# Modeling Large Scale e-Commerce Distribution Networks

by

Nelson Calero

B.S. in Industrial Engineering, Lehigh University (2013)

and

Yao Zhang

B.E. in Computer Science, Xiamen University (2008)

B.S. in Economics, Xiamen University (2008)

M.S. in Industrial Engineering, Georgia Institute of Technology (2009)

M.S. in Logistics and Supply Chain Management, National University of Singapore (2010)

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

JUNE 2019



Signature of Author:________________________________________

Department of Supply Chain Management
May 10, 2019

Signature of Author: ________________________________________

Department of Supply Chain Management
May 10, 2019

Certified by: ________________________________________

Milena Janjevic
Capstone Advisor
Postdoctoral Associate, MIT Megacity Logistics Lab

Accepted by:________________________________________

Dr. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

**Modeling Large Scale e-Commerce Distribution Networks**

by

Nelson Calero

and

Yao Zhang

Submitted to the Program in Supply Chain Management
on May 10, 2019 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

## ABSTRACT

As urbanization and e-commerce continue to grow at a rapid rate, companies are rethinking the way they operate their final-mile distribution networks to prepare for the scaling of their costs. How can companies leverage their existing e-commerce distribution networks to fulfill demand across multiple service offerings, and should they jointly consider inventory control decisions in those design decisions? We explore this question by examining where companies locate inventory, how they use their existing facilities, and how both are used to satisfy demand. This research uses the augmented routing cost estimation (ARCE) formula alongside mixed-integer linear programming to determine the most cost-optimal network design for an e-commerce retailer. Overall, we determined that inventory control decisions have a significant impact on e-commerce distribution networks and should be jointly considered when considering network design. Alongside the solution, the managerial insights derived from our findings support multiple strategies and offer considerations in implementing a redesigned distribution network, especially if a company is limited in their ability to reorder inventory or have limitations in their transhipment operations.

Capstone Advisor: Milena Janjevic
Title: Postdoctoral Associate, MIT Megacity Logistics Lab

## ACKNOWLEDGMENTS

There are many people we wish to thank: there are those who helped make this project possible in a literal sense, but there are also all of those who have guided and supported us on the long and arduous path to completion.

We would like to extend our sincere thanks and gratitude to Dr. Milena Janjevic for her guidance and expertise on this project. Thank you for challenging us to do our absolute best, and for your valuable input in helping shape this Capstone project.

We also thank the MIT Megacity Logistics Lab for the support they have given us on our project. We are proud to have brought your project to completion, and appreciate the support you have given us along the way.

We would also like to thank our friends and family for their unconditional support. It is not easy for two industry professionals to uproot their lives and pursue Master's degrees, but your faith helped us persevere. Thank you for sharing us with the MIT community.

We also extend our thanks to the entire MIT Supply Chain Management Master's Program. We are blessed to be surrounded by the greatest minds in the world, and are grateful to have shared this experience together. Thank you for always lending us your ears, and we hope we were able to make you feel just as supported.

And lastly, we would like to thank the faculty and staff of MIT's Center for Transportation and Logistics. The groundwork you laid out has allowed our ideas to flourish, and we thank you for always challenging us to be the best we can be.

## TABLE OF CONTENTS

**LIST OF FIGURES**

## LIST OF TABLES

# 1. INTRODUCTION

As online shopping penetration increases and continues to encroach on traditional offline retailers' markets globally, both e-commerce companies and traditional retailers compete fiercely with each other to retain or gain market share (Forrester, 2014). Online retailers must excel in all dimensions - product range, price competitiveness, cost efficiency, delivery speed, inventory control, etc. - to effectively convert shoppers from the traditional retailers and retain their business. Many companies have identified last-mile delivery services as a key differentiator as online shoppers increasingly demand time-sensitive delivery service options. Shoppers also value the perceived quality of the delivery services, which has become a major decision-making criteria for choosing where to shop (Joerss, Neuhaus, & Schröder, 2016).

Unlike traditional retail distribution, e-commerce distribution is characterized by a high number of SKUs and product categories, a large number of smaller shipments, and more and unpredictable geographically-dispersed delivery points. e-Commerce distribution networks are generally composed of multi-tier structures of logistics facilities that are interconnected through various transportation services. These logistics facilities include both "traditional" and "non-traditional" facilities such as central distribution hubs, e-fulfillment centers, transshipment hubs, local transshipment centers, etc. All together, they enable the extensive storage and fast distribution of thousands of SKUs within the supply chain network. By offering different delivery service types to end customers, online retailers reduce delivery times and enable immediate product access. This acts as a catalyst for growth in online sales (Hausmann, Herrmann, Krause, & Netzer, 2014).

For the global e-commerce parcel delivery market, same-day and instant delivery are expected to reach a combined share of 20% to 25% by 2025 (Joerss et al., 2016), and companies around the world are already adopting multi-tiered distribution networks to create unparallelled delivery experiences for customers in urban areas. In the United States, Amazon started free same-day delivery (order before noon and get the items by 9 p.m.) in 14 U.S. metropolitan areas on over a million items in May 2015 (Alba, 2015). It has also piloted “Prime Now” that provides an “ultra-fast” one-hour delivery service and is expanding its capability rapidly through nationwide Whole Foods outlets (Redman, 2018). In China, JD.com’s same day delivery service has already covered more than 25 metropolitan areas in China, and is expanding the same offering in Southeast Asian cities such as Bangkok. Both Alibaba and JD.com are competing head-to-head on 30-minute grocery deliveries through their Hema Supermarkets and 7 Fresh Stores in almost all tier-1 and 2 Chinese cities (Najberge, 2017; Liao, 2018).

Such recent development has made e-commerce companies very dependent on efficient and flexible supply chain distribution networks that enable large product ranges, high inventory availability, fast deliveries, and low delivery cost. Many have started to refine their last-mile strategies to redesign multi-echelon distribution networks with satellite facilities or existing retail outlets in close proximity to customers. Furthermore, heterogeneous fleets with new and diverse vehicle types (such as cargo-cycles, electric vehicles, mobile warehouses, and autonomous or semi-autonomous vehicles) are becoming increasingly popular to accommodate local service or policy requirement. These non-traditional methods have provided more optimal strategies to the demands they are faced with, but they do not come without their own set of challenges in determining optimal solutions.

As a consequence of the heavy reliance on supply chain distribution networks, e-commerce companies need to simultaneously consider an increasing number of decisions such as location planning, inventory control, and transportation. In designing these networks, the proposed model needs to be flexible enough to allow the exploration of as many variables and scenarios as possible, but not to the point that the model becomes too complex and intractable to solve.

In this project, our goal is to establish a quantitative model that supports the strategic e-commerce distribution network design by incorporating inventory control decisions into existing location-routing models. We consider the following to be our primary contributions:

1. We will formulate an integrated mixed integer programming (MIP) model considering location, routing, and inventory decisions;
2. The model will be built upon location-routing models with a multi-echelon setting by incorporating inventory control decisions;
3. We will investigate the impacts of these decisions on the overall cost of operations and highlight potential tradeoffs;
4. The results of this project will help the e-commerce industry to design efficient last-mile e-commerce distribution networks and be a basis for modeling and analysis of more advanced systems.

## 2. LITERATURE REVIEW

When designing supply chain distribution networks, three important decisions must be considered. Facilities location decisions determine the optimal number, type and locations of the facilities; inventory management decisions define inventory allocation and movement between facilities; and routing decisions deal with the size and composition of the required vehicle fleet and how vehicles are optimally routed to deliver products to the end customers.

These three decisions are highly interrelated and should be integrated in building a decision support system to avoid sub-optimal setup. Therefore, a model that focuses on only one or two of these factors might be insufficient for designing optimal distribution systems.

In this section on literature review, we start by reviewing the e-commerce last mile distribution strategies developed by leading e-commerce players around the world (Section 2.1), and then review the previous literature on network design models, and especially the location routing problems (Section 2.2). Next, in Section 2.3, we dive into the application of continuum approximation method in multi-echelon network design models. Lastly, we examine the previous literature on incorporating inventory control decisions in network design in Section 2.4. We conclude the literature review by identifying the literature gaps in Section 2.5.

### 2.1. e-Commerce Distribution Strategies for Last Mile Distribution

To begin tying together our model, we must first examine existing strategies employed in e-commerce supply chains for last mile distribution. Agatz, Fleischmann, and van Nunen (2008) provided a high level but comprehensive overview of e-commerce fulfillment models and reviewed supply chain management issues specific to multi-channel fulfillment. They highlighted the emerging trends of introducing 'bricks-and-clicks' to integrate a portfolio of multiple alternative distribution channels, but lacks optimization models that explicitly consider the important trade-offs in a multichannel setting. Through extensive literature review, Ghezzi, Mangiaracina, and Perego (2012) identified the logistics problems caused by product and service-driven complexities; and logistics strategies such as inventory ownership and location, order picking policy, order assembly policy, and order delivery policy. By analyzing 28 case studies of leading B2C e-commerce companies in Italy, they were able to observe a strong relationship between the features of the logistics problem and the logistics strategies adopted. For the more recent adoption of omnichannel retailing, Lim, Rabinovich, Rogers, and Lasester (2016) expanded the linearly "chain-centric" extended supply chain models and developed a new last-mile supply network typology; comprising four ideal configurations: Simple, Hyperlocal, One-Stop, and Protean. The four configuration types are described based on network structure, network flow, relationship governance, and service architecture.

Through analysis of 31 case studies and literature reviews, Winkenbach and Janjevic (2018) comprehensively explored five basic variables that characterize last-mile delivery models for e-commerce distribution, including order lead time, place of order preparation, route organization, intermediate transhipment, and product exchange point. From these five characteristic variables, they generated 10 archetypes that generally represent the models employed by e-commerce companies around the world. FIgure 1 shows the 10 archetypes and the variables that characterize each archetype.

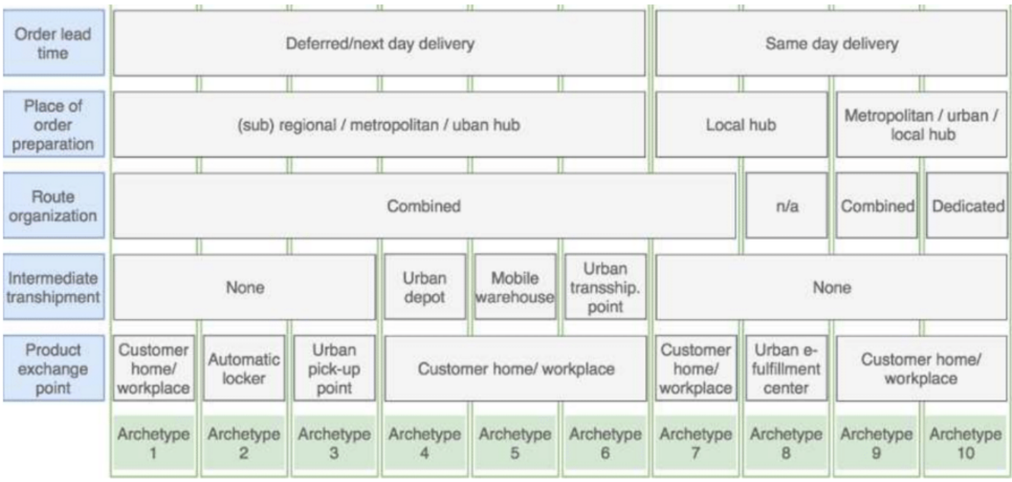


Figure 1: Classification of Delivery Model Archetypes across the Key Indicator Variables (Winkenbach & Janjevic, 2018)

In reviewing the previous research, we draw the below inferences. Generally speaking, to meet customer demand of deferred, next day, and same day deliveries, e-retailers have adopted solutions that minimize the distance between place of order preparation and product exchange point. This can be done by having facilities closer to Points of Demand (POD) or moving the product exchange point to a preset pick-up point. Based on the deployed archetypes, areas with a high demand density and low urban accessibility may require the use of intermediate transhipment points to achieve an optimal solution in our created model. This will have an impact on the Location-Routing Problem (LRP) as well as the Inventory Control Decisions in the model which will be discussed in depth in later sections.

### 2.2. Network Design Models and the Location-Routing Problem (LRP)

Designing distribution networks requires a combination of two hard combinatorial optimization problems. Facility Location Problem (FLP) considers which facilities, out of a finite or infinite set of potential ones, should be opened to minimize the total transportation distance to the sites. Multi-facility Vehicle Routing Problems (VRP) determine which vehicle routes should be built from which facilities in order to serve a set of given demand points. Both FP and VRP are NP-hard (Garey & Johnson, 1979). However, in settings where location and routing decisions are intertwined, making these two types of decisions independently of one another may result in highly suboptimal planning results (Salhi & Rand, 1989).

Location-Routing Problem (LRP) aims to bring the two problems together. Existing literature provides many different variations of LRPs and heuristics. In the following, we provide a overview of the more recent literature.

Lin and Lei (2009) formulated a strategic design model for three-echelon distribution systems with two-level routing considerations. They introduced a hybrid genetic algorithm embedded with a routing heuristic to find near-optimal solutions. The heuristic solution is proven to be within 1% of optimal for small problem set with one plant, 2 potential DC's, 7 big clients and 28 small clients.

Boccia, Sforza, and Sterle (2010) designed a bottom-up approach to solve a LRP model for a two-echelon freight distribution system. They decomposed the problem into two main location-routing problems for the depot-to-satellite transfer and satellite-to-customer delivery. Each is then further decomposed into a capacitated facility location problem (CFLP) and a multi-depot vehicle routing problem (MDVRP). They then applied the Tabu search heuristic on the sub problems and then iteratively built the overall solution combining the sub problem solutions.

Similarly, Crainic, Mancini, Perboli and Tadei (2011) also separated the first and second-level routing sub problems and solved them sequentially. They then applied Multi-Start heuristic to iteratively find better solutions through different diversification strategies and feasibility search rules. The new method was demonstrated to perform efficiently and accurately in the dataset of 1 CDC, 4 satellites and 50 customers.

LRP is a combination of two NP-hard problems, and therefore become challenging to scale in larger environments with larger network sizes. Existing literature has focused on two-staged metaheuristic approaches work well for small and medium-sized instances of up to 5 central facilities, 20 satellites, and 200 customers (Schneider & Drexl, 2017). However, large scale real-world problems, especially those faced by e-commerce companies, normally involve hundreds or even thousands of customers per square kilometer. The sheer size of the real-world problems render these metaheuristic methods computationally infeasible.

In addition, LRP in e-commerce distributions involves additional constraints depending on the local demand density and urban accessibility (Winkenbach & Janjevic, 2018). The topics of demand density and urban accessibility are discussed in Bektaş et al. (2017)'s study on city logistics networks. The main concepts related to LRP discussed in this paper are systems, planning, and business models. In reviewing the systems concept in City Logistics, key items in relation to the LRP are as follows (Winkenbach, Kleindorfer, & Spinler, 2016; Bektaş et al., 2017; Winkenbach & Janjevic, 2018; ).

1. Facilities: City Distribution Centers (CDC) are primarily concerned with the receiving of inbound parcels for operations such as storage, storing, and consolidation for the efficient delivery of shipments within cities. Satellites (also known as Intermediate Depots or IDs) act as a multi-purpose layer between CDC's and a Point-of-Demand (POD) that can act as either a transshipment point or a distribution center. Examples of these satellite locations can be bus stops, parking lots, retail and even small office spaces, etc.
2. Layout: Describes the number of tiers/echelons within the system considered in the LRP. These can either be a single-tier layout or a multi-tier layout. Single-tier layouts

have shipments move directly from a CDC to a POD which is well-suited for lower demand density, higher urban accessibility cities. Multi-tier layouts have shipments moving between multiple echelons, but typically a CDC to Satellite to POD layout is seen. Multi-tier layouts are more well-suited for higher demand density, lower urban accessibility cities.

3. Transport: Describes the vehicle selection in moving the shipments between the CDC, Satellites, and POD. Most City Logistics projects have relied on motorized vehicles such as varying size trucks, but with low urban accessibility studies have found that the use of bicycles and pedestrians have led to more optimal results for the overall network design.

### 2.3. Multi-Echelon Network Design Models with Continuum Approximation

Due to LRPs being NP-hard, many have explored the use of route length estimation (RLE) formulas to develop a closed-form expression to derive routing costs as a function of demand characteristics and geographical properties of demand areas, in order to speed up the routing part of the heuristic (Prodhon & Prins, 2014).

In order to solve a two-echelon LRP formulation for Swiss Post, Bruns, and Klose Stähly (2000) employed a route cost function to estimate the routing cost of delivering parcels from an intermediate depot *j* to a customer zone *i*. The route cost function estimates the share of customer zone *i* on the travel time and distance of a delivery tour which starts from the intermediate depot *j* and serves the customer zone *i*. Travel time and distance include these required for line haul, intra-stop, and service time for each stop.

Winkenbach et al. (2016) expanded previous work introduced by Smilowitz & Daganzo (2007) and introduced an augmented routing cost estimation (ARCE) formula that accounts for multiple vehicle options; mixed fleets; destination-specific vehicle capacities; a global maximum allowed service constraint; and joint pickup and delivery along the same routes. The ARCE formula is conceptualized in the city logistics context as below and is illustrated in Figure 2 with the below context.

- A facility location *d* needs to serve a city segment *i*, which has many Points-of-Delivery (PODs), and each POD has multiple deliveries and / or pickups;
- All PODs in the city segment *i* will be served by the facility location *d*, assuming unlimited number of vehicles and manpower, and unlimited capacity and inventory in the facility location *d*;
- Multiple vehicles run multiple tours within a day from the facility location *d* serving the PODs in the city segment *i*. For each vehicle tour, routing cost is calculated based on all the variable and fixed cost spent on the tour. Examples of such costs are a one-time vehicle setup, vehicle loading and unloading in the facility location, vehicle fixed cost, line-haul driving to the center of the city segment, driving and service time between all the PODs.

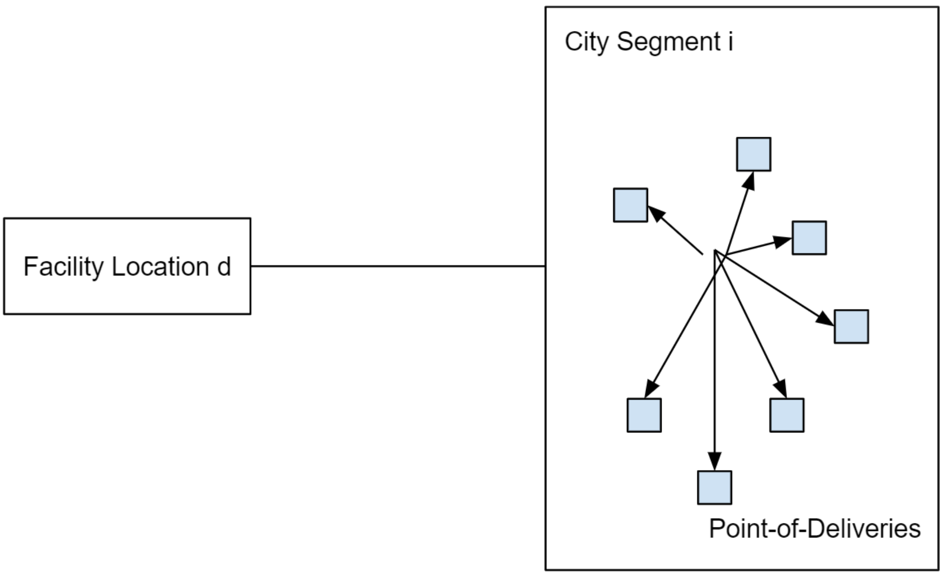


Figure 2: Augmented Routing Cost Estimation (ARCE) Formula Application Context

Quality of the above approximation was benchmarked against the TS meta-heuristic (Côté & Potvin, 2009), which has proven most effective in the past. Tested on a broad range of parameter settings, the routing cost obtained from the ARCE formula deviates by less than 10% from the results of the TS meta-heuristic. Built upon the ARCE formula, Winkenbach et al. (2016) also constructed a mixed-integer linear programming formulation for the capacitated two-echelon LRP (2E-CLRP) that can be solved within a reasonable time to produce high-quality approximations for large-scale real-world problems.

Merchan and Winkenbach (2018) extended the model of Winkenbach et al. (2016) and presented a data-driven method to incorporate road network features (i.e. road network obstacles and complications) by calculating zone-specific circuity factors for each city segment. They demonstrated the importance of deriving fine-grained circuity factors to increase the resolution of road-network data and to better calibrate the closed-form tour distance approximation.

Using demand uncertainty in the strategic network design decisions; Snoeck, Winkenbach, and Mascarino (2018) presented a large-scale stochastic MILP model for a two-echelon capacitated LRP based on the ARCE formulation for routing cost. The effect of including stochasticity was then evaluated on the resulting network design of facility location and fleet composition; the expected cost and level of service of the network; and the downside risk. The model was applied to the planning and delivery operations for Coca-Cola FEMSA in Bogota and resulted in more robust resulting network with lower expected operational cost and reduced downside risk.

### 2.4. Integrated Network Design Models with Inventory Control Decisions

The rationale for incorporating inventory control decisions (i.e. how much and how frequently to reorder and what level of safety stock to maintain,) into the network design model is that they are interrelated to facility locations and vehicle routing decisions. However, most of the research on location-routing problems deals only with facility locations and routing decisions.

Few research addresses inventory decisions simultaneously (Prodhon & Prins, 2014). Such problems are called inventory location-routing problems (ILRP).

Liu and Lee (2003) proposed a mathematical model for a single-echelon, single-product multi-depot location-routing problem that take inventory control decision decisions, such as order quantity and order-up-to level for replenishment, into consideration. They designed a two-phase heuristic method that consists of deriving an initial solution with a route-first locate-allocate-second heuristic and subsequent optimization through local search. The model assumes a common probability density function for customer demand during lead-time, and adopted an economic order quantity (EOQ) style formula to balance the ordering, holding, and shortage costs for an order-up-to-level replenishment policy by customers. Numerical experiments were conducted and the proposed heuristic method demonstrated lower overall system cost than two LRP metaheuristics that do not take inventory into consideration. Liu and Lin (2004) improved the same model by introducing global search heuristic methods such as tabu search and simulated annealing.

For a two-echelon distribution network that consists of distribution center, wholesalers, and retailers; Ma and Davidrajuh (2005) created an iterative approach where a strategic model (number and location of facilities) is determined first using mixed integer programming, and then the tactical model (inventory planning and routing) is worked out using genetic algorithm and probability theory based on the strategic model. The tactical model's parameters are then input back to the strategic model to re-optimize the distribution network. Such iterations continue until the design results converge. In the tactical model, continuous review (r, Q) model, i.e. reorder a fixed quantity Q when inventory level falls below reorder point r, is used by the retailers, wholesalers, and the distribution centers respectively as the inventory control policy. Fixed order quantity (Q) is calculated using EOQ while reorder level is determined based on a safety factor to achieve a certain service level, assuming normal distributed demand. With the given r and Q, the inventory holding cost considering both carrying cost and shortage cost is then calculated through a simulation of a regenerative process.

Shen and Qi (2007) formulated a nonlinear mixed integer programming model to simultaneously optimize the facility opening cost, transportation cost, and the expected inventory ordering and holding cost for a single-period, two-echelon stochastic supply chain with one supplier, DC's, and customers. Each customer has normally distributed demand, and orders at the beginning of the period. Each open, uncapacitated DC combines the orders of the customers it serves and orders from the supplier using a continuous review (r,Q) policy and maintains a safety stock for a given service level. The objective function is the sum of the inventory-location cost and continuous approximation-based routing cost. The problem is solved by Lagrangian relaxation and is tested with 40-320 customers.

Ahmadi-Javid and Azad (2010) extended Shen and Qi (2007) and presented a mixed integer convex model that simultaneously optimizes facility location, inventory and routing decisions without any approximation. They also devised a two-stage metaheuristic hybrid method with Tabu Search and Simulated Annealing to solve larger instances with 40-400 customers and up to 50 DC's. Similar to Shen and Qi (2007), facilities adopt (r,Q) inventory policy and use expected value and standard deviation of the aggregate demand to determine the reorder point and fixed ordering quantity. The results showed that the proposed heuristic is considerably efficient compared to the model in Shen and Qi (2007) with up to 27% of improvement for different problem sizes.

### 2.5. Literature Review Summary

e-Commerce companies are increasingly adopting multi-echelon distribution networks, and use heterogeneous fleets to serve hundreds of thousands of customers in major cities. To effectively optimize such network, we need an integrated supply chain network model that simultaneously considers facility location, routing, and inventory control decisions. This model must also be able handle real-world large scale multi-echelon problems, which we identify as a research gap based on the literature review. Table 1 provides a summary of the literature review and there is a clear gap in taking inventory control decisions in a multi-echelon LRP with continuum approximation.

Table 1: Related Literature Review Summary

| | Large Scale | Multi Echelon | LRP with Explicit Routing | LRP with Continuum Approximation | Inventory Control Decisions |
|---|---|---|---|---|---|
| Liu and Lee (2003) | | | ✓ | | ✓ |
| Lin and Lei (2009) | | ✓ | ✓ | | |
| Boccia et al. (2011) | | ✓ | ✓ | | |
| Croinic et al. (2011) | | ✓ | ✓ | | |
| Ma and Davidrajuh (2005) | | ✓ | ✓ | | ✓ |
| Shen and Qi (2007) | | ✓ | ✓ | ✓ | ✓ |
| Ahmadi-Javid and Azad (2010) | | ✓ | ✓ | | ✓ |
| Winkenbach et al. (2016) | ✓ | ✓ | | ✓ | |
| Merchan and Winkenbach (2018) | ✓ | ✓ | | ✓ | |
| Snoeck et al. (2018) | ✓ | ✓ | | ✓ | |

This project fulfills the gap by developing a multi-echelon capacitated location-routing model containing deterministic demand, diverse facility types and multiple delivery service types and incorporating inventory control decisions. We introduce a mixed-integer linear programming model with continuum approximation and inventory control decisions. To enable large-scale application, we adopt the ARCE formulation for approximating the optimal routing cost under various constraints such as vehicle's maximum service time constraint, vehicle capacity restrictions, etc, for different delivery service types. For inventory control decisions, we extend the existing continuum approximation-based LRP model by evaluating the satellite locations' reordering frequency, which is key to the tradeoff between inventory holding cost and replenishment cost.

## 3. DATA AND METHODOLOGY

In this chapter, we first introduce the distribution network and the network elements that we are solving through this research and define relevant notations and assumptions. Next, we describe the optimization model formulation and elaborate on the routing cost calculation for different service types. Finally, we propose an iterative approach to support the strategic design of urban last-mile distribution networks.

### 3.1. Description of Distribution Network

In the distribution network we consider, there are two levels of facilities and multiple delivery service types as illustrated in Figure 3. A fixed central distribution hub serves several satellite facilities using one type of first-echelon vehicles to deliver both packed customer orders to satellite facilities for transshipment and unpacked inventory to satellite facilities for storage and distribution. These satellite facilities can function either as a transshipment center, or a distribution center or both. They must also serve customers using one different type of second-echelon vehicle. Three types of delivery services are considered: Standard deliveries have delivery lead time of more than one day, while express deliveries are same-day deliveries. Instant deliveries need to be delivered within 2 hours upon order placement.

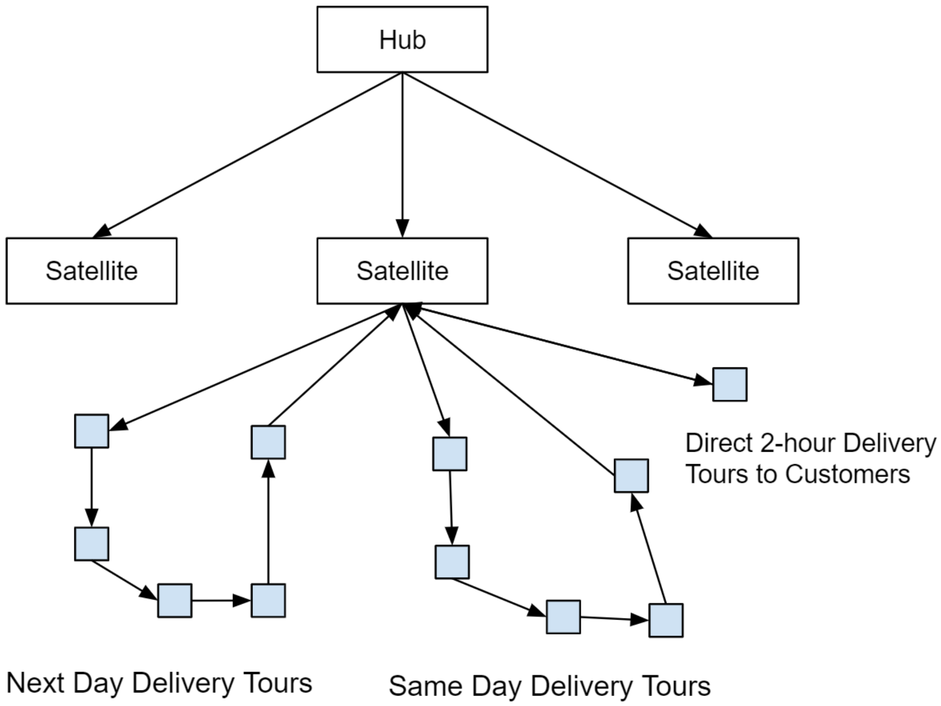


Figure 3: Multi-Tiered Distribution Network

Let J be the set of satellite locations in the two-echelon distribution network illustrated in Figure 3. A satellite location can provide a set of facility services U, i.e. distribution center service, transshipment center service, or both. Hence, demand in the distribution area can be directly fulfilled by a satellite location's on-site inventory (distribution center or DC facility service), or fulfilled by the hub and then transhipped via a satellite location (transhipment center or TC facility service).

The demand area is divided into a set of discrete segments I and demand within the segment is aggregated to the segment level for each segment $i \in I$. Each segment can be described by its area, circuity factor, density, drop size, and physical size of each delivery service types $s \in S$.

Tabel 2 to 8 lists all the index and parameters used in the model, covering facility locations, facility types, demand zone segments, and delivery service types.

Table 2: Model Index

| Notation | Explanation |
|---|---|
| J = {*j*} | set of satellite facility locations |
| U = {*u*} | set of facility service types, possible values are *tc* for transhipment center or *dc* for distribution center. |
| I = {*i*} | set of demand zone segments *i* |
| S = {*s*} | set of delivery service types, possible values are *e* for express,*ι*for instant, and *s* for standard. |

Table 3: Central Distribution Hub Parameters

| Notation | Explanation | Unit |
|---|---|---|
| $c^{H,dc}$ | unit handling cost for replenishment inventory | [$/item] |
| $c^{H,tc}$ | unit handling cost for customer fulfillment inventory | [$/item] |
| $c_H^{IH}$ | unit inventory holding cost at hub | [$/item/year] |

Table 4: Satellite Facility Parameters

| Notation | Explanation | Unit |
|---|---|---|
| $Z_j^{tc}$ | maximum daily handling capacity of satellite facility (in number of items) as a transhipment center | [items] |
| $Z_j^{dc}$ | maximum daily storage capacity of satellite facility (in number of items) as a distribution center | [items] |
| $c_{j,tc}^{F}$ | fixed cost to enable a satellite facility in location *j* as a transhipment center | [$] |
| $c_{j,dc}^{F}$ | fixed cost to enable a satellite facility in location *j* as a distribution center | [$] |
| $c_{j,tc}^{H}$ | handling cost at satellite facility *j* as a transhipment center | [$/item] |
| $c_{j,dc}^{H}$ | handling cost at satellite facility *j* as a distribution center | [$/item] |
| $c_j^{IH}$ | inventory holding cost at satellite facility j | [$/item] |
| $c_j^{IO}$ | Inventory reordering cost at satellite facility | [$/item] |

| | | |
|---|---|---|
| $b$ | reordering frequency of satellite facilities, i.e. Days Between Orders. | [days/reorder] |

Table 5: Demand Segment Parameters

| Notation | Explanation | Unit |
|---|---|---|
| $A_i$ | area of demand segment *i* | [km$^2$] |
| $\gamma_{is}$ | density of deliveries of a delivery service type *s* in demand segment *i* | [deliveries/ km$^2$] |
| $\rho_{is}$ | average number of items per delivery (drop size) for a delivery service type *s* in demand segment *i* | [items/ delivery] |
| $\theta_{is}$ | average physical size of a delivery item for a delivery service type *s* in demand segment *i* | [m$^3$/item] |
| $k_i^{\beta,s}$ | circuity factor for intra-stop distances by first echelon vehicles in segment *i* | [] |
| $s_i^{\beta,s}$ | average intra-stop speed for vehicles operating on the second echelon | [km/hour] |

Table 6: Parameters Describing Inbound Flows between Central Distribution Hub and Satellite Facilities

| Notation | Explanation | Unit |
|---|---|---|
| $d_j$ | distance between central distribution hub and satellite facility location *j* | [km] |
| $c_j^{T,tc}$ | first-echelon transportation cost of packed customer deliveries per distance unit to candidate satellite location *j* for transshipment | [$ /(item-km)] |
| $c_j^{T,dc}$ | first-echelon transportation cost of unpacked inventory per distance unit to candidate satellite location *j* for inventory holding | [$ /(item-km)] |

Table 7: Parameters Describing Outbound Flows between Satellite Facilities and Demand Segments

| Notation | Explanation | Unit |
|---|---|---|
| $r_{ij}$ | distance between candidate satellite facility location *j* and demand segment *i* | [km] |

Table 8: Vehicle Parameters

| **Notation** | **Explanation** | **Unit** |
|---|---|---|
| $T^m_{s,tc}$ | maximum allowed service time per vehicle per day for standard delivery service that is fulfilled from central distribution hub and transshipped from satellite facility | [hour] |
| $T^m_{s,dc}$ | maximum allowed service time per vehicle per day for standard delivery service that is fulfilled from directly from the inventory in satellite facility | [hour] |
| $T^m_{e,tc}$ | maximum allowed service time per vehicle per day for express delivery service that is fulfilled from central distribution hub and transshipped from satellite facility | [hour] |
| $T^m_{e,dc}$ | maximum allowed service time per vehicle per day for express delivery service that is fulfilled from directly from the inventory in satellite facility | [hour] |
| $\xi_v$ | volumetric carrying capacity of second-echelon vehicles | [$m^3$] |
| $c^f$ | per-day fixed cost of vehicle type *v* | [$/day] |
| $s^{\beta,l}$ | average line-haul speed for vehicles operating on the second echelon | [km/hour] |
| $k^{\beta,l}$ | circuity factor for line-haul distances by second echelon vehicles | [] |
| $w$ | wage per hour | [$/hour] |
| $t^{C,f}$ | fixed time per delivery at customers' locations | [hour/delivery] |
| $t^{L,f}$ | fixed time for vehicle loading and tour setup [hour] | [hour/tour] |
| $c^{ins}$ | per-distance cost of an instant delivery | [$/delivery/km] |

Below lists the assumptions made for the distribution network.

1. The entire demand area under consideration is divided into very small demand segments. Each demand segment is considered in its entirety and all demand of the same delivery service type within a demand segment will be served by one facility's one facility service type only.

2. Depending on the actual geographical and demand profile, a demand segment is represented by a unique set of attributes, e.g. density of deliveries of a delivery service type, distance metric factor for driving within the city segment (in-tour distances), average number of items per delivery service type, and average physical size of a delivery item per delivery service type.
3. Demand of each delivery service type in each demand segment is known and therefore deterministic.
4. The location for the central distribution hub is fixed with unlimited capacity. Only one type of first-echelon vehicles are used to deliver both packed customer deliveries to satellite facilities for transshipment and unpacked inventory to satellite facilities for storage and distribution. Two different unit transportation cost will be applied - this is because we can transport unpacked inventory in full pallets, which are more efficient in handling and space utilization compared with packed customer deliveries.
5. Satellite facilities can function either as a transshipment center or a distribution center, and they serve customers using one type of second-echelon vehicle. The fixed costs required for opening the satellite facility as a transhipment point are different than opening the satellite facility as a distribution center.
6. Satellite distribution centers place a fixed order of Q units before inventory drops to 0. A satellite distribution hub requires a fixed cost for placing each replenishment order and a unit cost for holding inventory.
7. Replenishment from central distribution hub to satellite distribution hubs is immediate. This assumption can be relaxed by allowing an L-day lead time, however, since we are dealing with deterministic demand, even if there is a lead time for replenishment, the satellite facilities just need to place an order L days before inventory runs to 0.
8. Three second-echelon delivery service types are modeled - standard, express, and instant deliveries. The rationale for fulfilling a customer order from a satellite location's inventory directly compared with fulfilling it from the central distribution hub is the time saved from driving from the central distribution hub to transhipment center. Fulfilling

directly from a satellite location's inventory could potentially be advantageous for express deliveries due to their short lead time and the unpredictable traffic conditions during the day. For modeling, we will allow vehicles that deliver orders fulfilled directly from satellite distribution centers to have longer allowable service time, and therefore to have higher utilization on a daily basis.

9. Instant deliveries will be considered as point-to-point, meaning that each delivery will be fulfilled and delivered directly from the satellite locations to the demand zone, with no consolidation. In addition, Instant deliveries can only be fulfilled from satellite facilities that function as distribution centers.
10. Unlike standard and express deliveries whose cost will be determined using the ARCE formulation, a per-distance cost function will be used for instant deliveries.

### 3.2. Optimization Model

We design a mathematical model to support strategic decision making, such as what facilities to be activated for what types of facility service (DC or TC) and which demand segments' each delivery service types to be served by which facility. We define the below three sets of decision variables:

- $Y_{j,tc}$= 1, if satellite facility at location *j* is active as a transhipment center; 0, otherwise
- $Y_{j,dc}$= 1, if satellite facility at location *j* is active as a distribution center; 0, otherwise
- $X_{ijs,tc}$= 1, if satellite facility at location *j* serves delivery service type s in demand segment *i as a transhipment center*; 0, otherwise
- $X_{ijs,dc}$= 1, if satellite facility at location *j* serves delivery service type s in demand segment *i as a distribution center*; 0, otherwise

Total network cost is defined as $K_C^H + K^F + K_S^H + K^{IH} + K^{IO} + K^T + K^D$, where:

1. The handling cost at central distribution hub,

$$K_C^H = c^{H,tc} \sum_{j\in J} \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_iX_{ijs,tc} + c^{H,dc} \sum_{j\in J} \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_iX_{ijs,dc} \quad (1)$$

2. The fixed cost of enabling satellite facilities as transhipment centers and distribution centers,

$$K^F = \sum_{j\in J} c_{j,tc}^F Y_{j,tc} + \sum_{j\in J} c_{j,dc}^F Y_{j,dc} \quad (2)$$

3. The handling cost at each satellite facility as transhipment centers and distribution centers,

$$K_J^H = \sum_{j\in J} c_{j,tc}^H \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_iX_{ijs,tc} + \sum_{j\in J} c_{j,dc}^H \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_iX_{ijs,dc} \quad (3)$$

4. Additional cost due to inventory holding cost at each satellite facility,

$$K^{IH} = \sum_{j\in J} \left(c_j^{IH} - c_H^{IH}\right)/365 * \left(b \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_iX_{ijs,dc}\right)/2 \quad (4)$$

The inventory level here is dependent on satellite locations' reordering quantity, which given a Days Between Orders *b,* is modelled as *b* * the daily demand allocated to the facility for distribution center services.

$$b \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_iX_{ijs,dc} \qquad j \in J \quad (5)$$

5. The inventory reordering cost at each satellite facility,

$$K^{IO} = \sum_{j\in J} Y_{j,dc}\, c_j^{IO}/b \quad (6)$$

6. The transportation cost of unpacked inventory and packed customer deliveries from central distribution hub to satellite facilities,

$$K^T = \sum_{j\in J} c_j^{T,tc} \sum_{i\in I} \sum_{s\in S} \gamma_{sz}\rho_{sz}A_z X_{ijs,tc} +$$

$$\sum_{j\in J} c_j^{T,dc} \sum_{i\in I} \sum_{s\in S} \gamma_{is}\rho_{is}A_i X_{ijs,dc} \quad (7)$$

7. The routing cost from satellite facilities to demand segments,

$$K^D \sum_{j\in J} \sum_{i\in I} \sum_{s\in S} f_{ijs,dc}X_{ijs,dc}+$$

$$\sum_{j\in J} \sum_{i\in I} \sum_{s\in S} f_{ijs,tc}X_{ijs,tc} \quad (8)$$

Here, $f_{ijs,dc}$and $f_{ijs,tc}$is the cost of serving demand segment i from facility j for a specific delivery service type s through distribution center or transshipment center. These costs are further specified in Section 3.4 to 3.5, and equations (15) through (23), which are based on the work of Winkenbach et al (2016).

The problem formulation is as follows.

$$\text{min: } K_C^H + K^F + K_S^H + K^{IH} + K^{IO} + K^T + K^D \quad (9)$$

Subject to:

1. All non-negative demand segments must be serviced, and by only one satellite facility:

$$\sum_{j\in J} (X_{ijs,dc} + X_{ijs,tc}) = 1 \ \forall i \in I, \forall s \in S \quad (10)$$

2. Instant deliveries can only be fulfilled from satellite facilities that function as distribution centers:

$$X_{ijs,tc} = 0, if\ s = instant \quad \forall i \in I,\ \forall j \in J \quad (11)$$

3. Allocation of segments is restricted to active satellite facilities and total volume assigned to each satellite facility is within its capacity:

$$\sum_{i\in I} \sum_{s\in S} \gamma_{is}\,\rho_{is}A_i X_{ijs,tc} \le Z_j^{tc}\, Y_{j,tc} \ \forall j \in J \quad (12)$$

$$\sum_{i\in I} \sum_{s\in S} \gamma_{is}\,\rho_{is}A_i X_{ijs,dc} \le Z_j^{dc}\, Y_{j,dc} \ \forall j \in J \quad (13)$$

4. DC function of a facility must be activated if a demand segment is fulfilled by a facility location's distribution center service:

$$\sum_{i \in I} \sum_{s \in S} \gamma_{is} \rho_{is} A_i X_{ijs,dc} \leq Z_j^{dc} Y_{j,dc} \quad j \in J \tag{14}$$

5. Decision domain of the decision variables:

$$Y_{j,tc}, Y_{j,dc}, X_{ijs,tc}, X_{ijs,dc} \in 0,1 \tag{15}$$

### 3.3. Routing Cost Approximation - Standard and Express Deliveries

Built upon the ARCE formula of Winkenbach et al. (2016), we approximate the routing cost of a satellite location *j* serving a demand segment *i*'s service type *s* either as a transshipment center or distribution center as follows:

$$f_{ijs} = wc_{ij}(t^{L,f} + \frac{2r_{ij}}{s^{\beta,l}} + n_{ij}(t^{C,f} + \frac{k_i k^b}{s_i^{\beta,s}\sqrt{\gamma_i}})) + c^f q_{ij}$$

$$\forall i \in I, j \in J, s \in S\{s,e\} \tag{16}$$

where

$$\zeta_i = \frac{\xi}{\theta_i^C \rho_i}, \forall i \in I \tag{17}$$

$$T_{ij}^f = t^{L,f} + \frac{2r_{ij}}{s^{\beta,l}}, \forall i \in I, j \in J \tag{18}$$

$$T_i^v = t^{C,f} + \frac{k_i k^b}{s_i^{\beta,s}\sqrt{\gamma_i}}, \forall i \in I \tag{19}$$

$$\begin{aligned} n_{ij} = \ & \zeta_i, \ if\ T_{s,tc}^m \geq T_{ij}^f + \zeta_i T_i^v \\ & \frac{T_{s,tc}^m - T_{ij}^f}{T_i^v}, if\ T_{ij}^f + \zeta_i T_i^v \geq T_{s,tc}^m \geq T_{ij}^f \\ & 0,\ if\ T_{s,tc}^m \leq T_{ij}^f \qquad \forall i \in I, j \in J \end{aligned} \tag{20}$$

$$m_{ij} = \frac{T^m_{s,tc}}{T^f_{ij} + \zeta_i T^v_i}, \qquad if\ T^m_{s,tc} \geq T^f_{ij} + \zeta_i T^v_i$$
$$1, if\ T^f_{ij} + \zeta_i T^v_i \ \geq \ T^m_{s,tc} \geq \ T^f_{ij}$$
$$0,\ if\ \ T^m_{s,tc} \ \leq \ T^f_{ij} \qquad \forall i \in I, j \in J \tag{21}$$

$$q_{ij} = \frac{\gamma_i A_i}{n_{ij} m_{ij}}, if\ T^m_{s,tc} \geq \ T^f_{ij}$$
$$\infty, otherwise \qquad \forall i \in I, j \in J \tag{22}$$

$$c_{ij} \ = \ m_{ij} q_{ij} \qquad \forall i \in I, j \in J \tag{23}$$

Equation (16) provides the total cost of using a specific route. It encompasses the number of routes, the duration of a route, and the inter-stop travel duration within a segment. Equations (17) through (23) are the individual calculations for the components of equation (16). Equation (17) represents the number of customers that can be served on a route with vehicle capacity being taken into consideration. Equation (18) to (19) define the fixed route time and the variable duration per customer respectively. Equation (20) compares the result of Equation (17) with respect to Tm in order to determine an effective number of customers that can be served. Equation (21) then determines the number of routes a vehicle can make with respect to Tm. Equation (22) calculates the number of vehicles needed to service demand. Equation (23) provides the total number of vehicle tours needed to service demand.

### 3.4. Routing Cost Approximation - Instant Deliveries

For instant deliveries, the ARCE formulation is not used because the deliveries are handled point-to-point with no consolidation. Instead, a per-distance cost is used in the model formulation to determine the cost of fulfilling instant deliveries directly from satellite locations.

$$f_{ijs} = c^{ins} r_{ij} \gamma_{is} A_i, \forall i \in I, j \in J, s \in S\{i\} \tag{24}$$

### 3.5 Model Solution

The optimization model proposed is non-linear because inventory reordering frequency b appears as both the numerator in equation (4) and denominator in equation (5) in the objective function.

To avoid non-linearity, we propose a two-staged solution. First, we assign a given value to the satellite facilities' reordering frequency *b* (i.e. Orders once every 1, 2, 3, … or 7 days), and then use the given value as input to the network models and solve the mixed integer models through Gurobi. We iterate different Days Between Orders *b* values to obtain the optimal network setup and distribution cost for each value. Next, we compare the impact of the value to the different network setup and total cost. The modeling approach is illustrated in Figure 4.

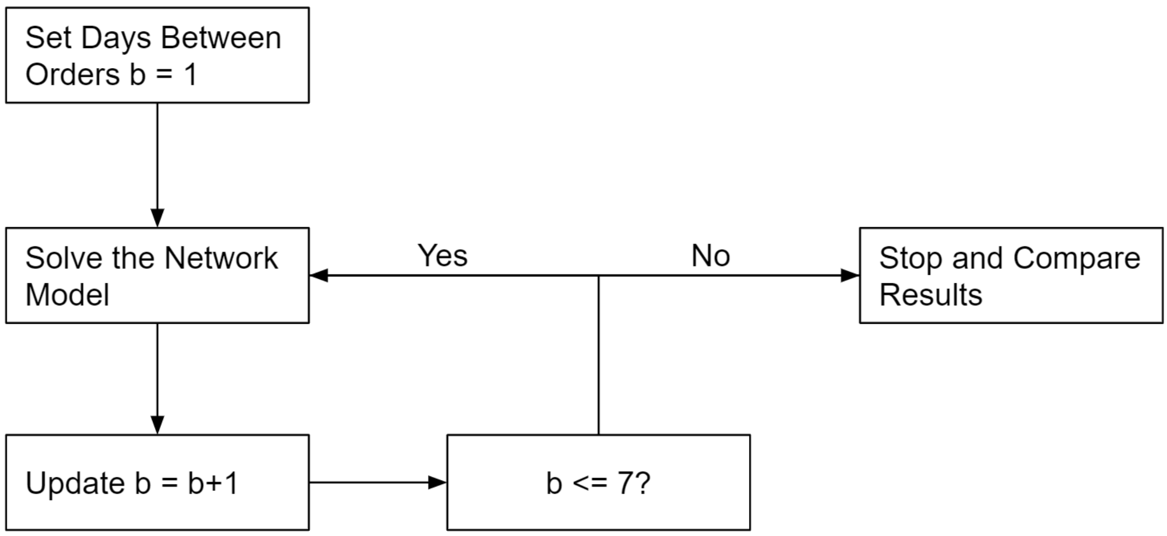


Figure 4: Modeling Approach

## 4. RESULTS AND ANALYSIS

### 4.1. Case Study

We apply the model to a synthetic dataset inspired by real-life operations of a Brazilian e-commerce company's last-mile delivery network in metropolitan area of São Paulo, which represents the company's largest and most relevant market. In the synthetic dataset, the company makes about 18,000 customer deliveries a day through a two-echelon network that consists of a large central hub outside of the city and seven satellite facilities in the city.

We aggregate the daily customer deliveries of different service types into 2,400 demand segments to be served by the network. We treat the current seven satellite facilities as candidate locations that can act either as a transshipment center, or a distribution center, or both. Our model then calculates the optimal allocation of each demand zone's different deliveries to a satellite location, acting as either a transshipment center or a distribution center. In addition, the model also informs us the right capacity of each satellite location.

### 4.2. Data and Parameters

The e-commerce retailer utilizes seven satellite facilities and one central distribution center within the São Paulo territory. In Figure 5, the satellites are denoted by a flag icon, and the central distribution center is by a star . This central distribution center can also service the demand zones instead of utilizing a satellite facility. The demand for a specific latitude and longitude is known, and the individual demand points are consolidated into 1 km$^2$ demand zones with an associated demand density. These are represented in Figure 5 as square blocks colored by the density of orders in that specific demand zone.

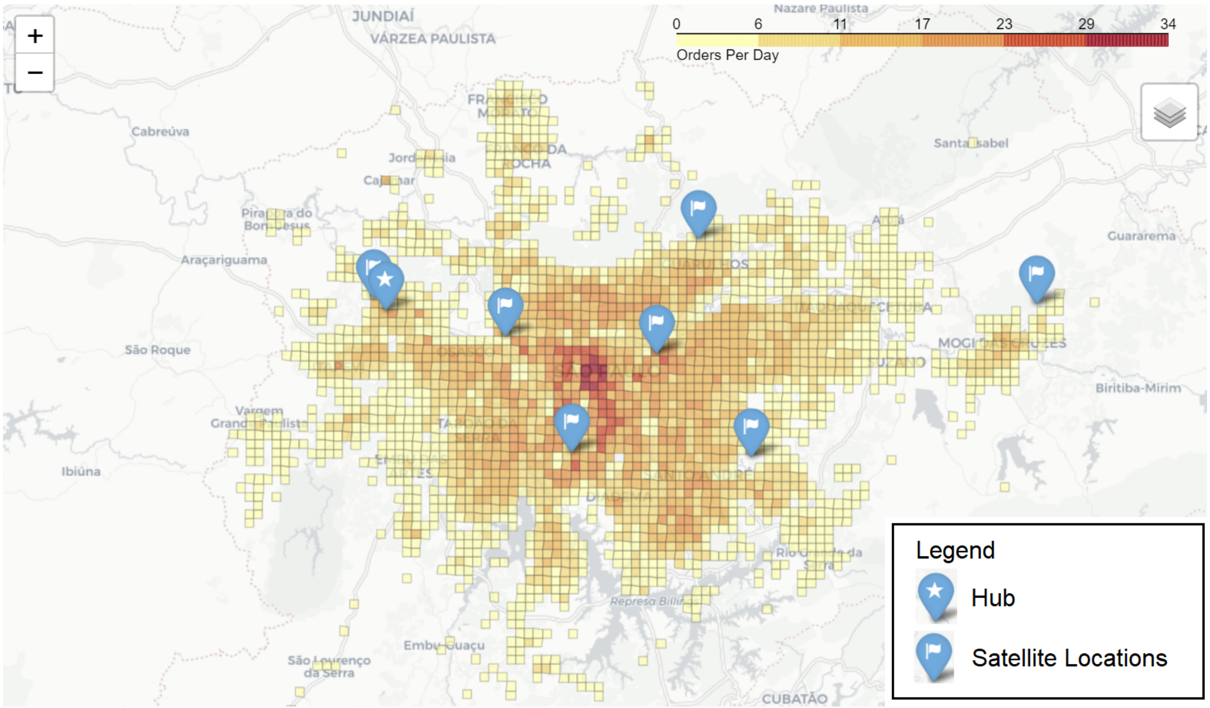


Figure 5: Map of the Demand Area with Facilities and Demand Distribution and Density

### 4.3. Model Implementation

The proposed MILP model is coded in Python 3.7.1 and Gurobi 8.1.0 and executed on a Intel(R) Core(™) i7 CPU 1.90GHz PC with 16GB memory. For a problem set of 1 hub, 8 satellite locations, and 2,400 demand zones, the program takes about 30 seconds to complete each iteration with a given inventory reordering frequency.

### 4.4. Results and Insights

We constructed 19 unique scenarios and ran the model on each scenario to examine the effect of different parameters and delivery service types' density on the overall network design and cost. Table 9 details the parameters adjusted for different scenarios.

Table 9: Scenario Breakdown

| Scenario # | Standard % | Express % | Instant % | Tmax Express | Days Between Orders | Satellite Inventory Holding Cost |
|---|---|---|---|---|---|---|
| 1 | 70% | 20% | 10% | 5 | 1 to 7 | 0.6 |
| 2** | 70% | 20% | 10% | 6 | 1 to 7 | 0.6 |
| 3 | 70% | 20% | 10% | 7 | 1 to 7 | 0.6 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | 70% | 20% | 10% | 6 | 1 to 7 | 0.7 |
| 5 | 70% | 20% | 10% | 6 | 1 to 7 | 0.8 |
| 6 | 100% | 0% | 0% | 6 | 1 to 7 | 0.6 |
| 7 | 0% | 100% | 0% | 5 | 1 to 7 | 0.6 |
| 8 | 0% | 100% | 0% | 6 | 1 to 7 | 0.6 |
| 9 | 0% | 100% | 0% | 7 | 1 to 7 | 0.6 |
| 10 | 0% | 0% | 100% | 6 | 1 to 7 | 0.6 |
| 11 | 90% | 10% | 0% | 5 | 1 to 7 | 0.6 |
| 12 | 90% | 10% | 0% | 6 | 1 to 7 | 0.6 |
| 13 | 90% | 10% | 0% | 7 | 1 to 7 | 0.6 |
| 14 | 80% | 20% | 0% | 5 | 1 to 7 | 0.6 |
| 15 | 80% | 20% | 0% | 6 | 1 to 7 | 0.6 |
| 16 | 80% | 20% | 0% | 7 | 1 to 7 | 0.6 |
| 17 | 70% | 30% | 0% | 5 | 1 to 7 | 0.6 |
| 18 | 70% | 30% | 0% | 6 | 1 to 7 | 0.8 |
| 19 | 70% | 30% | 0% | 7 | 1 to 7 | 1 |

**As a sample, summary results for Scenario 2 can be found in Appendix A

- **Standard %, Express %, Instant %:** These parameters represent the percentage of total demand that the service level encompasses.
- **Tmax Express:** This parameter represents the maximum allowable service time for express deliveries utilizing transhipment centers.

- **Days Between Orders:** This parameter represents the Days Between Orders (b = 1 to 7). All scenarios are run with all seven values of b.
- **Satellite Inventory Holding Cost:** This parameter represents the per-item daily inventory holding cost of a demand unit if serviced through a distribution center.

For the purpose of establishing baseline parameters for comparison across the different scenarios, we used a Tmax Express value of 6 (hours) and Satellite Inventory Holding Cost of $0.60 (per item per day). Under the baseline parameters, we compare the cost per parcel across the different demand mixes (for the optimal b) in Figure 6.

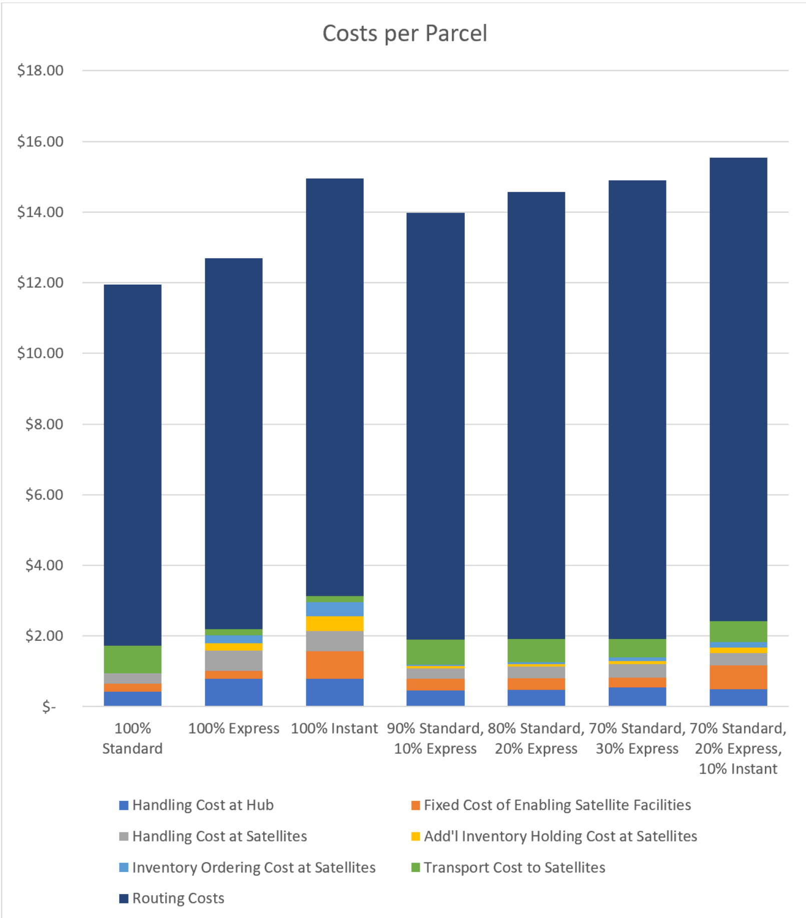


Figure 6: Cost per Parcel Breakdown

For the the 100% demand mixes (100% Standard, 100% Express, 100% Instant), the cost per parcel increases as we move from standard to instant delivery service types. The routing costs are the highest driver of total cost, and the higher tier service levels offset the increase in routing cost by utilizing distribution centers instead of transhipment centers. This introduces higher fixed and variable handling costs at the satellites, but it lowers the overall magnitude of the routing cost increases.

When we examine the mixed delivery service types, we see that as percentage of demand allocated to express and instant becomes higher, total overall cost turns out to be higher. The increase in costs with the added express and instant mainly come from the increased routing costs, fixed facility costs, and the inventory costs (ordering and holding).

In the below sections, we present some of the managerial insights we discover through detailed analysis of the model results.

#### 4.4.1. Impact of Demand Mix on Network Configuration

When looking at three runs of the model with 100% standard, 100% express, and 100% instant demand at the most optimal value of b with baseline parameters, we see 100% standard delivery scenario solely utilizing TC's, while the 100% express delivery scenario and 100% instant delivery scenario solely utilizing DC's as illustrated in Figure 7. In addition, the 100% instant scenario utilizes three more DC's than express to offset instant deliveries' higher routing costs due to the very strict on-demand service requirement.

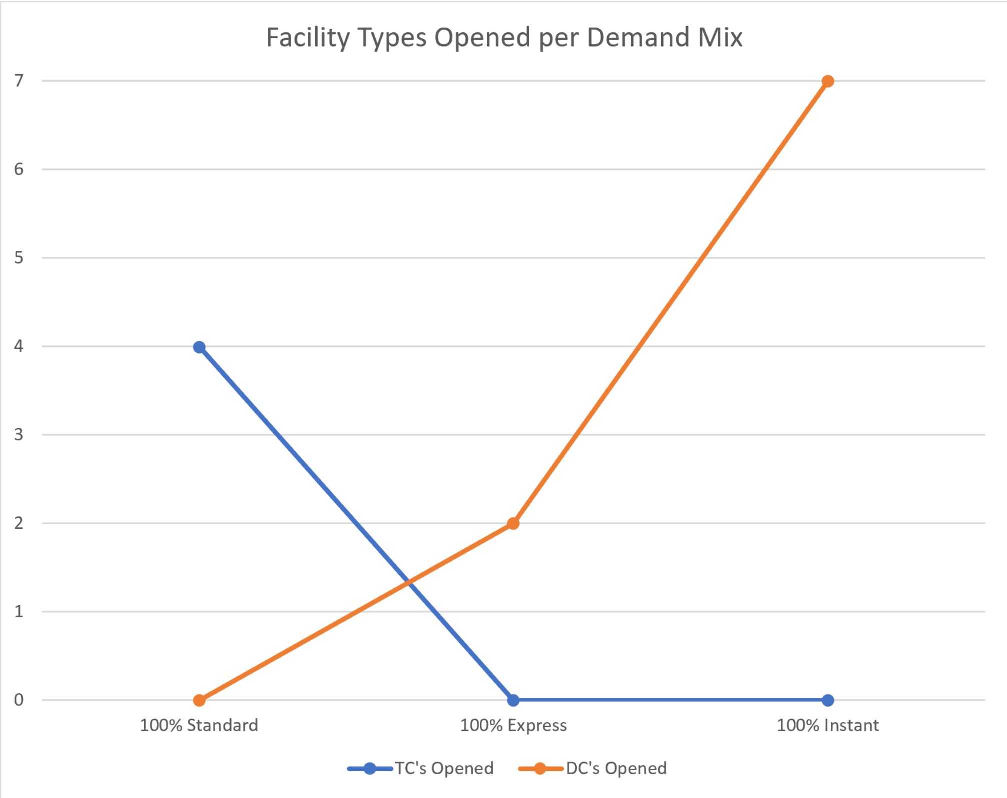

Figure 7: Facility Types Opened per Demand Mix

To better model actual e-commerce demand operations, we use “Mixed” delivery service scenarios to understand how introducing higher tier service levels to standard deliveries impact the overall network design. Figure 8 shows that, overall, as more express and instant demand is introduced into the network, we see less utilization of TC’s and more utilization of DC’s. When introducing instant demand, it has a higher impact on DC utilization in the network overall. This is because of the high routing cost associated with instant deliveries that is offset by having more DC’s opened within the network.

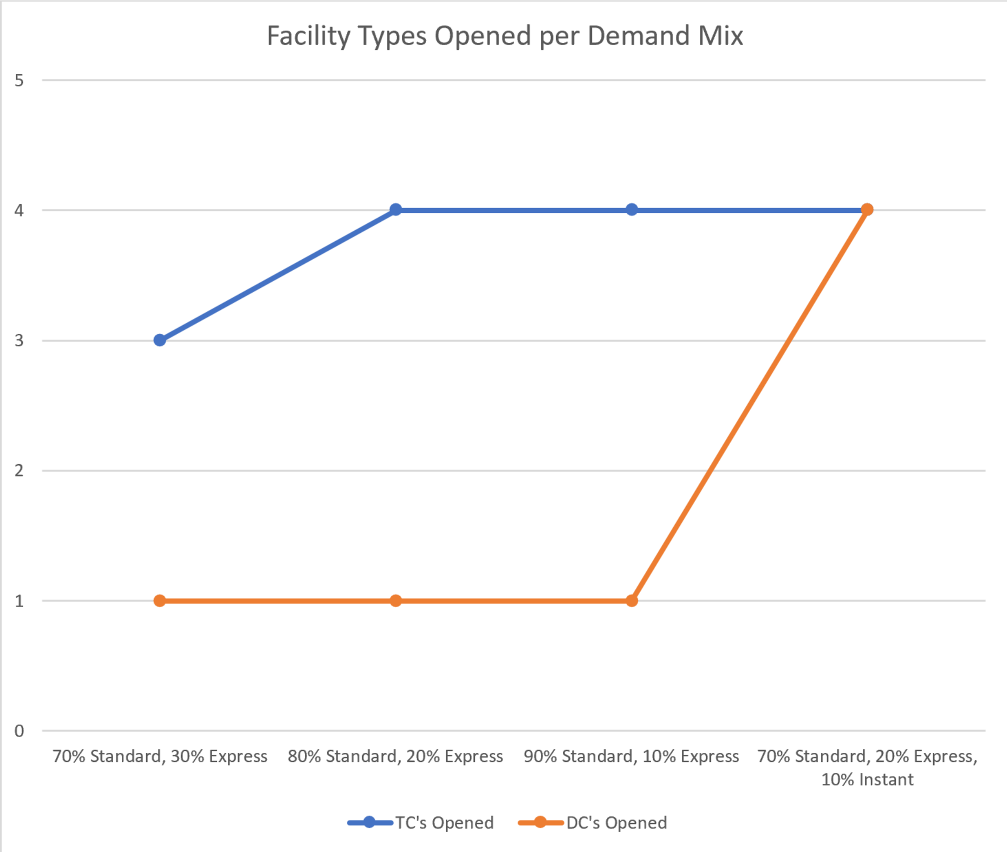


Figure 8: Facility Types Opened per Demand Mix (Mixed Demand)

#### 4.4.2. Impact of Service Time on Network Configuration

One important benefit of using a DC versus a TC is that added service time is allowed for routing due to the elimination of transhipment transit between the hub and the satellites. While there is increased facility and inventory holding costs associated with this, many times routing cost savings from satellite to demand segment outweigh such additional costs. We explore adjusting the value of the global service time for Express shipments serviced through TC's in order to determine the impact to the overall network configuration. To measure this impact, we use a Tmax value of 5, 6 (the baseline parameter), and 7. We do not use a Tmax of 8 or greater as this would negate cost savings from using a DC and we do not use a value lower than 5 due to no trade offs occuring after this point.

When observing the change in Tmax between the Standard/Express mixes, the network design favors the use of more TC's (Figure 9). For the 80%/20% mix and the 90%/10% mix, we see that at a Tmax of 7, DC's are no longer used to service any demand. For the 70%/30% mix, the higher volume of express shipments still warrant the use of a single DC, but an additional TC is opened to service Express demand in certain city segments.

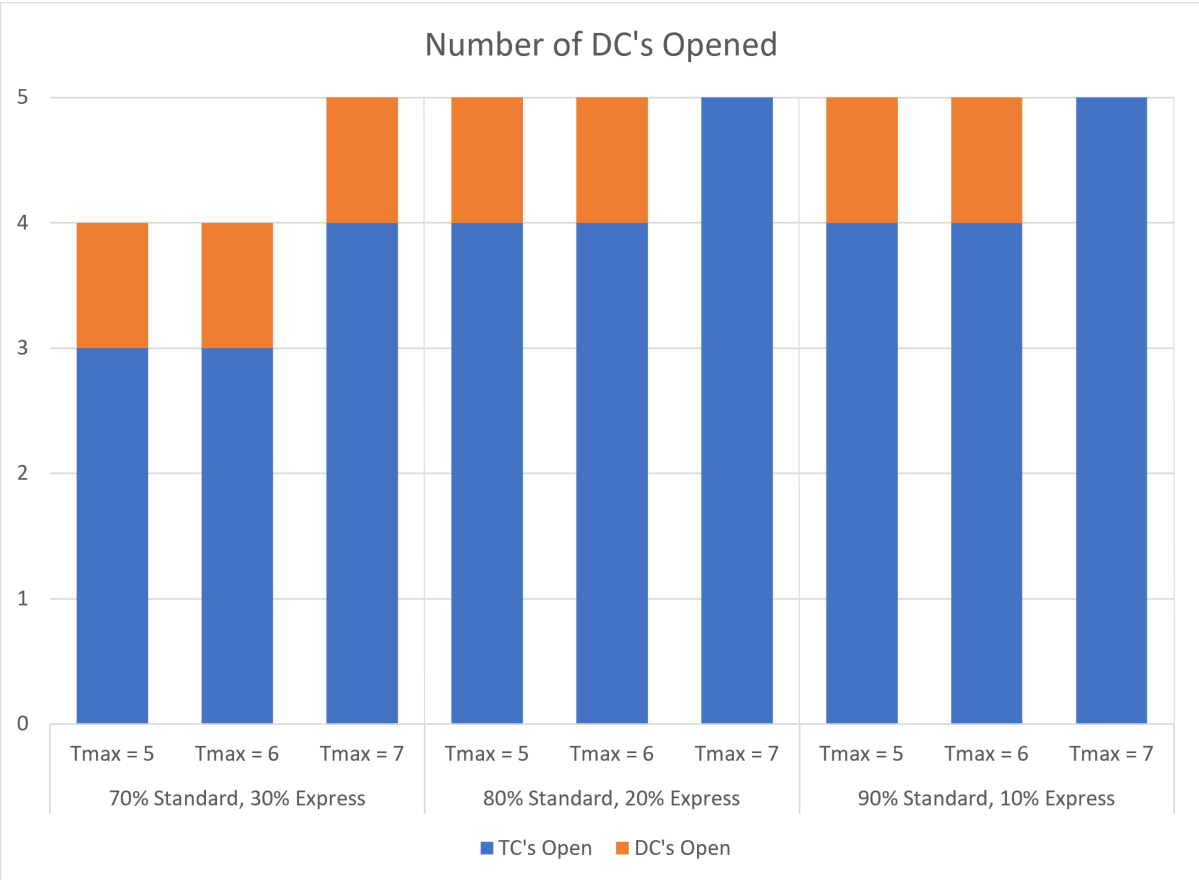


Figure 9: Number of DC's Opened per Tmax Change

When we look at the 100% Express Mix and the 70%/20%/10% Mix, there is no change in the number of facilities opened for either demand mix as shown in Figure 10. For the 100% Express mix, there is no change in cost to service demand when Tmax increases from 6 to 7. For the 70%/20%/10% mix, the increased routing cost does not offset the inventory ordering cost and is therefore not impactful to the network design.

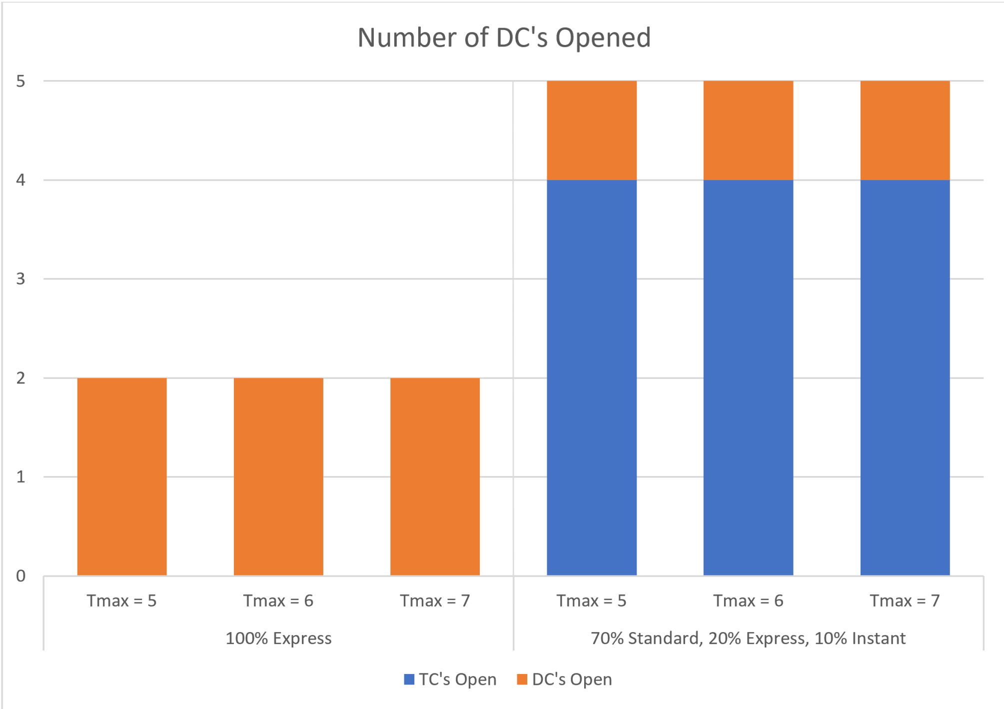

Figure 10: Number of DC's Opened per Tmax (No Change)

When exploring any Tmax Express values smaller than 6, the routing costs increase and DC's will be the chosen. We start seeing the benefits of opening TC's at the level of Tmax = 7, but a company must be able to perform the delivery from hub to satellite and the transhipment operations within an hour for this to be a feasible network design.

### 4.4.3. Impact of Reordering Frequency on Network Configuration

When determining inventory decisions in network design, we use the Days Between Orders (b) to calculate both the inventory held at a satellite facility (as a DC) as well as the reordering cost per facility. We run each model using Days Between Orders values ranging from 1 to 7, where b=1 represents a daily replenishment policy, while b = 7 represents a once a week replenishment policy.

In Figure 11 below, we examine the impact of b on a 100% Standard demand mix. Because of the sole use of transhipment centers for 100% Standard demand, the Days Between Orders has no impact on the network configuration. This is because utilizing DC's does not provide any benefits for routing cost savings due to the Tmax value for both TC's and DC's being the same; but DC's trigger additional fixed cost, inventory holding cost and satellite handling cost. Therefore, the optimal solution always utilizes TC's for the lowest total cost.

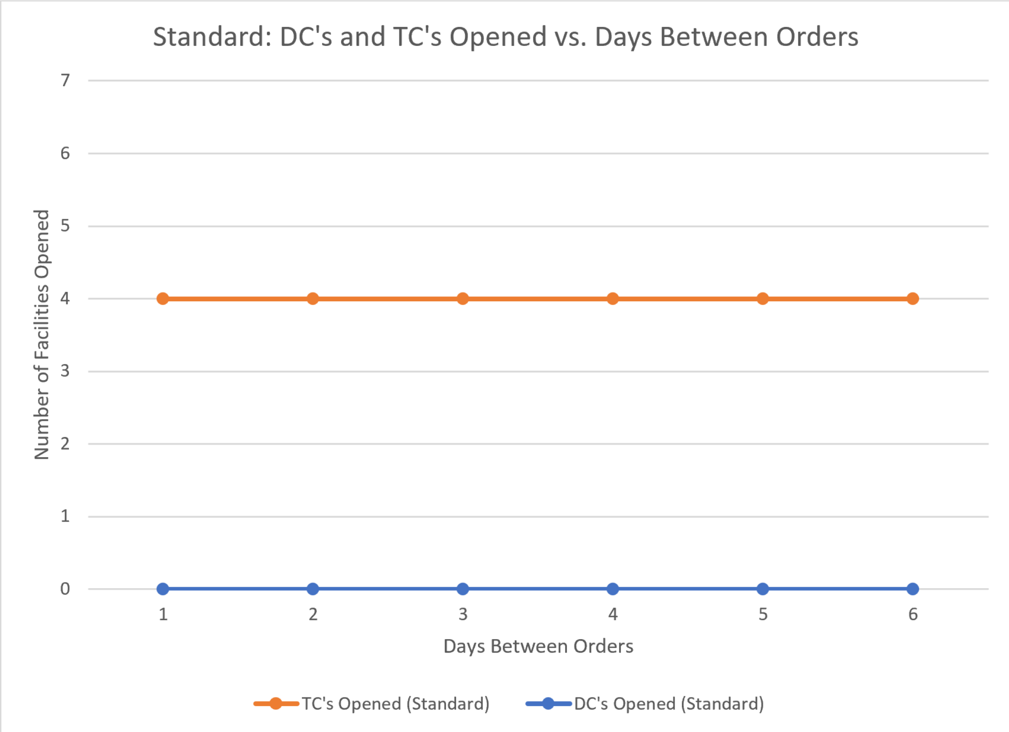


Figure 11: Standard DC's and TC's Opened vs Days Between Orders

As mentioned in the assumption, we only allow DC's to fulfill instant deliveries. For the 100% Instant delivery scenario, the routing cost is extremely high and therefore the model chooses to open up all the DC's to offset the routing cost with relatively insignificant DC fixed cost, inventory holding, and inventory handling. As Days Between Orders increases, each DC triggers higher inventory holding cost but lower reordering cost, as illustrated in Figure 12.

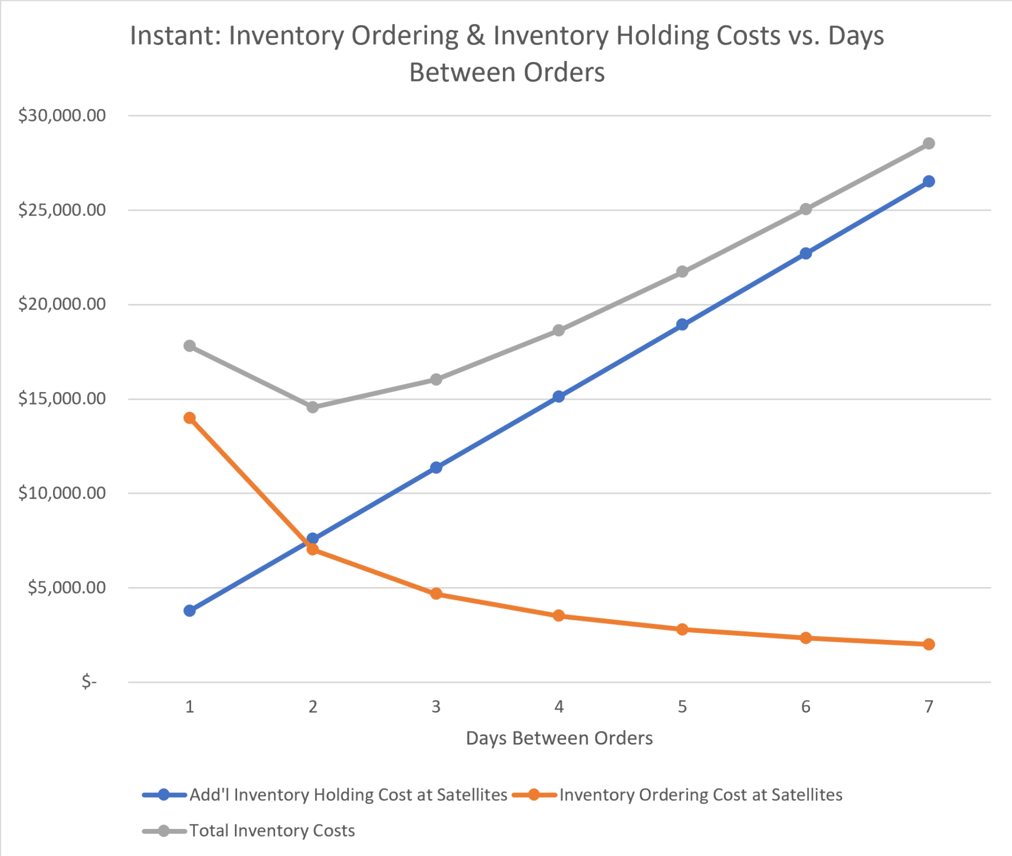


Figure 12: Instant Inventory Reordering, Inventory Holding Costs vs. Days Between Orders

In examining the 100% Express mix, the two additional Tmax hours from using a DC provides a huge cost advantage that is large enough to offset the additional fixed costs, inventory holding costs, and reordering costs. As Days Between Orders increases and inventory holding cost goes up, the model chooses to open an additional DC and use the savings in routing cost to offset the fixed cost (Figure 13).

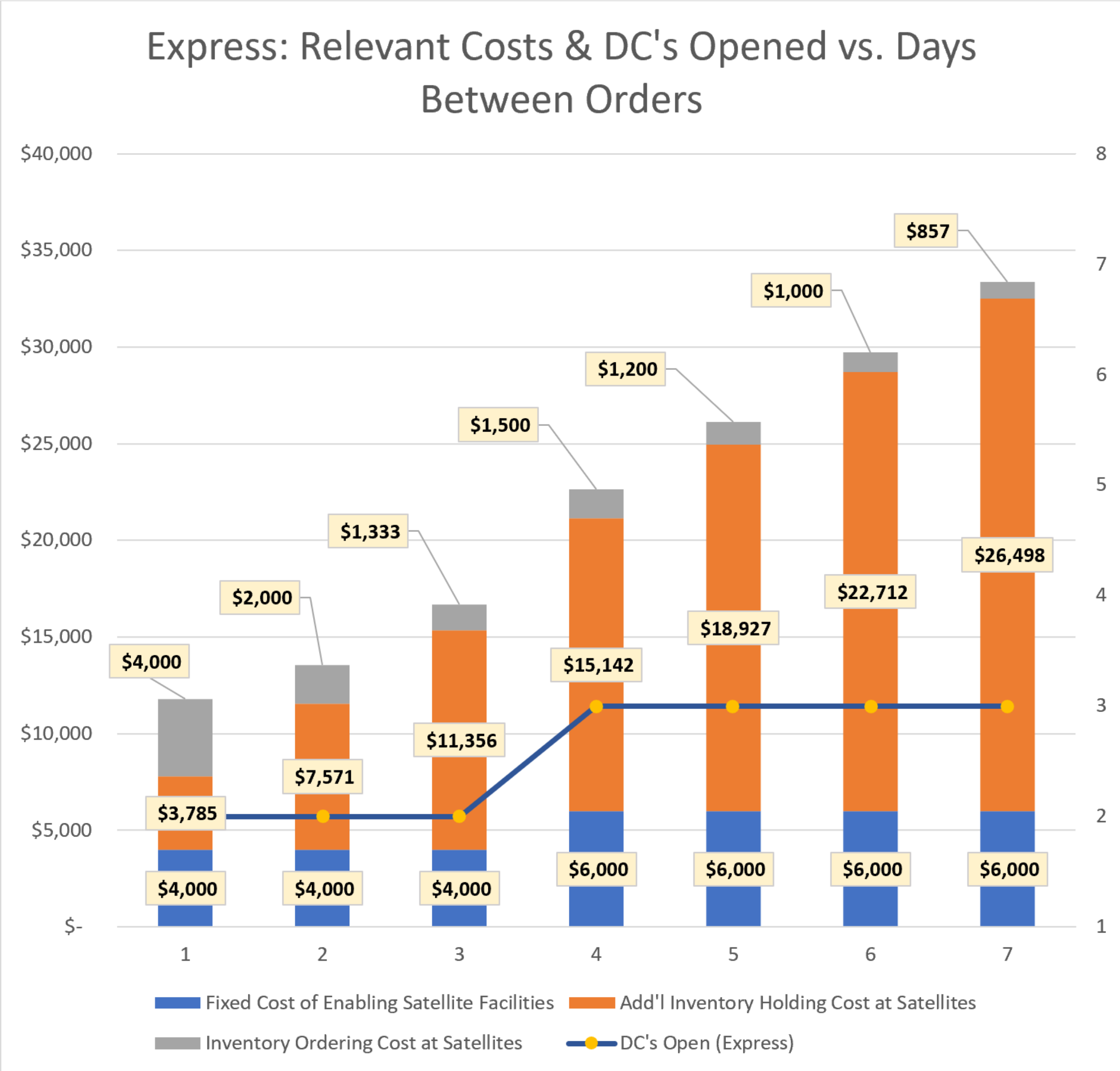


Figure 13: Express - Relevant Costs & DC's Opened vs. Days Between Orders

Figure 14 shows that when all delivery service types are blended and more express demand is introduced into the network, the Days Between Orders for the minimum total cost solution is b = 1. This is due the increased inventory holding cost associated with the higher units of express demand. At the 90%/10% demand mix, the Days Between Orders for the minimum total cost solution becomes b = 3, because of the higher ordering costs compared to the inventory holding costs.

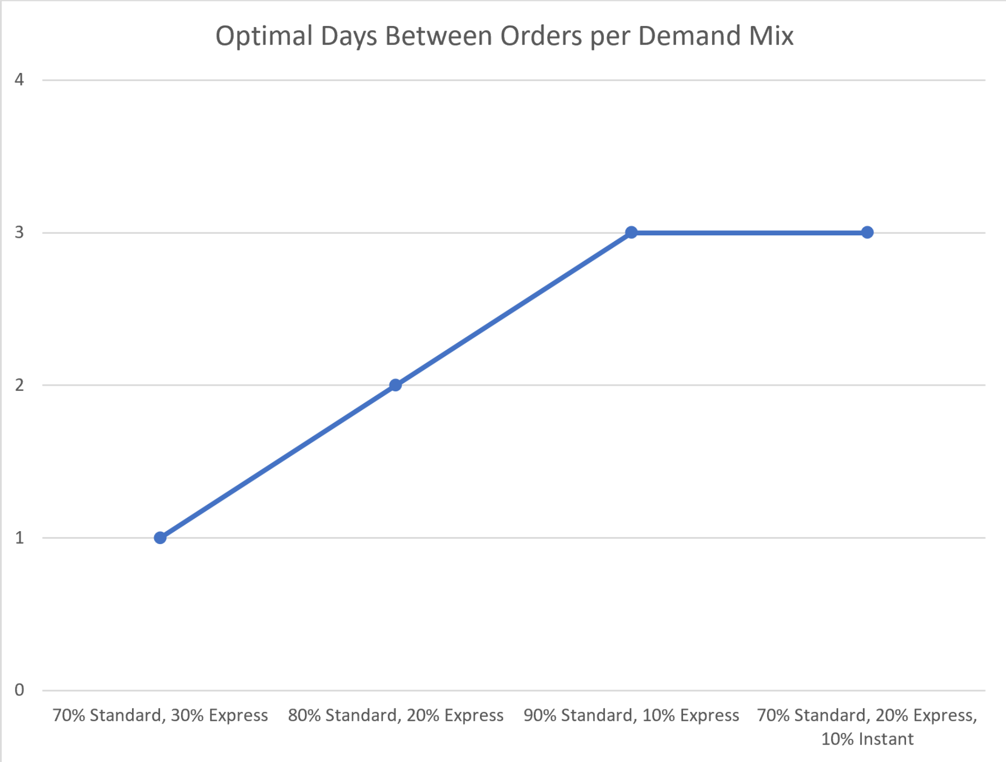


Figure 14: Optimal Days Between Orders for Different Demand Mixes

When we start incorporating the full demand mix (70%/20%/10%), we see clear trade offs as Days Between Orders increases. The minimum total cost is achieved when the Days Between Orders is 3 days (Figure 14). As Days Between Orders increases, inventory holding costs increase while the inventory reordering costs begin to decrease (Figure 15) Because of the lower inventory ordering costs as Days Between Orders increases, we see more facilities opening up to offset the higher routing costs.

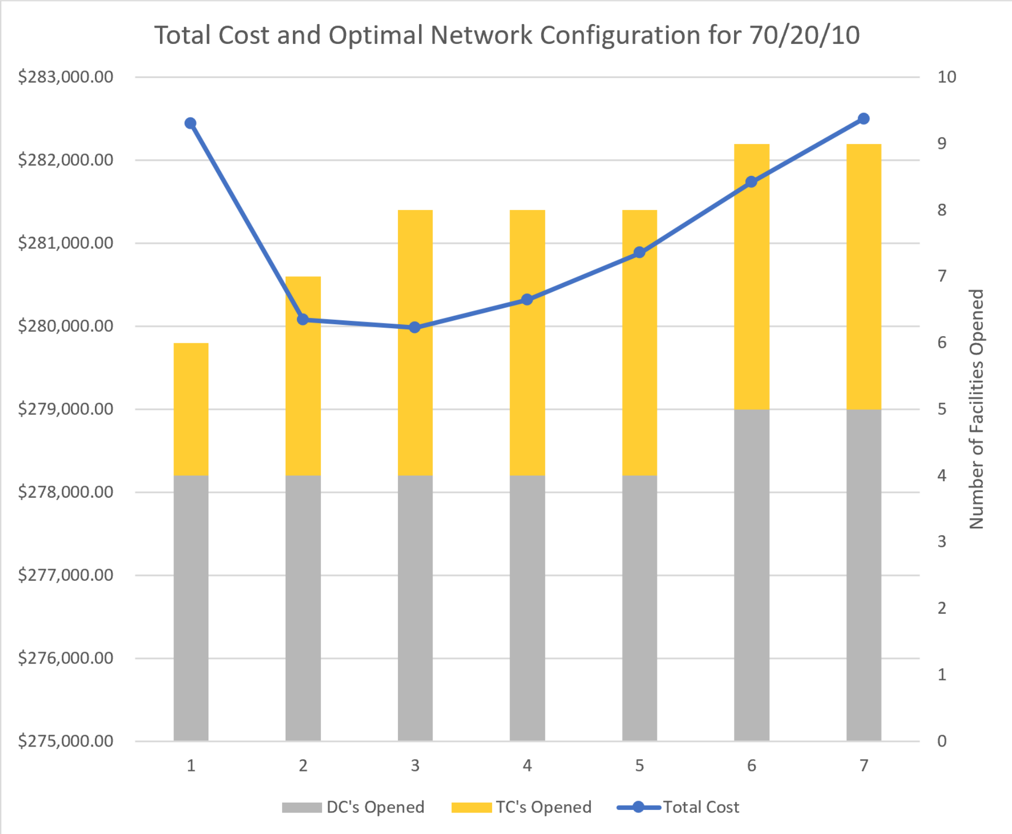


Figure 15: Full Demand Mix: Days Between Orders vs. Total Cost and Network Configuration

### 4.4.4. Impact of Inventory Holding Cost on Network Configuration

The last parameter we examine is the inventory holding cost. When considering the full demand mix (70%/20%/10%), increasing inventory holding cost does not impact the optimal network configuration for instant as a large number of DC's are still required to keep the instant deliveries' routing costs low. As inventory holding costs go up, the model chooses to use less number of DC's for express and pushes some express deliveries to TC's to offset the higher inventory holding cost.

In terms of the total cost, we observe that when inventory holding cost is low, the optimal network setup is achieved with b = 3. As inventory holding cost increases, the optimal setup is achieved at lower Days Between Orders, meaning we need to replenish the satellite locations more frequently to lower total cost, as shown in Figure 16.

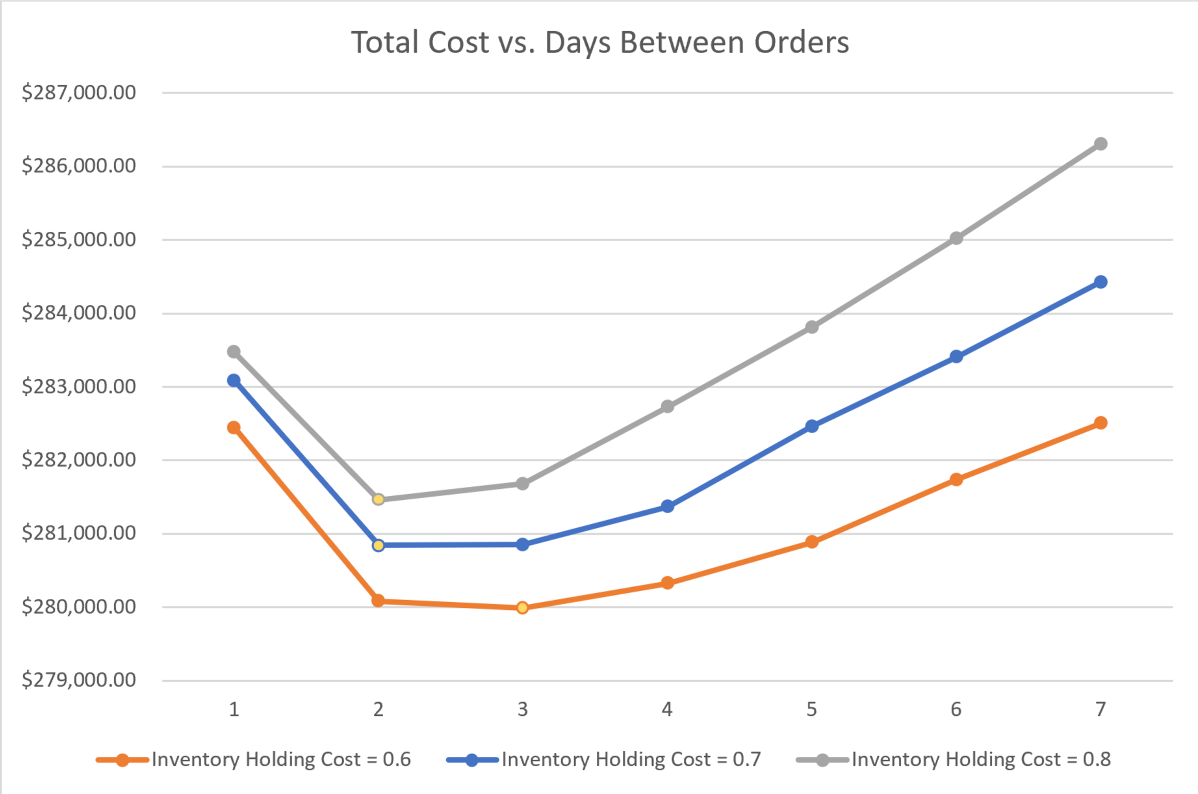


Figure 16: Total Cost vs Days Between Orders for Three Different Holding Costs

## 5. CONCLUSION

To summarize, our goal for this project was to establish a quantitative model that supports strategic e-commerce distribution network design by incorporating inventory control decisions into existing location-routing models with continuum approximation. By creating this model, we sought to highlight the significance of jointly considering facility location, vehicle routes, and inventory control decisions in a large-scale e-commerce setting. Our two primary motivations behind pursuing the project were:

1. The combined effect of increasing e-commerce market share, and the growing trend of urbanization will scale the current costs of operating a last-mile e-commerce distribution network.
2. Within the current breadth of research, there were no models that jointly consider inventory control decisions in large-scale e-commerce distribution networks.

### 5.1. Contribution

The results of our modeling have shown that the overall system costs and network configuration are significantly impacted by the inventory control decisions deployed by an e-commerce retailer. When examining the different demand mixes e-commerce retailers faces, it is advantageous to hold inventory in facilities closer to the demand segments for both express and instant deliveries in order to service customers timely and to reduce overall total system costs. In situations where demand for these express or instant services is smaller, the network design changes to move away from keeping inventory in facilities close to the customer and instead suggest the use of transhipment centers.

When examining the maximum service time a vehicle can support, we see that there are instances when express demand can be supported through a transhipment network, but only with a very small window of time allowed for transhipment operations. Depending on the overall geography of the region as well as the e-commerce retailer's operational capabilities, one hour may not be a feasible amount of time for all transhipment operations to occur and considerations will need to be made in how inventory is stored.

The Days Between Orders (b) in our model highlights the importance of selecting an optimal inventory policy based on the e-commerce retailer's expected demand mix. In situations where more express demand is introduced into the system, distribution centers servicing customers should have a frequent inventory replenishment to drive down the cost of holding inventory. As the level of express demand decreases, distribution centers are still utilized to service demand but at a less frequent replenishment policy. Introduction of instant demand favors less frequent replenishment policies as well, but that is due to the high cost of ordering inventory to the higher amount of distribution locations.

While our modeling and results present recommended strategies for e-commerce companies, these considerations may not be possible in all situations, depending on how the company has their business structured. Using the 70% Standard, 20% Express, and 10% Instant delivery type mix as an example: if a company wishing to adopt a similar distribution network cannot support the specified Days Between Orders, there are alternative approaches suggested by the model. In the highest value of b solved by the model (b=7, which assumes a company is replenishing inventory once a week,) the suggestion to the company would be to transship nearly all of their standard deliveries while only using distribution centers to service same-day (express) or two-hour (instant) deliveries. The other solutions (b = 2 through b=6) provide a blended approach, but highlight how companies can expect their distribution network to change depending on how much inventory is being stored at satellite facilities. Before moving forward with a similar set-up highlighted by the model, companies need to consider the facility and process designs, the labor used in the operation, the training associated with running the smaller distribution centers, the information technology utilized by the facility, and the overall operational management in order to fully assess the feasibility of implementation.

### 5.2. Managerial Insights

An important consideration from a managerial standpoint is the additional responsibilities assumed when opening these smaller distribution centers. In addition to the managing the current inventory at the hub, a company must closely monitor the inventory levels at the satellite facilities to ensure the cost savings of keeping stock close to their demand segments and remain cost effective. This has a direct impact on the daily operating procedures of the company, and requires a different skill set for all of the associated job functions.

When considering different service offerings, it is important to understand how these will be serviced throughout the network. Keeping inventory close to the customer will drive down the final-mile transportation costs, but a satellite facility will need to potentially manage two distinct operations simultaneously, i.e. transhipment operations for one service offering and distribution operations for another. Combining two types of facility services in the same facility will increase the impact to operations management and complexities as well.

In summary, the managerial implications can be distilled as below:

- To obtain the lowest total system cost when offering all delivery services, it is recommended that Standard service be serviced through transhipment centers while all other services be serviced through distribution centers.
- If there are constraints on how often inventory can be ordered, it is recommended to use transhipment options versus holding larger amount of inventory at satellite facilities.
- Facility, process, labor, training, informational technology, and operational management need to be further considered in the network design decisions.

### 5.3. Future Research

A number of extensions to the model can be considered for future work. First, we assume deterministic demand for the network, and do not consider the impact of safety stock in handling daily demand fluctuations. As inventory is distributed across the satellite locations, total amount of safety stock in the system increases and requires additional working capital, potentially impacting the network decisions. Second, our current model only considers a single product, and future work can evaluate multiple products as well as a constraint on total maximum allowable inventory at a satellite location. Third, it will be useful to extend the model to consider a larger number of flexible satellite locations that in practice can be outsourced to third party logistics companies. Fourth, our model currently considers uniform inventory ordering and holding costs across each facility, which may be different in practice depending on the facility type (third party logistic facility, brick and mortar store, dark store, etc.) Lastly, we avoid non-linearity by treating the Days Between Orders as an input to the model and iteratively solve multiple models to find the optimal inventory policy. Implementation of additional solution methods could potentially handle the non-linearity as well.

**REFERENCES**

Agatz, N. A. H., Fleischmann, M., & van Nunen, J. A. E. E. (2008). E-fulfillment and multi-channel distribution – A review. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2007.04.024

Ahmadi-Javid, A., & Azad, N. (2010). Incorporating location, routing and inventory decisions in supply chain network design. *Transportation Research Part E: Logistics and Transportation Review*, *46*(5), 582–597.

Alba, D. (2015, May 28). Free Same-Day Delivery Is Amazon's Gambit to Own All Retail. Retrieved March 11, 2019, from https://www.wired.com/2015/05/free-day-delivery-amazons-gambit-retail/

Kearney, A. T. (2015). Global Retail E-Commerce Keeps On Clicking - Full Report - A.T. Kearney - A.T. Kearney. Retrieved May 8, 2019, from https://www.atkearney.com/consumer-goods/article?/a/global-retail-e-commerce-keeps-on-clicking

Bektaş, T., Crainic, T. G., & Van Woensel, T. (2017). From Managing Urban Freight to Smart City Logistics Networks. In *Series on Computers and Operations Research* (pp. 143–188).

Boccia, M., Crainic, T. G., Sforza, A., & Sterle, C. (2010). A Metaheuristic for a Two Echelon Location-Routing Problem. *Experimental Algorithms*. https://doi.org/10.1007/978-3-642-13193-6_25

Bruns, A., Klose, A., & Stähly, P. (2000). Restructuring of Swiss parcel delivery services. *OR Spektrum*. https://doi.org/10.1007/s002910050106

Côté, J.-F., & Potvin, J.-Y. (2009). A tabu search heuristic for the vehicle routing problem with private fleet and common carrier. *European Journal of Operational Research*, *198*(2), 464–469.

Crainic, T. G., Mancini, S., Perboli, G., & Tadei, R. (2011). Multi-start Heuristics for the Two-Echelon Vehicle Routing Problem. *Evolutionary Computation in Combinatorial Optimization*. https://doi.org/10.1007/978-3-642-20364-0_16

Forrester. (2014, May 29). European Online Retail Forecast: 2013 To 2018. Retrieved May 8, 2019, from https://www.forrester.com/report/European+Online+Retail+Forecast+2013+To+2018/-/E-RES115752#

Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-completeness*. W.H. Freeman.

Ghezzi, A., Mangiaracina, R., & Perego, A. (2012). Shaping the E-Commerce Logistics Strategy: A Decision Framework. *International Journal of Engineering Business Management*. https://doi.org/10.5772/51647

Hausmann, L., Herrmann, N.-A., Krause, J., & Netzer, T. (2014, March). Same-day delivery: The next evolutionary step in parcel logistics. Retrieved May 8, 2019, from https://www.mckinsey.com/industries/travel-transport-and-logistics/our-insights/same-day-delivery-the-next-evolutionary-step-in-parcel-logistics

Joerss, M., Neuhaus, F., & Schröder, J. (2016, October). How customer demands are reshaping last-mile delivery. Retrieved May 8, 2019, from https://www.mckinsey.com/industries/travel-transport-and-logistics/our-insights/how-customer-demands-are-reshaping-last-mile-delivery

Liao, R. (2018, November 16). Alibaba and Amazon move over, we visited JD's connected grocery store in China. Retrieved March 11, 2019, from http://social.techcrunch.com/2018/11/15/jd-7fresh-supermarket/

Lim, S. F. W. T., Stanley Frederick W, Rabinovich, E., Rogers, D. S., & Lasester, T. M. (2016). Last-mile Supply Network Distribution in Omnichannel Retailing: A Configuration-Based Typology. *Foundations and Trends® in Technology, Information and Operations Management*. https://doi.org/10.1561/0200000045

Lin, J.-R., & Lei, H.-C. (2009). Distribution systems design with two-level routing considerations. *Annals of Operations Research*. https://doi.org/10.1007/s10479-009-0628-y

Liu, S. C., & Lee, S. B. (2003). A two-phase heuristic method for the multi-depot location routing problem taking inventory control decisions into consideration. *International Journal of Advanced Manufacturing Technology*, *22*(11-12), 941–950.

Liu, S. C., & Lin, C. C. (2004). A heuristic method for the combined location routing and inventory problem. *International Journal of Advanced Manufacturing Technology*, *26*(4), 372–381.

Ma, H., & Davidrajuh, R. (2005). An iterative approach for distribution chain design in agile virtual environment. *Industrial Management & Data Systems*, *105*(6), 815–834.

Max Shen, Z.-J., Shen, Z.-J. M., & Qi, L. (2007). Incorporating inventory and routing costs in strategic location models. *European Journal of Operational Research*, *179*(2), 372–389.

Merchan, D., & Winkenbach, M. (2018). High-Resolution Last-Mile Network Design. In E. T. E. Thompson (Ed.), *City Logistics 3: Towards Sustainable and Liveable Cities* (p. 201).

Najberge, A. (2017, July 17). Freshippo Supermarket Offers Shoppers a "New Retail" Experience. Retrieved March 11, 2019, from https://www.alizila.com/hema-supermarket-offers-shoppers-new-retail-experience/

Prodhon, C., & Prins, C. (2014). A survey of recent research on location-routing problems. *European Journal of Operational Research*, *238*(1), 1–17.

Redman, R. (2018, August 29). Amazon, Whole Foods extend Prime Now delivery. Retrieved from https://www.supermarketnews.com/online-retail/amazon-whole-foods-extend-prime-now-delivery

Salhi, S., & Rand, G. K. (1989). The effect of ignoring routes when locating depots. *European Journal of Operational Research*. https://doi.org/10.1016/0377-2217(89)90188-4

Smilowitz, K. R., & Daganzo, C. F. (2007). Continuum approximation techniques for the design of integrated package distribution systems. *Networks. An International Journal*, *50*(3), 183–196.

Snoeck, A., Winkenbach, M., & Mascarino, E. E. (2018). Establishing a Robust Urban Logistics Network at FEMSA through Stochastic Multi-Echelon Location Routing. *City Logistics 2*. https://doi.org/10.1002/9781119425526.ch4

Winkenbach, M., & Janjevic, M. (2018). Classification of Last-Mile Delivery Models for e-Commerce Distribution: A Global Perspective. In *City Logistics 1* (pp. 209–229).

Winkenbach, M., Kleindorfer, P. R., & Spinler, S. (2016). Enabling Urban Logistics Services at La Poste through Multi-Echelon Location-Routing. *Transportation Science*, *50*(2), 520–540.

## APPENDIX

Appendix A: Results Summary for Scenario 2 with Baseline Parameters

| Days Between Orders | Satellite Inventory Holding Cost | Total Cost | Handling Cost at Hub | Fixed Cost of Enabling Facility | Handling Cost at Satellites | Additional Inventory Holding Cost at Satellites | Inventory Ordering Cost at Satellites | Transport Cost to Satellites | Routing Cost to Demand Segments | TC's Opened (Standard) | TC's Opened (Express) | TC's Opened (Instant) | DC's Opened (Standard) | DC's Opened (Express) | DC's Opened (Instant) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.6 | $282,441 | $ 11,323 | $ 10,000 | $ 8,072 | $ 2,290 | $ 8,000 | $ 6,814 | $ 235,941 | 2 | 0 | 0 | 3 | 4 | 4 |
| 2 | 0.6 | $280,081 | $ 9,762 | $ 11,000 | $ 6,823 | $ 2,758 | $ 4,000 | $ 9,417 | $ 236,321 | 3 | 0 | 0 | 2 | 4 | 4 |
| 3 | 0.6 | $279,988 | $ 9,095 | $ 12,000 | $ 6,289 | $ 2,969 | $ 2,667 | $ 10,529 | $ 236,439 | 4 | 0 | 0 | 1 | 4 | 4 |
| 4 | 0.6 | $280,323 | $ 9,063 | $ 12,000 | $ 6,264 | $ 3,886 | $ 2,000 | $ 10,581 | $ 236,528 | 4 | 0 | 0 | 1 | 4 | 4 |
| 5 | 0.6 | $280,885 | $ 9,033 | $ 12,000 | $ 6,240 | $ 4,770 | $ 1,600 | $ 10,631 | $ 236,610 | 4 | 0 | 0 | 1 | 4 | 4 |
| 6 | 0.6 | $281,738 | $ 9,122 | $ 14,000 | $ 6,311 | $ 6,034 | $ 1,667 | $ 10,484 | $ 234,121 | 4 | 0 | 0 | 2 | 5 | 5 |
| 7 | 0.6 | $282,504 | $ 9,117 | $ 14,000 | $ 6,307 | $ 7,019 | $ 1,429 | $ 10,492 | $ 234,140 | 4 | 0 | 0 | 2 | 5 | 5 |