program at each iteration, as demonstrated by our computational experiments.

Variable Fixing

This section outlines two methods aimed at enhancing the convergence of the Benders decomposition algorithm. We illustrate techniques for reducing the dimension of the master problem through variable fixing methods. These techniques contribute to reducing the total computational time of solving the master problem, in our case through solution space reduction, by eliminating variables that are known not to appear in an optimal solution. These were first employed by Contreras et al. (2011a), later used by many others including Taherkhani et al. (2020), where hubs were eliminated by using information obtained during iterations of the Benders decomposition algorithm. In our case, we utilize these tests to eliminate hub arcs from the model.

Our first improvement strategy involves the linear programming relaxation of the Ben-

ders master problem (BMP). Let $\hat{z}_{LR}$ be the value of the linear relaxation of the BMP at a given iteration of the Benders decomposition algorithm, and rc_{af} represent the reduced cost corresponding to the variable z_{af} . Let UB denote the currently known, best upper bound of the optimal value of the BMP. The expression $\hat{z}_{LR} + rc_{af}$ provides an upper bound on the BMP solution. We can eliminate hub arcs $a \in A$ (and their associated variables) for which $\hat{z}_{LR} + rc_{af} > UB$, as those arcs cannot be part of an optimal solution. By fixing the z -variables in this manner, we effectively reduce the solution space of the problem.

The second strategy to fix variables involves reduction to the *BMP* by a subset $A' \subset A$ of hub arcs which are closed in the optimal solution. This is tested by modifying the master problem by adding a constraint and checking the obtained solution against the known upper bound. Let $BMP(A')$ be the model at an iteration obtained by adding constraint $\sum_{a \in A' f \in F} z_{af} \geq 1$. The solution to this model, $\hat{z}_{BMP(A')}$, gives an upper bound on $BMP(A')$. If $\hat{z}_{BMP(A')} > UB$, that constraint is violated by the preceding solution. Hence none of the hub arcs in A' can be in the optimal solution. We fix variables z_{af} to 0, for $a \in A'$ and $f \in F$.

The effectiveness of the second variable-fixing technique is significantly influenced by the choice of the subset $A' \subset A$. The hub arcs in A' should have the least likelihood of being included in the optimal solution. If such arcs can be identified in the early iterations of the Benders decomposition algorithm, that algorithm is more likely to converge faster. We adopt the following policy for the construction, updating, and re-initiation of the set A' , inspired by the works of [Contreras et al. \(2011a\)](#) and [Taherkhani et al. \(2020\)](#).

To construct A' , we initially assume another set $A'' = A$. This set is modified based

on information obtained from each iteration, and then used to construct A' . At every iteration, we eliminate from A'' the hub arcs that are open in the optimal solution (as of that iteration), and those eliminated by the first variable fixing approach. We then sort the hub arcs in A'' in descending order of the total flow originating at the first node of each hub arc. We set A' to a certain random proportion $p_{A'}$ of A'' . The quantity $p_{A'}$ is set to at most 75%, so that at least 2 (or more) hub arcs are in A' in each iteration. With this $p_{A'}$, we compute $\hat{z}_{BMP(A')}$, check if $\hat{z}_{BMP(A')} > UB$, and fix variables in the BMP accordingly. If A' becomes empty in any iteration, we reset $A'' = A$ for the subsequent iterations.

The z variables correspond to each potential hub arc and each frequency that may be operated on those arcs. The variable fixing technique described in the previous paragraph eliminates all variables v_{af} if an arc $a \in A'$ is identified for elimination in a particular iteration. This approach does not consider that a variable on an arc for a specific frequency may be part of the optimal solution, while other variables on the same arc (for other frequencies) may be eliminated. Based on this observation, we enhance the elimination test, as follows.

Initially, we construct another set A_1'' containing tuples $(a, f), \forall a \in A, f \in F$. At each iteration, we eliminate from this set the tuples corresponding to variables z_{af} that are open in the optimal solution, as well as those eliminated by the first variable-fixing approach. We then sort the hub arcs in A_1'' in non-descending order of the total flow originating at the first node of each hub arc. Subsequently, we construct $A_1' \subset A_1''$ in a similar manner to the construction of $A' \subset A''$. We randomly select an $f \in F$ to perform the second elimination test by adding to $BMP(A_1')$ the constraint $\sum_{a \in A_1'} z_{af} \geq 1$ for that f . Based on this test, we fix variables in the BMP for specific arcs and frequencies. If A_1' becomes empty, it is

reset using the same policy applied in the case of A' .

In summary, to enhance the computational performance of the Benders decomposition algorithm, we propose the utilization of elimination tests to fix decision variables on candidate hub arcs, and the implementation of Pareto-optimal cuts during the subproblem solutions at each iteration. Subsequently, we evaluate the performance and effectiveness of these proposed methodologies.

4.3 Computational Analysis

We present some numerical results in this section, using the model and algorithm presented in Sections 4.1 and 4.2. We employed an IBM computer with Intel Core-i7 processor and 32 GB RAM for running the experiments. The formulation and algorithm presented have been implemented in CPLEX 12.9 and Java (IntelliJ IDEA 21.2.1 using JDK 15.0.2). Section 4.3.1 exhibits the data and parameters used for the experiments, Section 4.3.2 highlights some optimal results and insight into the results. Section 4.3.3 presents the computational efficiency of the proposed model and algorithms.

4.3.1 Data and Parameters

To test our model and methodology, 10 and 15-city instances, with cities corresponding to those of the Civil Aeronautics Board (CAB) data set (O’Kelly, 2014) were used. All pairs of arcs between these cities were considered as hub arcs. The demand quantities between pairs of cities were generated from 2022 data of The Freight Analysis Framework

(FAF) (CFTA, 2023) in a similar way to that of the previous chapter. These demands were randomly distributed over the different time periods of our planning horizon, utilizing a uniform distribution.

As in Chapter 3, we used the same transportation cost on allocation links for the year 2022 (Leslie and Murray, 2022). We also considered cargo planes B-747-F and B-747-100 for transportation on the hub arcs, as employed by Masaeli et al. (2018). For fixed and variable costs and capacities of these two aircraft, the same figures as reported in the previous chapter were applied.

To test our model and algorithm against ranges of data parameters, we have varied the holding costs and hub arc costs. Our holding cost categories of high, medium-high, medium-low, and low, respectively correspond to \$286, \$124, \$19, and \$9 per ton per day, as in Chapter 3. Hub arc establishment costs were varied in three categories - low, medium and high, corresponding to \$200, \$400, \$1,000. These costs were considered on the basis of the annualized cost of a hub (again as reported in the previous chapter), and assuming that a hub may be shared by multiple hub arcs. We consider a planning horizon of one week, with each period corresponding to one day.

We have utilized 2, 3, and 5 frequency options for operating hub arcs.

- **2-Frequency Option:** This includes either operating two B-757-F flight on Mondays, one on Wednesday and Friday, or alternatively two B-757-100 flights on each of Tuesday and Thursday.
- **3-Frequency Option:** This involves choosing between operating one B-757-F flight

on Monday, Wednesday, and Saturday; one B-757-100 flight on Thursday and Saturday; or one B-757-100 flight on Monday and Friday.

- **5-Frequency Option:** This includes the following choices:
 - one B-757-F flight on Monday and Thursday
 - one B-757-100 flight on Tuesday and Friday
 - one B-757-100 flight on Wednesday and Saturday
 - one B-757-F flight on Monday, Wednesday, and Friday
 - one B-757-100 flight on Tuesday, Thursday, and Saturday

4.3.2 Impact of Frequency Selection on Solutions

In this section, we study the impact of using different choices of frequencies in the optimal design of hub arc networks. We have used 2, 3 and 5-frequency choices (as mentioned above) to generate optimal networks for 10 and 15-city instances. We allowed a run time of 4 and 8 hours (for 10 and 15-city instances, respectively) per instance to get optimal results for all cases reported. We report those results in Tables 4.1 and 4.2. Each row in these two tables reports costs of transportation on allocation links and hub arcs, and total inventory cost of a solution, with number of hub arcs established and numbers of the two types of aircraft used in a solution. This is done for each choice of frequency (F), hub arc establishment cost (ζ) and unit holding cost (η). Here in this section we only look into the impact of frequency on the optimal results. A comparative study of computational

performance between running the proposed model in CPLEX and our proposed Benders algorithm is given in next section.

Table 4.1: Optimal results for 10-city instances using 2, 3 and 5 frequencies

$ F $	ζ	η	Cost (1000\$)				No. of Hub Arcs		No. of B-747-F	No. of B-747-100
			Alloc.	Transp.	Inv.	Total				
2	Low	low	139	123	2	263	2	8	-	-
		med-low	139	123	4	265	2	8	-	-
		med-high	139	123	24	285	2	8	-	-
		high	139	123	55	317	2	8	-	-
	Med	low	139	123	2	263	2	8	-	-
		med-low	139	123	4	265	2	8	-	-
		med-high	139	123	24	286	2	8	-	-
		high	139	123	55	317	2	8	-	-
	High	low	139	124	2	265	2	8	-	-
		med-low	139	124	4	267	2	8	-	-
		med-high	139	124	24	287	2	8	-	-
		high	139	124	55	318	2	8	-	-
3	Low	low	71	102	3	176	4	-	-	8
		med-low	71	102	6	179	4	-	-	8
		med-high	56	124	17	197	5	-	-	10
		high	56	124	38	219	5	-	-	10
	Med	low	71	103	3	177	4	-	-	8
		med-low	71	103	6	180	4	-	-	8
		med-high	56	125	17	198	5	-	-	10
		high	56	125	38	220	5	-	-	10
	High	low	71	105	3	179	4	-	-	8
		med-low	71	105	6	182	4	-	-	8
		med-high	56	128	17	201	5	-	-	10
		high	56	128	38	223	5	-	-	10
5	Low	low	55	131	2	188	5	2	-	8
		med-low	55	131	4	191	5	2	-	8
		med-high	61	133	17	211	5	2	-	8
		high	55	146	29	230	6	2	-	10
	Med	low	55	132	2	189	5	2	-	8
		med-low	55	132	4	192	5	2	-	8
		med-high	61	134	17	212	5	2	-	8
		high	55	147	29	231	6	2	-	10
	High	low	55	135	2	192	5	2	-	8
		med-low	55	135	4	195	5	2	-	8
		med-high	61	137	17	215	5	2	-	8
		high	55	151	29	235	6	2	-	10

As we see in Table 4.1, the network structure does not vary while using $|F| = 2$. Two hub arcs (from Chicago to Atlanta, and Atlanta to Chicago) were selected in all solutions, whatever the variation of unit holding cost and hub arc costs. In total, eight B-747-F aircraft were employed to carry goods on the hub arcs; no B-747-100 aircraft were utilized in these solutions. Transportation cost on allocation links remained the same for all parameter variations, while the transportation cost on hub arcs increased just a little as the hub arc cost increased from low to high. The total inventory costs were increasing with η , for each of the three cases of hub arc costs. This was the major contributor to the difference in total costs obtained. Across variation of the mentioned parameters, the two-frequency choice resulted in the highest average total cost for network establishment and operations for the 10-city instances. Understandably, with a limited number of choices for frequencies, the cost is higher.

Employing $|F| = 3$ for 10-city instances resulted in networks that have 4 to 5 hub arcs, where 8 or 10 B-747-100 aircraft were utilized. No B-747-F aircraft were flown in those networks. The costs of transportation on allocation links and on hub arcs were strictly increasing in the unit holding cost, for all four values of the latter parameter; this happened for all cases of hub arc costs. Total inventory costs in the networks also grew significantly with an increase in that unit holding cost. However, variations in hub arc establishment cost did not impact the network design in this case. For low and med-low unit holding costs, four hub arcs were selected, while for med-high and high unit holding costs, five hub arcs were selected. In the former case, hub arcs were between Atlanta to Chicago, Chicago to Atlanta, Chicago to Dallas, and Dallas to Chicago. The latter case had the same, plus one more hub arc from Chicago to Knoxville. We also observe that additional flights were

required when the unit holding cost was med-high or high.

The case of $|F| = 5$, as reported at the bottom of Table 4.1, shows that networks were established with 5 to 6 hub arcs, operating two B-747-F and eight or ten B-747-100. These networks also demonstrated that variations in hub arc establishment cost did not significantly affect design of the network. Rather, the network structure was impacted by variations in unit holding cost. All components of the costs increased as the unit holding cost grew from low, med-low, med-high to high. The only exception is that the transportation cost on allocation links decreased as the unit holding cost increased from med-high to high. Six hub arcs were used in the cases of high unit holding cost; these cases utilized 2 additional B-747-100 aircraft in the fleet. All networks here had hub arcs between Atlanta and Chicago, Chicago and Dallas, and Chicago to Knoxville. The networks with high unit holding cost had an additional hub arc from Atlanta to Birmingham.

$|F| = 3$ provided us with the lowest mean-cost networks amongst the three frequency setups we used for 10-city networks. On average, the resultant networks with $|F| = 3$ observed respective improvements of 46% and 6% in total cost over networks with $|F| = 2$ and 5. Compared to $|F| = 2$, the lowest total cost networks obtained had a 119% improvement in transportation cost on allocation links, 7% improvement in the cost of transportation on hub arcs, and a 33% total-inventory-cost improvement in the average case, resulting in an overall 46% saving. The improvement is not as great if we compare the lowest total-cost networks with networks for $|F| = 5$. In this case, a 20% savings in transportation cost on hub arcs was obtained with 11% and 19% extra cost of transportation on allocation links and inventory handling – thus resulting an overall improvement of 6%. The networks with $|F| = 3$ employed only B-747-100 aircraft.

Table 4.2: Optimal results for 15-city instances using 2, 3 and 5 frequencies

$ F $	ζ	η	Cost (1000\$)				No. of Hub Arcs	No. of B-747-F	No. of B-747-100
			Alloc.	Transp.	Inv.	Total			
2	Low	low	338	256	2	596	4	16	-
		med-low	338	256	4	598	4	16	-
		med-high	338	256	27	620	4	16	-
		high	340	256	59	654	4	16	-
	Med	low	338	257	2	596	4	16	-
		med-low	338	257	4	599	4	16	-
		med-high	338	257	27	621	4	16	-
		high	340	257	59	655	4	16	-
	High	low	338	259	2	599	4	16	-
		med-low	338	259	4	601	4	16	-
		med-high	338	259	27	624	4	16	-
		high	340	259	59	658	4	16	-
3	Low	low	148	233	9	390	9	-	18
		med-low	148	233	18	399	9	-	18
		med-high	158	253	64	475	9	3	16
		high	180	254	111	545	9	3	16
	Med	low	148	235	9	391	9	-	18
		med-low	148	235	18	401	9	-	18
		med-high	158	255	64	477	9	3	16
		high	180	255	111	546	9	3	16
	High	low	148	240	9	397	9	-	18
		med-low	148	240	18	406	9	-	18
		med-high	158	261	64	482	9	3	16
		high	180	261	111	552	9	3	16
5	Low	low	167	214	6	388	8	2	14
		med-low	167	214	13	395	8	2	14
		med-high	172	230	55	458	8	3	14
		high	185	265	57	507	9	3	17
	Med	low	167	216	6	390	8	2	14
		med-low	167	216	13	397	8	2	14
		med-high	172	232	55	459	8	3	14
		high	185	266	57	509	9	3	17
	High	low	167	221	6	394	8	2	14
		med-low	167	221	13	401	8	2	14
		med-high	172	237	55	464	8	3	14
		high	185	272	57	514	9	3	17

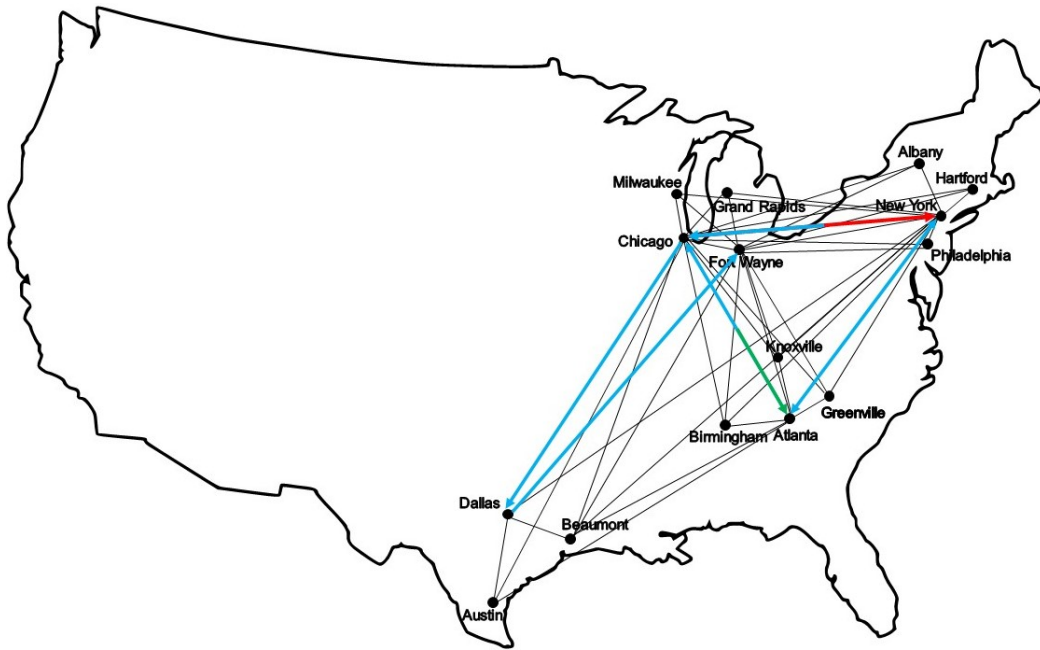
The 15-city instances naturally had more nodes and demand quantities than the 10-city instances. Consequently, the resultant networks on these larger instances were more interesting. Table 4.2 shows that the network structure for 15-city instances does not vary much against variations in parameter settings in case of $|F| = 2$. The solutions here always had four hub arcs operating 16 B-747-F aircraft. (The arcs used in these networks is discussed in Figure 4.2 later in this section.) The transportation cost on hub arcs increased just a little as the hub arc cost increased from low to high. This cost was not affected by variations in unit holding cost. However, transportation cost on allocation links remained unchanged as unit holding cost increased from low to med-high, but increased a little bit for high. As in the 10-city networks with $|F| = 2$, the major contributor to differences in overall total costs was the total inventory cost in a solution. This cost naturally increased in the unit holding cost parameter. Similar to the 10-city networks, employing $|F| = 2$ in 15-city instances yielded the highest total-average-cost networks.

The case $|F| = 3$ in 15-city instances resulted in networks that utilize nine hub arcs and 18 or 19 aircraft. With low and med-low unit holding cost, 18 B-747-100 aircraft were used. When that parameter increased to med-high and high, a fleet of three B-747-F and 16 B-747-100 aircraft was employed in the optimal solutions. The same happened for all variations in hub arc establishment cost. Transportation cost on allocation links and total inventory cost of these networks were not influenced by changes in ζ . However, the cost of transportation on allocation links grew as η increased from med-high to high. The total inventory cost of course increased in η .

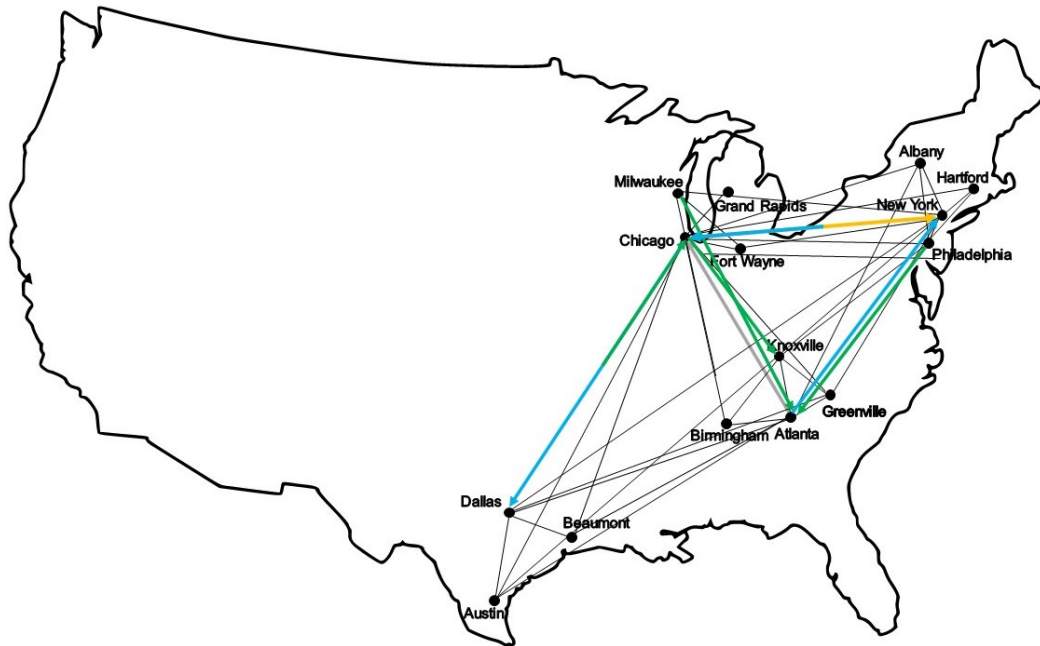
The case $|F| = 5$, for 15-city instances, provided us with the average-lowest-cost networks amongst the three frequency setups tested. Against variation of parameters, eight

or nine hub arcs were established and 16-20 aircraft flown. Transportation cost on allocation links and total inventory cost were unaffected by variations in hub establishment cost. However, transportation cost on allocation links was generally increasing in the unit holding cost η . Similarly, total inventory cost increased with η for all variations in hub arc establishment cost. Transportation costs on hub arcs were affected by changes in both parameters ζ and η . With low and med-low unit holding cost, a fleet composed of two B-747-F and 14 B-747-100 was used. One B-747-F was added as unit holding cost grew to med-high, and 3 additional B-747-100 were added to the fleet as that parameter increased further to high. This was done to avoid holding inventory when unit holding cost was med-high and high, by accepting an increase in transportation cost on hub arcs. We present instances of these networks in Figures 4.1 and 4.2 below.

$|F| = 5$ provided us with the lowest-mean-cost networks amongst the three frequency setups tested for 15-city networks. On average, the resultant networks with $|F| = 5$ observed respectively 41% and 4% improvement in total cost over networks with $|F| = 2$ and $|F| = 3$. Compared to $|F| = 2$, the lowest-total-cost networks obtained had a 96% improvement in the cost of transportation on allocation links, and 10% improvement in that cost on hub arcs in the average case. The mean total inventory cost here was 31% above that of $|F| = 2$. Thus, by holding more inventory, transportation cost savings was possible in the case of $|F| = 5$, resulting in an overall 41% cost improvement relative to $|F| = 2$. Compared to the average case of networks with $|F| = 3$, the total cost improvement is only 4%. In this case, an 8% increase in transportation cost on allocation links allowed respectively 5% and 52% savings in the cost of transportation on hub arcs and inventory handling – thus resulting in an overall improvement of 4%.



(a) 15-city instance with *low* arc cost of and *low* holding cost.



(b) 15-city instance with *low* arc cost of and *high* holding cost.

Figure 4.1: Resulting hub networks for *low* vs *high* holding cost using 5-frequencies (colors are explained in the text)

Two resultant optimal networks with $|F| = 5$ and 15-city instances are compared in Figure 4.1. Figure 4.1a represents the network obtained for low hub arc establishment cost and low unit holding cost. In total, eight hub arcs were established in this case: six between Chicago, New York and Atlanta, and additional two from Chicago to Dallas, and Dallas to Fort Wayne. Except for the arcs between Chicago to New York and Atlanta, all other hub arcs in this figure are marked in light blue. Those represent B-747-100 aircraft operating on these arcs – one on Wednesday, the other on Saturday. The hub arc from Chicago to New York, marked in red, operates one B-747-F flight on each Monday and Thursday. The hub arc from Chicago to Atlanta, marked in green, operates B-747-100 flights: one on Tuesday, the other on Friday.

Considering the origin node of a hub arc to which demand is allocated, Chicago played the most central role, followed by Atlanta and New York. The same cities hold the majority of inventory. Demands from Chicago, Fort Wayne and Milwaukee were routed through hub arcs originating in Chicago. Cities such as Birmingham, Atlanta, Greenville, Knoxville, Beaumont were allocated to hub arcs originating in Atlanta and Chicago. Some of Beaumont’s demands were allocated to hub arcs originating in Dallas. All demands from Dallas and Austin were routed along the same hub arc. Demands from New York were allocated to hub arcs originating in New York. Demands from Albany, Hartford, Grand Rapids and Philadelphia were also routed through hub arcs originating in New York and Chicago.

In Figure 4.1b, the hub arc establishment cost remained low, while the unit holding cost was high. In total, nine hub arcs were selected in this solution: two between Chicago and New York, two between Chicago and Dallas, and one each from Chicago to Knoxville, Atlanta to New York and Chicago, Philadelphia to Atlanta, and Milwaukee to Atlanta.

As noted previously, light blue arcs in this network represent B-747-100 flights – one on Wednesday, the other on Saturday; green arcs denote B-747-100 flights - one on Tuesday, the other on Friday. The arc from Chicago to New York, marked in Yellow, represents one B-737-F flight operating on each Monday, Wednesday and Friday. The grey arc, from Atlanta to Chicago, indicates single B-747-F flights operating each Tuesday, Thursday and Saturday. Since the unit holding cost is higher in this scenario, this network operates one additional B-747-F and three additional B-747-100 flights to avoid holding inventory. Total cost is thereby reduced, by increasing transportation cost on hub arcs to lower the inventory cost.

In this network, too, Chicago played the most central part, as the origin of a hub arc being responsible for the greatest inventory holding. New York and Atlanta followed Chicago in the same function in Figure 4.1a, but in Figure 4.1b, that role is distributed to Philadelphia and Milwaukee. Dallas was again an important hub arc origin in the west, as in the previous case. For the majority of freight, Atlanta, Birmingham, Greenville, and Knoxville were allocated to hub arcs originating in Chicago and Atlanta. Grand Rapids, Fort Wayne and Chicago were allocated to Chicago, while Albany, Hartford, Philadelphia were allocated to New York, Philadelphia and Chicago. New York was allocated to itself and Philadelphia. The hub arcs originating in Milwaukee got freight allocations from itself, Chicago, Birmingham and Fort Wayne. Beaumont and Austin were allocated to Atlanta, Chicago and Dallas, while some flow from Dallas and Milwaukee got routed through Chicago, in addition to being routed through themselves. In comparison to the low-holding-cost scenario described above, this network with the above allocations to the nine hub arcs, where four additional flights operated, could reduce the inventory quantities

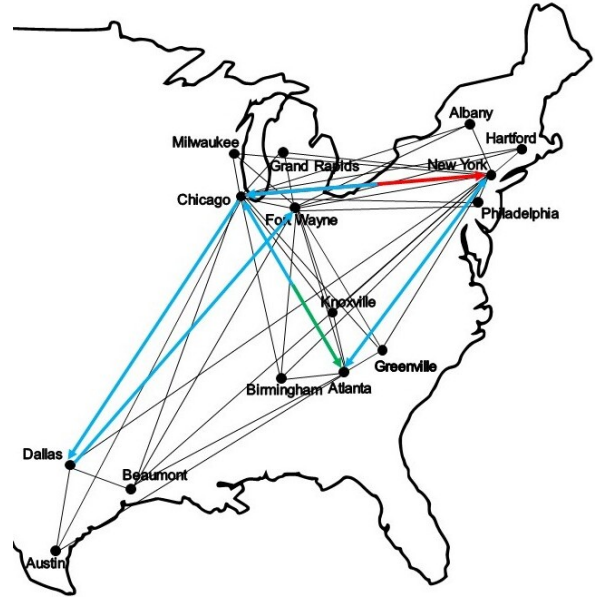
by 67%.

Figure 4.2 depicts contrast between $|F| = 2$ and 5, under low and high unit holding cost. Figures 4.2a and 4.2c present optimal networks for $|F| = 2$, with the former depicting the network for low unit holding cost η and the latter the case of high η . We observe that a total of four hub-ars were selected in both networks between Chicago and Atlanta, and New York and Chicago. Each network employs the same set of allocation arcs, with only one exception: the allocation arc between New York and Atlanta in Figure 4.2c. All these hub arcs, marked in purple, represent four flights of B-747-F operating on an arc: two on Monday, one on Wednesday and one on Friday. Thus, in total, 16 flights operated on the network in both cases. Since each network uses a very similar allocation pattern, the difference between the total cost of transportation on allocation and hub arcs is negligible. The total cost of establishment and operation does differ, mainly due to the impact of total inventory cost. However, in terms of the inventory quantity held, the network in Figure 4.2c only saved 7% compared to that of Figure 4.2a.

The contrast between the networks in Figure 4.2a and 4.2b is more evident. The former has four hub arcs, while the latter has eight. Both networks use 16 aircraft on hub arcs. However, in the first one, all aircraft are B-747-F, while in the latter, only two are B-747-F, and the remaining 14 are B-747-100. Out of two frequency choices, the earlier network used only one; while the second employed three of its five frequency choices. Thus, by flying a heterogeneous fleet of aircraft and distributing flights over a wider spectrum, utilizing $|F| = 5$ made a 35% savings on total cost, in spending about three times on inventory, but saving 51% and 16% on the costs of transportation on allocation and hub arcs, respectively.



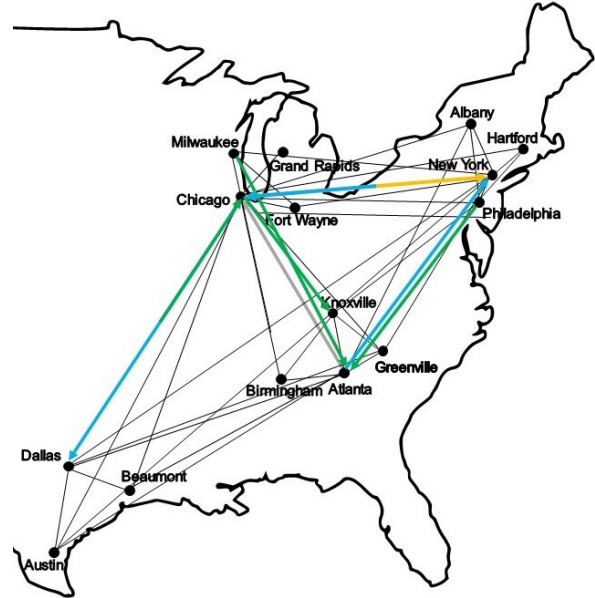
(a) Low holding cost with 2 frequencies



(b) Low holding cost with 5 frequencies



(c) High holding cost with 2 frequencies



(d) High holding cost with 5 frequencies

Figure 4.2: Impact of using different frequencies (colors are explained in the text)

A similar observation can be made of the contrast between the networks in Figures 4.2c and 4.2d. Both networks assume high holding cost. The former one used four hub arcs, while the latter used nine. The former network employed 16 B-747-F, while the latter had three B-747-F and 17 B=747-100. The former network utilized just a single frequency choice, while the second employed four different frequencies on its arcs. The network obtained with $|F| = 5$ yielded 22% savings on total cost of network establishment and operation over the one with $|F| = 2$. This was made possible by respectively saving 46% and 3% in transportation cost on allocation links and on inventory cost. There was also a compromise of spending 4% extra on transportation cost on hub arcs.

4.3.3 Algorithmic Performance

In this section, we evaluate the run time and computational performance of the proposed model and algorithm. The model was run in CPLEX under the default settings. The tables in this section thus report performance of the models under the respective heading “CPLEX” and the performance of our proposed Benders algorithm under column head, “Proposed Methodology”. We report run times and optimality gaps for CPLEX and for the proposed methodology. We also report numbers of iterations, and numbers of different types of cuts added.

Table 4.3 presents detailed comparison between the model’s performance in CPLEX against the performance of the proposed Benders methodology, for 10-city instances. The first three columns describe parameter settings, i.e. number of frequencies to choose from, hub arc establishment cost, and unit holding cost. The next two columns describe the

Table 4.3: Algorithmic Performance for 10 city instances

$ F $	ζ	η	CPLEX			Proposed Methodology						
			Time (s)	Gap (%)	Time (s)	#Iter.	#BD-C	#PO-C	#VF-1	#VF-2	Gap (%)	
2	Low	low	8	0	1009	3146	3119	0	114	90	0	
		med-low	17	0	3600	3689	3642	1	17	90	0	
		med-high	14	0	336	1394	1372	2	153	90	0	
		high	9	0	142	548	536	3	268	66	0	
		low	16	0	3600	3848	3826	0	96	90	0	
		med-low	16	0	3600	4298	4243	1	31	90	0	
	Med	med-high	14	0	811	2599	2574	2	104	90	0	
		high	8	0	139	545	534	2	267	70	0	
		low	14	0	430	1680	1658	1	111	90	0	
	High	med-low	16	0	3600	4261	4204	1	18	90	0	
		med-high	15	0	165	683	665	1	148	90	0	
		high	8	0	145	561	543	4	271	68	0	
3	Low	low	627	0	3601	3693	3586	0	142	92	0	
		med-low	1220	0	163	563	535	1	267	68	0	
		med-high	133	0	165	565	529	5	264	72	0	
		high	108	0	178	564	533	1	266	74	0	
		low	940	0	3600	3360	3280	0	136	93	0	
		med-low	743	0	160	555	523	1	261	66	0	
	Med	med-high	227	0	173	560	533	2	266	67	0	
		high	84	0	164	583	544	2	272	66	0	
		low	514	0	3600	3544	3490	0	110	91	0	
	High	med-low	614	0	171	585	552	3	272	56	0	
		med-high	144	0	157	559	531	2	265	72	0	
		high	124	0	171	578	546	2	273	70	0	
5	Low	low	3606	0	1597	1657	1492	1	746	78	0	
		med-low	3607	0	1672	1657	1484	1	742	82	0	
		med-high	3601	1.7	1639	1646	1499	1	749	83	0	
		high	3601	2.4	1745	1709	1508	1	754	84	0	
		low	3605	0	1738	1714	1492	1	746	77	0	
		med-low	3608	0	1585	1542	1490	1	745	83	0	
	Med	med-high	3602	0.4	1646	1632	1498	1	749	84	0	
		high	3602	2.3	1721	1688	1523	1	761	86	0	
		low	3605	4.5	1715	1602	1501	1	750	69	0	
	High	med-low	3605	0	1558	1543	1479	1	739	83	0	
		med-high	3601	1.4	1602	1614	1502	1	751	78	0	
		high	3602	1	1683	1638	1505	1	752	84	0	

run time (in seconds) and optimality gap of the solution using CPLEX. The solutions reported for this table were generated using a 1-hour run-time limit given to all models and algorithms. The optimality gaps were computed using the results of Table 4.1. The next columns indicate number of iterations (#Iter), number of Benders and Pareto Optimal cuts added (#BD-C and #PO-C, respectively), and numbers of the two types of variable fixing cuts (#VF-1 and # VF-1, respectively) added.

For 10-city instances, all models and algorithms could reach optimal solutions within the given time limit, with the exception of a few instances. With $|F| = 5$, optimal results were not obtained for seven instances while running the model in CPLEX. The optimality gap in these cases ranged between 0.4% to 4.5%. All instances reached optimal results while running the model with our proposed Benders methodology. Our proposed Benders methodology had higher run times than CPLEX for all instances with $|F| = 2$, but lower run times than that of CPLEX for all instances with $|F| = 5$. For $|F| = 3$ cases, the proposed Benders algorithm outperformed CPLEX's run time in 33% of the instances. On average, the $|F| = 2$ cases had highest number of iterations, and $|F| = 3$ had the lowest. The fact that the $|F| = 3$ cases had the lowest number of iterations is consistent, in that these cases needed on average to add the smallest total number of cuts, and required the lowest average run time, as compared to results obtained using other values for $|F|$. Similarly, the $|F| = 2$ cases had a greater number of iterations, and consistently also required adding the largest total number of cuts. Amongst the types of cuts, Pareto Optimal cuts were added fewer times than other cuts; the number of cuts added ranged from 0 to 4. On average, the largest number of Benders cuts were added in cases of $|F| = 2$, and the smallest for $|F| = 3$. More VF-1 cuts were added, as compared to the number of

VF-2 cuts.

Table 4.4 presents information analogous to that in the table for the 15-city instances. Solutions shown in this table were generated using a 1-hour runtime limit, given to all models and algorithms, and optimality gaps were calculated using results of Table 4.1. For these instances, CPLEX could not reach optimality within the given 1-hour run time. The lowest optimality gap reported here is 12.3%, while the highest is 52.5%. For $|F| = 2$, the optimal solutions were found by the proposed Benders algorithm. For others, i.e. $|F| = 3$ and 5, the smallest and largest optimality gaps reported respectively are 0.1% and 8.7%. This algorithm, within a 1-hour runtime, outperformed the solutions obtained by CPLEX in all cases.

Instances for $|F| = 2$ had the greatest number of iterations, and the largest number of total cuts added, in the average case. However, since for $|F| = 3$ and 5, no optimal results could be obtained, the number of iterations and cuts are not comparable with those of $|F| = 2$. Similar to the case of the 10-city instances, out of the total number of cuts added the majority were Benders cuts; only a few Pareto Optimality cuts were added. The number of VF-1 cuts was significantly larger than the number of VF-2 cuts added in all cases.

4.4 Summary

In this chapter, we introduced a hub arc location problem that integrates the ideas of shipment consolidation and inventory holding. This is in a setup where the freight carrier

Table 4.4: Algorithmic Performance for 15 city instances

F	ζ	η	CPLEX			Proposed Methodology					
			Time (s)	Gap (%)	Time (s)	#Iter.	#BD-C	#PO-C	#VF-1	#VF-2	Gap (%)
2	Low	low	3601	28.6	3606	1416	1376	1	688	119	0
		med-low	3602	27.2	3609	1472	1356	3	678	124	0
		med-high	3601	12.3	3609	1342	1303	2	651	149	0
		high	3602	22.1	3608	1340	1305	13	652	155	0
	Med	low	3603	27.2	3604	1409	1339	1	669	126	0
		med-low	3601	22.1	3602	1414	1353	4	676	122	0
		med-high	3601	27.8	3606	1361	1333	1	666	134	0
		high	3601	22.1	3603	1331	1301	13	160	155	0
	High	low	3601	14.9	3605	1401	1324	2	662	134	0
		med-low	3601	30.9	3600	1420	1353	1	676	131	0
		med-high	3601	23.1	3600	1375	1326	3	663	136	0
		high	3601	26.9	3601	1368	1320	8	660	133	0
3	Low	low	3602	42.4	3602	1229	1208	1	604	161	5.1
		med-low	3603	34.2	3602	1155	1131	1	565	158	3.9
		med-high	3603	35.8	3605	1305	1265	1	632	154	3.8
		high	3602	47.2	3604	1284	1255	1	627	180	1.6
	Med	low	3604	33.9	3603	1202	1173	1	586	153	2.2
		med-low	3602	39.4	3603	1180	1147	1	573	159	4.3
		med-high	3603	40.6	3608	1288	1258	1	629	167	2.6
		high	3602	44.7	3606	1314	1274	1	637	174	1.8
	High	low	3604	39	3603	1174	1148	1	574	147	5.5
		med-low	3602	33.7	3602	1135	1112	1	556	165	5
		med-high	3606	37.7	3603	1293	1258	1	629	171	0.1
		high	3602	18.4	3601	1297	1259	1	629	169	1.5
5	Low	low	3606	49.8	3603	1335	1331	1	665	106	4.5
		med-low	3604	39.5	3607	1321	1311	1	655	100	6
		med-high	3605	38.9	3604	1349	1317	1	658	136	5.4
		high	3604	47.6	3604	1482	1289	1	644	198	8.7
	Med	low	3605	39.9	3609	1364	1332	1	666	103	8.4
		med-low	3604	34.8	3607	1230	1195	1	597	139	3.4
		med-high	3605	32.4	3607	1147	1131	1	565	133	5
		high	3604	41.5	3604	1222	1106	1	553	176	8
	High	low	3605	34.6	3607	1268	1257	1	628	107	6.9
		med-low	3605	37.5	3603	1293	1279	1	639	121	5.9
		med-high	3605	57.1	3603	1189	1090	1	545	148	4.4
		high	3604	52.7	3609	1557	1240	1	620	205	7.9

decides the frequencies of vehicles (from its heterogeneous fleet) to be operated on a set of hub arcs. We presented a mixed integer formulation that models the problem. We also proposed a Benders decomposition algorithm to solve larger instances. This algorithm's performance was enhanced by incorporating Pareto Optimal cuts, and by two variable-fixing techniques.

We performed an extensive computational study of the proposed model and algorithm by using a dataset generated from CAB and FAF, described in previous sections, and as employed in the last chapter. In view of our computational results, the proposed Benders Decomposition algorithm seemed to outperform the model run in CPLEX. Additionally, we observed that consideration of further frequency choices lead to a network with lower total cost of establishment plus operation.

Chapter 5

Shipment Consolidation in Hub Arc Location Problems under Uncertainty

Shipment consolidation in hub arc location involves network design problems that require strategic decisions to be maintained over a significant time frame. During this period, various unpredictable factors, such as fluctuations in demand, may influence operational decisions. Consequently, deterministic models may fail to yield optimized decisions under such conditions. That highlights the necessity of incorporating uncertainty into the decision-making process. In this chapter, we assume random demand, and seek to determine optimal decisions under the anticipated uncertainty.

We thus develop a model that considers stochastic demand. This is a two stage model

of the decision-making process, and considers when information is revealed. In the first stage of this model, we determine the location of arcs, as well as the types and frequencies of vehicles that will operate on them. Subsequently, we allocate demands to the arcs, perform optimal routing, and make decisions regarding inventory holding. Thus, the first-stage decisions are long-term strategic in nature. The second-stage decisions are operational ones that vary according to the particular realization of the uncertain demand.

To solve the stochastic model, we propose an algorithm based on the Sample Average Approximation. We assess the algorithm's performance through extensive computational experiments under various parameter settings. The outline of this chapter is as follows: Section 5.1 presents the problem and its mathematical formulation. Sections 5.2 and 5.3 propose an SAA algorithm and present computational analysis using that algorithm. Section 5.4 provides concluding remarks.

5.1 Mathematical Formulation

We adopt the same notations for graph, sets of nodes and arcs, endpoints of an arc, planning horizon and its intervals, vehicles, their fixed and variable costs, frequencies, capacities, and inventory cost as used in Section 4.1. In this chapter, we explore a scenario where demand is unpredictable, but follows a known probability distribution. Introducing stochastic demand into our model, we define additional notation $w_k(\xi)$ are random variables representing the anticipated flow starting from time $t^0(k) \in T$ at the origin node $o(k) \in N$ and ending at the destination node $d(k) \in N$. It is assumed that each $w_k(\xi)$ adheres to the same, known probability distribution, i.e. these demands are independent of each other

and identically distributed. We assume these demands are divisible at the first hub along the destination path. We use E_ξ to symbolize the expected value in relation to $\xi \in \Xi$, with Ξ denoting the support for ξ .

To introduce the stochastic program, we use the same z_{af} variable as defined in Section 4.1 as a first-stage variable here. We also define $x_{ak}^t(\xi)$ as the second-stage decision variable to represent the fraction of flow on arc $a \in A$, starting at the beginning of time $t \in T$ with respect to the demand $k \in K$ and for realization $\xi \in \Xi$. Assuming stochastic demand, shipment consolidation in the hub arc location problem can be modeled by extending formulation 4.1 - 4.5 as the following two-stage program:

$$\min_{\mathbf{z} \in \mathcal{U}} E_\xi \left[\varphi(\mathbf{z}, \xi) \right] + \sum_{a \in A} \sum_{f \in F} C_{af} z_{af} \quad (5.1)$$

where E_ξ is the expectation w.r.t. ξ and the domain of variable $\mathbf{z}$ is defined as $\mathcal{U} = \{\mathbf{z} : z_{af} \in \{0, 1\} \forall f \in F \forall a \in A\}$. The function $\varphi(\mathbf{z}, \xi)$ represents the costs on allocation arcs plus inventory cost, for a particular realization of $\xi \in \Xi$ and a given $\mathbf{z}$. Computation of the cost $\varphi(\mathbf{z}, \xi)$ requires solving the second stage of the stochastic problem with probabilistic demand, as follows:

$$\begin{aligned} \varphi(\mathbf{z}, \xi) = \min & \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} c_{ak} w_k(\xi) x_{ak}^t(\xi) \\ & + \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} \eta(t - t^0(k) - \tau_{ak}) w_k(\xi) x_{ak}^t(\xi) \end{aligned} \quad (5.2)$$

subject to

$$\sum_{a \in A} \sum_{t \in T: t \geq t^0(k)} x_{ak}^t(\xi) = 1 \quad \forall k \in K, \xi \in \Xi \quad (5.3)$$

$$\sum_{k \in K: t^0(k) \geq t - \tau_{ak}} w_k(\xi) x_{ak}^t(\xi) \leq \sum_{f \in F} \delta_f^t \Gamma_f \mathbf{z}_{af} \quad \forall a \in A, o(a) \neq d(a), \forall t \in T, \xi \in \Xi, \quad (5.4)$$

$$x_{ak}^t(\xi) \geq 0 \quad \forall k \in K, \forall a \in A \forall t \in T, \xi \in \Xi, \quad (5.5)$$

The objective function (5.1) represents the total cost, which is calculated by adding two components - the deterministic cost of operating on hub arcs, and the expectation of costs on allocation arcs plus inventory cost. The second component is detailed in equation (5.2). Constraints (5.3) ensure that each demand can be partially satisfied through different hub arcs, but its flow must be routed via hub arcs. Constraint (5.4) ensures that flow on an arc in each period is bounded by the total capacity of vehicles operating on that arc. We note that multiple vehicle frequencies may be activated on any arc. The constraints (5.2) - (5.4) incorporate realizations ξ of uncertain scenerios. The domains of the decision variables are defined by constraints (5.5).

The described model outlines a two-stage stochastic programming approach. The objective function of the model (5.2) features a deterministic component, accounting for the cost associated with hub arcs, alongside the expected value of the second stage objective. This second component incorporates the expected value of cost related to allocation arcs and inventory holding.

The initial stage tackles a long-term strategic problem, focusing on selecting hub arcs, along with the corresponding vehicles and their frequencies. These decisions remain stable

and unaffected by fluctuations in demand. Consequently, the variables z_{af} are determined during the first phase. However, the allocation of flows to arcs and routing decisions are subject to change, based on demand fluctuations. These tactical choices are made in the second stage, tailored to specific instances of the vector ξ within its defined space Ξ . As such, the variables x_{ak}^t are established during this latter stage.

5.2 Solution Methodology for Stochastic Model

According to [Kleywegt et al. \(2002\)](#), the primary difficulty in addressing the stochastic model (5.2)–(5.5) lies in calculating the expectation of the objective function. The Sample Average Approximation technique tackles this by producing $n \in \mathcal{N}$ random samples of the demand-vector, thereby estimating the second-stage expectation of the objective (5.2):

$$E_{\xi} \left[\sum_{k \in K} \sum_{a \in A} \sum_{t \in T} c_{ak} w_k(\xi) x_{ak}^t(\xi) + \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} \eta(t - t^0(k) - \tau_{ak}) w_k(\xi) x_{ak}^t(\xi) \right]$$

through the sample average function in the following manner:

$$\frac{1}{|\mathcal{N}|} \left[\sum_{n \in \mathcal{N}} \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} c_{ak} w_k^n x_{ak}^{tn} + \sum_{n \in \mathcal{N}} \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} \eta(t - t^0(k) - \tau_{ak}) w_k^n x_{ak}^{tn} \right]$$

In this context, w_k^n is the flow initiated at the at start of time $t^0(k) \in T$ from node $o(k) \in N$, heading towards node $d(k) \in N$ under sample $n \in \mathcal{N}$, while the variable x_{ak}^{tn} denotes the fraction of flow on arc $a \in A$ commencing at the start of time $t \in T$, relevant to demand $k \in K$ within the same scenario. The SAA method approximates the stochastic problem (5.2)–(5.5) as follows:

$$\begin{aligned}
& \text{minimize} && \sum_{a \in A} \sum_{f \in F} \mathcal{C}_{af} z_{af} \\
& && + \frac{1}{|\mathcal{N}|} \left[\sum_{n \in \mathcal{N}} \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} c_{ak} w_k^n x_{ak}^{tn} \right. \\
& && \left. + \sum_{n \in \mathcal{N}} \sum_{k \in K} \sum_{a \in A} \sum_{t \in T} \eta(t - t^0(k) - \tau_{ak}) w_k^n x_{ak}^{tn} \right] \tag{5.6}
\end{aligned}$$

subject to

$$\sum_{a \in A} \sum_{t \in T: t \geq t^0(k)} x_{ak}^{tn} = 1 \quad \forall k \in K, n \in \mathcal{N} \tag{5.7}$$

$$\begin{aligned}
& \sum_{k \in K: t^0(k) \geq t - \tau_{ak}} w_k^n x_{ak}^{tn} \leq \sum_{f \in F} \delta_f^t \Gamma_f z_{af} \\
& \quad \forall a \in A, o(a) \neq d(a), \forall t \in T, n \in \mathcal{N}, \tag{5.8}
\end{aligned}$$

$$x_{ak}^{tn} \geq 0 \quad \forall k \in K, \forall a \in A \forall t \in T, n \in \mathcal{N}, \tag{5.9}$$

$$z_{af} \in \{0, 1\} \quad f \in F, a \in A \tag{5.10}$$

The SAA formulation (5.6) - (5.10) will be utilized as the mathematical model for our stochastic problem henceforth. We repeat the process of randomly generating $\mathcal{N}$ samples and solving the formulation in (5.6) - (5.10) $\mathcal{M}$ times, taking the average of the objective function values of the solutions generated in these $\mathcal{M}$ replications. Thus, SAA estimates the optimal value of the stochastic program (5.2) - (5.5).

Given values for the first-stage decision variables z_{af} , SAA generates a larger set of samples $\mathcal{N}'$, for the second-stage problem to be solved. We then take the average of the objective function values of the second-stage problem, providing an estimation of the upper bound for our problem. Algorithm 4 details the procedure of the SAA algorithm for the

stochastic problem presented above.

Algorithm 4 SAA Algorithm

- 1: **for** $j \in \mathcal{M}$ **do**
 - 2: Generate sample of size $|\mathcal{N}|$
 - 3: Solve formulation (5.6) - (5.10) for each sample $\mathcal{N}_j$ to optimality
 - 4: Store the objective value of this solution in $\mathcal{V}_{\mathcal{N}_j}$
 - 5: **end for**
 - 6: Calculate $\bar{\mathcal{V}}_{\mathcal{M}}^{\mathcal{N}} \leftarrow \frac{1}{|\mathcal{M}|} \sum_j \mathcal{V}_{\mathcal{N}_j}$
 - 7: $UB \leftarrow \bar{\mathcal{V}}_{\mathcal{M}}^{\mathcal{N}}$
 - 8: Choose a feasible solution $\hat{z}_{af}$ to formulation (5.2) - (5.5)
 - 9: Generate a sample of size $\mathcal{N}'$, such that $|\mathcal{N}'| \gg |\mathcal{N}|$
 - 10: Solve formulation (5.6) - (5.10) for $\hat{z}_{af}$ and sample $\mathcal{N}'$
 - 11: Store the objective value of this solution to $\bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})$
 - 12: $LB \leftarrow \bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})$
 - 13: Calculate $SAA_{Gap}(\mathcal{N}, \mathcal{M}, \mathcal{N}', \hat{z}_{af}) \leftarrow \frac{|\bar{\mathcal{V}}_{\mathcal{M}}^{\mathcal{N}} - \bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})|}{\bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})}$
-

Algorithm 4 initially generates $\mathcal{M}$ independent samples, each of size $|\mathcal{N}|$. It then solves the SAA formulation (5.6) - (5.10) for each sample. $\mathcal{V}_{\mathcal{N}_j}$ represent the objective values. This is done in steps 1-5 of the algorithm. Next, in steps 6-7, we calculate the average of all optimal objective values in values from previous step.

In steps 8-12, we select a feasible solution formulation (5.2) - (5.5) and solve (5.6) - (5.10) to estimate the objective function value $\bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})$ using this solution. Finally, the SAA gap is computed in step 13.

The Sample Average Approximation provides a mechanism for computing statistical estimators of both upper and lower bounds on the optimal value of our stochastic problem. In this algorithm, $\mathcal{V}_{\mathcal{N}_j}$ represents the optimal solution derived from solving the SAA problem defined by Equations (5.6) - (5.10) for each sample $\mathcal{N}_j$. The symbol $\bar{\mathcal{V}}_{\mathcal{M}}^{\mathcal{N}}$ denotes the average of all these optimal solutions obtained from the SAA problems. $\bar{\mathcal{V}}_{\mathcal{M}}^{\mathcal{N}}$ serves as

a statistical upper bound for the optimal value of the original problem.

Given values for the first-stage decision variables, SAA generates a much larger set of samples $\mathcal{N}'$, for which the second-stage problem will be solved. The average of the objective function values of the second-stage problem, taken over this larger set of scenarios, is then used as an estimator of a lower bound. The term $\bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})$ denotes the objective function value of the original stochastic problem when employing the solution $\hat{z}_{af}$ and a sample size of $\mathcal{N}'$. Considering that the first-stage variables are fixed, the number of scenarios $\mathcal{N}'$ that one can generate is significantly larger than the sample size $\mathcal{N}$ employed in solving the SAA problems. The estimator $\bar{\mathcal{V}}_{\mathcal{N}'}(\hat{z}_{af})$ acts as a statistical lower bound for the optimal objective function value. With these two bounds, the SAA gap, $SAA_{Gap}(\mathcal{N}, \mathcal{M}, \mathcal{N}', \hat{z}_{af})$, can be computed.

In stochastic programming, complete recourse ensures that for any first-stage decision and realization of uncertainty, there exists a feasible solution in the second stage, thus avoiding infeasibility altogether. In contrast, relative recourse occurs when second-stage feasibility or performance depends heavily on the chosen first-stage decisions and the uncertain parameters, which may lead to infeasibility or suboptimal outcomes in certain scenarios. [Crainic et al. \(2016\)](#) employed Benders decomposition combined with SAA for a large-scale problem where uncertainty significantly impacts the recourse structure to effectively handle relative recourse. This approach was implemented to address relative recourse challenges by iteratively refining first-stage decisions using the Benders feasibility and optimality cuts derived from second-stage evaluations. Similarly, in this research, we employ a Benders decomposition implementation in CPLEX, which refines first-stage decisions and ensures feasibility in the second stage.

Beyond the use of Benders decomposition, the literature has explored various strategies for handling relative recourse. [Kleywegt et al. \(2002\)](#) utilized adaptive sampling that focuses on identifying critical scenarios dynamically during the optimization process, thereby refining the approximation of the stochastic problem. [Chen and Fukushima \(2005\)](#) utilized modeling techniques that introduce a penalty term in the objective function for infeasible second-stage scenarios, and thereby incentivizes first-stage decisions that minimize the impact of infeasibility. [Pagnoncelli et al. \(2009\)](#) managed relative recourse by ensuring second-stage feasibility under a high-probability constraint, which balances solution quality and robustness. [Alnaggar et al. \(2020\)](#) utilized a set of additional variables (in their problem using a “spot market carrier” if demand exceeds capacity) to ensure that the problem is feasible regardless of the realization of demand, thereby handling relative recourse. [Ahmed and Shapiro \(2002\)](#) applied a hybrid strategy for more comprehensive handling of relative recourse.

5.3 Computational Analysis

In this section, we analyze the impact of demand uncertainty on the solutions. To conduct the computational experiments, we employ the same parameter settings as outlined in Section [4.3.1](#). The experiments are performed on 10-city instances. Initially, we focus on assessing the practical convergence of the SAA scheme in the context of the stochastic model. This is followed by an estimation of the value of employing stochastic solutions.

The SAA algorithm was implemented in Java (IntelliJ IDEA 21.2.1 using JDK 15.0.2) and integrated with CPLEX 12.9. The CPLEX solver was utilized with its built-in Benders

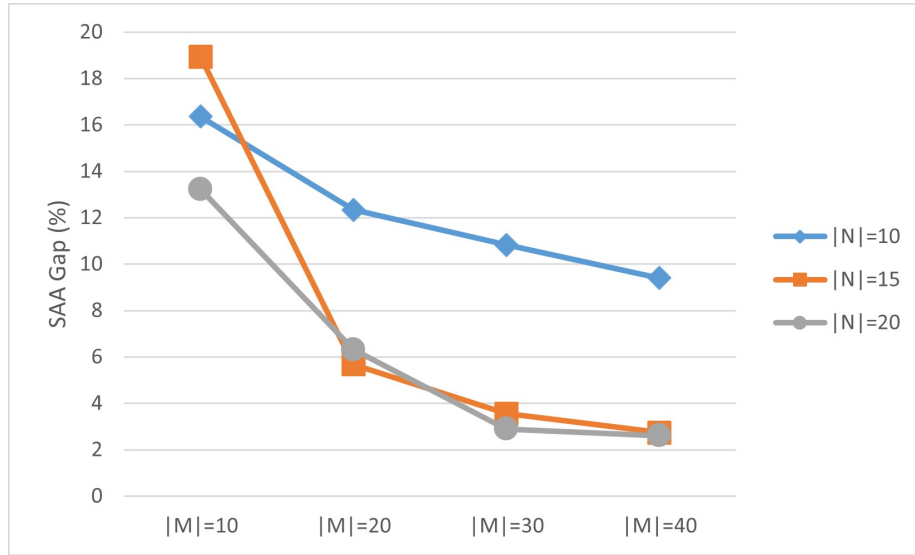
decomposition implementation. All instances were executed on an IBM machine equipped with an Intel Core i7 processor and 64 GB of RAM. The SAA gap was evaluated using a sample size of $|\mathcal{N}'| = 500$.

5.3.1 Sample Generation

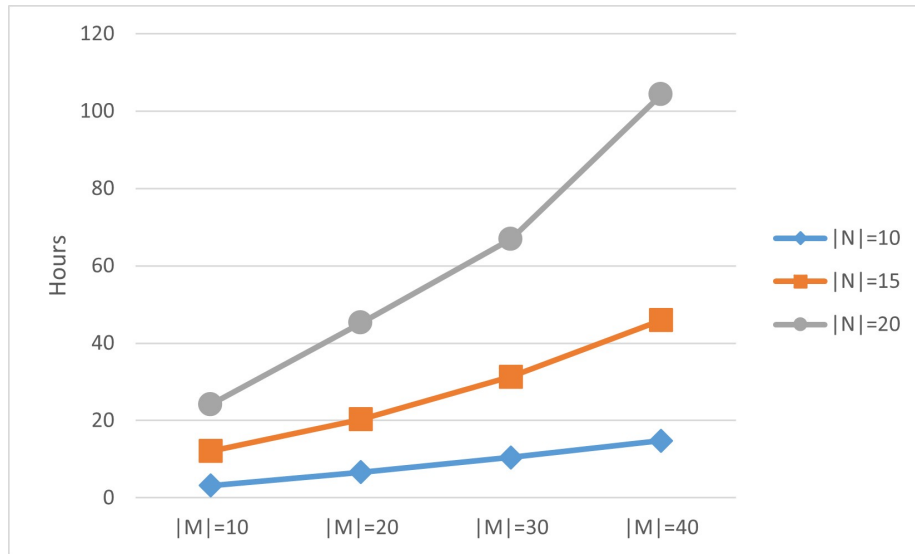
We generate independent samples for O-D demands using a normal distribution. The respective mean values (μ) of this distribution are estimated from the deterministic O-D demands outlined in Section 4.3.1. Using the same notation from that section, w_k represents the flow originating at the start of time $t^0(k) \in T$ from node $o(k) \in N$, destined for node $d(k) \in N$. We compute the total demand between $o(k)$ and $d(k)$ as $\bar{w} = \sum_{t^0(k) \in T} w_k$, and this value is used as the mean demand, i.e., $\mu = \bar{w}$ for $o(k)$ and $d(k)$. The standard deviation, σ , is a percentage of the mean value and is set as $\sigma = \nu \bar{w}$, where ν is the coefficient of variation. We use $\nu = \{0.2, 0.5, 0.7, 1.0\}$ to randomly generate demand samples in our experiments. Using this distribution, $\mathcal{N}(\bar{w}, \nu \bar{w})$, we generate demands for flows between $o(k)$ and $d(k)$ across different time periods.

5.3.2 Convergence of SAA-results

This section aims to evaluate the practical convergence of the SAA scheme to determine the appropriate sample size, $|\mathcal{N}|$, and number of replications, $|\mathcal{M}|$, that provide an optimal balance between solution quality and computational time. We conducted numerical experiments with sample sizes $|\mathcal{N}| = 10, 15, 20$ and replication counts $|\mathcal{M}| = 10, 20, 30, 40$. A 10-city instance was used, with *low* hub arc establishment cost and *low* unit holding



(a) Optimality Gap



(b) CPU Time

Figure 5.1: Optimality gap and Total CPU Time for the SAA algorithm, across varying sample sizes and number of replications

cost, and the five-frequency option $|F| = 5$ was tested. The SAA gap was assessed using a sample size of $|\mathcal{N}'| = 500$.

Figure 5.1 illustrates the optimality gap and total computational time required for the SAA algorithm across various sample sizes $|\mathcal{N}|$ and replications $|\mathcal{M}|$. As shown in Figure 5.1a, increasing the sample size generally leads to a reduced optimality gap. However, as the parameters $|\mathcal{N}|$ and $|\mathcal{M}|$ increase, the corresponding computation time rises significantly, as illustrated in Figure 5.1b. In general, as the $|\mathcal{M}|$ grows, the best average SAA gaps are obtained for sample sizes $|\mathcal{N}| = 20$ and 30. However, those differ in computational time. Overall, the computational time for $|\mathcal{N}| = 30$ is considerably greater than that of $|\mathcal{N}| = 20$. Based on these observations, we selected a configuration of fifteen scenarios ($|\mathcal{N}| = 15$) and thirty replications ($|\mathcal{M}| = 30$) for the subsequent numerical analysis, aiming to strike a balance between computational efficiency and solution quality.

5.3.3 Quality of Solutions

In this section, we evaluate the computational performance of the proposed SAA algorithm using 10-city instances. We focus on assessing how variations in parameters — such as the number of frequency options available, hub arc establishment costs, and unit holding costs — affect the solution. Specifically, we analyze the stochastic solution costs (calculated across fifteen samples and thirty replications), the number of hub arcs selected, the numbers employed of each aircraft type (B-747-F and B-747-100) in the optimal scenario, the SAA gap, and the computational time required.

Table 5.1 presents the results. The first three columns detail the number of frequency

options ($|F|$) and the specific parameter variations, while the fourth column reports the optimal stochastic solution cost ($\bar{\mathcal{V}}_{\mathcal{M}}^{\mathcal{N}}$). The subsequent three columns list the number of hub arcs, and the numbers of B-747-F and B-747-100 aircraft used in the lowest-cost scenario. The eighth column provides the SAA gap; the final column reports the total computational time in hours.

Despite some variations in hub arcs and in the frequencies assigned to hub arcs for individual replications, the results are generally consistent across different parameter settings. Notably, changes in parameters, such as hub arc establishment costs and unit holding costs, do not significantly impact the number of hub arcs selected or the types of aircraft used. The one exception is the case $|F| = 5$. In this instance, when the unit holding cost increased from *med-high* to *high*, the model opted for two additional B-747-100 aircraft and two fewer B-747-F aircraft in the best-case scenario. This consistency across results is largely attributable to the inclusion of multiple scenarios, which allows the model to adapt the hub arcs and frequency choices to accommodate demand fluctuations effectively. Additionally, we observe that the optimal cost tends to decrease as the number of available frequency choices increases. This suggests that frequency choices have a more pronounced effect on total cost than variations in other parameters (such as hub arc establishment or holding costs).

On average, across different parameter and frequency settings, the instances demonstrated an estimated optimality gap of 3%, with a minimum gap of 1.9% and a maximum of 3.8%. The lowest average gap of 2.6% occurred when $|F| = 2$, followed by 2.9% for $|F| = 3$, and 3.6% for $|F| = 5$. In terms of computational time, the cases with $|F| = 2$ reported the shortest runtime, while runtimes increased significantly for the other two-

Table 5.1: Computational results of the stochastic model for 10-node instances

$ F $	ζ	η	Cost (1000\$)	No. of Hub Arcs	No. of B-747-F	No. of B-747-100	SAA Gap (%)	Time (hrs)
2	Low	low	1054	5	20	0	2.53	2.5
		med-low	1056	5	20	0	2.3	3.1
		med-high	1072	5	20	0	3.01	1.8
		high	1095	5	20	0	3.66	4.6
	Med	low	1057	5	20	0	2.46	1.9
		med-low	1059	5	20	0	2.42	1.8
		med-high	1075	5	20	0	2.94	3.6
		high	1098	5	20	0	3.7	2.4
	High	low	1068	5	20	0	2.61	1.8
		med-low	1069	5	20	0	2.6	2.2
		med-high	1085	5	20	0	3.13	3.9
		high	1108	5	20	0	3.71	4.6
3	Low	low	842	8	0	18	2.1	18.6
		med-low	843	8	0	18	1.91	20.1
		med-high	858	8	0	18	2.63	18.1
		high	883	8	0	18	3.34	45
	Med	low	847	8	0	18	1.99	16.8
		med-low	849	8	0	18	2.14	19.5
		med-high	864	8	0	18	2.73	18.3
		high	888	8	0	18	3.35	47.7
	High	low	866	8	0	18	2.25	15.3
		med-low	867	8	0	18	2.2	17.2
		med-high	883	8	0	18	3.01	23
		high	906	8	0	18	3.32	47
5	Low	low	799	6	13	2	3.55	31.4
		med-low	801	6	13	2	3.44	33.8
		med-high	821	6	13	2	3.73	35
		high	848	6	11	4	3.8	77.1
	Med	low	803	6	13	2	3.45	28.3
		med-low	805	6	13	2	3.53	29
		med-high	825	6	13	2	3.68	34.7
		high	854	6	11	4	3.74	81.9
	High	low	816	6	13	2	3.51	79.2
		med-low	817	6	13	2	3.58	63.4
		med-high	837	6	13	2	3.72	87.5
		high	863	6	11	4	3.79	86.3

frequency settings. The $|F| = 3$ instances required a moderate amount of time, with the $|F| = 5$ cases taking the longest to solve.

These findings confirm the efficiency of the traditional SAA algorithm. The results emphasize the algorithm’s practical applicability in real-world settings, since computation times remain within acceptable limits.

5.3.4 Value of Stochastic Solution

[Birge \(1982\)](#) introduced a statistical measure known as the Value of the Stochastic Solution (VSS), which provides bounds on the benefit gained from employing a stochastic, rather than a deterministic, solution. VSS is particularly useful in assessing the advantage of solving a stochastic program compared to solving its deterministic optimization counterpart, where expected values are substituted for random variables. The VSS metric allows us to quantify the potential improvement that arises from explicitly modeling uncertainty in the system. In this section, we apply the VSS to evaluate the effectiveness of incorporating stochastic elements into the two-stage model proposed in Equations (5.1) through (5.5). By doing so, we aim to determine whether addressing uncertainty within the model offers substantial benefits in terms of solution quality and cost-effectiveness, compared to relying on a deterministic approach that neglects the inherent randomness of demand data. This analysis helps justify the additional computational effort required for stochastic programming, and highlights the added value derived from accounting for variability.

We now outline the procedure for calculating the VSS for the model described in equa-

Table 5.2: Value of Stochastic Solution for 10-node instances

F	ζ	η	Stochastic					Deterministic					Gap (%)
			Cost (1000\$)	No. of Hub Arcs	No. of B-747-F	No. of B-747-100	Cost (1000\$)	No. of Hub Arcs	No. of B-747-F	No. of B-747-100			
2	Low	low	825	13	52	0	886	15	60	0	6.9		
		med-low	827	13	52	0	888	15	60	0	6.89		
		med-high	844	13	52	0	905	15	60	0	6.74		
		high	867	13	52	0	926	15	60	0	6.41		
		low	828	13	52	0	889	15	60	0	6.8		
		med-low	829	13	52	0	891	15	60	0	6.89		
	Med	med-high	847	13	52	0	907	15	60	0	6.66		
		high	870	13	52	0	930	15	60	0	6.48		
		low	835	13	52	0	898	15	60	0	7		
	High	med-low	838	13	52	0	900	15	60	0	6.85		
		med-high	854	13	52	0	917	15	60	0	6.82		
		high	878	13	52	0	938	15	60	0	6.5		
3	Low	low	675	19	6	40	752	19	6	46	10.25		
		med-low	677	18	6	40	754	19	6	46	10.25		
		med-high	694	17	6	40	773	18	6	46	10.14		
		high	715	16	6	40	799	19	6	46	10.42		
		low	680	19	6	40	757	19	6	46	10.25		
		med-low	682	19	6	40	759	19	6	46	10.21		
	Med	med-high	698	18	6	40	778	18	6	46	10.21		
		high	720	17	6	40	804	19	6	46	10.43		
		low	693	18	6	40	772	19	6	46	10.31		
	High	med-low	694	18	6	40	774	19	6	46	10.37		
		med-high	711	18	6	40	793	19	6	46	10.31		
		high	733	17	6	40	819	19	6	46	10.46		
5	Low	low	635	14	36	0	743	18	42	0	14.56		
		med-low	637	14	36	0	746	18	42	0	14.54		
		med-high	658	13	35	0	772	17	41	0	14.78		
		high	687	13	35	0	804	16	40	0	14.52		
		low	638	14	36	0	747	18	42	0	14.59		
		med-low	640	14	36	0	749	18	42	0	14.62		
	Med	med-high	661	13	35	0	775	17	41	0	14.81		
		high	690	13	35	0	808	16	40	0	14.58		
		low	646	14	36	0	758	18	42	0	14.66		
	High	med-low	648	14	36	0	760	18	42	0	14.72		
		med-high	668	13	35	0	786	17	41	0	14.93		
		high	698	13	35	0	817	16	40	0	14.63		

tions (5.1) through (5.5). We utilize $|\mathcal{N}| = 15$ samples derived from a single replication of the dataset. This allows us to assess the potential improvement gained by solving the stochastic model in comparison to a deterministic approach.

Using these samples, the process unfolds in two key steps. In the *first step*, we solve the SAA formulation presented in Equations (5.6) through (5.10), yielding the objective function value $\mathcal{V}_{\mathcal{N}}$, which represents the expected costs incurred by solving the stochastic program. In the *second step*, we focus on demands of the samples in $\mathcal{N}$ and calculate a single aggregated demand sample $\bar{w}_k$. This is done by averaging the demands across all samples $n \in \mathcal{N}$, i.e., $\bar{w}_k = \frac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} w_k^n$. Using this average demand sample, we then solve the deterministic model as defined in Equations (4.1) through (4.5), obtaining the first-stage decision variables z . With those z values now fixed, we proceed to solve the second-stage subproblem for each of the $n \in \mathcal{N}$ samples to compute $\varphi(\mathbf{z}, \xi)$.

Given the selected values of z , if all the samples in $\mathcal{N}$ are feasible, we compute the deterministic value $\mathcal{V}_{Det}$ using the formula:

$$\mathcal{V}_{Det} = \sum_{n \in \mathcal{N}} \frac{1}{|\mathcal{N}|} \varphi(\mathbf{z}, \xi) + \sum_{a \in A} \sum_{f \in F} \mathcal{C}_{af} z_{af}.$$

Conversely, if some samples are found to be infeasible for the chosen first-stage variables, we revise the set $\mathcal{N}$ in the *second step* by removing those infeasible samples identified. We then iterate through the first and second steps until we arrive at a subset $\mathcal{N}$ for which we have successfully computed $\mathcal{V}_{\mathcal{N}}$ in the first step, and all remaining samples are feasible in the second step. At this point, we can compute the corresponding $\mathcal{V}_{Det}$.

When all samples utilized in the first step are found to be feasible in the second step,

we compute the VSS using the following formula, and express this gap as a percentage, as presented in the last column of Table 5.2:

$$\text{VSS} = (\mathcal{V}_{Det} - \mathcal{V}_N) / \mathcal{V}_{Det}.$$

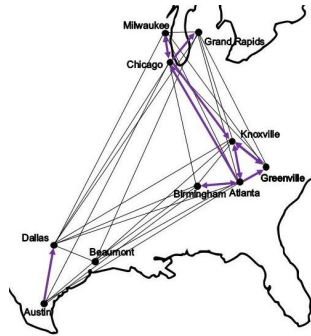
That table presents the VSS across various parameter configurations. The first three columns specify the number of frequency options ($|F|$), along with the hub arc establishment cost and unit holding cost. Columns four through seven detail the total cost, the number of hub arcs, and the numbers of B-747-F and B-747-100 aircraft in the stochastic solution. Similarly, columns eight through eleven provide the corresponding information for the deterministic (i.e., expected value) solution. Column twelve reports the percentage gap between the costs of the stochastic and deterministic solutions, representing the VSS for each parameter configuration. For this table, calculations were based on the respective worst-case scenarios across the three frequency configurations. The initial set consisted of 15 samples, but after two iterations, six scenarios were excluded, resulting in the final computational results being based on nine samples.

For the three frequency choices, respectively $|F| = 2, 3, 5$, the reported VSS were 6.75%, 10.3%, and 14.7%. These values reflect the average gains of the two-stage stochastic model (Equations (5.6) through (5.10)) compared to the deterministic model (Equations (4.1) through (4.5)). In the case of $|F| = 2$, the deterministic solution required two additional hub arcs and eight more B-747-100 aircraft, although neither the stochastic nor deterministic solutions utilized any B-747-F aircraft. For $|F| = 3$, the stochastic solution employed 16 to 19 hub arcs, with six B-747-F and 40 B-747-100 aircraft in operation. In contrast, the deterministic solution utilized 18 or 19 hub arcs and required six additional

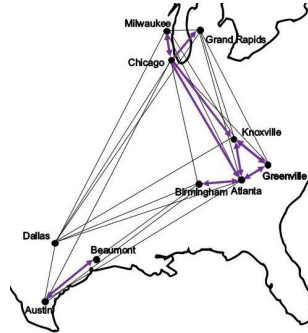
B-747-100 aircraft. For $|F| = 5$, the stochastic solutions established 13 or 14 hub arcs, while the deterministic solutions required 16 to 18 hub arcs. Regarding aircraft usage, the deterministic solution showed an average 16% increase in B-747-F aircraft compared to the stochastic case. No B-747-100 were used in these cases.

To facilitate a comparison between the deterministic and stochastic solutions, selected resultant networks are illustrated in Figure 5.2. From the outcomes presented in Table 5.2, we highlight the stochastic and deterministic solution pairs for instances with *medium* hub arc costs and *high* holding costs across different values of $|F|$. In all diagrams, directed arrows indicate the hub arcs, with the arrow's direction corresponding to that of flow. The frequency of vehicles on each hub arc is represented by different colors. For cases where $|F| = 2$ and 5, solid lines denote the hub arcs, while for $|F| = 3$, dashed lines are used. The color codes for the frequencies will be detailed in the subsequent paragraphs that describe each pair of solutions for different values of $|F|$.

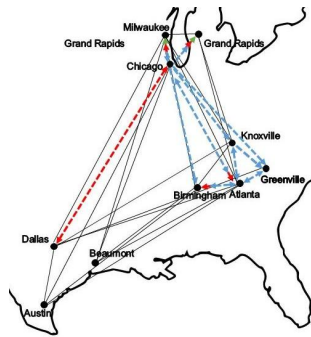
Figures 5.2a and 5.2b depict the network configurations for the stochastic and deterministic solutions, respectively, for $|F| = 2$. In both figures, the hub arcs are represented in purple, indicating the deployment of B-747-F aircraft on these routes: two on Monday, one on Wednesday, and one on Friday. The alternative frequency option, which involved the use of B-747-100 aircraft, was not employed in either solution. Of the hub arcs utilized, 11 were consistent between both configurations. However, the stochastic solution included two additional hub arcs: from Austin to Dallas and from Atlanta to Chicago, which were absent in the deterministic configuration (Figure 5.2b). Conversely, the deterministic solution featured four extra hub arcs connecting Austin to Beaumont, Beaumont to Austin, Atlanta to Chicago and Greenville to Atlanta.



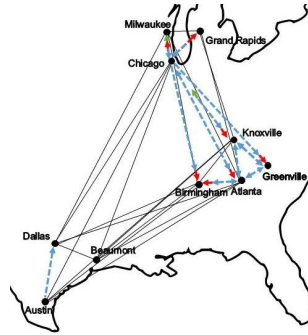
(a) Stochastic with two frequencies, Cost=\$870K



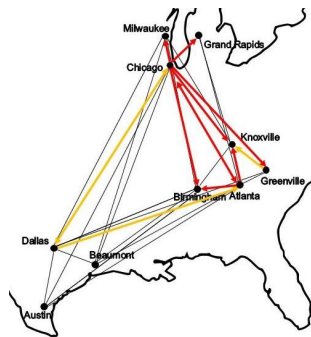
(b) Deterministic with two frequencies, Cost=\$930K



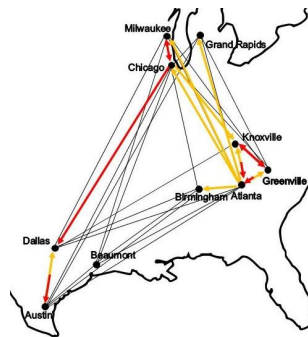
(c) Stochastic with three frequencies, Cost=\$720K



(d) Deterministic with three frequencies, Cost=\$804K



(e) Stochastic with five frequencies, Cost=\$690K



(f) Deterministic with five frequencies, Cost=\$808K

Figure 5.2: Resultant networks for *med* hub arc cost and *high* holding cost (colors are explained in the text)

Figures 5.2c and 5.2d illustrate the network configurations for the stochastic and deterministic solutions for $|F| = 3$. In both figures, hub arcs are depicted using dashed lines, with different colors representing distinct frequencies: red indicates one B-757-100 flight on Monday and Friday; green denotes one B-757-F flight on Monday, Wednesday, and Saturday; and blue corresponds to one B-757-100 flight on Thursday and Saturday. Some arcs in these figures have several directional arrows, indicating that they accommodate multiple frequencies or, in some cases, different aircraft types on the same arc.

The stochastic solution operated 22 frequencies across 17 distinct hub arcs, while the deterministic solution involved 24 frequencies across 19 distinct hub arcs. Of these, 18 hub arcs and their corresponding frequencies were identical in both configurations. The stochastic solution, however, featured additional blue and green hub arcs from Chicago to Grand Rapids, and two red hub arcs (from Chicago to Dallas and Dallas to Chicago), which were absent in the deterministic configuration. Conversely, the deterministic solution included six additional hub arcs: three red arcs from Chicago to Birmingham, Greenville, and Knoxville; two blue arcs, from Grand Rapids to Chicago and Austin to Dallas; and one green arc from Knoxville to Chicago. Notably, both solutions employed an equal number of green arcs, though the deterministic solution used one extra blue and red hub arc each.

The cases with $|F| = 5$ are depicted in Figures 5.2e and 5.2f. There red, green, blue, yellow, and gray solid lines were used to represent hub arcs, each corresponding to different operating frequencies. Interestingly, in both the stochastic and deterministic solutions for that scenario, only two frequencies were utilized. The red arcs denote one B-757-F flight on Monday and Thursday, while the yellow arcs indicate one B-757-F flight on Monday, Wednesday, and Friday. The stochastic solution employed 13 distinct hub arcs, while the

deterministic solution used 16. In the stochastic configuration, yellow arcs accounted for 69% of the total hub arcs, whereas in the deterministic case, yellow arcs comprised 50% of the total. Six hub arcs and their associated frequencies were identical in both solutions. However, the stochastic solution included seven additional arc-frequency pairs not present in the deterministic configuration, while the deterministic solution featured ten such pairs not found in the stochastic one.

Overall, Chicago and Atlanta consistently served as major end-nodes for hub arcs across all solutions. These two nodes were followed by Knoxville, which played a similar role in the cases where $|F| = 2$ and $|F| = 5$. However, for $|F| = 3$, Greenville took on a more prominent role as an end-node of a hub arc, surpassing Knoxville in this capacity.

5.4 Summary

The hub arc location problem integrates shipment consolidation and inventory management, in a setting where a freight carrier determines the vehicle frequencies for a heterogeneous fleet operating on a designated set of hub arcs. In this chapter, we introduced demand uncertainty. We formulated the resulting problem using a two-stage stochastic mixed-integer programming model. The first stage of the model focuses on strategic decisions, such as hub arc locations and the corresponding frequencies and types of vehicles assigned to each arc. The second stage addresses operational decisions, including origin-destination flows, inventory holding, and shipment consolidation, based on the realized demands. To solve this complex stochastic formulation, we developed an SAA algorithm.

We conducted an extensive computational study by generating independent random samples, assuming normally distributed demands. For sample sizes of $\mathcal{N} = 15$ and 30 replications ($\mathcal{M} = 30$), our results demonstrated that the proposed SAA algorithm converged. Performance was evaluated using metrics such as the SAA gap and the Value of the Stochastic Solution, both of which underscored the effectiveness of the approach. Additionally, we analyzed specific problem instances to derive managerial insights regarding the impact of uncertainty on decision-making. For instance, we found that the two-stage stochastic model provides measurable benefits over its deterministic (expected value) counterpart. Overall, our study not only highlights the performance of the proposed SAA algorithm, but also offers practical guidance for freight carriers in managing complex, uncertainty-driven logistics problems.

Chapter 6

Conclusion

This chapter summarizes the key findings of the thesis, highlights its contributions, and outlines potential future research directions. In this work, we investigated the integration of shipment consolidation and inventory management within hub location problems using various modeling approaches. Our focus was on designing hub networks that enhance operational efficiency by holding freight inventory at hubs when advantageous, so as to dispatch several smaller shipments together as a combined load. By pooling shipments from different periods and distinct origins, the models leveraged economies of scale. The research addressed both strategic and tactical challenges, particularly examining the role of inventory holding in the design of optimal hub networks. The thesis' main contributions are organized into three core areas, which correspond to Chapters 3, 4, and 5.

Chapter 3 introduced a mixed-integer linear programming formulation aimed at minimizing the total costs associated with hub establishment, inventory holding at hubs, inter-

hub transportation, and transportation costs on allocation links. This hub location model addressed multi-period demand, ensuring that each demand met a promised delivery time at the destination node, while demand nodes were assigned to hubs using a single allocation strategy. The model determined optimal hub locations, allocations of non-hub nodes to hubs, the timing and locations for inventory holding, and the number and types of vehicle dispatches on inter-hub links for each period. To incorporate shipment consolidation into the network design, we considered a heterogeneous fleet for inter-hub transportation. However, numerical experiments showed that the model was computationally intensive. To address this, a VNS (Variable Neighborhood Search)-based solution methodology was proposed to efficiently solve larger instances. The algorithm decomposed the problem into two stages: first, determining hub locations and allocations, followed by solving a shipment consolidation subproblem for each hub pair. The first-stage solutions were iteratively improved by generating neighbors and conducting a search through the VNS algorithm.

The computational experiments revealed important cost savings when shipment consolidation and inventory management were optimized simultaneously. Notably, the results showed a considerable reduction in the number of vehicles required, and by optimizing inventory decisions, the transportation cost structure of the network improved significantly. Additionally, the experiments provided valuable managerial insights, highlighting the sensitivity of cost structures, network configurations, and vehicle utilization to changes in key parameters, such as hub establishment and inventory holding costs. These findings underscore the importance of considering both shipment consolidation and inventory management in the design of efficient hub networks.

To the best of our knowledge, this is the first attempt to study the integration of

shipment consolidation decisions in hub location problems that account for multi-period demand, multi-modal transportation, and a promised time for delivery. The contributions of this chapter include (i) formulation of a hub location and shipment consolidation problem that considers the dynamic nature of demands and a heterogeneous fleet for inter-hub transportation, (ii) formulation of a mixed-integer optimization model for this problem, (iii) development of a solution methodology that incorporates modeling of a new shipment consolidation subproblem, (iv) design of a VNS algorithm that obtains better solutions than a commercial solver, spending only a fraction of the computational time, and (v) furnishing managerial insights via the preparation of a new data set that reflects real-life situations.

Chapter 4 introduced another novel modeling approach called hub arc location, where *hub links*, rather than hub nodes, are strategically placed. We extended this concept to a hub arc location problem that incorporated multi-period demand, multiple allocations of demand nodes to hubs, and inventory holding at the origin node of hub arcs. A mixed-integer programming model was developed to optimize the placement of hub arcs and determine the frequency of vehicles from a heterogeneous fleet, operating on those arcs. The objective was to minimize the total costs associated with hub arc establishment, the respective vehicle frequencies, inventory holding at origin nodes, and transportation on allocation links. Various frequency choices were introduced to study their effect on network design. To solve larger instances efficiently, we proposed an exact solution method based on Benders Decomposition, which was further enhanced through the incorporation of Pareto-optimal cuts and two variable-fixing techniques to reduce computational time.

The computational analysis validated the effectiveness of the proposed approach, demon-

strating that, in many cases, the Benders decomposition-based method outperformed a commercial solver such as CPLEX. Results offered valuable insights into optimal hub arc locations, network design, preferred vehicle frequencies, inventory holding strategies, and flow routing decisions. Furthermore, the analysis revealed that increasing the number of frequency options led to improved shipment consolidation, which in turn reduced the total costs associated with both network establishment and operational efficiency. These findings underscore the significance of integrating frequency and inventory decisions to achieve cost-effective logistics solutions.

The contributions of Chapter 4 are as follows: (i) formulation of a multiple-allocation hub arc location and shipment consolidation problem that accounts for demand distributed over multiple periods within the time horizon, utilizing a heterogeneous fleet for inter-hub transportation; (ii) the development of an MILP model to address this problem; (iii) design of an exact solution methodology, based on the Benders decomposition algorithm; and (iv) enhancement of this algorithm through incorporation of Pareto-optimal cuts and two variable-fixing techniques. To the best of our knowledge, this is the first study to integrate shipment consolidation decisions into hub arc location problems, while simultaneously considering multi-period demand, multiple allocations of end nodes to hub arcs, and multi-modal transportation options in freight transportation.

Chapter 5 built upon the work presented in Chapter 4 by introducing demand randomness to the problem of shipment consolidation for multiple allocation hub arc location. To address this uncertainty, we adopted a two-stage stochastic programming framework. In this approach, the first stage involved decisions regarding hub arc locations and vehicle frequencies, while the second stage handled operational decisions such as inventory hold-

ing and flow routing. To efficiently solve this problem, we employed the Sample Average Approximation method.

Computational results demonstrated that the stochastic model offered significant advantages over its deterministic counterpart, particularly in enhancing cost efficiency under conditions of uncertainty. The SAA algorithm showed effective convergence with ample sizes of 15 and 30 replications. Furthermore, the Value of the Stochastic Solution highlighted the critical role of incorporating demand uncertainty into decision-making processes. The findings underscore the importance of integrating uncertainty into long-term strategic planning, particularly when managing fluctuating demand across multiple periods.

The contributions of Chapter 5 are threefold: (i) formulation of a hub arc location and shipment consolidation problem that incorporates multiple allocations, accounts for stochastic demand distributed over multiple time periods, and considers the use of a diverse fleet for inter-hub transportation, (ii) development of a two-stage stochastic programming model to address this problem, and (iii) the application of an SAA algorithm to solve the model, providing valuable insights through analysis of practical convergence, SAA gaps, and the value of the stochastic solution.

Overall, our investigation into the benefits of freight inventory holdover revealed a significant reduction in total costs. In particular, scenarios where hubs held inventory resulted in substantial savings in transportation expenses. Additionally, in some instances, the inclusion of inventory holding reduced the number of hubs or hub arcs required in the network. Holding freight inventory also decreased the number of flights necessary to meet

customer satisfaction levels. These findings underscore the importance of simultaneously optimizing shipment consolidation and hub location decisions.

While this thesis primarily focused on Less-than-Truckload freight transportation, the models and methodologies developed here are equally applicable to other areas, such as small parcel and postal delivery. We demonstrated that the simultaneous consideration of shipment consolidation and inventory holding in hub location problems can lead to significant operational cost savings for a transportation common carrier. Furthermore, by reducing the total number of vehicles in operation, these approaches contribute to environmentally sustainable logistics practices. Any common carrier can benefit from our work when designing their logistics networks. For cases with low demand variability, the models and methodologies discussed in Chapters 3 and 4 are particularly suitable. For newer or smaller firms, the hub location model and VNS algorithm presented in Chapter 3 are more appropriate. In contrast, larger companies handling high volumes of demand will find the hub arc location model and the Benders decomposition-based methodology from Chapter 4 more relevant, particularly for networks where frequent vehicle operations between hubs are necessary. The stochastic model is most applicable in situations where there is high variability in demand.

There are several potential extensions and promising future research directions in the broader area of shipment consolidation within hub location problems. In this thesis, we have followed the classical assumption that there are no direct shipments between demand nodes, and all demand must be served when determining optimal hub locations. As a result, the dispatch of large vehicles between hubs has been influenced by small origin-destination demands. However, excluding a subset of these small demands from inter-hub

transportation could potentially further optimize the total cost of hub establishment and operations. Two key research directions emerge here. First, we could relax the assumption of no direct shipments between demand nodes, and explore the possibility of outsourcing such direct shipments to a third-party logistics provider at a cost per unit weight per unit distance. Second, we could shift from a cost-minimization to a profit-maximization perspective by relaxing the assumption that all demand must be served. The first approach would necessitate modifications to the existing models, while the second would require entirely new modeling frameworks. Both approaches would require development of novel solution methods.

In Chapter 3, we employed a deterministic demand pattern distributed across half-day periods over a seven-day planning horizon, with the assumption that these demands, on average, would recur and represent future horizons. However, real-world demand variability can significantly influence decisions related to inventory, transportation, and most importantly, the strategy of hub location choices. As a direction for future research, the analysis can be enhanced to incorporate demands spread over a lengthened time scale, utilizing a rolling horizon approach to evaluate the implications of our model for strategic decision-making in the long run.

Another promising avenue for future research in shipment consolidation is the integration of volume discounts in transportation on allocation links. When pickups, deliveries, and last-mile movements are handled by an LTL common carrier, transportation costs could be further reduced by invoking a “bumping clause”. This will depend on the timing and/or frequency of pickups and drop-offs, and involves the possibility of over-deceleration of the shipment weights, to obtain a favorable all-unit quantity discount.

Additionally, the multi-period O-D demand considered in our models is based on data from the Freight Analysis Framework, with the assumption that freight is available at the supply point at the start of the planning horizon. This assumption conflates two factors—demand quantity and its arrival time. In contrast, existing shipment consolidation models often treat these as separate variables, using distinct distributions for ‘orders per day’ and ‘weight per order.’ Adopting similar assumptions in our models would present a challenging and intriguing avenue for future research.

From a hub-location perspective, the model presented in Chapter 5 could be extended by incorporating more advanced uncertainty modeling techniques, such as robust optimization. This approach would enable the exploration of more conservative solutions, offering deeper insights into managing unpredictability in freight logistics. Such an extension could enhance the model’s ability to account for uncertainty and improve decision-making under variable conditions.

Expanding the model to account for environmental considerations, such as carbon emissions, would make valuable contributions to the field of sustainable logistics. In this case, the objective would still be total cost minimization, where the cost of emissions is now included. We may, however, incorporate cost as a constraint in the model by imposing a predetermined budget cap. Finally, all the results generated by our models and algorithms could serve as valuable inputs for addressing other operational challenges, such as aircraft rebalancing, flight scheduling, and determining optimal flight timings.

References

- Adulyasak, Y., Cordeau, J.-F., and Jans, R. (2015). Benders decomposition for production routing under demand uncertainty. *Operations Research*, 63(4):851–867.
- Ahmed, S. and Shapiro, A. (2002). The sample average approximation method for stochastic programs with integer recourse. Technical report, School of Industrial & Systems Engineering, Georgia Institute of Technology, Atlanta GA.
- Albareda-Sambola, M., Fernández, E., Hinojosa, Y., and Puerto, J. (2009). The multi-period incremental service facility location problem. *Computers & Operations Research*, 36(5):1356 – 1375.
- Alnaggar, A., Gzara, F., and Bookbinder, J. H. (2020). Distribution planning with random demand and recourse in a transshipment network. *EURO Journal on Transportation and Logistics*, 9(1):100007.
- Alumur, S. A., Campbell, J. F., Contreras, I., Kara, B. Y., Marianov, V., and O’Kelly, M. E. (2021). Perspectives on modeling hub location problems. *European Journal of Operational Research*, 291(1):1–17.

- Alumur, S. A. and Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research*, 190(1):1–21.
- Alumur, S. A., Kara, B. Y., and Karasan, O. E. (2009). The design of single allocation incomplete hub networks. *Transportation Research Part B: Methodological*, 43(10):936 – 951.
- Alumur, S. A., Kara, B. Y., and Karasan, O. E. (2012a). Multimodal hub location and hub network design. *Omega*, 40(6):927 – 939.
- Alumur, S. A., Nickel, S., Saldanha-da Gama, F., and Segerdin, Y. (2016). Multi-period hub network design problems with modular capacities. *Annals of Operations Research*, 246(1-2):289–312.
- Alumur, S. A., Nickel, S., Saldanha-da Gama, F., and Verter, V. (2012b). Multi-period reverse logistics network design. *European Journal of Operational Research*, 220(1):67–78.
- Auad-Perez, R. and Van Hentenryck, P. (2022). Ridesharing and fleet sizing for on-demand multimodal transit systems. *Transportation Research Part C: Emerging Technologies*, 138:103594.
- Aydin, N. and Murat, A. (2013). A swarm intelligence based sample average approximation algorithm for the capacitated reliable facility location problem. *International Journal of Production Economics*, 145(1):173–183.
- Basciftci, B. and Van Hentenryck, P. (2023). Capturing travel mode adoption in designing on-demand multimodal transit systems. *Transportation Science*, 57(2):351–375.

- Bashiri, M., Rezanezhad, M., Tavakkoli-Moghaddam, R., and Hasanzadeh, H. (2018). Mathematical modeling for a p-mobile hub location problem in a dynamic environment by a genetic algorithm. *Applied Mathematical Modelling*, 54:151–169.
- Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4:238–252.
- Birge, J. R. (1982). The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical programming*, 24:314–325.
- Bookbinder, J. H., Cai, Q., and He, Q.-M. (2011). Shipment consolidation by private carrier: The discrete time and discrete quantity case. *Stochastic Models*, 27(4):664–686.
- Bookbinder, J. H., Elhedhli, S., and Li, Z. (2015). The air-cargo consolidation problem with pivot weight: Models and solution methods. *Computers & Operations Research*, 59:22 – 32.
- Bookbinder, J. H. and Higginson, J. K. (2002). Probabilistic modeling of freight consolidation by private carriage. *Transportation Research Part E: Logistics and Transportation Review*, 38(5):305 – 318.
- Brimberg, J., Salhi, S., Todosijević, R., and Urošević, D. (2023). Variable neighborhood search: The power of change and simplicity. *Computers & Operations Research*, 155:106221.
- Cai, Q., He, Q.-M., and Bookbinder, J. H. (2014). A tree-structured markovian model of the shipment consolidation process. *Stochastic Models*, 30(4):521–553.

- Campbell, J. F. (1990). Locating transportation terminals to serve an expanding demand. *Transportation Research Part B: Methodological*, 24(3):173 – 192.
- Campbell, J. F. (1994a). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research*, 72(2):387–405.
- Campbell, J. F. (1994b). A survey of network hub location. *Studies in Locational Analysis*, 6:31–49.
- Campbell, J. F. (2009). Hub location for time definite transportation. *Computers & Operations Research*, 36(12):3107–3116.
- Campbell, J. F., Ernst, A., and Krishnamoorthy, M. (2002). Hub location problems. In Drezner, Z. and Hamacher, H., editors, *Facility Location: Applications and Theory*. Springer.
- Campbell, J. F., Ernst, A. T., and Krishnamoorthy, M. (2005a). Hub arc location problems: Part II—formulations and optimal algorithms. *Management Science*, 51(10):1556–1571.
- Campbell, J. F., Ernst, A. T., and Krishnamoorthy, M. (2005b). Hub arc location problems: Part I—introduction and results. *Management Science*, 51(10):1540–1555.
- Campbell, J. F. and O’Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Science*, 46(2):153–169.
- Campbell, J. F., Stiehr, G., Ernst, A. T., and Krishnamoorthy, M. (2003). Solving hub arc location problems on a cluster of workstations. *Parallel Computing*, 29(5):555–574.

- Çetiner, S., Sepil, C., and Süral, H. (2010). Hubbing and routing in postal delivery systems. *Annals of Operations Research*, 181(1):109–124.
- Çetinkaya, S. (2005). Coordination of inventory and shipment consolidation decisions: A review of premises, models, and justification. In Geunes, J., Akçali, E., Pardalos, P. M., Romeijn, H. E., and Shen, Z.-J. M., editors, *Applications of Supply Chain Management and E-Commerce Research*, pages 3–51. Springer US, Boston, MA.
- CFTA (2023). Freight Analysis Framework, Center for Transportation Analysis. Available at https://faf.ornl.gov/faf5/dtt_total.aspx (last accessed on 2023/08/01).
- Chadha, S. S., Ülkü, M. A., and Venkatadri, U. (2022). Freight delivery in a physical internet supply chain: an applied optimisation model with peddling and shipment consolidation. *International Journal of Production Research*, 60(16):4995–5011.
- Chen, X. and Fukushima, M. (2005). Expected residual minimization method for stochastic linear complementarity problems. *Mathematics of Operations Research*, 30(4):1022–1038.
- Contreras, I. (2021). Hub network design. In Crainic, T. G., Gendreau, M., and Gendron, B., editors, *Network Design with Applications to Transportation and Logistics*, pages 567–598. Springer International Publishing, Cham.
- Contreras, I., Cordeau, J.-F., and Laporte, G. (2011a). Benders decomposition for large-scale uncapacitated hub location. *Operations Research*, 59(6):1477–1490.
- Contreras, I., Cordeau, J.-F., and Laporte, G. (2011b). The dynamic uncapacitated hub location problem. *Transportation Science*, 45(1):18–32.

- Contreras, I., Cordeau, J.-F., and Laporte, G. (2011c). Stochastic uncapacitated hub location. *European Journal of Operational Research*, 212(3):518–528.
- Contreras, I., Cordeau, J.-F., and Laporte, G. (2012). Exact solution of large-scale hub location problems with multiple capacity levels. *Transportation Science*, 46(4):439–459.
- Contreras, I. and Fernández, E. (2014). Hub location as the minimization of a supermodular set function. *Operations Research*, 62(3):557–570.
- Contreras, I. and O’Kelly, M. (2019). Hub location problems. In Laporte, G., Nickel, S., and Saldanha da Gama, F., editors, *Location Science*, pages 327–363. Springer International Publishing.
- Correia, I., Nickel, S., and da Gama, F. S. (2018). A stochastic multi-period capacitated multiple allocation hub location problem: Formulation and inequalities. *Omega*, 74:122 – 134.
- Cortes, J. D. and Suzuki, Y. (2020). Vehicle routing with shipment consolidation. *International Journal of Production Economics*, 227:107622.
- Crainic, T. G., Rei, W., Hewitt, M., and Maggioni, F. (2016). *Partial Benders decomposition strategies for two-stage stochastic integer programs*, volume 37. CIRRELT.
- CSCMP (2024). 35th Annual State of Logistics Report by Council of Supply Chain Management Professionals. Technical report, CSCMP, Nashville TN.
- Cunha, C. B. and Silva, M. R. (2007). A genetic algorithm for the problem of configuring

- a hub-and-spoke network for a LTL trucking company in Brazil. *European Journal of Operational Research*, 179(3):747 – 758.
- Dalmeijer, K. and Van Hentenryck, P. (2020). Transfer-expanded graphs for on-demand multimodal transit systems. In Hebrard, E. and Musliu, N., editors, *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 167–175, Cham. Springer International Publishing.
- de Camargo, R. and Miranda, G. (2012). Single allocation hub location problem under congestion: Network owner and user perspectives. *Expert Systems with Applications*, 39(3):3385–3391.
- de Camargo, R., Miranda, G., and Luna, H. (2008). Benders decomposition for the uncapacitated multiple allocation hub location problem. *Computers & Operations Research*, 35(4):1047–1064.
- de Camargo, R. S., de Miranda, G., and Luna, H. P. L. (2009). Benders decomposition for hub location problems with economies of scale. *Transportation Science*, 43(1):86–97.
- de Camargo, R. S., de Miranda, G., and Løkketangen, A. (2013). A new formulation and an exact approach for the many-to-many hub location-routing problem. *Applied Mathematical Modelling*, 37(12):7465–7480.
- de Camargo, R. S., de Miranda, G., O’Kelly, M. E., and Campbell, J. F. (2017). Formulations and decomposition methods for the incomplete hub location network design problem with and without hop-constraints. *Applied Mathematical Modelling*, 51:274 – 301.

- de Sá, E. M., de Camargo, R. S., and de Miranda, G. (2013). An improved Benders decomposition algorithm for the tree of hubs location problem. *European Journal of Operational Research*, 226(2):185 – 202.
- Emelogu, A., Chowdhury, S., Marufuzzaman, M., Bian, L., and Eksioglu, B. (2016). An enhanced sample average approximation method for stochastic optimization. *International Journal of Production Economics*, 182:230–252.
- Ernst, A. T. and Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location Science*, 4(3):139 – 154.
- Eydi, A., Khaleghi, A., and Barzegar, K. (2022). Ring hierarchical hub network design problem: Exact and heuristic solution methods. *EURO Journal on Transportation and Logistics*, 11:100096.
- Fan, Y., Behdani, B., Bloemhof-Ruwaard, J., and 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.
- Farahani, R. Z., Hekmatfar, M., Arabani, A. B., and Nikbakhsh, E. (2013). Hub location problems: A review of models, classification, solution techniques, and applications. *Computers & Industrial Engineering*, 64(4):1096 – 1109.
- Fotuhi, F. and Huynh, N. (2018). A reliable multi-period intermodal freight network expansion problem. *Computers & Industrial Engineering*, 115:138–150.
- Gelareh, S. (2008). *Hub location models in public transport planning*. PhD thesis, Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern.

- Gelareh, S., Neamatian Monemi, R., and Nickel, S. (2015). Multi-period hub location problems in transportation. *Transportation Research Part E: Logistics and Transportation Review*, 75:67 – 94.
- Gelareh, S. and Nickel, S. (2011). Hub location problems in transportation networks. *Transportation Research Part E: Logistics and Transportation Review*, 47(6):1092–1111.
- Gelareh, S., Nickel, S., and Pisinger, D. (2010). Liner shipping hub network design in a competitive environment. *Transportation Research Part E: Logistics and Transportation Review*, 46(6):991–1004.
- Goldman, A. J. (1969). Optimal locations for centers in a network. *Transportation Science*, 3(4):352–360.
- Guastaroba, G., Speranza, M. G., and Vigo, D. (2016). Intermediate facilities in freight transportation planning: A survey. *Transportation Science*, 50(3):763–789.
- Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459.
- Hanbazazah, A. S., Abril, L., Erkoc, M., and Shaikh, N. (2019). Freight consolidation with divisible shipments, delivery time windows, and piecewise transportation costs. *European Journal of Operational Research*, 276(1):187 – 201.
- Hanbazazah, A. S., Castro, L. E., Erkoc, M., and Shaikh, N. I. (2020). In-transit freight consolidation of indivisible shipments. *Journal of the Operational Research Society*, 71(1):37–54.

- Hansen, P., Mladenović, N., and Moreno Pérez, J. A. (2010). Variable neighbourhood search: methods and applications. *Annals of Operations Research*, 175:367 – 407.
- Higginson, J. K. and Bookbinder, J. H. (1994). Policy recommendations for a shipment-consolidation program. *Journal of Business Logistics*, 15(1):87–112.
- Higginson, J. K. and Bookbinder, J. H. (1995). Markovian decision processes in shipment consolidation. *Transportation Science*, 29(3):242–255.
- Hoff, A., Peiró, J., Ángel Corberán, and Martí, R. (2017). Heuristics for the capacitated modular hub location problem. *Computers & Operations Research*, 86:94 – 109.
- Horner, M. W. and O’Kelly, M. E. (2001). Embedding economies of scale concepts for hub network design. *Journal of Transport Geography*, 9(4):255 – 265.
- Huang, K., Lee, Y.-T., and Xu, H. (2020). A routing and consolidation decision model for containerized air-land intermodal operations. *Computers & Industrial Engineering*, 141:106299.
- Ilić, A., Urošević, D., Brimberg, J., and Mladenović, N. (2010). A general variable neighborhood search for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research*, 206(2):289 – 300.
- Jara-Díaz, S. R., Cortés, C. E., and Morales, G. A. (2013). Explaining changes and trends in the airline industry: Economies of density, multiproduct scale, and spatial scope. *Transportation Research Part E: Logistics and Transportation Review*, 60:13 – 26.

- Johansson, L., Sonntag, D. R., Marklund, J., and Kiesmüller, G. P. (2020). Controlling distribution inventory systems with shipment consolidation and compound Poisson demand. *European Journal of Operational Research*, 280(1):90 – 101.
- Karimi, H. and Bashiri, M. (2011). Hub covering location problems with different coverage types. *Scientia Iranica*, 18(6):1571 – 1578.
- Karimi, H. and Setak, M. (2018a). A bi-objective incomplete hub location-routing problem with flow shipment scheduling. *Applied Mathematical Modelling*, 57:406 – 431.
- Karimi, H. and Setak, M. (2018b). Flow shipment scheduling in an incomplete hub location-routing network design problem. *Computational and Applied Mathematics*, 37(2):819–851.
- Kartal, Z., Hasgul, S., and Ernst, A. T. (2017). Single allocation p-hub median location and routing problem with simultaneous pick-up and delivery. *Transportation Research Part E: Logistics and Transportation Review*, 108:141 – 159.
- Khaleghi, A. and Eydi, A. (2021). Robust sustainable multi-period hub location considering uncertain time-dependent demand. *RAIRO-Operations Research*, 55(6):3541–3574.
- Khaleghi, A. and Eydi, A. (2022a). Multi-period hub location problem: a review. *RAIRO-Operations Research*, 56(4):2751–2765.
- Khaleghi, A. and Eydi, A. (2022b). Sustainable multi-period hub location under uncertainty. *Soft Computing*, 27(4):2009–2040.

- Khaleghi, A. and Eydi, A. (2023). Multi-period hub location problem considering polynomial time-dependent demand. *Computers & Operations Research*, 159:106357.
- Khaleghi, A. and Eydi, A. (2024). Hybrid solution methods for a continuous-time multi-period hub location problem with time-dependent demand and sustainability considerations. *Journal of Ambient Intelligence and Humanized Computing*, 15(1):115–155.
- Kiani Mavi, R., Goh, M., Kiani Mavi, N., Jie, F., Brown, K., Biermann, S., and A Khanfar, A. (2020). Cross-docking: A systematic literature review. *Sustainability*, 12(11):4789.
- Kim, S., Pasupathy, R., and Henderson, S. G. (2015). A guide to sample average approximation. In Fu, M. C., editor, *Handbook of Simulation Optimization*, pages 207–243. Springer New York, New York, NY.
- Kimms, A. (2006). Economies of scale in hub & spoke network design models: We have it all wrong. In *Perspectives on Operations Research*, pages 293–317. DUV, Wiesbaden.
- Kınay, Ö. B., Gzara, F., and Alumur, S. A. (2021). Full cover charging station location problem with routing. *Transportation Research Part B: Methodological*, 144:1–22.
- Kiya, F. and Davoudpour, H. (2012). Stochastic programming approach to re-designing a warehouse network under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 48(5):919–936. Selected papers from the 14th ATRS and the 12th WCTR Conferences, 2010.
- Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502.

- Krishnan, D. R. and Liu, T. (2019). Freight consolidation problem with time windows, pickup and delivery sequence. In *MATEC Web of Conferences*, volume 271, page 06005. EDP Sciences.
- Leslie, A. and Murray, D. (2022). An Analysis of the Operational Costs of Trucking: 2022 Update, published by American Transportation Research Institute. Technical report, ATRI, Arlington VA.
- Li, Z., Bookbinder, J. H., and Elhedhli, S. (2012). Optimal shipment decisions for an airfreight forwarder: Formulation and solution methods. *Transportation Research Part C: Emerging Technologies*, 21(1):17 – 30.
- Lin, C.-C. and Lin, J. S.-J. (2007). The multistage stochastic integer load planning problem. *Transportation Research Part E: Logistics and Transportation Review*, 43(2):143–156.
- Lin, C.-C. and Wu, Y.-C. (2001). The multiple frequency delivery operations in the time-definite delivery industry. *Journal of the Operational Research Society*, 52(11):1215–1224.
- Linderoth, J. and Wright, S. (2003). Decomposition algorithms for stochastic programming on a computational grid. *Computational Optimization and Applications*, 24:207–250.
- Lopes, M. C., de Andrade, C. E., de Queiroz, T. A., Resende, M. G. C., and Miyazawa, F. K. (2016). Heuristics for a hub location-routing problem. *Networks*, 68(1):54–90.
- Lu, J., Trasatti, A., Guan, H., Dalmeijer, K., and Van Hentenryck, P. (2024). The impact of congestion and dedicated lanes on on-demand multimodal transit systems. *Travel Behaviour and Society*, 36:100772.

- Ma, H., Miao, Z., Lim, A., and Rodrigues, B. (2011). Crossdocking distribution networks with setup cost and time window constraint. *Omega*, 39(1):64 – 72.
- Magnanti, T. L. and Wong, R. T. (1981). Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29(3):464–484.
- Mahéo, A., Kilby, P., and Van Hentenryck, P. (2019). Benders decomposition for the design of a hub and shuttle public transit system. *Transportation Science*, 53(1):77–88.
- Mak, W.-K., Morton, D. P., and Wood, R. (1999). Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56.
- Marasco, A. (2008). Third-party logistics: A literature review. *International Journal of Production Economics*, 113(1):127 – 147.
- Marin, A., Cánovas, L., and Landete, M. (2006). New formulations for the uncapacitated multiple allocation hub location problem. *European Journal of Operational Research*, 172(1):274–292.
- Martínez, J. A. S., Román, D., and Ozuna, L. (2020). Mixed integer programming model for facility location problems: Case study for consolidation centers. In Chlamtac, I., editor, *Mobile Networks and Applications*, volume 25, pages 2118–2125. Springer.
- Martins de Sá, E., Contreras, I., and Cordeau, J.-F. (2015). Exact and heuristic algorithms for the design of hub networks with multiple lines. *European Journal of Operational Research*, 246(1):186 – 198.

- Martins de Sá, E., Morabito, R., and de Camargo, R. S. (2018). Benders decomposition applied to a robust multiple allocation incomplete hub location problem. *Computers & Operations Research*, 89:31–50.
- Marufuzzaman, M. and Eksioglu, S. D. (2014). Developing a reliable and dynamic inter-modal hub and spoke supply chain for biomass. In *IIE Annual Conference. Proceedings*, page 2417. Institute of Industrial and Systems Engineers (IISE).
- Masaeli, M., Alumur, S. A., and Bookbinder, J. H. (2018). Shipment scheduling in hub location problems. *Transportation Research Part B: Methodological*, 115:126–142.
- Meier, J. F. (2017). An improved mixed integer program for single allocation hub location problems with stepwise cost function. *International Transactions in Operational Research*, 24(5):983–991.
- Melo, M., Nickel, S., and Saldanha da Gama, F. (2006). Dynamic multi-commodity capacitated facility location: a mathematical modeling framework for strategic supply chain planning. *Computers & Operations Research*, 33(1):181 – 208.
- Mikić, M., Todosijević, R., and Urošević, D. (2019). Less is more: General variable neighborhood search for the capacitated modular hub location problem. *Computers & Operations Research*, 110:101 – 115.
- Miranda, P. A., Garrido, R. A., and Ceroni, J. A. (2009). e-work based collaborative optimization approach for strategic logistic network design problem. *Computers & Industrial Engineering*, 57(1):3–13.

- Mokhtar, H., Krishnamoorthy, M., and Ernst, A. T. (2018). A modified Benders method for the single- and multiple allocation p-hub median problems. In Klierer, N., Ehmke, J. F., and Borndörfer, R., editors, *Operations Research Proceedings 2017*, pages 135–141, Cham. Springer International Publishing.
- Mokhtar, H., Redi, A. P., Krishnamoorthy, M., and Ernst, A. T. (2019). An intermodal hub location problem for container distribution in indonesia. *Computers & Operations Research*, 104:415 – 432.
- Mutlu, F., Çetinkaya, S., and Bookbinder, J. H. (2010). An analytical model for computing the optimal time-and-quantity-based policy for consolidated shipments. *IIE Transactions*, 42(5):367–377.
- Muñoz-Villamizar, A., Velázquez-Martínez, J. C., Mejía-Argueta, C., and 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.
- Neamatian Monemi, R., Gelareh, S., Nagih, A., Maculan, N., and Danach, K. (2021). Multi-period hub location problem with serial demands: A case study of humanitarian aids distribution in lebanon. *Transportation Research Part E: Logistics and Transportation Review*, 149:102201.
- Newbourne, M. J. (1976). *A guide to freight consolidation for shippers*. Traffic Service Corporation.

- Nguyen, C., Dessouky, M., and Toriello, A. (2014). Consolidation strategies for the delivery of perishable products. *Transportation Research Part E: Logistics and Transportation Review*, 69:108 – 121.
- Nickel, S. and Saldanha-da Gama, F. (2019). Multi-period facility location. In Laporte, G., Nickel, S., and Saldanha da Gama, F., editors, *Location Science*, pages 303–326. Springer International Publishing, Cham.
- O’Kelly, M. (2014). Available at https://www.researchgate.net/publication/269396247_cab100_mok (last accessed on 2023/08/01).
- O’Kelly, M. E. (1986a). Activity levels at hub facilities in interacting networks. *Geographical Analysis*, 18(4):343–356.
- O’Kelly, M. E. (1986b). The location of interacting hub facilities. *Transportation Science*, 20(2):92–106.
- O’Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research*, 32(3):393–404.
- O’Kelly, M. E. (1992). Hub facility location with fixed costs. *Papers in Regional Science*, 71(3):293–306.
- O’Kelly, M. E. and Miller, H. J. (1994). The hub network design problem: A review and synthesis. *Journal of Transport Geography*, 2(1):31–40.
- Oliveira, F. A., de Sá, E. M., and de Souza, S. R. (2022). Benders decomposition applied

- to profit maximizing hub location problem with incomplete hub network. *Computers & Operations Research*, 142:105715.
- O’Kelly, M. E. and Bryan, D. (1998). Hub location with flow economies of scale. *Transportation Research Part B: Methodological*, 32(8):605 – 616.
- Pagnoncelli, B. K., Ahmed, S., and Shapiro, A. (2009). Sample average approximation method for chance constrained programming: theory and applications. *Journal of optimization theory and applications*, 142(2):399–416.
- Pourghader Chobar, A., Adibi, M. A., and Kazemi, A. (2021). A novel multi-objective model for hub location problem considering dynamic demand and environmental issues. *Journal of Industrial Engineering and Management Studies*, 8(1):1–31.
- Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2017). The benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817.
- Rahmati, R., Neghabi, H., Bashiri, M., and Salari, M. (2023). Stochastic regional-based profit-maximizing hub location problem: A sustainable overview. *Omega*, 121:102921.
- Rahmati, R., Neghabi, H., Bashiri, M., and Salari, M. (2024). Stochastic green profit-maximizing hub location problem. *Journal of the Operational Research Society*, 75(1):99–121.
- Rodríguez-Martín, I. and Salazar-González, J. J. (2008). Solving a capacitated hub location problem. *European Journal of Operational Research*, 184(2):468–479.

- Romauch, M. and Hartl, R. F. (2005). Dynamic facility location with stochastic demands. In Lupanov, O. B., Kasim-Zade, O. M., Chaskin, A. V., and Steinhöfel, K., editors, *Stochastic Algorithms: Foundations and Applications*, pages 180–189, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Romero-Silva, R. and Mujica Mota, M. (2022). Trade-offs in the landside operations of air cargo hubs: Horizontal cooperation and shipment consolidation policies considering capacitated nodes. *Journal of Air Transport Management*, 103:102253.
- Rostami, M. and Jabbarzadeh, A. (2021). Integrated hub location and flow processing schedule problem under renewable capacity constraint. *Computational and Applied Mathematics*, 40(5):1–18.
- Rothenbächer, A.-K., Drexl, M., and Irnich, S. (2016). Branch-and-price-and-cut for a service network design and hub location problem. *European Journal of Operational Research*, 255(3):935–947.
- Rouberol, B. (2022). Haversine 2.7.0. Available at <https://pypi.org/project/haversine/> (last accessed on 2023/08/01).
- Santoso, T., Ahmed, S., Goetschalckx, M., and Shapiro, A. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167(1):96–115.
- Sasaki, M., Campbell, J. F., Ernst, A. T., and Krishnamoorthy, M. (2009). Hub arc location with competition. In *Technical Report of the Nanzan Academic Society - Information*

- Sciences and Engineering (NANZAN-TR-2009-02)*, page 1–41. Nanzan University, Seto, Japan.
- Sasaki, M. and Fukushima, M. (2001). Stackelberg hub location problem. *Journal of the Operations Research Society of Japan*, 44(4):390–402.
- Satır, B., Erenay, F. S., and Bookbinder, J. H. (2018). Shipment consolidation with two demand classes: Rationing the dispatch capacity. *European Journal of Operational Research*, 270(1):171–184.
- Schütz, P., Tomasgard, A., and Ahmed, S. (2009). Supply chain design under uncertainty using sample average approximation and dual decomposition. *European Journal of Operational Research*, 199(2):409–419.
- Şen, H., Detienne, B., Sadykov, R., and Vanderbeck, F. (2016). A review of algorithmic enhancements for benders decomposition. In *ISCO 2016-4th International Symposium on Combinatorial Optimization*.
- Serper, E. Z. and Alumur, S. A. (2016). The design of capacitated intermodal hub networks with different vehicle types. *Transportation Research Part B: Methodological*, 86:51 – 65.
- Shapiro, A. and Homem-de Mello, T. (1998). A simulation-based approach to two-stage stochastic programming with recourse. *Mathematical Programming*, 81(3):301–325.
- Silver, E. A., Pyke, D. F., and Thomas, D. J. (2017). *Inventory and production management in supply chains*. CRC Press, 4th edition.

- Taghipourian, F., Mahdavi, I., Mahdavi-Amiri, N., and Makui, A. (2012). A fuzzy programming approach for dynamic virtual hub location problem. *Applied Mathematical Modelling*, 36(7):3257–3270.
- Taherkhani, G. and Alumur, S. A. (2019). Profit maximizing hub location problems. *Omega*, 86:1 – 15.
- Taherkhani, G., Alumur, S. A., and Hosseini, M. (2020). Benders decomposition for the profit maximizing capacitated hub location problem with multiple demand classes. *Transportation Science*, 54(6):1446–1470.
- Taherkhani, G., Alumur, S. A., and Hosseini, M. (2021). Robust stochastic models for profit-maximizing hub location problems. *Transportation Science*, 55(6):1322–1350.
- Tan, P. Z. and Kara, B. Y. (2007). A hub covering model for cargo delivery systems. *Networks*, 49(1):28–39.
- Tanash, M., Contreras, I., and Vidyarthi, N. (2017). An exact algorithm for the modular hub location problem with single assignments. *Computers & Operations Research*, 85:32 – 44.
- Torkestani, S. S., Seyedhosseini, S. M., Makui, A., and Shahanaghi, K. (2018). The reliable design of a hierarchical multi-modes transportation hub location problems (HMMTHLP) under dynamic network disruption (DND). *Computers & Industrial Engineering*, 122:39–86.
- Tyan, J. C., Wang, F.-K., and Du, T. C. (2003). An evaluation of freight consolidation policies in global third party logistics. *Omega*, 31(1):55 – 62.

- Verweij, B., Ahmed, S., Kleywegt, A. J., Nemhauser, G., and Shapiro, A. (2003). The sample average approximation method applied to stochastic routing problems: a computational study. *Computational Optimization and Applications*, 24:289–333.
- Vidović, M., Zečević, S., Kilibarda, M., Vlajić, J., Bjelić, N., and Tadić, S. (2011). The p-hub model with hub-catchment areas, existing hubs, and simulation: A case study of Serbian intermodal terminals. *Networks and Spatial Economics*, 11(2):295–314.
- Wandelt, S., Dai, W., Zhang, J., and Sun, X. (2022). Toward a reference experimental benchmark for solving hub location problems. *Transportation Science*, 56(2):543–564.
- Wei, L., Jasin, S., and Kapuscinski, R. (2017). Shipping consolidation with delivery deadline and expedited shipment options. *Ross School of Business Paper*, 1375.
- Wei, L., Kapuscinski, R., and Jasin, S. (2021). Shipping consolidation across two warehouses with delivery deadline and expedited options for e-commerce and omni-channel retailers. *Manufacturing & Service Operations Management*, 23(6):1634–1650.
- Wu, H., Herszterg, I., Savelsbergh, M., and Huang, Y. (2023). Service network design for same-day delivery with hub capacity constraints. *Transportation Science*, 57(1):173–287.
- Wu, T. (2024). Exact method for production hub location. *INFORMS Journal on Computing*, 36(5):1287–1315.
- Yaman, H. and Carello, G. (2005). Solving the hub location problem with modular link capacities. *Computers & Operations Research*, 32(12):3227–3245.

- Yaman, H., Karasan, O. E., and Kara, B. Y. (2012). Release time scheduling and hub location for next-day delivery. *Operations Research*, 60(4):906–917.
- Zhang, C., Xie, F., Huang, K., Wu, T., and Liang, Z. (2017). MIP models and a hybrid method for the capacitated air-cargo network planning and scheduling problems. *Transportation Research Part E: Logistics and Transportation Review*, 103:158–173.
- Zhang, H., Yang, K., Gao, Y., and Yang, L. (2022). Accelerating benders decomposition for stochastic incomplete multimodal hub location problem in many-to-many transportation and distribution systems. *International Journal of Production Economics*, 248:108493.
- Zhang, Y., Lin, W.-H., Huang, M., and Hu, X. (2021). Multi-warehouse package consolidation for split orders in online retailing. *European Journal of Operational Research*, 289(3):1040–1055.
- Zhang, Y., Sun, L., Hu, X., and Zhao, C. (2019). Order consolidation for the last-mile split delivery in online retailing. *Transportation Research Part E: Logistics and Transportation Review*, 122:309 – 327.

APPENDICES

Appendix A

List of Cities in the Data Sets

Table A.1: Cities in different instances

Instance Size	Cities
10	Birmingham, Atlanta, Chicago, Grand Rapids, Greenville, Knoxville, Austin, Beaumont, Dallas, Milwaukee
15	Birmingham, Hartford, Atlanta, Chicago, Fort Wayne, Grand Rapids, Albany, New York, Philadelphia, Greenville, Knoxville, Austin, Beaumont, Dallas, Milwaukee
20	Birmingham, Hartford, Jacksonville, Miami, Orlando, Tampa, Atlanta, Chicago, Fort Wayne, Grand Rapids, Albany, New York, Charlotte, Philadelphia, Greenville, Knoxville, Austin, Beaumont, Dallas, Milwaukee
30	Birmingham, Los Angeles, Sacramento, San Diego, San Francisco, Fresno, Hartford, Jacksonville, Miami, Orlando, Tampa, Atlanta, Chicago, Fort Wayne, Grand Rapids, Albany, New York, Charlotte, Portland, Philadelphia, Greenville, Knoxville, Austin, Beaumont, Corpus Christi, Dallas, Houston, San Antonio, Seattle, Milwaukee
45	Birmingham, Los Angeles, Sacramento, San Diego, San Francisco, Fresno, Hartford, Washington, Jacksonville, Miami, Orlando, Tampa, Atlanta, Chicago, Indianapolis, Fort Wayne, Louisville, Baltimore, Detroit, Grand Rapids, St. Louis, Albany, Buffalo, New York, Rochester, Charlotte, Greensboro, Raleigh, Cincinnati, Cleveland, Portland, Philadelphia, Pittsburgh, Greenville, Memphis, Knoxville, Austin, Beaumont, Corpus Christi, Dallas, Houston, San Antonio, Richmond, Seattle, Milwaukee
65	Birmingham, Mobile, Phoenix, Tucson, Los Angeles, Sacramento, San Diego, San Francisco, Fresno, Denver, Hartford, Washington, Jacksonville, Miami, Orlando, Tampa, Atlanta, Chicago, Indianapolis, Fort Wayne, Wichita, Louisville, Baton Rouge, New Orleans, Baltimore, Boston, Detroit, Grand Rapids, Minneapolis, Kansas City, St. Louis, Omaha, Las Vegas, Albany, Buffalo, New York, Rochester, Charlotte, Greensboro, Raleigh, Cincinnati, Cleveland, Columbus, Dayton, Oklahoma City, Tulsa, Portland, Philadelphia, Pittsburgh, Charleston, Greenville, Memphis, Nashville, Knoxville, Austin, Beaumont, Corpus Christi, Dallas, El Paso, Houston, San Antonio, Salt Lake City, Richmond, Seattle, Milwaukee