ABSTRACT

Title of Thesis: A PLANNING MODEL FOR FLEXIBLE-ROUTE FREIGHT DELIVERIES IN RURAL AREAS BASED ON ADJUSTED TOUR LENGTH ESTIMATIONS

Zheyu Li, Master of Science, 2023

Thesis Directed By: Dr. Paul M. Schonfeld, Professor
Department of Civil and Environmental Engineering

Addressing the issue of delivery efficiency and transportation service quality in rural areas, this thesis presents an analysis of total cost of delivery services in regions with low demand density and low road network density. It focuses on designing a cost-effective and efficient freight delivery system, which is crucial for promoting a vibrant rural economy. A flexible-route service model is developed, aiming to improve farm products and other deliveries by optimizing the service zone size and frequency to minimize the average cost per delivered package. The model is tailored for a potential truck operation scenario in the central Appalachian region, serving as a representative case study, with a general formulation of total cost that can be adapted to similar cases elsewhere. Considering the influence of dead-end roads in rural area, this study presents an adjusted formulation of length estimation for Traveling Salesman Problem (TSP) tours based on the

literature review and regression on multiple graphs with road network, and develops a mathematical formulation of total cost, integrating operation and user costs, supported by reasonable assumptions and system constraints. The results from the baseline study suggest that one truck can serve a large service area by exceeding the maximum working hours constraint. This observation is made without considering the potential expansion into a multi-zone system, which might be necessary due to the combined factors of road network complexity and the perishability of farm products. The results from our sensitivity analysis show that a system with a single large truck will have the lowest average cost per package when demand is low. Considering an actual road network, this study also explores the possibility of combining the flexible-route delivery service with self-deliveries and the extension of Vehicle Routing Problem (VRP) with maximum working hour constraint. The study concludes with suggested future research directions in this important domain.

A PLANNING MODEL FOR FLEXIBLE-ROUTE FREIGHT DELIVERIES IN RURAL AREAS BASED ON ADJUSTED TOUR LENGTH ESTIMATIONS

by

Zheyu Li

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2023

Advisory Committee:
Professor Paul Schonfeld, Chair
Professor Cinzia Cirillo
Dr. Xianfeng Terry Yang

Acknowledgements

Throughout my journey towards a Master's degree, I have been fortunate to receive immense support. Foremost, I extend my profound gratitude to my esteemed advisor, Dr. Paul Schonfeld, whose academic insights, encouragement, remarkable personality, and patience have been invaluable to my both academic and life journey. His effort and warm care are deeply appreciated.

I am also immensely thankful to my advisory committee members, Dr. Cinzia Cirillo and Dr. Xianfeng Terry Yang, for their insightful comments and thought-provoking perspectives on my research. My sincere thanks go to the Appalachian Regional Commission (ARC) and its dedicated staff for their steadfast support during my studies. A special note of gratitude is due to Jason Wang at ARC for his generosity and readiness to assist and my senior Dr. Youngmin Choi for providing suggestions and ideas of an extended research topic based on his previous work experience on the topic at ARC.

Lastly, but most importantly, I wish to convey my deepest appreciation to my family for their unwavering support and love. Their support, encouragement, and sacrifices have been the cornerstone of my ability to complete my Master's degree and inspired me to aspire for higher academic pursuits.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background and Motivation

Currently, in many rural areas such as Appalachia, low demand density, long travel distances, and detours due to terrain constraints are major factors limiting freight delivery services provided by major operators. Rural communities face transportation insecurity due to factors such as remoteness, high travel costs, lack of efficient public transportation, and safety issues. Several delivery companies, such as FedEx and UPS, are suspending or requesting extra charges for deliveries in low-density rural markets, which significantly disadvantages rural communities. Under the Justice40 Initiative, the USDOT is prioritizing projects aimed at eliminating transportation insecurity (Justice40 Initiative). Thus, designing a cost-effective freight delivery system is important for a vibrant rural economy.

In addressing delivery challenges in rural areas, it is important to balance route optimization with cost considerations for both users and operators. Traditional methods such as high frequency delivery with fixed route designed for urban setting often struggle in rural contexts due to distinct topographical and demand variables. Advanced route optimization and tour length estimation can offer a substantial economic benefit for an known area with low demand based on approximate tour, eliminating the need for specific point details. The Traveling Salesman Problem (TSP) has emerged as a central topic in related research. The TSP provides a methodological foundation for determining the most efficient delivery routes with minimal driving distance and travel time, making it relevant to rural delivery challenges. The adaptability of the TSP offers significant potential benefits for optimizing rural deliveries.

The TSP is formulated to determine the shortest tour distance for one vehicle visiting all the given *n* points and returning to the origin. It is widely used for modeling delivery and public transit service with small or medium numbers of served points. Solving a TSP problem for each exact route requires a large computational effort since the computing time for finding the global optimal path increases exponentially with number of points. In most scenarios, the demand and exact locations are unknown during the planning and design stage for logistics and public transportation (Choi, 2021). Therefore, estimating typical TSP tour lengths rather than applying TSP algorithms for each exact route is useful for planning vehicle operations, although not for real-time decisions. The approximation of tour length also solidly supports the economic analysis of total cost, which helps reveal the relations among the decision variables, including service headway and size of delivery area, and the demand density in the delivery area.

For rural areas, compared to the suburban and urban areas with denser road networks, a tour length with the same distribution of nodes in same size of service area is longer due to the sparse road network and dead-end roads. Based on the literature reviews, existing tour length models are largely focused on the distribution of nodes, density of nodes, few points, and urban road networks. A proposed adjusted tour length model considering the influence of dead-ends and topology information is more suitable for rural area road networks.

This thesis analyzes the total cost of delivery services in rural areas with low demand density and explores possible solutions based on adjusted tour length approximations. The flexible-route service can help potential delivery scenarios, such as farm product deliveries, by optimizing service frequency and service zone size to minimize costs. The proposed economic model is used to analyze a potential truck operation scenario in the central Appalachian region as a case study. The general formulation of total cost and tour length approximation can be adapted to other similar cases with proper adjustments.

1.2 Research Objectives

The overall objective of this thesis is to develop a tour length estimation relation for rural areas and a flexible-route service revealing various factors affecting service headway, size of service area and delivery cost for rural freight transportation. The methodology has the following features:

1. It refines the distance approximation models developed by Beardwood et al. (1959) by recomputing the circuity factor and focusing on tours of small number of visit points with sparse road network and dead-end roads.
2. It compares the accuracy of length approximation results with multiple heuristic algorithms.
3. It develops the economic model of flexible-route service focusing on minimizing the total cost (user cost + operation cost), while considering perishability of farm products.

4 It jointly optimizes the headway and size of service area for flexible-route deliveries by minimizing the average cost per package through the partial derivative of the cost with size of service area.

5 It explores the sensitivity of relationships between headway, size of service area and average cost in relation to influential input parameters, which represent the range of operating environments.

The remaining parts are organized as follows.

Chapter 2 is the literature review summarizing existing studies in 1) Truck operation and cost 2) Flexible-route service and its application, 3) Traveling Salesman Problem (TSP) 4) the development and summary of tour length approximation for TSP, 5) heuristic algorithm for solving the TSP. The literature focuses on overview of the approximation method and consideration of real-world constraints with proper assumptions. Chapter 3 first develops the TSP approximations. Experiment settings are discussed, including the points distribution, simulation of road networks with dead-ends, sample size, and ordinary least squares (OLS) regression analysis. Then the results of regression are compared with those of the Lin-Kernighan-Helsgaun (LKH) algorithm, which was selected based on a comparison among multiple heuristic algorithms. The optimal TSP length estimation is determined based on the adjusted formulation using statistical analysis. Chapter 4 explores a mathematical formulation of total cost, considering both the operation cost and user cost, with proper assumptions and system constraints. The optimization of minimizing the average cost per package through the partial derivative of the cost with respect to the size of service area is displayed. Chapter 5 presents the case study from the central Appalachian area. Based on the baseline study and sensitivity analysis, the results reveal how various factors affect the service headway, size of service area, and delivery cost. The possible advantage of flexible-route service compared to existing self-delivery service can then be assessed. Chapter 6 summarizes the tasks completed in the thesis and potential extensions for future research.

Chapter 2: Literature Review

The thesis selectively reviewed papers about flexible-route services and papers and book chapters approximating TSP tour lengths based on regressions or heuristic algorithms. Multiple factors are considered for length estimation including 1) number of points 2) distribution of points in rural areas 3) the road network.

2.1 Truck Operation and Costs

Trucks are the main mode of freight transportation in the United States. In 2017, the value of shipments by truck is 12.017 trillions in 2017 dollars, which is about 63.6% of the value of freight transported by all modes (U.S. Bureau of Transportation Statistics, 2023). Such a vast reliance on trucks for transportation has stimulated many studies on their operation and costs, including **basic operating cost** such as fuel cost (Sdoukopoulos et al., 2015; Winebrake et al., 2015; Wu et al., 2021), salary of driver, maintenance of fleet (Topal and Ramazan, 2012), and road condition (Barnes and Langworthy, 2004), **type of vehicle** (Barnes and Langworthy, 2004), **route optimization with cost-effectiveness** (Kulović, 2004).

The cost of fuel can be reduced through the utilization of specific vehicles (e.g. eco-driving) based on the estimated operation cost for road freight transportation in Greece (Sdoukopoulos et al., 2015). When the price of fuel is fluctuating, the estimated elasticity of fuel price can help in computing the cost of single-unit truck activity based on the data from 1980 to 2012 (Winebrake et al., 2015). Besides the vehicles using clean energy (e.g. charging and hydrogenation), the method of battery-swapping for specific truck types is also reducing the cost of fuel, according to a case study in China

(Wu et al., 2021). However, the cost of fuel saved by using new types of trucks requires updates of the entire truck delivery system in the region. Topal and Ramazan (2012) used stochastic integer programming (SIP) for annually scheduled maintenance for a fixed fleet of mining trucks based on the case study of delivering the gold mine in Australia. They proposed a stochastic parameter for maintenance cost due to the significant level of uncertainty and scheduled the available fleet to meet the annual production targets. The output indicated that the maintenance cost value changes non-linearly depending on the size, age and the truck type.

Barnes and Langworthy (2004) based on a fuel price of $1.50 per gallon and other costs in 2003 dollars, found the difference in operation cost between personal vehicles and trucks where personal vehicles average 17.1 cents per mile to operate and trucks average 43.4 cents per mile, not counting costs associated with the driver or travel time. When considering extremely rough pavement and city driving conditions involving frequent stops and starts, the cost also increased proportionally. For delivery service on rural roads, the consideration of dwell time per stop and the vehicle operation speed can be inputted as a part of vehicle operation cost.

Considering the factors affecting the cost-effectiveness, Kulović (2004) computed the coefficients of utilization rate of fleet, time, and path to minimize the loss of time for different organizational activities, suggesting the low demand pattern will reduce the cost-effectiveness of delivery. The estimation of coefficients enabled the change of input parameters, but for a delivery system at planning stage, it is infeasible to obtain detailed operation data sets.

In general, operational costs play a crucial role in management. It is essential to account for both the potential risks associated with increased costs and the opportunities for cost reductions. The relevant components of operational costs can be integrated into the model as a fixed hourly cost with truck and a fixed hourly cost with its capacity. Overall, the unit truck operating cost is assumed to be a linear function of the truck's capacity.

2.2 Flexible-route Services

The characteristics of rural delivery include low demand density, long travel distances, and detours due to terrain characteristics. Flexible-route services can provide a high service quality where the service is needed with low demand density, which is not profitable for setting a fixed route service. Compared to conventional transit service, which requires a higher density of both demands and terminals, the cost of running and maintaining flexible-route service is also lower. In Figure 1, the black line represents the optimized route connecting all the demand points in the service zone with the consideration of the road network, since the shortest Euclidean distance or rectilinear distance is not always available between any two points.

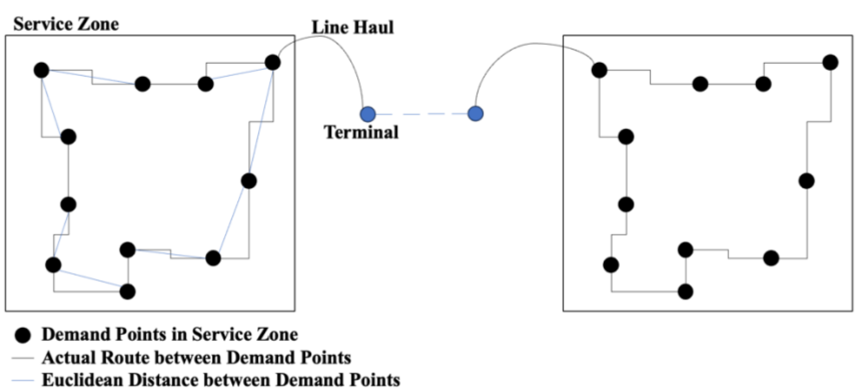

Figure 1. Illustration of Flexible-Route Delivery System in Rural Areas

For national delivery companies, such as UPS, using an On-Road Integrated Optimization and Navigation (ORION) algorithm, provides an optimized route for each driver based on the packages to be picked up and delivered on that day (Holland et al., 2017). Although the details of the algorithm are confidential, they are based on the travelling salesman problem (TSP) with time windows adding practical constraints, such as practical routes avoiding frequent zigzags and initial bounds from delivery history. The side constraints are apparent. For example, some groups of stops must follow a strict, predetermined delivery order and the preference for right-turns over left-turns can reduce the driving time for deliveries (Holland et al., 2017).

Considering the size of vehicles as a decision variable in the total cost function, Chang and Schonfeld (1991) found that for a conventional service for passengers the optimal number of zones and vehicle size should be kept in constant proportion regardless of demand density. For a mixed transportation system, rather than a purely fixed or flexible system, the performance is sensitive to the threshold of the demand densities.

For both users and operators, the cost and the service quality are the essential measures of deliveries, and the time efficiency is the first consideration (Kallio et al., 2000). The freight delivery environment varies between urban and rural areas. The density of transfer points and service frequency based on the demand density is much higher in urban than rural areas. The difference and importance of optimizing the service zone and frequency is highlighted. In flexible route research, Adebisi pointed out that the average demand is small but has considerable variance for doorstep service (Adebisi, 1980). The total cost of variable, flexible service is lower compared to pure fixed or flexible service, which suggests that these variable-type services can perform well for areas with low demand over a long time (Kim and Schonfeld, 2012). To make the objective function responsive to changes in demand, the elasticities of the factors should be considered when computing the net benefit of flexible route service (Kim and Schonfeld, 2015), solving the problem of one terminal connecting multiple regions. The principal objectives are the optimization of service areas and frequencies, as well as the implementation of various types of services, all aimed at reducing overall costs.

However, finding the global optimum for a substantial problem may be quite difficult. Kim and Schonfeld (2015) optimized the service headway and the zone size using the approximate lengths of Travelling Salesman Problem (TSP) tours during the planning model for freight delivery. Zone sizes could be optimized based on capabilities of single trucks while multi-zone systems were also optimized (Kim and Schonfeld, 2012). Han et al. (2022) explored flexible-route bus operations with financial constraints for both break-even and subsidized cases. They verified the global optimizations from the formulation of net benefit function, assuming fluctuating demand, by checking the Karush-Kuhn-Tucker (KKT) condition and used the Sequential Quadratic Programming (SQP) method for optimization. Their significant finding was that increased financial subsidies do not increase welfare proportionally, while lower fixed-cost components of transit operation cost can significantly improve the net benefit of transit operations. Kim and Schonfeld (2022) optimized the service headway and the zone size using travelling salesman problem (TSP) tours in a planning model for freight deliveries based on the Lagrangian relaxation to the integer fleet size. Their numerical analysis indicated that one truck can serve a smaller zone with low demand, and suggested that a multi-zone system with multiple periods is a potential research direction for last-mile delivery.

2.3 Tour Length Estimation

2.3.1 TSP Tour Length Estimation

The average distance between two points in both Euclidean and rectilinear space can be mathematically derived (Choi, 2021). The Euclidean space allows the movements in straight lines between any pair of points and rectilinear space allows the movements in two orthogonal coordinates.

Beardwood et al. (1959) showed that the shortest path through n points is asymptotically proportional to $\sqrt{nA}$ for large n in formulation of $L \approx \beta\sqrt{nA}$ with $0.62 \leq \beta \leq 0.92$ and $\beta \approx 0.75$ for Euclidean space, where L is the estimated TSP tour length of n points in the area with the size of A. Then, Stein (1978) estimated the value of β at 0.765 for Beardwood's formula through Monte Carlo experiments and concluded the minimum number of points for TSP length estimation in a fairly compact and fairly convex area should not less than 5. Multiple researchers estimated the value of coefficient β for TSP tour estimation using regression for small number of n ranging 5 to 50 (Choi and Schonfeld, 2022) and large number of n over 500 (Brunetti et al., 1991), and over 10,000 (Johnson et al., 1996; Lee and Choi, 1994). The range of β is reduced to values between 0.7 and 0.9. Vig and Palekar (2008) derived regression equations for predicting the first four moments of distribution of TSP tour length and concluded the empirical results with the Beta distribution showing excellent fitness for small to moderately sized TSP problems. Besides the number of points, other factors also influence the length of TSP tours, including point distribution, shape of area, and geometry (Chien, 1992; Choi, 2021; Daganzo, 1984). Choi (2021) considered the adjustment factor β and provided a guideline of β for different conditions including 1) whether the depot is in a square area 2) random-to-center conversion factor of points 3) point distribution 4) length-to-width ratio of area.

Transformations of formula $L \approx \beta\sqrt{nA}$ are widely used for estimation. Daganzo (1984) introduced the circuity factor *c* for road network in the formulation of $L \approx \beta c\sqrt{nA}$. Based on this formulation, Merchán and Winkenbach (2019) focused on last-mile problems in urban area instead of Euclidean space and rectilinear space with fixed $\beta = 0.9$. Kou et al. (2022) used standard deviation as a predictor for both Euclidean space and rectilinear space. They emphasized that the shortest Euclidean distance between two nodes might not accurately represent the actual driving distance due to the characteristics of road networks. In a subsequent study in 2023, the same authors introduced the Sammon map (Sammon, 1969) to address non-convex service areas. The technique transforms the size of non-convex service area A to a convex-shaped area A_{Sammon} based on minimizing the error function of Sammon map projection. By preserving pairwise distances of vertices, this transformation to A_{Sammon} indicates a stronger computation time advantage of $\beta\sqrt{nA_{Sammon}}$ relative to LKH or alternative VRP heuristic solvers (Kou et al., 2023).

2.3.2 Approximation Formulations

Besides Beardwood's formulation of approximation $L \approx \beta\sqrt{nA}$, other accurate estimations have been presented in the literature. Chien (1992) argued that $L^* = 2.1\bar{R} + 0.67\sqrt{(n-1)A_1}$ can generate a robust estimation after the comparison of seven estimated models, where $\bar{R}$ represents the average distance between all other vertices and the starting vertex, and A_1 represents the area of smallest rectangle that can cover all vertices. Basel and Willemain (2001) ran the regression on 17 Euclidean instance from TSPLIB (1991) and introduced a log-linear regression formulation of $\ln(L^*) = 1.798 + 0.927\ln\,(std_{RT})$, where std_{RT} is the standard deviation of tour lengths for 20,000 random tours. Their results indicated that the computation of std_{RT} does not require the space to be Euclidean or rectilinear or require any known metric. Cavdar and Sokol (2015) found a distribution-free model for estimation of TSP tour:

$$L^* = 2.791\sqrt{N\left(cstdev_x * cstdev_y\right)} + 0.2669\sqrt{N\left(stdev_x * stdev_y\right)\frac{A}{\bar{c}_x\bar{c}_y}} \qquad (1)$$

where $cstdev_x$ and $cstdev_y$ are the measurement of standard deviation of the absolute distances to the vertices from the horizontal and vertical midpoint lines of the rectangular space, $stdev_x$ and $stdev_y$ are the standard deviations in x and y coordinates, A represents the size of the rectangular space, $\bar{c}_x$ and $\bar{c}_y$ are the average distance of vertices to the horizontal and vertical midpoint lines of the space. Generally, most of the correlations observed between L and $\sqrt{n}$ estimations are still based on the assumption of Euclidean space.

TSPLIP (1991) provides the data in a specific format, which includes information about the problem type (e.g., TSP, Asymmetric TSP, etc.), the type of weight (e.g., Euclidean distance, Manhattan distance, etc.), and the explicit distances between the cities or points. The maps from TSPLIP are widely used for testing the efficiency of tour length estimation model or the performance of heuristic model, and becomes the standard benchmark for TSP solvers. Unfortunately, the official dataset lacks details on network structures and dead-end roads. Its available distance matrix is not adequate for the scope of this research. Consequently, there is a pressing need to create a dataset tailored to meet the specific demands of this study or similar research.

2.3.3 Circuity Factor

A circuity factor c represents the average ratio of actual travel distance to Euclidean distance, based on the computation below:

$$c = \frac{\sum D_n}{\sum D_e} \tag{2}$$

where $\sum D_n$ is the summary of network distances of two points within the road network, and $\sum D_e$ is the summation of Euclidean distances between successive points pairs. For specific cities or regions, the value of c will varies due to road network density and connectivity (Ballou et al., 2002; Levinson and El-Geneidy, 2009), the curve of the Earth for large area within the same hemisphere (Ballou et al., 2002), and vertical height due to mountainous areas (Kweon, 2019). Although tables of average circuity factor with standard deviation are provided in some studies, including (Ballou et al., 2002; Levinson and El-Geneidy, 2009), it will be beneficial to recompute the circuity factor for a specific region.

2.4 Heuristic Algorithms

Heuristic algorithms can be used for solving TSP problems and find a near-optimal solution within required time. Choi and Schonfeld (2022) made a review for TSP tour length estimation of heuristic algorithm. The review in the thesis mainly focuses on algorithms including Greedy Algorithms, Dynamic Programming (DP), Simulated Annealing (SA), Tabu Search (TS), Lin-Kernighan-Helsgaun Algorithm (LKH) and Concorde server as a supplement of their research.

The performance of heuristic algorithms greatly depends on the problem scale and solution quality requirements. A greedy algorithm employs a short-sighted heuristic approach, making locally optimal decisions at each stage, and thus yields globally suboptimal solutions (Cormen et al., 2022). The result is acceptable for the Euclidean TSP but poor for the general symmetric and asymmetric TSP and it can help find the worst tour (Bang-Jensen et al., 2004; Gutin et al., 2002). Both Simulated Annealing and Tabu Search are methods that can potentially yield near-optimal TSP solutions but are computationally intensive. Specifically, they require substantial computational resources due to their iterative and complex nature (Glover and Laguna, 1998; Kirkpatrick, 1983). The Concorde solver uses a branch and bound search strategy, as well as cutting planes to reduce the search space, which is time-consuming but can most closely approach a TSP's global optimum (Concorde TSP Solver). The research based on Concorde improves the difference of runtime cost, such as using partition crossover (Sanches et al., 2017). Compared to the Concorde solver, the LKH algorithm is a highly successful refinement of the 2-opt and 3-opt techniques. It generally outperforms other algorithms for larger problems due to its ability to conduct deeper explorations, but at the cost of increased complexity (Helsgaun, 2000). Helsgaun (2000) also pointed out that providing an initial solution for algorithms as a starter instead of starting from randomly generated one can reduce the computation time.

2.5 Research Gaps

From the literature review, it is found that for rural areas with low demand density, flexible-route services can be a cost-effective method for freight deliveries. Besides the objective function of cost or revenue function, most studies about deliveries focus on the waiting cost of users and the operation cost. Some factors are missing, such as the perishability of farm products and the in-vehicle cost from long driving times in rural area. The relation between headway and size of service area can be revealed with the consideration of all decision variables and baseline inputs.

Most studies on tour length estimation focus on square, rectangular, and circle-based shapes without considering scenarios where points have the same distribution (e.g. normal distribution) across different road networks. For the estimation of TSP tour in rural areas, the sparseness of road network and existence of dead-end roads increases the TSP tour length, which cannot be reflected in existing estimation formulas. For practical applications, estimating average tour length for road network with multiple dead-ends in irregular convex shape becomes of paramount importance for delivery services by vehicles with limited capacity in rural areas. The effectiveness of flexible-route service is highly linked with the estimation model.

Instead of a precise route with minimal length for specific points, the goal of TSP tour length approximation is to provide a general view of tour length for a number of points and basic road network in an area. Approximated tour lengths are useful for the joint optimization of multiple variables, where a quantified length can be applied for revealing the relation between variables. Therefore, the formulation of TSP tour length estimation introduced in chapter 3 will consider the dead-end roads and circuity factor for the accuracy of estimation and economic model in chapter 4 will use the result from the chapter 3. The definitions and units of input variables and other factors are listed as Table 1.

Table 1 Variable Definitions and Baseline Values

Symbol	Variable	Units	Value	Reference
	Decision Variables			
A	Size of Delivery Area	mi^2		
h	Headway	hr		
	Output Variables			
C_1	System Cost of TSP Tour	\$/hr		
C_2	User Cost of Self-delivering	\$/hr		
C_A	Average Cost per Package	\$/package		
C_o	Operating Cost	\$/hr		
C_P	Perishability Cost per Tour	\$/hr		
C_V	In-vehicle Cost per Tour	\$/hr		
C_W	Waiting Cost per Tour	\$/hr		
L	Tour Distance	mi		
N_t	Number of Trucks		Integer	
T	Round-trip Time	hr		
t_p	Total Time of Perishability per Tour	hr		
t_v	In-vehicle Time per Tour	hr		
t_w	Waiting Time per Tour	hr		
	Input Variables			
a	Parameter for Truck Operating Cost	\$/(vehicle·hr)	30	
b	Parameter for Truck Operating Cost	\$/(package·hr)	0.3	
B	Truck Operating Cost ($B = a + bS$)	\$/(vehicle·hr)		
c	Circuity Factor	dimensionless	1.3579	
d	Dwell Time	hr/stop	0.1	
L_X	Line-haul Distance per Tour	mi	10	
N	Modified Number of Stops per Route			Equation 8
n	Number of Stops per Route			

p	Perishability rate	per hr	0.2%	(Blackburn and Scudder, 2009)
Q	Demand Density	Packages / ($mi^2 \cdot hr$)	$2.7*10^{-4}$	
q	Demand	Packages/hr		
S	Truck Capacity	Packages	100	
v	Value of Package	$/package	100	(U.S. Bureau of Labor Statistics, 2023)
v_t	Value of Freight Time	$/(vehicle·hr)	0.415	(Kawamura, 2000)
V	Vehicle Operation Speed	mph	60	
Z	Size of Current Service Area	mi^2	56442.69	(Choi, 2020)

Chapter 3: Methodology

This chapter presents a model of TSP tour length estimation model with dead-ends roads using ordinary least squares (OLS) as the basis. First, the reason of adapting Lin-Kernighan-Helsgaun (LKH) Algorithm is discussed. Then, the simulation setting and factors influencing the approximation are presented. Lastly, a solution procedure based on the improvement of current research is discussed.

3.1 Solution Method

3.1.1 Formulation of Traveling Salesman Problem

The algorithm for solving the exact solution of TSP is formulated as the following integer program:

$$Minimize \sum_{i}\sum_{j} d_{ij}x_{ij}$$

Subject to

$$\sum_{j=1}^{n} x_{ij} = 1 \; \forall \; i = 1, 2, \dots, n \tag{3}$$

$$\sum_{j}^{n} x_{ji} = 1 \; \forall \; i = 1, 2, \dots, n \tag{4}$$

$$u_i - u_j + nx_{ij} \leq n - 1 \; \forall \; i = 2, 3, \dots, n; \; \forall \; j = 2, 3, \dots, n; i \neq j \tag{5}$$

$$x_{ij} = 0 \; or \; 1 \tag{6}$$

$$u_i > 0 \; \forall \; i = 1, 2, \dots, n \tag{7}$$

where *n* is the number of *n* points should be visited, d_{ij} is the distance between points *i* and *j* measuring in Euclidean or rectilinear space, x_{ij} is the binary variables indicating whether the sub-loop route from *i* to *j* is formed in the tour, u_i is the sequence number or order number in which point *i* is visited.

The objective function is minimizing the tour distance for one vehicle visiting all the nodes. Constraints 3 and 4 specify that every node must be visited once. Constraint 5 represents subtour elimination and prohibits solutions consisting of several disconnected tours; it is named the Miller-Tucker-Zemlin constraint (Miller et al., 1960). Constraints 6 and 7 define variables x_{ij} and u_i. When $x_{ij} = 0$, the difference of non-adjacent nodes i, j is computed where the upper bound $\Delta_{ij} \leq n - 1$. When $x_{ij} = 1$, the difference of adjacent nodes i, j is computed where the lower bound $\forall \Delta_{ij} \geq 1$. The design of this constraints avoids enumeration with low efficiency and prevents the inconsistent subtours. However, no loop will form under this constraint since the edge between the last visiting point with order n and departure points with order 1, where $\Delta_{n1} = 1 - n < 0$. Therefore, the dummy node for the departure node is generated which helps the formulation of the enclosed loop for TSP solution.

3.1.2 Lin-Kernighan-Helsgaun (LKH) Algorithm and Comparison of Solution Methods

LKH algorithm is one of the state-of-the-art heuristic algorithms for solving the TSP and its variants, building upon the Lin-Kernighan (LK) method. The LK method, introduced by Shen Lin and Brian Kernighan in the 1970s (Kernighan and Lin, 1970), was one of the first effective heuristics for the TSP. Helsgaun later improved this method, which resulted in the LKH algorithm.

The most up-to-date version of LKH code version is 2.0.10 (November 2022) which can be found from its official website (LKH website). The information about nodes can be edited in a text file and saved as tsp format before generating the optimal results. Multiple parameters (more than 20) can be adjusted, core parameters including 1) **RUNS**: the number of times the algorithm should be run, 2) **MAX_TRIALS**: the maximum number of iterations of the algorithm in each run, 3) **MAX_CANDIDATES**: Controls the number of candidate edges for each city, 4) **MOVE_TYPE**: Defines the k value for k-opt search, and 5) **SUBPROBLEM_SIZE**: Defines the size of the subproblems of the split problem.

The effectiveness of LKH can be reflected from a computational comparison based on replicating the codes from the literature for the best-known algorithms is presented below. Figure 2 indicates the computation time for solving TSP's with same 11, 13, 15, 17, 19, 21, 23, 50, and 100 nodes within the service zone. Using the result from the solver Concorde as the baseline value, Table 2 indicates the % difference between the near-optimal solution of each algorithm and the result from the solver, in which the bold number indicates the smallest tour distance among the outputs.

Table 2 Tour Length (in Miles) Results from Heuristic Algorithms and the Concorde Solver

Nodes	Greedy Algorithm	Dynamic Programming	Simulated Annealing	Tabu Search	LKH	Concorde
11	203	**162**	**162**	**162**	**162**	**162**
13	190	**186**	**187**	**187**	**187**	**187**
15	242	**238**	**238**	242	**238**	**238**
17	269	**234**	236	267	236	236
19	250	**226**	**226**	250	**226**	**226**
21	299	**216**	218	228	**216**	**216**
23	405	**248**	255	299	**248**	**248**
50	591	-	368	405	**368**	**368**
100	1344	-	507	599	**486**	**486**

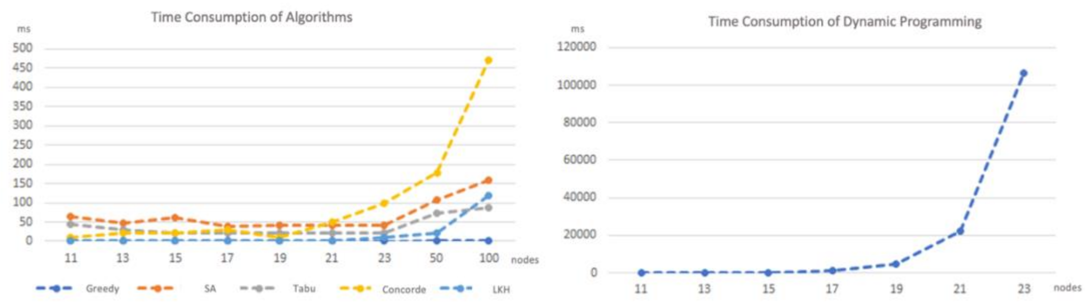

Figure 2. Time Consumption of Each Heuristic Algorithm

In Table 2, due to the Java platform limit on the maximum length of arrays, the use of dynamic planning algorithms for larger datasets is restricted. The efficiency of heuristic algorithms, especially Tabu Search, depends on the expertise of the coder. The suboptimal performance of the Tabu Search in this context does not indicate that it is inferior to other algorithms such as Simulated Annealing. The reproduction of corresponding codes comes from the several studies (Cormen et al., 2022; Glover & Laguna, 1998; Helsgaun, 2000; Kirkpatrick, 1983).

It can be found that when the problem is small, there is almost no difference in the speeds of the six algorithms. As the problem size increases, the difference in the speed of the algorithms become more evident. Dynamic programming requires the longest time among all the algorithms tested here. The greedy algorithm is fast but only reaches locally optimal solutions. Compared to the Concorde solver, except for dynamic programming, heuristic algorithms are faster.

The accuracy of dynamic programming and simulated annealing is satisfactory for smaller datasets. However, when dealing with large datasets, the potency of the LKH algorithm and the Concorde solver becomes noticeable. The operational effectiveness of an algorithm is influenced by several factors, such as design philosophy, implementation process, programming language, computer performance, among others. A rough comparison based on a single case from literature reviews is performed. Based on the review of similar studies (Choi, 2021) and the output of this comparison, the LKH algorithm-based methodology is used for processing the TSP route in this paper's case study. Therefore, the output of LKH algorithm is determined as the near-optimal solution for simulation settings.

3.2 Simulation Settings

Most studies on the TSP in Euclidean space focus on two types of graphs: complete graphs, where every node is directly connected to every other node, and incomplete graphs. In the latter, a 'big M' constraint is often introduced to limit movement between specific corresponding nodes (Boccia et al., 2021). The definition of dead-end road in the graph starts from a node with one edge, passes through the nodes with two edges (if exist), and ends at a node with more than two edges. Thus, to visit the nodes at the end of dead-ends, some nodes must be visited more than once, which avoids the TSP constraint that each node only can be visited once. Nodes visited more than once are only being served when first visited during the entire tour. Therefore, the consideration of dead-end road does not violate a TSP.

The distribution of points is assumed to follow normal distributions considering the low demand in large rural area. Considering the delivery frequent and vehicle capacity, the number of nodes should not be too large. The estimation of $\sqrt{nA}$ is suitable for a single vehicle visiting 5 or more stops in one service area. Coupled with the need to balance operational feasibility and comprehensive data analysis, the range of n in this thesis is from 5 to 150.

A depot or terminal of delivery service is a facility where vehicles start and end their tours, which may in or near a city or a location with a quite convenient road network. The shape of the delivery area *A* should be fairly compact and fairly convex. Thus, the determination of *A* is also displayed.

When the dead-ends are added into the network, the overall network is the combination of tree-structure network and complete graph. With the consideration of dead-end roads in real world, the simulation of a road network based on a graph is as follows:

Step 1: Generate a complete graph with 5-25 points following normal distribution, representing the road network near terminal. Node 0 represents the terminal of departure and return. The gray links represents the Euclidean distance between two nodes. The red dashed line around the graphs represents the minimal convex area covering all the nodes based on the Convex Hull function from Scipy (Virtanen et al., 2020). The area consisting of red dashed line is regarded as the area *A* in the formulation. Figure 3 is the sample visualization of complete graph with convex area.

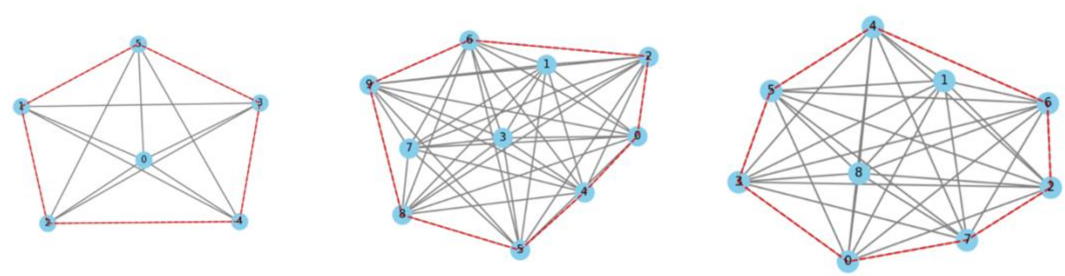

Figure 3. Complete graph for step 1

Step 2: Set the number of added nodes following the rule introduced at the end of this paragraph. In accordance with the established rule governing the relation between new nodes and the initial graph, there is a 30% probability that a new node will be part of a tree-like network or a dead-end structure, while there is a 70% probability that it will become a child node, thereby not leading to a dead-end. It is possible for number of nodes exceeding the preset boundary with 150 nodes or generating the graph to not be consistent with actual road network due to too many dead-ends. Therefore, the number of new nodes in the dead-end should not exceed the number of nodes in the initial complete graph, and the total number of new nodes should not exceed six times the number of nodes in the initial complete graph.

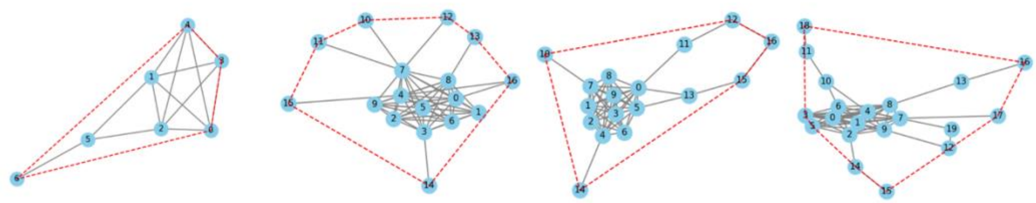

Figure 4. Road network with dead-ends for step 2

Step 3: For each initial graph, iterate 100 times following the description of step 2 and save the information of the graph into the csv file. Iterations are performed 100 times to ensure data diversity while also not consuming excessive computational resources. The definition of each column includes:

1. Number of nodes $\boldsymbol{n}$: the total number of nodes in the graph.
2. Number of edges $\boldsymbol{e}$: the total number of edges in the graph.
3. Total dead-end length $\boldsymbol{L_d}$: the total length of dead-end roads. The definition of a dead-end road in the graph is starting from a node with one edge, passing through nodes with two edges (if they exist), and ending at a node with more than two edges.
4. Total graph length $\boldsymbol{L_g}$: the total length of all edges from the graph.

5. Dead-end length ratio $\boldsymbol{r} = {}^{\boldsymbol{L_d}}/_{\boldsymbol{L_g}}$: the ration between total dead-end length and total graph length.
6. Average degree $d = {}^{2e}/_{n}$: the average degree for each node in the graph
7. Area $\boldsymbol{A}$: the size of the convex area by the definition above.

Step 4: Compute the TSP tour length $\boldsymbol{L}$ based on LKH algorithm for each graph and add to corresponding row of csv file.

Following the steps above, the process begins by generating a complete graph with n=5 nodes 100 times. New nodes are then added with the condition that their number does not surpass six times the original node count, and the count of new dead-end nodes remains below the initial node count. Then the same process for generating a complete graph with $n + 1$ nodes will stop when $n = 25$.

3.3 Improvements of Current Research

3.3.1 Circuity Factor

From Ballou et al. (2002), the circuity factor for the U.S road network at east of the Mississippi River of U.S. is between 1.20 and 2.10, depending on road network density and connectivity.

To compute the circuity factor in a rural area with a sparse rural road network, 500 random points are generated within rural zones of West Virginia and Virginia, specifically in the Appalachian Region. These points are merged with the 2016 shapefile for these states obtained from the U.S. Census Bureau. After eliminating points that fall outside the road network, distances for 436 origin-destination (O-D) pairs are extracted for subsequent computations. A circuity factor of 1.3579 with a standard deviation of 0.4708 is generated based on Equation 2.

3.3.2 Consideration of Dead-ends in the Rural Area Road Network

The adjusted length estimation L is based on Beardwood's formula $\sqrt{nA}$. The coefficient β is reconsidered as a part of number of points n. Due to the existence of dead-end roads, several nodes are traversed more than once. This results in the count of traversals, denoted as N, being higher than the number of nodes n. Based on the information collected from the simulation setting, a relation between N and n is revealed. The goal of the model is building a function of $N(n)$ to estimate the TSP tour length $L \approx \sqrt{NA}$ with $N(n)$.

3.4 Process of Tour Length Approximation

In the final date set, information from 97,400 graphs is collected. To avoid multicollinearity among the variables for regression analysis, it is essential to check the relation between input parameters and carefully choose appropriate ones. In the model, N is derived from or equivalent to n; hence, a strong correlation between N and n is anticipated. Due to this relation, variables such as e and d, which are related to n, will be removed from our considerations. L_d represents the cumulative length of all dead-end roads, while r denotes its proportion in relation to L_g. Given that r encapsulates information about both the total dead-end length and the overall graph length, we incorporate either r or L_d—but not both—as a variable in the model. Therefore, two models are considered: one is the prediction of N using n and r as explanatory variables, and the other is the prediction of N using n and L_d as explanatory variables. To reveal the relation among variables, the ordinary least square (OLS) method is considered. For model 1, it can be expressed as $N_1 = \beta_0 + \beta_1 n + \beta_2 r + \epsilon$. For model 2, it can be expressed as $N_2 = \beta_0 + \beta_1 n + \beta_2 L_d + \epsilon$. The parameter and the statistical information can be found in Tables 3 and 4.

Table 3 Results of Parameters from OLS Regression for Model 1

	Coefficient	Std error	t	$P > \|t\|$	[0.025	0.975]
constant	-30.8001	0.311	-99.077	0.000***	-31.409	-30.191
n	2.9905	0.004	796.660	0.000***	2.983	2.998
r	-8.2495	1.523	-5.416	0.000***	-11.235	-5.264

Table 4 Results of Parameters from OLS Regression for Model 2

	Coefficient	Std error	t	$P > \|t\|$	[0.025	0.975]
constant	-30.8001	0.311	-99.077	0.000***	-31.409	-30.191
n	2.9905	0.004	796.660	0.000***	2.983	2.998
r	-8.2495	1.523	-5.416	0.000***	-11.235	-5.264

Figure 5 shows the QQ plots indicating the residual for normality of both models and Figure 6 shows the histogram of residuals.

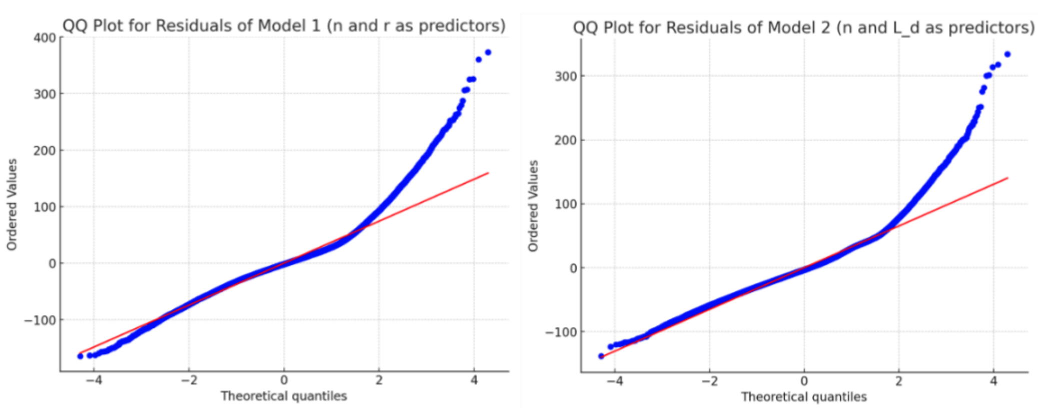

Figure 5 QQ Plot for residual of two models

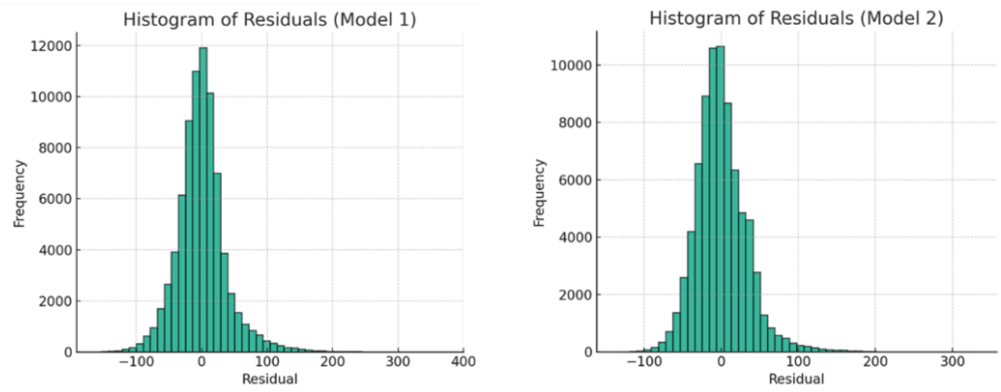

Figure 6 Histogram of residuals of two models

The R^2 is 0.884 for model 1 and 0.914 for model 2. All explanatory variables from both models are statistically significant as their p-values are much less than 0.05. The skewness of model 1 is 0.8653 with the kurtosis of 3.9033 and Durbin-Watson statistic of 1.9942. The skewness of model 2 is 0.9295 with the kurtosis of 3.2590 and Durbin-Watson statistic of 2.0014. The residuals from both models appear to exhibit positive skewness and kurtosis. The value of the Durbin-Watson statistic is around 2, which usually means that the residuals have no autocorrelation.

The QQ plot indicates that the distribution of the residuals is not completely normal, especially in the tail of the QQ Plot. This may indicate some form of heteroskedasticity, nonlinearity, or extreme values in the model. The linear regression function from Tables 3 and 4 for two models cannot describe several conditions when n is small (less than 10) with no dead-end length due to the negative value under the square root $\sqrt{NA}$. For example, considering the case with 10 nodes and no dead-end roads, the value of N is negative.

Based on the literature (Basel Iii and Willemain, 2001), the log transformation of N is used to stabilize variance and handle skewed distributed data to improve the residual distribution of TSP tour estimation. Therefore, the log-transformed model is expressed as $\ln(N_1) = \beta_0 + \beta_1 n + \beta_2 r + \epsilon$ for model 1 and $\ln(N_2) = \beta_0 + \beta_1 n + \beta_2 L_d + \epsilon$ for model 2. The parameter and the statistical information can be found from Tables 5 and 6.

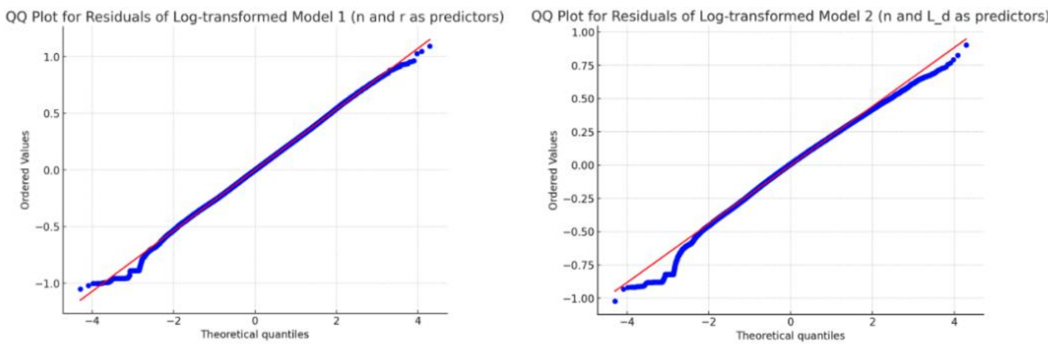

Figure 7 QQ Plot for residual of two log-transformed models

Table 5 Results of Parameters from OLS Regression for Log-transformed Model 1

	Coefficient	Std error	t	$P > \|t\|$	[0.025	0.975]
constant	3.4530	0.0024	1414.14	0.000***	3.4483	3.4578
n	0.0384	0.000135	283.893	0.000***	0.0381	0.0386
r	10.2213	0.0512	199.748	0.000***	10.121	10.322

Table 6 Results of Parameters from OLS Regression for Log-transformed Model 2

	Coefficient	Std error	t	$P > \|t\|$	[0.025	0.975]
constant	3.4229	0.0018	1941.68	0.000***	3.4194	3.4263
n	0.0125	0.000042	296.496	0.000***	0.0124	0.0126
L_d	0.1237	0.0006	206.964	0.000***	0.1225	0.1249

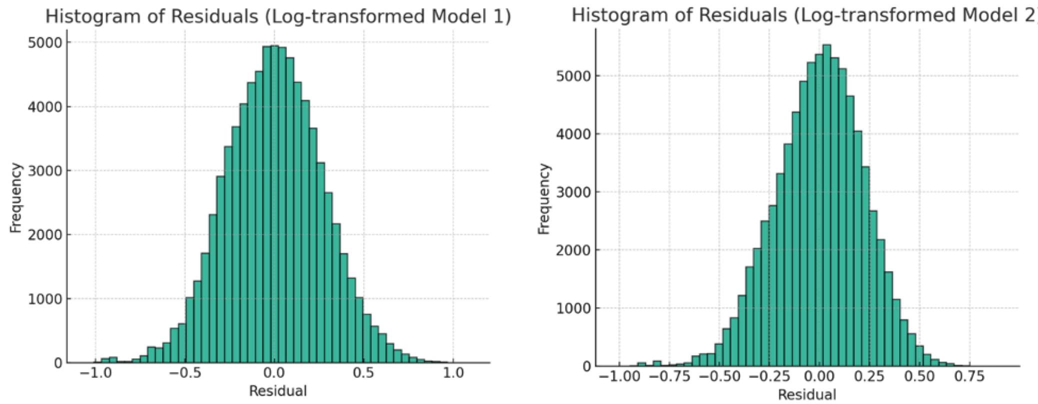

Figure 8 Histogram of residuals of two log-transformed models

After the log transformation, the R^2 is 0.8729 for model 1 and 0.9136 for model 2, i.e., slightly lower than for the linear relation. All explanatory variables from both models are statistically significant as their p-values are much less than 0.05. The skewness of model 1 is -0.0107 with the kurtosis of 0.1088 and Durbin-Watson statistic of 2.0022. The skewness of model 2 is -0.2353 with the kurtosis of 0.2119 and Durbin-Watson statistic of 2.0109. From the above results, after logarithmic transformation of N, the skewness and kurtosis of the residuals of the model are close to 0, which means that the residuals are closer to a normal distribution.

Table 7 R^2 and MAPE of Training Set and Test Set for Log-transformed Model

	Log-transformed Model 1 N and n, r	Log-transformed Model 2 N and n, L_d
R^2 (Training Set)	0.8732	0.9135
R^2 (Test Set)	0.8714	0.9139
MAPE (Training Set)	21.906%	18.070%
MAPE (Test Set)	21.896%	17.941%

After splitting the data into 80% training set and 20% test set, the R^2 and MAPE of both training set and test set can be found in Table 7. The results provide the predictive performance of the two models on the training and test sets. From these results, it is found that the second model predicts better. Compared to the linear relation between N and n, L_d, the log-transformed model can always generate the positive value of N with a wider versatility for length estimation with known network information. Also, compared to the log-transformed model 1, no information about L_g is required for model 2 since the computation of variable r is based on L_d and L_g. Overall, the relation between N and n, L_d can be expressed as follows:

$$\ln(N) = 3.4229 + 0.0125n + 0.1237L_d \tag{8}$$

3.5 Summary

In this chapter, Beardwood's formulation $\beta\sqrt{nA}$ is reconsidered as $\sqrt{NA}$ with a function indicating the relation among N, number of nodes n and dead-end length L_d. A generation function of N is established in equation 8. The models consider the various scenarios, including networks with or without dead-ends for between 5 and 150 nodes. Combining equation 8 with the recomputed circuity factor c, provides a contribution to TSP tour length estimation of rural roads, using the formulation below.

$$L \approx c\sqrt{NA} \tag{9}$$

Chapter 4: Economic Analysis of Flexible-route Freight Deliveries

This chapter presents a mathematical formulation of total cost, combining the operation cost and user cost, with proper assumptions and system constraints.

4.1 Assumptions for Delivery System

The following simplified assumptions are made for a baseline model. These can be revised to fit specific cases.

1. The demand is uniformly distributed within the service area.
2. All demands are served.
3. Packages are homogeneous in size, weight, and density.
4. The value of farm products decreases linearly with time while stored in a refrigerated truck or warehouse with 0 Celsius degrees environment (Blackburn and Scudder, 2009).
5. Truck drivers drive a limited number of hours per day and round trips may extend beyond one day.
6. The truck size is uniform for each self-delivery and flexible-route delivery service.

For assumption 1, the demands of customers are assumed to be non-stochastic and known in advance. Assumption 3 simplifies a real case considered for the planning model. Also, combined with assumption 1, a general formula of in-vehicle time and waiting time can be used during the analysis. Assumption 4 reflects the perishability of farm products. Since the value of farm products decreases over time, a scenario at 0

Celsius degrees is considered, where the value decays linearly overtime, unlike the exponential decay that may be expected in environments exceeding 10 Celsius degrees (Blackburn and Scudder, 2009). However, fairly long driving distances may be acceptable due to the small fleet size and maximum allowable driving time per day, which are reflected in assumption 5. According to assumption 6, the truck size is the same for both self-delivery and multiple-destinations delivery tours.

4.2 Baseline Values

The demand or delivery service is determined by the demand density Q, the size of service area Z, the number of stops n, and the service headway h. Based on the research of chapter 3, the number of stops n will be replaced with N if the information of road networks is known. Currently, the gathering of farm products relies on self-deliveries by farm owners. The demand density Q fluctuates with the peak and off-peak periods for harvesting farm products.

Equation 9 approximates the shortest TSP tour that connects randomly located points n in the service area A.

$$T = \frac{2L_X + L}{V} + nd \tag{10}$$

$$t_v = \frac{T}{2} \tag{11}$$

$$t_w = \frac{h}{2} \tag{12}$$

The truck operation cost varies linearly with the truck capacity S, while labor cost is and related to working time, and fuel consumption is related to both time and truck size. Equation 10 formulates the round-trip time. The dwell time is for loading and unloading at each stop. Due to the assumptions of uniformly distributed demand and homogenous packages, the in-vehicle time and waiting time of packages can be expressed in Equations 11 and 12. To keep the low operation cost, the minimal required number of trucks is:

$$N_t = \frac{T}{h} \tag{13}$$

Based on the 20 farms listed on the website of ASD, the three main farm products are dairy, livestock, and eggs. In the baseline model, the value of a package is estimated from the average price data of U.S. Bureau of Labor Statistics (2023). Using the finding from (Kawamura, 2000) , the value of time for waiting for the shipment is 41.5 \$/hr in 2023 prices for large commercial vehicles. Since that freight time value is based on large shipments, when considering the freight time value for self-delivery, this value is proportionally reduced. The weight per package is set at 50 pounds, based on what one person can carry. With trucks of 5,000 pounds capacity providing flexible-route service, the maximum capacity of one truck is equal to 100 packages and the value of time spent waiting for freight v_t is 0.415 \$/(package·hr). For a practical application, the U.S. Department of Transportation (USDOT) developed values of travel time by modes and business types (USDOT, 2014).

4.3 Cost Function

4.3.1 Cost Function for Flexible-route Delivery Service

The cost function consists of operator and user cost. The operator cost C_o considers the operating hours of trucks and the number of trucks. The user cost consists of perishability cost C_P, in-vehicle cost C_V, and waiting cost C_W. The total cost of flexible-route delivery service is expressed as:

$$C_1 = operator\ cost + user\ cost = C_o + C_P + C_V + C_W \tag{14}$$

The cost components in Equation 14 are formulated in Equations 15-18.

$$C_o = N_t \cdot B = N_t \cdot (a + bS) \tag{15}$$

$$C_P = qpvt_p = QApv(\alpha t_v + \beta t_w) \tag{16}$$

$$C_V = qv_t t_v = QAv_t t_v \tag{17}$$

$$C_W = qv_t t_w = QAv_t t_w \tag{18}$$

Equation 15 indicates that the cost of operating trucks depends on the number of trucks N and the unit cost of operating truck, where the number of trucks is determined from Equation 13. Equation 16 includes the in-vehicle time and waiting time for computing the farm products perishability, where α and β range from 0 to 1 representing the influence of time on freshness. Based on the assumption of the baseline model, α is 1 and β is 1. Equations 17 and 18 express the cost of in-vehicle and waiting time, which depend on the demand density Q, size of delivery area A, value of freight time v_t, and respective spent time t_v and t_w.

4.3.2 Cost Function for Self-Delivery Service

The cost function component of the self-delivery service closely resembles that of the flexible-route delivery service. However, instead of considering the operator cost, the cost of self-driving is considered, where the TSP tour is replaced by round-trip between delivery points and terminal. Accordingly, the truck capacity S, size of current service area Z, and other input variables change with the mode of delivery.

$$C_2 = user\ cost = C_o + C_P + C_V + C_W \tag{19}$$

4.4 System Constraints

In the delivery system, the total number of packages should not exceed the maximum capacity of the truck. Based on the input variables, the relation among the number of stops n, the number of packages per stop u (from the assumption 2 and 3, value of u is equal to 1) and the demand q with headway h is established in Equation 22. Lastly, the number of trucks should be integer for practical applications, but this constraint can be relaxed during the sensitivity analysis.

$$h \leq \frac{S}{QA} \tag{20}$$

$$h = \min\left\{h_{design}, \frac{S}{QA}\right\} \tag{21}$$

$$n \cdot u = qh = QAh \tag{22}$$

$$N_t = integer \tag{23}$$

4.5 Optimization of Flexible-route Freight Deliveries

This section will discuss the mathematical proof of the existence of minimal value of average cost per package within range of value, which is the base of the joint optimization of headway and size of service area from the objective function.

4.5.1 Optimization of Flexible-route Delivery Service of Planning Stage

The optimal service headway h can be found by differentiating the objective function with respect to h. The headway h is replaced with the maximum allowable headway $h = \frac{S}{QA}$ to deal with the low demand from the baseline input variable (12, 13, 15). The average cost per package C_A is:

$$C_A = \frac{C_1}{QA} = \left(\frac{(a+bS)c}{VS} + \frac{(pv+v_t)c}{2V}\right)\sqrt{NA} + (pv+v_t)\frac{S}{2QA}$$
$$+ \frac{2L_X(a+bS)}{S} + \frac{Nd(a+bS)}{S} + (pv+v_t)(\frac{L_X}{V} + \frac{Nd}{2}) \qquad (24)$$

Replacing $k_1 = (\frac{(a+bS)c}{VS} + \frac{(pv+v_t)c}{2V})\sqrt{N}$, $k_2 = (pv_n + v_t)\frac{S}{2Q}$, $X = A^{-\frac{1}{2}}$, the differentiating process is expressed in Equations 17 and 18.

$$\frac{\partial C_A}{\partial A} = \frac{1}{2}\left(\frac{(a+bS)c}{VS} + \frac{(pv+v_t)c}{2V}\right)\sqrt{\frac{N}{A}} + (-1)(pv+v_t)\frac{S}{2QA^2} = X\left(\frac{1}{2}k_1 - k_2X^3\right) \quad (25)$$

$$\frac{\partial^2 C_A}{\partial A^2} = -\frac{1}{4}\left(\frac{(a+bS)c}{VS} + \frac{(pv+v_t)c}{2V}\right)\sqrt{N}A^{-\frac{3}{2}} + 2(pv+v_t)\frac{S}{2Q}A^{-3} = X^3\left(-\frac{1}{2}k_1 + 2k_2X^3\right) \quad (26)$$

Since k_1, k_2, X are positive, by setting Equation 25 $\frac{\partial C_A}{\partial A} = 0$ and Equation 26 $\frac{\partial^2 C_A}{\partial A^2} > 0$, the global optima of the size of service zone can be determined, as Equation (19) expresses.

$$A = \left(\sqrt[3]{\frac{k_1}{2k_2}}\right)^{-2} = \left(\sqrt[3]{\frac{\left(\frac{(a+bS)c}{VS} + \frac{(pv+v_t)c}{2V}\right)\sqrt{N}}{(pv+v_t)\frac{S}{Q}}}\right)^{-2} \qquad (27)$$

The headway h can be found based on Equation 20. Based on the result, the optimal size of service zone and headway is determined for the planning stage. To verify this optimal value is within the range of solvable values, a discussion about necessary condition of the global optimality is needed. By setting Equation 26 to be positive, the value of the Equation 25 is increasing as X^3 increasing, where its increase is due to the decrease of the service zone A. Therefore, the upper limit (maximum) of the service zone A can be determined when the Equation 26 is equal to zero.

$$A_{MAX} = \left(\sqrt[3]{\frac{k_1}{4k_2}}\right)^{-2} = \left(\sqrt[3]{\frac{\left(\frac{(a+bS)c}{VS} + \frac{(pv+v_t)c}{2V}\right)\sqrt{N}}{(pv+v_t)\frac{2S}{Q}}}\right)^{-2} \tag{28}$$

By comparing the result of maximum size of service area and optimal size of service area, the existence of an optimal value is guaranteed, since $\frac{A_{MAX}}{A} = \sqrt[3]{16} > 1$ always hold regardless of the input parameters.

4.5.2 Optimization of Flexible-route Delivery Service with Self-Delivery

In analyzing practical cases, given information on delivery points (i.e., geocoordinates data, demand), it is straightforward to obtain an actual Traveling Salesman Problem (TSP) tour within the service area using the Lin-Kernighan-Helsgaun (LKH) algorithm. Due to the characteristics of the rural road network, significant detours may be required to reach demand nodes located at dead-ends or far from other existing nodes when considering the TSP tour visiting the nodes in the service area. Therefore, the algorithm merging the current self-delivery service and flexible-route service, considering that the number of trucks is a fixed value for existing rural communities, aims to minimize the total cost of the delivery service. Figure 9 illustrates this algorithm.

Since the proposed flexible-route delivery service does not account for users' willingness to switch delivery methods, the selection of delivery method for each user is primarily based on economic considerations. This algorithm also can indicate the marginal cost of adding or dropping one or more nodes in the flexible-route service based on the baseline inputs and geocoordinates.

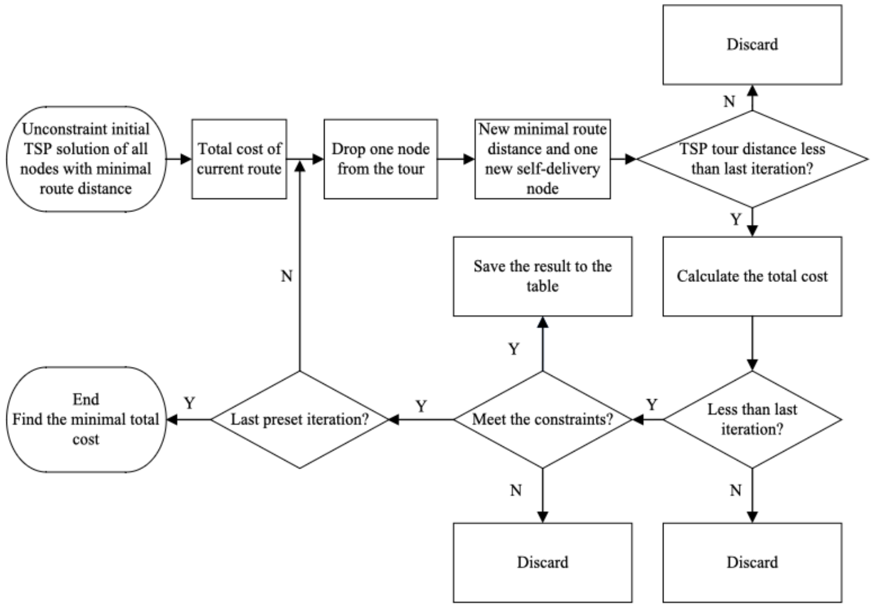

Figure 9 Algorithm combining flexible-route service and self-delivery

With the results of initial optimal TSP tour length and the total cost of flexible-route service, to avoid detours to points outside the optimal TSP tour and find a better solution, the algorithm starts by dropping one random point of the TSP tour. If total cost decreases and satisfies the constraints, the result is saved. The iterations end when the stopping criteria are satisfied, e.g., by reaching the maximum available number of vehicles or when the remaining nodes in the TSP tour drop below a preset number.

Chapter 5: Case Study and Sensitivity Analysis

In this chapter, a case study evaluates the performance of flexible-route delivery service mixing with current self-delivering service and estimates the difference between the planning stage model and real delivery points in the rural road network. A truck delivery system is specified with a mathematical formulation, which can be used for similar farm product logistic scenarios in rural areas.

5.1 Dataset

The Appalachian Region of the U.S., as defined by Congress in 1965 through the Appalachian Regional Development Act (ADRA) which established the Appalachian Regional Commission (ARC) to oversee economic development in the Region, is a 205,000 mile^2 region covering West Virginia and parts of 12 other states that follows the spine of the Appalachian Mountains from southern New York to northern Mississippi. This region has encompassed numerous economically distressed areas where accessibility to transportation and economic opportunities remains a pivotal factor in fostering an equitable and sustainable economic growth (Shayanfar et al., 2019). In 2016, the Central Appalachian Food Enterprise Corridor Project was launched by the Appalachian Sustainable Development (ASD) organization through a grant funded by ARC. ASD is a non-profit organization with a focus on sustainable agriculture development in the central Appalachian region (Choi et al., 2022). The project seeks to develop a coordinated local foods distribution network throughout Central Appalachia which will connect established and emerging producers in Ohio, West Virginia, Tennessee, Southwest Virginia, and Eastern Kentucky to wholesale distribution markets. The goal of the ASD project is to bridge the gap between producers and wholesale or retail outlets seeking local farm products. This endeavor aims to spark economic opportunities within the food and agriculture sector in underserved communities (Choi et al. 2022). According to the project data provided by ASD in 2020, two refrigerated trailers were used to provide the cross-state routes weekly. Meanwhile, ASD is exploring potential routes to expand its delivery service within the Appalachian area, aiming to fill the delivery gaps in the existing routes. Taking into account the maximum available service time for both current over-state delivery services and potential new ones, it is possible to operate short delivery trips for local communities within the existing service gaps.

5.2 Existing Delivery Services

There are 6 aggregation points, including farm market and workshops, collecting the farm products from 90 growers within the Appalachian region. The demand in the entire service area can be computed from the baseline inputs in Table 1.

$$q = QZ = 15.24\, packages/hour \qquad (28)$$

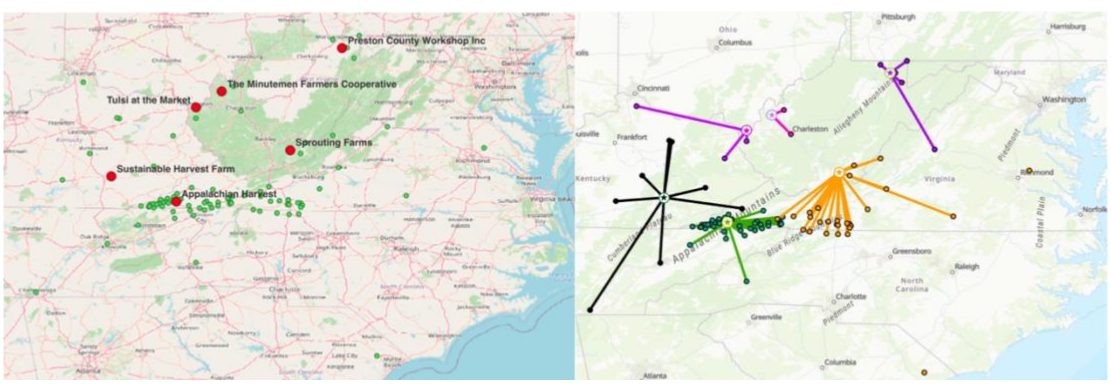

Figure 10 Illustration of location of points and self-delivery Euclidean distance

Figure 10 illustrates the locations of both the aggregation points and growers, as well as the current association of growers with their corresponding aggregation points. Appalachian Harvest (green point with 46 self-delivery routes) and Sprouting Farms (Orange points with 27 self-delivery route) cover most of the points within the service area. The assumptions of Stein's formula include a fairly compact and fairly convex service area with more than five points that are fairly uniformly distributed. Otherwise, that approximation is unreliable (Stein, 1978). The case study focuses on these two aggregation points, which satisfy Stein's assumptions. Both aggregation points are situated within a reasonably compact and convex service zone, including far more than five stop points. Prior to assessing the performance of the flexible-route delivery service, a discussion on point attribution, grounded on the isochrones of the aggregation points, is necessary.

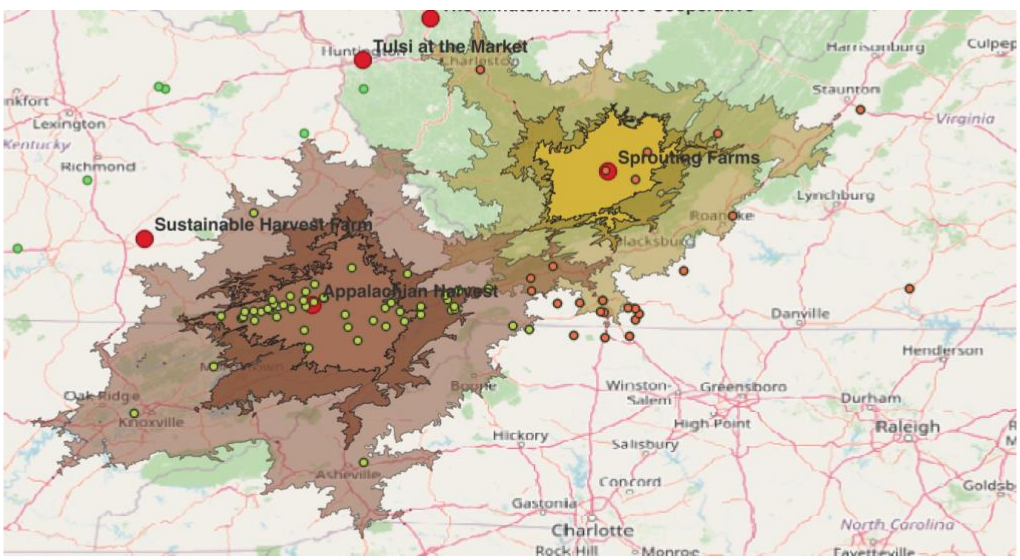

Figure 11 Illustration of 1-hour, 2-hour, and 3-hour isochrone of aggregation points

The visualization of the isochrone is based on the QGIS, using data accessed from Travel Time API. For Appalachian Harvest, located at the intersection of US-23 and US-58, there are 36 points within the 1-hour isochrone, 7 additional points between 1-hour and 2-hour isochrone, and 9 additional points between 2-hour and 3-hour isochrone. For Sprouting Farms, situated in the mountainous area of central Appalachia, there are 3 points within the 1-hour isochrone, no additional points between 1-hour and 2-hour isochrone, 10 points between 2-hour and 3-hour isochrone (including 4 points overlapping with another 3-hour isochrone), and 9 points located close but outside the 3-hour isochrone. It is sensible to maintain the current attribution of growers, taking into account the transportation convenience for nearby growers and preserving the diversity of farm products for both aggregation points. This consideration also incorporates growers within a 3-hour isochrone. Following this, the process of optimizing the total cost of both the flexible-route delivery service and self-delivery service is performed. In the case study, Area 1 represents Appalachian Harvest region (suburban area). Area 2 represents Sprouting Farms region (mountainous area)

The driving distance between aggregation points and each corresponding grower can be computed based on the distance matrix from Google Map API. The maximum allowable service headway of self-driving can be calculated through $h = \frac{S}{q}$. Accordingly, the total user cost of perishability cost C_P, in-vehicle cost C_V, and waiting cost C_W are determined based on the Equation 15-18.

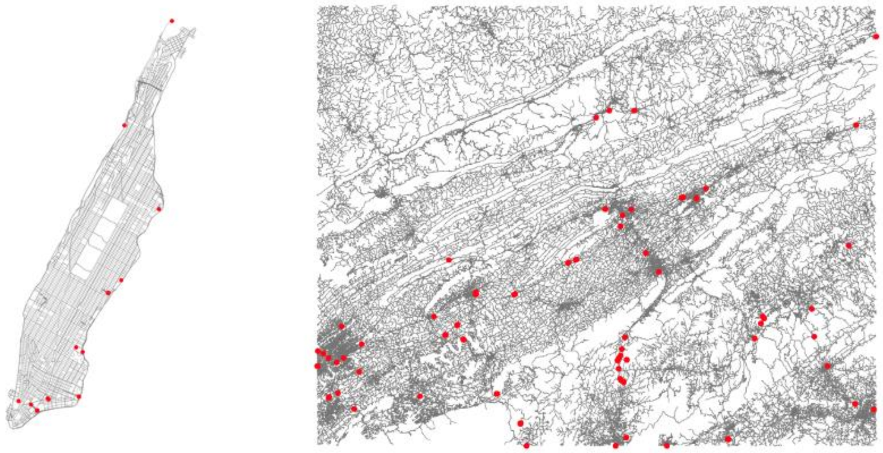

Figure 12 Visualization of dead-ends in urban and rural area (Area 1 of case study)

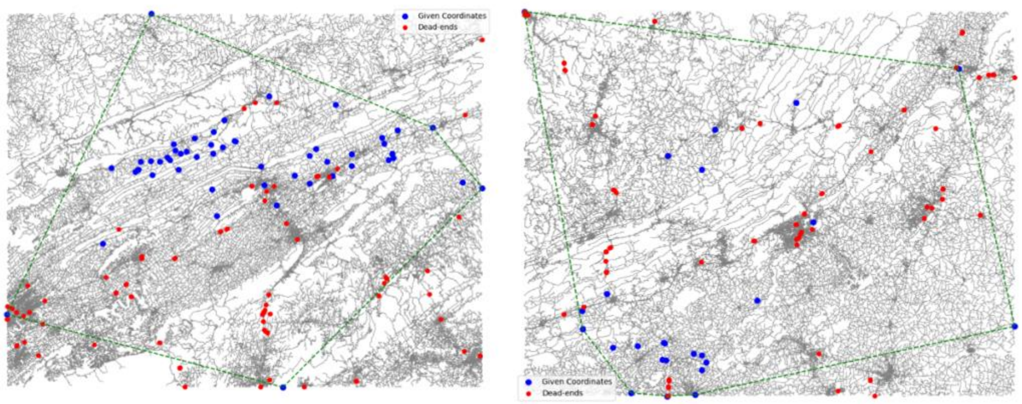

Figure 13 Visualization of coordinates, dead-ends and convex area 1 and area 2

The transformation of number of nodes from *n* to *N* can be found through the road network from the API provided by Open Street Map. Based on the definition of dead-ends roads, figure 12 is the visualization of dead-ends of Manhattan Island, New York and an example of Appalachian Harvest region, a suburban area which will be mentioned in the case study. The red dots represent the dead-ends in the road network of possible edge of TSP tour, where only some of them will be visited if one TSP tour is determined. In other words, this value can be used for determining the upper boundary of an approximated TSP tour. Then, we must convert the real road network with *n* nodes (shown in blue) into its minimal convex area (indicated by green dashed lines) and document all the related information. Figure 13 is the visualization of the results.

For area 1 with 52 delivery points, its service area size is 10,694.68 square mile and the length of dead-end road is 9.97 miles. For area 2 with 25 delivery points, the size of service area is 15,706.64 square mile and the length of dead-end road is 10.84 miles. The lower boundary of estimated TSP tour length is not considering the existence of dead-end roads. Therefore, the range of approximated tour length can be computed with Equation 9 based on the length of dead-end roads within a service area with known number of demand points. To indicate the advantage of flexible-route service, the longest dead-end length is considered for optimization to reflect the worst case, such as visiting all the nodes in the dead-end roads.

Based on Equation 9, without considering the dead-end roads, the TSP tour length of current service size for area 1 is 1069.8 miles and for area 2 is 1216.9 miles. The maximum of TSP tour length of current service size, considering all the dead-end roads, with same number of delivery points for area 1 is 1958.2 miles and 1860.4 miles for area 2.

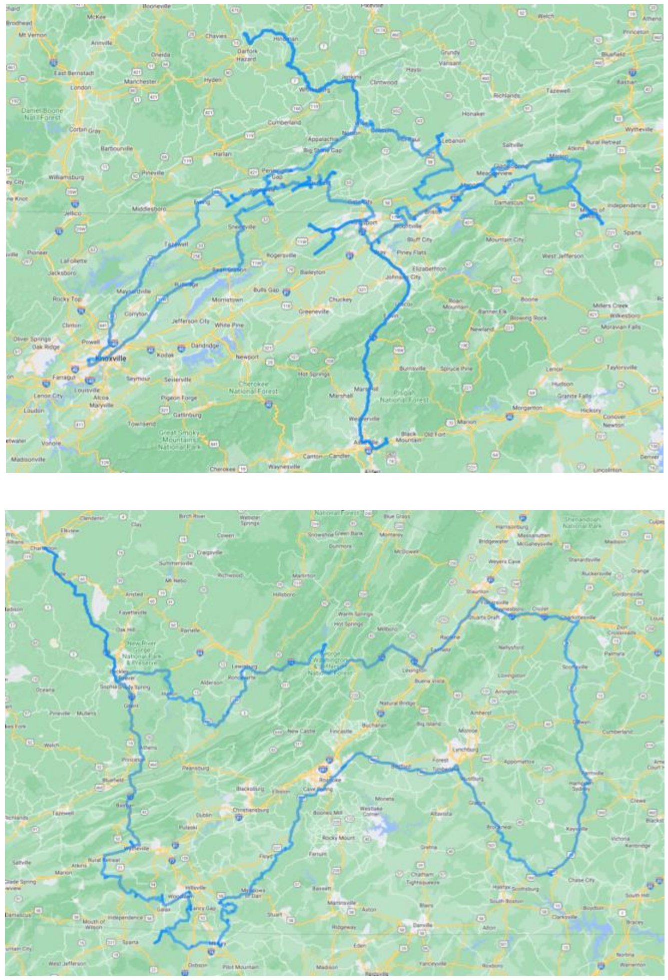

Figure 14 Optimal TSP tour visualization

The visualization of optimized TSP tour by LKH algorithm generated using the Google Map API indicates that several route segments are dead-end roads. Within the context of this thesis, dead-end roads are defined as routes or edges connected to nodes that have only a single segment leading in or out. Consequently, nodes with two or more connecting segments are not classified as dead-ends, even if their impact on the TSP tour is analogous to that of dead-end roads. Based on the result, the minimal tour length with current nodes is 1040.8 miles for area 1 and 877.5 miles for area 2.

5.3 Results of the Planning Model

Using the baseline inputs listed in Table 1, the results for flexible-route delivery service are shown in Tables 8 and 9.

Table 8 presents the results for the baseline case with integer fleet size and size of service zone for two cases. Two aggregation points feature different transportation environments: one is located in a suburban area while the other is situated in a mountainous area with a sparse road network. Area 1 represents Appalachian Harvest region (suburban area). Area 2 represents Sprouting Farms region (mountainous area). Based on the same methodology introduced in the chapter 3.3.1, the resulting circuity factor is 1.35 for the service area 1 and 1.5 for the service area 2.

While the total and average costs of self-delivery are lower than those of flexible-route service, the number of trucks in real-world scenarios tends to exceed the figures obtained from our comparison. This is often the case when the number of trucks matches the number of stops, especially as most self-delivery services are operated by the users themselves. This situation significantly increases operating costs. The results presented are based on the assumption that the minimum number of trucks is being used. Consequently, Table 8 displays the minimum cost for self-delivery. In real-world scenarios, achieving a long total delivery time with small trucks fleet in order to minimize costs is challenging. The result also indicates the optimized average cost and total cost of flexible-route service compared to the current self-delivery.

Compared to flexible-route service, the total delivery time is significantly higher for self-delivery service, but the average in-vehicle cost is less. This result inspires the idea of combining self-delivery service and flexible-route delivery service, to avoid unnecessary detour of flexible-route service for access the nodes far from the terminal.

Table 8 Baseline Case and Optimized Results for Delivery Routes

	Existing Self-Delivery with Minimal Operating Trucks		Flexible-route Service Integer Fleet with Current Area Constraint	
	Area 1	Area 2	Area 1	Area 2
Headway, h * (hr)	41.8*	42.5*	38.6	33.8
Size of Service Zone, A * (mi^2)	10694.68*	15706.64*	10694.68	15706.64
Total delivery distance, $\sum L$(mi)	2,178.1	2,381.6	1982.2	1860.4
Total delivery time T(hr)	41.8	42.5	38.6	33.8
Package Demand q (packages/hr)	2.89	4.24	2.88	4.24
Number of trucks, N	1**	1**	1	1
Operating Cost, C_O ($/hr)	60	60	60	60
Perishability Cost, C_P ($/hr)	12.54	19.48	22.27	28.70
In-vehicle Cost, C_V ($/hr)	0.96	2.99	23.11	29.78
Waiting Cost, C_W ($/hr)	25.07	37.42	23.11	29.78
Total Cost, C($/hr)	98.57	119.89	128.49	148.26
Average Cost, C_A ($/package)	34.14	28.27	44.50	34.96

* The headway and size of self-delivery service zone is used for determining the package demand; therefore, the same size as flexible-route service zone is considered.
** The minimal number of trucks is calculated based on the round-delivery time over the headway.

Table 9 Optimal Solution of Baseline Study

	Optimal Solution		Optimal Solution with Integer Constraint	
	Area 1	Area 2	Area 1	Area 2
Headway, h * (hr)	43.5	41.7	35.0	30.5
Size of Service Zone, A * (mi^2)	8514.8	8882.3	8514.8	8882.3
Total delivery distance, $\sum L$(mi)	1768.7	1660.0	1768.7	1660.0
Total delivery time T(hr)	35.0	30.5	35.0	30.5
Package Demand q (packages/hr)	2.30	2.40	2.30	2.40
Number of trucks, N	0.80	0.73	1	1
Operating Cost, C_O ($/hr)	48.29	43.88	60	60
Perishability Cost, C_P ($/hr)	18.05	17.31	16.10	14.63
In-vehicle Cost, C_V ($/hr)	16.70	15.18	16.70	15.18
Waiting Cost, C_W ($/hr)	20.75	20.75	16.70	15.18
Total Cost, C($/hr)	103.79	97.13	109.50	104.98

Average Cost, C_A (\$/package)	45.15	40.50	47.63	43.78

A very low demand density Q of 2.7×10^{-4} package/sq. mile/hour is used here, which is roughly 0.01 pound/sq. mile/hour, representing the actual non-peak demand in the rural area. To find the influence of peak demand due to agricultural harvest, a sensitivity analysis should be performed. The case of Appalachian Harvest characterizes a service area in a suburban location, featuring a dense road network and a large number of stops. In contrast, the Sprouting Farms scenario represents a mountainous area where the road network is less dense and there are fewer stops. The number of trucks in the unconstrained case indicates that the fleet with more than one truck may be beneficial for finishing the delivery service within one day, even if the total cost in \$/hr is higher than for the one-truck fleet. In this case, another algorithm about Capacitated Vehicle Routing Problem (CVRP) should be adapted for route design.

5.4 Sensitivity Analysis of Case Study

5.4.1 Effect of Demand Density (in package/sq. mile/hr)

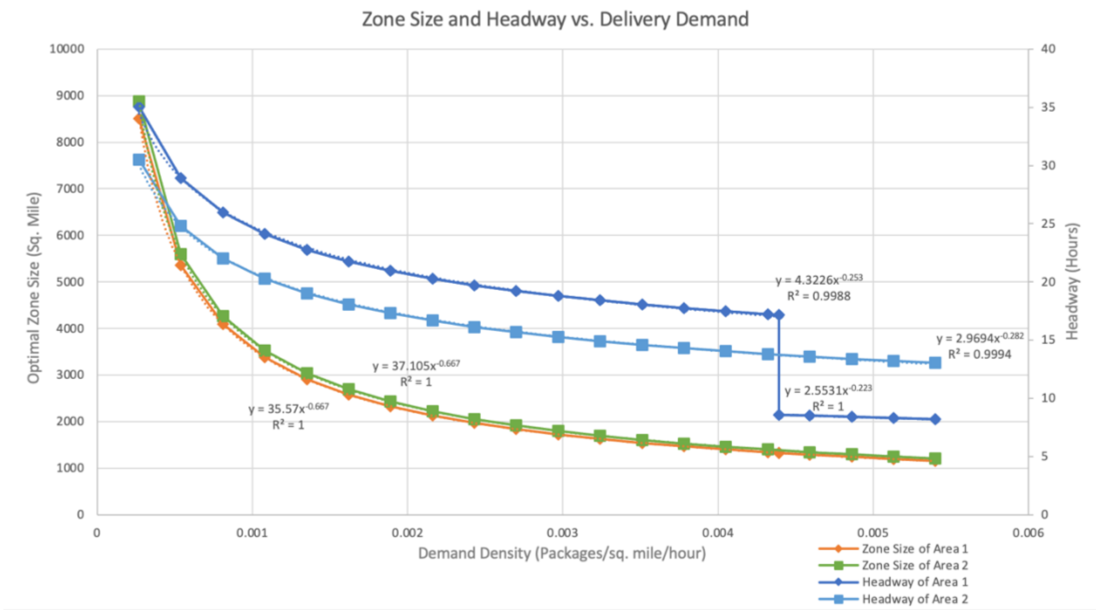

Figure 15 Zone size and headway vs. delivery demand

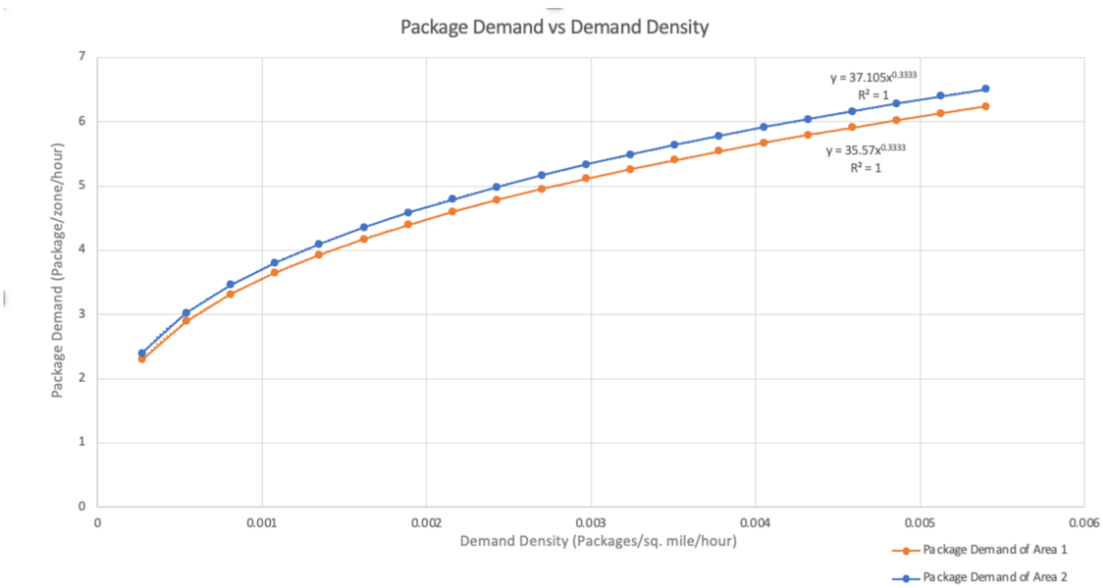

Figure 16 Package demand vs. demand density

In the sensitivity analysis, the integer fleet constraint is strictly applied to reflect the operation of flexible-route service in real cases. In Figure 15, the optimal service zone size and headway decrease as demand increases, following the exponential trendline, indicating that the rate of decrease is faster for optimal size of service area than the headway. The demand model is simplified as a power function. Although the package demand is larger for area 2 than area 1, as indicated in Figure 16, due to more stops in area 1, the sudden decrease of headway occurs when the fleet size increases from 1 to 2. This result suggests that it is possible to operate more than one truck to service high demand area or low demand area with occasional peak demand such as harvesting.

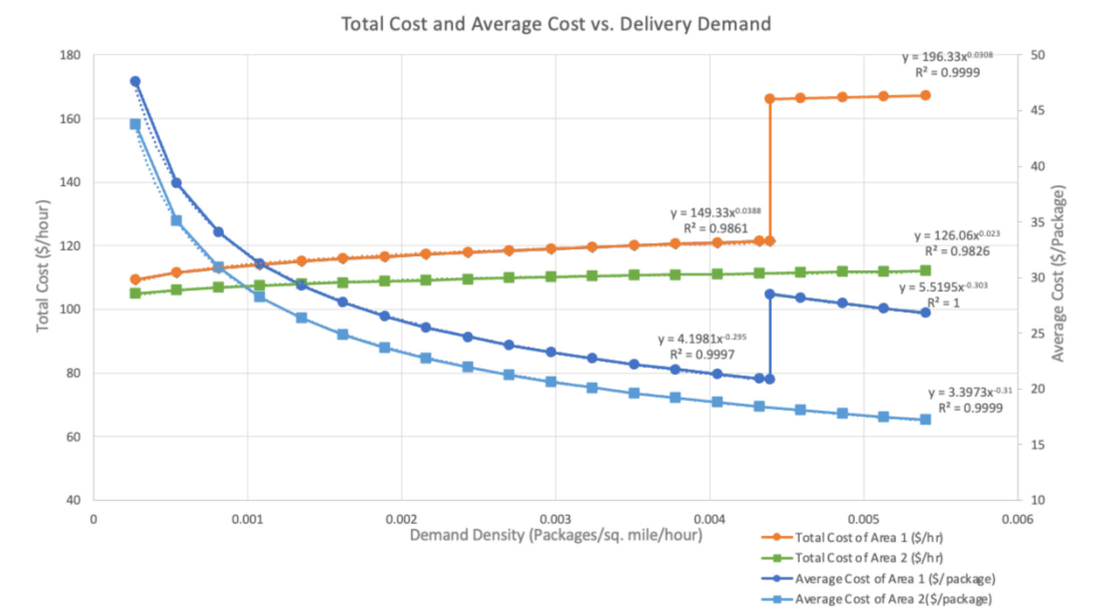

Figure 17 Total cost and average cost vs. demand density

Figure 17 illustrates that the rate of decrease for the average cost per package surpasses the rate of increase for the total cost per hour as demand rises. This indicates a high cost per package when the demand is low in the rural areas. Due to the integer constraint of fleet size, when the cumulated total demand within the round-trip time in the service area exceeds the capacity of one truck, additional truck should be used. When more than one truck is operating, both total cost and average cost increase significantly. This reveals the reason why the result about mixing self-delivery and flexible-route service is keeping a pure flexible-route service based on the baseline input. As Figure 14 indicates, the headway reduced by half with the increases of total cost. Figure 18 indicates the consist of user cost and operation cost for area 1. When an additional truck is placed into operation, the decrease of waiting cost and perishability cost is smaller than the extra cost of truck operation.

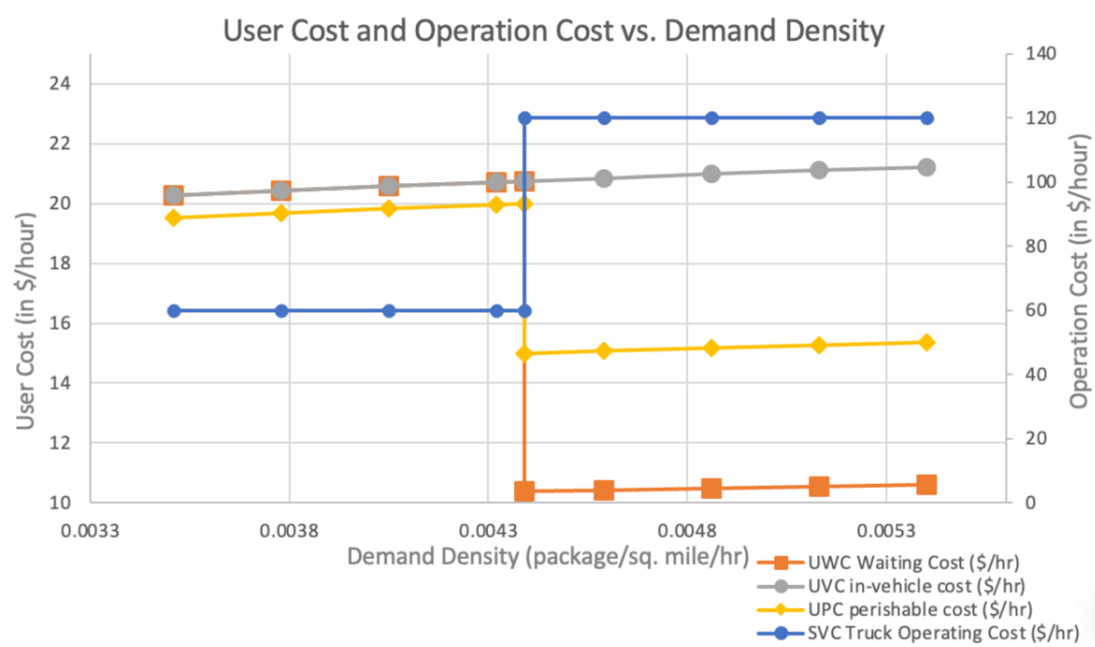

Figure 18 User Cost and Operation Cost vs. Demand Density

5.4.2 Effect of Demand (in package/hr)

To provide a more general view of optimal headway and zone size, the package demand is varied from 2 packages/hr up to 40 packages/hr in Fig. 18?. The demand density is the combination of demand density in packages/sq. miles/hr and zone size in sq. miles. The required fleet size changes as Figure 19 indicates. For area 1, the fleet size increases from 1 to 2 at 5.82 packages/hr and increases from 2 to 3 at 23.90 packages/hr. For area 2, the fleet size increases from 1 to 2 at 8.42 packages/hr.

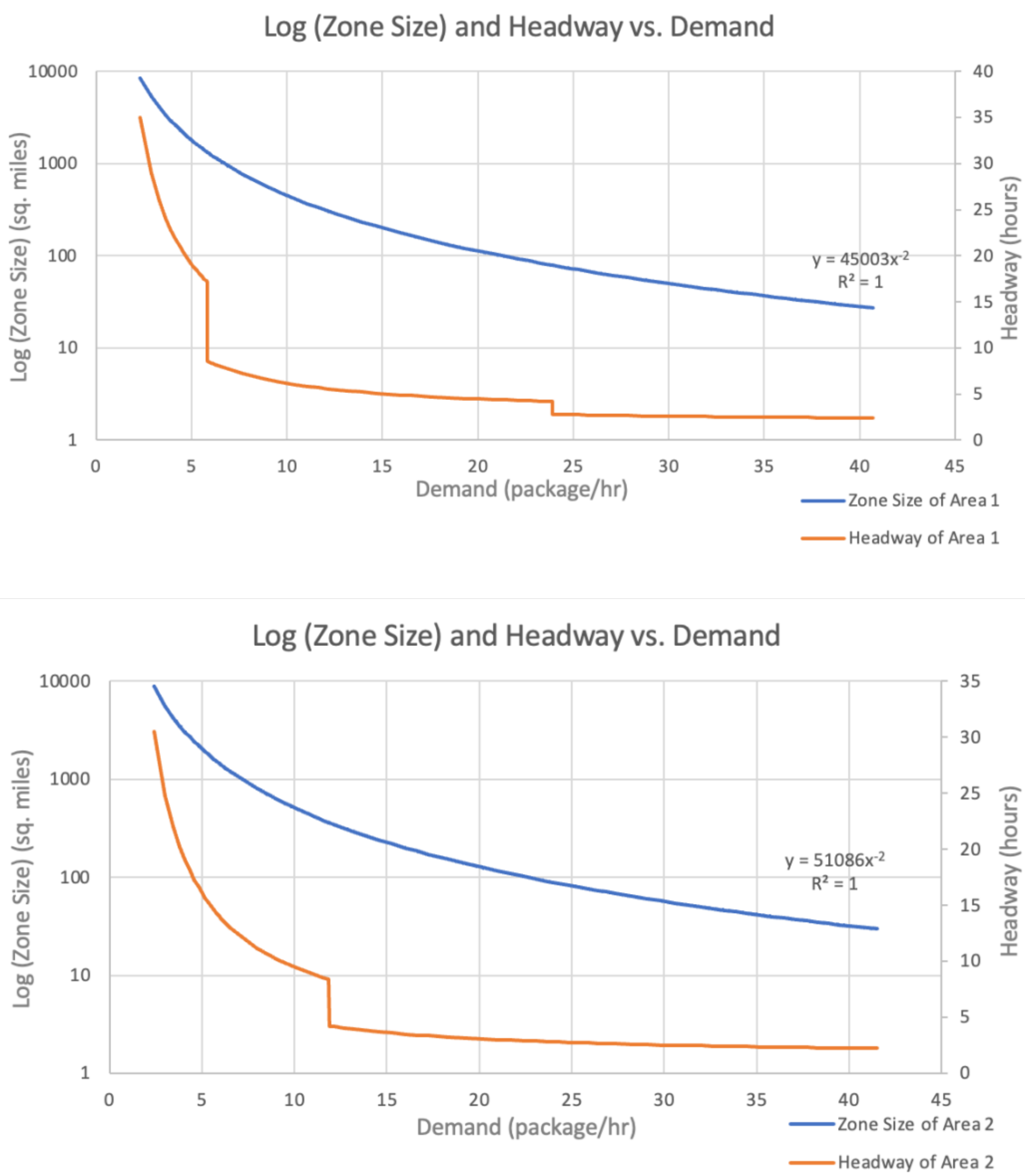

Figure 19 Log zone size and headway vs. demand density

5.4.3 Effect of Truck Size (in packages/truck)

Figure 19 suggests that, at low demand levels, the average cost per package is minimized when larger trucks with a capacity of 120 packages are used, compared to the existing trucks with a capacity of 100 packages, based on the baseline inputs. This occurs when the demand density is low, the larger trucks will be more cost-effective compared to small trucks, as Table 10 suggests. The headway and zone size increase as truck capacity increasing.

Table 10 Effect of Truck Size

Truck Capacity of Area 1 (packages/truck)	60	80	100*	**120**	140
Headway (hours)	28.8	32.2	35.0	**37.5**	39.6
Zone Size (sq. miles)	5304.1	6959.5	8514.8	**9985.5**	11384.6
Package Demand (packages/zone/hr)	1.43	1.88	2.30	**2.70**	3.07
Operating Cost, C_O ($/hr)	48.00	54.00	60.00	**66.00**	72.00
Perishability Cost, C_P ($/hr)	8.56	12.55	16.70	**20.95**	25.27
In-vehicle Cost, C_V ($/hr)	8.56	12.55	16.70	**20.95**	25.27
Waiting Cost, C_W ($/hr)	8.25	12.09	16.10	**20.20**	24.36
Total Cost, C($/hr)	73.36	91.19	109.50	**128.10**	146.89
Average Cost, C_A ($/package)	51.23	48.23	47.63	**47.52**	47.79
% Change of Average Cost	7.56	1.89	0	**-0.24**	0.33

Truck Capacity of Area 2 (packages/truck)	60	80	100*	**120**	140
Headway (hours)	24.7	27.8	30.5	**32.8**	34.8
Zone Size (sq. miles)	5533.0	7259.9	8882.3	**10416.5**	11876.0
Package Demand (packages/zone/hr)	1.49	1.96	2.40	**2.81**	3.21
Operating Cost, C_O ($/hr)	48.00	54.00	60.00	**66.00**	72.00
Perishability Cost, C_P ($/hr)	7.65	11.33	15.18	**19.14**	23.17
In-vehicle Cost, C_V ($/hr)	7.65	11.33	15.18	**19.14**	23.17
Waiting Cost, C_W ($/hr)	7.37	10.92	14.63	**18.45**	22.33
Total Cost, C($/hr)	70.67	87.57	104.99	**122.72**	140.67
Average Cost, C_A ($/package)	47.30	44.67	43.78	**43.64**	43.87
% Change of Average Cost	8.05	2.04	0	**-0.33**	0.21

* Baseline results

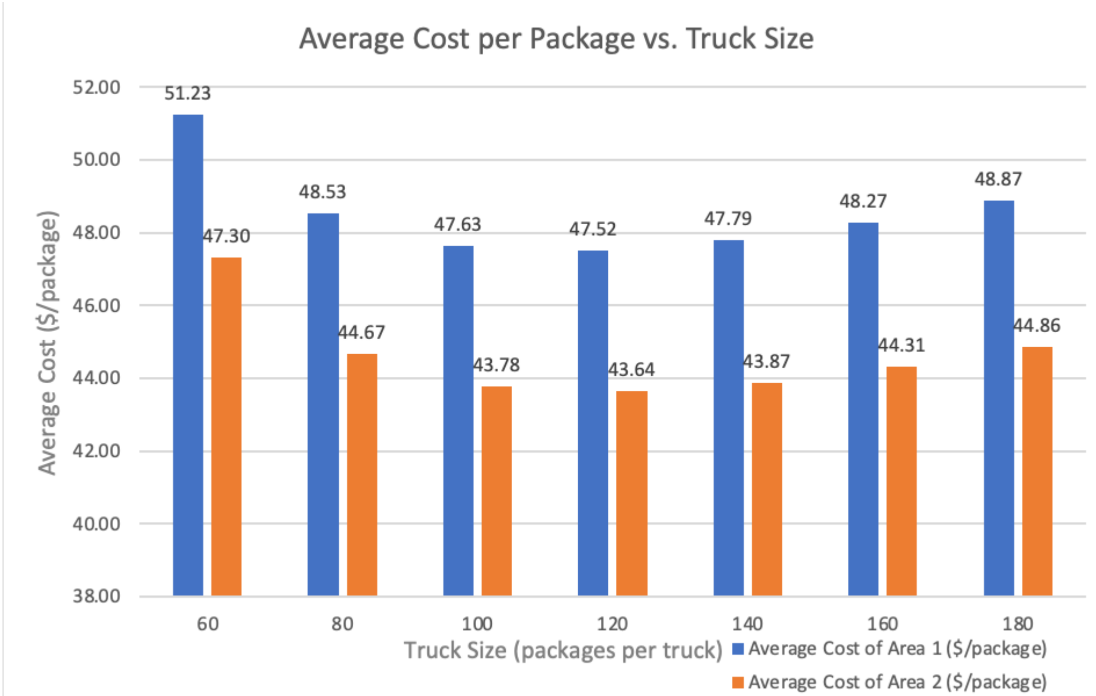

Figure 20 Average cost per package vs. truck size

5.4.4 Effect of Value of Time (in $/package/hr)

As the value of time for users becomes more important in total cost, more trucks are dispatched to deliver packages sooner since the value of time influencing in-vehicle cost, waiting cost and perishable cost. When the value of time increases tenfold over the baseline from 0.415 $/(package·hr) to 4.15 $/(package·hr), for area 1, the optimal headway decreases significantly from 35.0 to 22.3 h, the optimal zone size increases greatly from 8514.8 to 14892.6 sq. miles, while the truck capacity is kept at 100 packages/truck. A similar relation holds for area 2.

5.4.5 Effect of the Perishability Coefficient (in %/hr)

The perishability rate is used for farm products. Keeping other variables same as in the baseline inputs, besides adjusting the value of p from 0.2 to 1, the effect is summarized in Table 11. When the farm product is harder to preserve during the waiting and delivery time, the perishability cost increases significantly and dominates the user cost.

Table 11 Effect of Perishable Coefficient

Perishable Coefficient of Area 1 (%/hr)	**0.2***	0.4	0.6	0.8	1.0
Headway (hours)	**35.0**	36.8	19.1	19.6	20.0
Zone Size (sq. miles)	**8514.8**	9581.4	10412.8	11081.9	11633.1
Fleet Size (trucks)	**1**	1	2	2	2
Package Demand (packages/zone/hr)	**2.30**	2.59	2.81	2.99	3.14
Operating Cost, C_O ($/hr)	**60.00**	60.00	120.00	120.00	120.00
Perishability Cost, C_P ($/hr)	**16.70**	19.76	11.12	12.16	13.03
In-vehicle Cost, C_V ($/hr)	**16.70**	19.76	22.24	24.31	26.06
Waiting Cost, C_W ($/hr)	**16.10**	38.08	48.24	70.31	94.20
Total Cost, C($/hr)	**109.50**	137.59	201.61	226.78	253.29
Average Cost, C_A ($/package)	**47.63**	53.19	71.71	75.79	80.64
% Change of Average Cost	**0**	11.67	50.56	59.24	69.31

Perishable Coefficient of Area 1 (%/hr)	**0.2***	0.4	0.6	0.8	1.0
Headway (hours)	**30.5**	32.2	33.4	17.2	17.6
Zone Size (sq. miles)	**8882.3**	9995.0	10862.3	11560.2	12135.3
Fleet Size (trucks)	**1**	1	1	2	2
Package Demand (packages/zone/hr)	**2.40**	2.70	2.93	3.12	3.28
Operating Cost, C_O ($/hr)	**60.00**	60.00	60.00	120.00	120.00
Perishability Cost, C_P ($/hr)	**15.18**	18.02	20.34	11.14	11.96
In-vehicle Cost, C_V ($/hr)	**15.18**	18.02	20.34	22.28	23.91
Waiting Cost, C_W ($/hr)	**14.63**	34.74	58.82	64.42	86.43
Total Cost, C($/hr)	**104.99**	130.78	159.51	217.83	242.30
Average Cost, C_A ($/package)	**43.78**	48.46	54.39	69.79	73.95
% Change of Average Cost	**0**	10.69	24.23	59.41	68.91

* Baseline results

5.4.6 Effect of Truck Operation Speed (in mph)

he baseline value of truck operation speed is 60 mph. The result from Table 12 indicates that within the range from 30 to 70, the average cost decreases withhigher speed, but the large value may not fit the operation scenario since it is hard for vehicles to maintain high speeds on rural roads. The marginal benefit from increasing speed is decreasing since the % change of average cost decreases at a slower rate.

Table 12 Effect of Truck Operation Speed

Truck Operation Speed of Area 1 (%/hr)	30	40	50	60*	**70**
Headway (hours)	52.7	44.3	38.9	35.0	**32.1**
Zone Size (sq. miles)	5364.0	6498.0	7540.2	8514.8	**9436.3**
Fleet Size (trucks)	1	1	1	1	**1**
Package Demand (packages/zone/hr)	1.45	1.75	2.04	2.30	**2.55**
Operating Cost, C_O ($/hr)	60.00	60.00	60.00	60.00	**60.00**
Perishability Cost, C_P ($/hr)	15.83	16.14	16.43	16.70	**16.96**
In-vehicle Cost, C_V ($/hr)	15.83	16.14	16.43	16.70	**16.96**
Waiting Cost, C_W ($/hr)	15.25	15.55	15.83	16.10	**16.35**
Total Cost, C($/hr)	106.90	107.83	108.69	109.50	**110.27**
Average Cost, C_A ($/package)	73.81	61.46	53.39	47.63	**43.28**
% Change of Average Cost	54.97	29.04	12.09	0	**-9.13**

Truck Operation Speed of Area 2 (%/hr)	30	40	50	60*	**70**
Headway (hours)	47.1	39.3	34.1	30.5	**27.8**
Zone Size (sq. miles)	5595.5	6778.5	7865.7	8882.3	**9843.7**
Fleet Size (trucks)	1	1	1	1	**1**
Package Demand (packages/zone/hr)	1.51	1.83	2.12	2.40	**2.66**
Operating Cost, C_O ($/hr)	60.00	60.00	60.00	60.00	**60.00**
Perishability Cost, C_P ($/hr)	14.76	14.91	15.05	15.18	**15.30**
In-vehicle Cost, C_V ($/hr)	14.76	14.91	15.05	15.18	**15.30**
Waiting Cost, C_W ($/hr)	14.23	14.37	14.50	14.63	**14.75**
Total Cost, C($/hr)	103.75	104.18	104.59	104.99	**105.36**
Average Cost, C_A ($/package)	68.67	56.92	49.25	43.78	**39.64**
% Change of Average Cost	56.86	30.02	12.49	0	**-9.45**

* Baseline results

5.4.7 Effect of Truck Operation Cost (in \$/hr)

Since the truck operation cost $B = a + bS$ consists of fixed cost a in \$/(vehicle·hr) and cost b in \$/(package·hr) relating to truck capacity. The range of a and b is set to correspondingly span from -25% to 25%. With the increase of either parameter, the cost increases linearly and the headway decreases linearly.

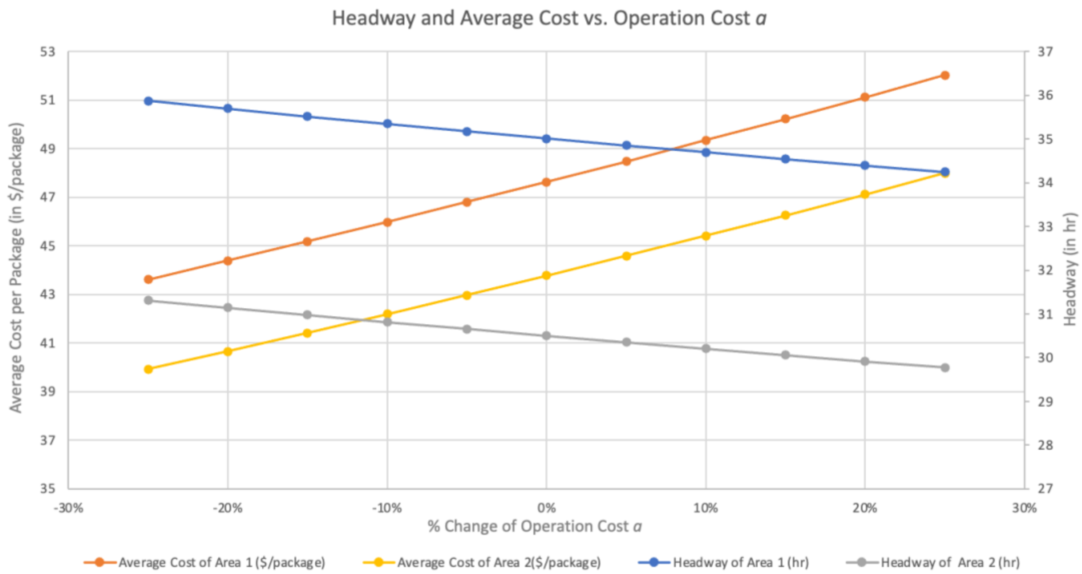

Figure 21 Headway and average cost per package vs. truck operation cost a

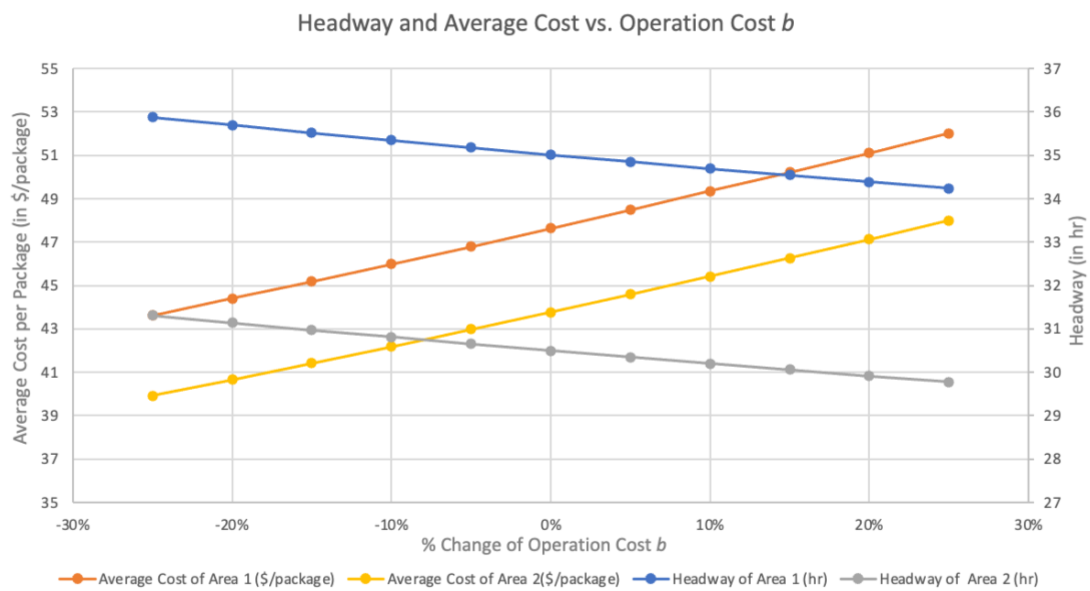

Figure 22 Headway and average cost per package vs. truck operation cost b

5.5 Results of Optimization Algorithm of Real Case

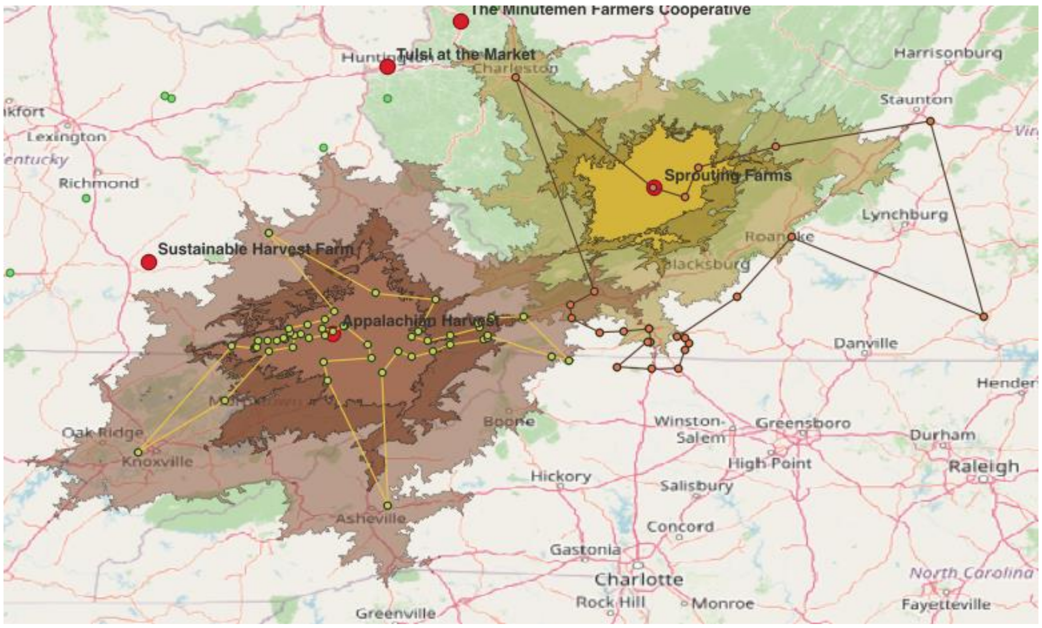

Figure 23 Optimal service route for area 1 and 2

Relying on the algorithm presented in Figure 9, the LKH algorithm is employed to compute a near-optimal route for the flexible-route service. This computation is based on geocoordinate inputs and is used to determine the minimal total cost of both flexible-route services and self-delivery services. Both constraints on maximum allowable delivery time and integer fleet size are strictly followed. The optimal delivery route covering all the nodes can be found in Figure 14 or Figure 20. The algorithm attempts to remove one node from the TSP tour and assign another truck to deliver the package separately, with the output of TSP tour covering all demand nodes with one truck. There is no output about mixing service and keep the same route as flexible-route service. One possible explanation is that when more than one truck is operating, both total cost and average cost increase significantly due to the large proportion of operating cost. Consequently, the cost saved by removing one node from TSP tour is less than the cost of dispatching another truck.

The results from baseline and sensitivity study all suggest that when demand is low and the zone size is large, it is impossible for a driver to finish the tour within the maximum allowable working time in one day (e.g., 10 hours maximum per day). To finish the tour, a driver has to stop for resting or a Vehicle Routing Problem (VRP) should be solved. Since the detailed constraints for a VRP algorithm are not discussed, the same constrains from TSP is applied with additional flow balance where the node only will be visited by one vehicle. Using the nodes from case 2 as an example, the near-optimal route for the flexible-route service of two vehicles with maximum allowable working time and balanced driving distance can be found in Figure 21. In this scenario, one vehicle completes its route in 9.0 hours covering 497 miles, while the other finishes in 9.8 hours, traveling 482 miles.

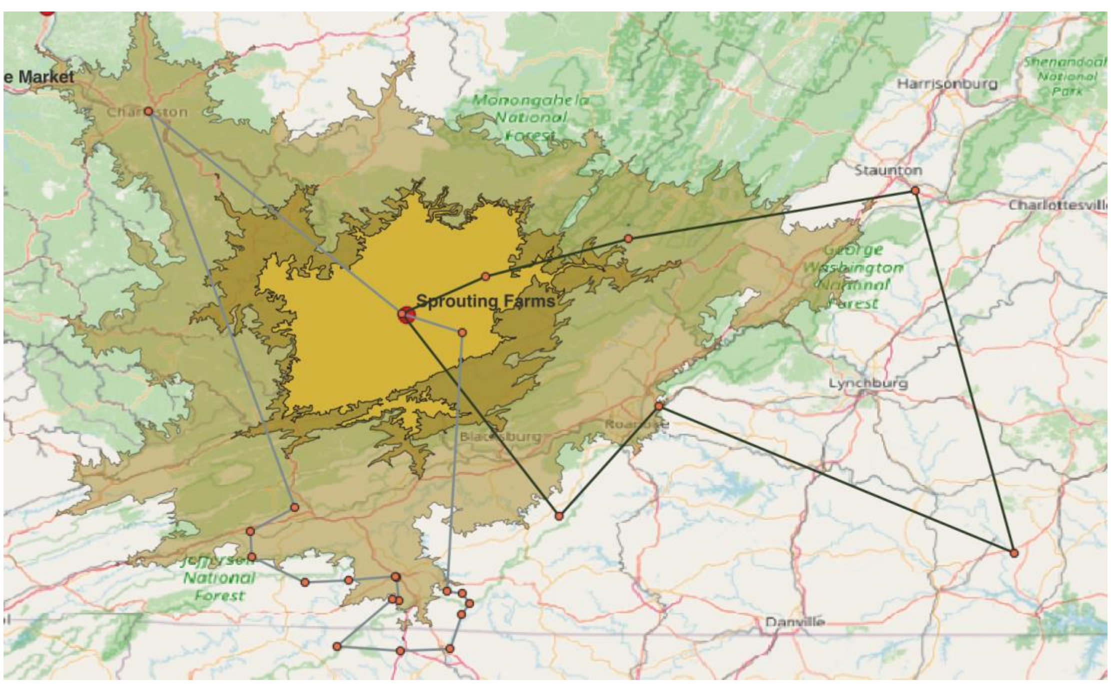

Figure 24 Optimal service route with two vehicles for area 2

5.6 Summary

This chapter discusses the application of economic planning model of total cost (user cost + operation cost) and finds the minimal average cost per package with optimized service headway and size of service zone in both ideal unconstrained case and in a realistic application operating jointly with existing self-delivery service. AVRP application would probably improve the practicalities of flexible-route service in a large area with low demand.

Chapter 6: Conclusion and Future Research

6.1 Research Summary and Contributions

Efficiency and low demand constitute significant challenges in achieving a cost-effective delivery system in rural areas. Based on literature reviews, this research focuses on integrating various methods and addressing several research gaps. These gaps include computing the circuity factor in rural areas, defining a new estimation formulation of TSP tour considering the influence of dead-end road and sparseness of the rural road network, incorporating the perishability factor into the cost function formulation, and applying a heuristic method to solve the Traveling Salesman Problem (TSP) tour distance in the case study, as opposed to estimating the distance.

This study focuses on the optimization of a planning stage model which minimizes the total cost and the average cost per package. Several findings reveal the characteristic of delivery in rural areas. First, the adjusted tour length estimation model based on log-transformation is discussed based on literature review and statistical support. Then, based on the estimated input parameters, in reducing the average cost per package, it is more beneficial to have a fleet with multiple trucks rather than a single-truck fleet. Furthermore, the optimal size of the service area can be determined by identifying the size that yields the lowest average cost per package. Lastly, it is observed that the same fluctuation in an input variable can lead to varying impacts on service areas depending on their road network densities and demand characteristics.

6.2 Future Research

Some potential extensions of this study are considered, including:

- Compiling more similar case studies and relaxing the simplifying assumptions in the planning model, such as variable demand per point.
- Developing a more realistic model for the perishability of products, such as an exponential decay model.
- Considering the influence of heterogeneous packages and trucks, as well as the integration of potential new delivery mediums like unmanned vehicles or drones.
- Using a better model for tour estimation in the planning model, such as one based on more raw data from real world cases.
- Using heuristic vehicle routing problem (VRP) algorithms for similar cases with greater numbers of trucks in the service zone.
- Considering probabilistic variations in demand and service times.

References

Adebisi, O. (1980). A theoretical travel-time model for flexible-route buses. *Transportation Research Part B: Methodological*, *14*(4), 319-330.

Ballou, R. H., Rahardja, H., & Sakai, N. (2002). Selected country circuity factors for road travel distance estimation. *Transportation Research Part A: Policy and Practice*, *36*(9), 843-848.

Bang-Jensen, J., Gutin, G., & Yeo, A. (2004). When the greedy algorithm fails. *Discrete optimization*, *1*(2), 121-127.

Barnes, G., & Langworthy, P. (2004). Per mile costs of operating automobiles and trucks. *Transportation Research Record*, *1864*(1), 71-77.

Basel Iii, J., & Willemain, T. R. (2001). Random tours in the traveling salesman problem: analysis and application. *Computational Optimization and Applications*, *20*(2), 211-217.

Beardwood, J., Halton, J. H., & Hammersley, J. M. (1959). The shortest path through many points. Mathematical Proceedings of the Cambridge Philosophical Society,

Blackburn, J., & Scudder, G. (2009). Supply chain strategies for perishable products: the case of fresh produce. *Production and Operations Management*, *18*(2), 129-137.

Boccia, M., Masone, A., Sforza, A., & Sterle, C. (2021). A column-and-row generation approach for the flying sidekick travelling salesman problem. *Transportation Research Part C: Emerging Technologies*, *124*, 102913.

Brunetti, R., Krauth, W., Mézard, M., & Parisi, G. (1991). Extensive numerical simulations of weighted matchings: Total length and distribution of links in the optimal solution. *Europhysics Letters*, *14*(4), 295.

Cavdar, B., & Sokol, J. (2015). A distribution-free TSP tour length estimation model for random graphs. *European journal of operational research*, *243*(2), 588-598.

Chang, S. K., & Schonfeld, P. M. (1991). Integration of fixed-and flexible-route bus systems. *Transportation Research Record*(1308).

Chang, S. K., & Schonfeld, P. M. (1991). Optimization models for comparing conventional and subscription bus feeder services. *Transportation Science*, *25*(4), 281-298.

Chien, T. W. (1992). Operational estimators for the length of a traveling salesman tour. *Computers & Operations Research*, *19*(6), 469-478.

Choi, Y. (2021). *Adjustment Factors and Applications for Analytic Approximations of Tour Lengths* University of Maryland, College Park].

Choi, Y., & Schonfeld, P. M. (2022). Review of length approximations for tours with few stops. *Transportation Research Record*, *2676*(3), 201-213.

Choia, Y., Schonfelda, P., & Wangb, J. Optimization of Farm Product Logistics: A Case Study in Central Appalachia.

Concorde TSP Solver, https://www.math.uwaterloo.ca/tsp/concorde.html

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms*. MIT press.

Daganzo, C. F. (1984). The length of tours in zones of different shapes. *Transportation Research Part B: Methodological*, *18*(2), 135-145.

DOT, U. (2014). TIGER Benefit-Cost Analysis (BCA) Resource Guide. United States Department of Transportation. Washington, DC. www. transportation. gov/sites/dot. gov/files/docs/TIGER_BCA_ResourceGuide2014. pdf. Last accessed December, 4, 2015.

Glover, F., & Laguna, M. (1998). *Tabu search*. Springer.

Gutin, G., Yeo, A., & Zverovich, A. (2002). Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. *Discrete Applied Mathematics*, *117*(1-3), 81-86.

Han, D., Kim, M. E., & Kim, E. (2022). A Planning Model for Flexible-Route Bus Operations with Financial Constraints. *KSCE Journal of Civil Engineering, 26*(7), 3114-3135.

Helsgaun, K. (2000). An effective implementation of the Lin–Kernighan traveling salesman heuristic. *European journal of operational research*, *126*(1), 106-130.

Holland, C., Levis, J., Nuggehalli, R., Santilli, B., & Winters, J. (2017). UPS optimizes delivery routes. *Interfaces*, *47*(1), 8-23.

Johnson, D. S., McGeoch, L. A., & Rothberg, E. E. (1996). Asymptotic experimental analysis for the Held-Karp traveling salesman bound. Proceedings of the 7th Annual ACM-SIAM Symposium on Discrete Algorithms,

Kallio, J., Saarinen, T., Tinnilä, M., & Vepsäläinen, A. P. (2000). Measuring delivery process performance. *The International Journal of Logistics Management*, *11*(1), 75-88.

Kawamura, K. (2000). Perceived value of time for truck operators. *Transportation Research Record*, *1725*(1), 31-36.

Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, *49*(2), 291-307.

Kim, M., & Schonfeld, P. (2012). Conventional, flexible, and variable-type bus services. *Journal of Transportation Engineering*, *138*(3), 263-273.

Kim, M. E., & Schonfeld, P. (2015). Maximizing net benefits for conventional and flexible bus services. *Transportation Research Part A: Policy and Practice*, *80*, 116-133.

Kim, M. E., Schonfeld, P., Roche, A., & Raleigh, C. (2022). Optimal service zones and frequencies for flexible-route freight deliveries. *Transportation Research Part A: Policy and Practice*, *159*, 182-199.

Kirkpatrick, S. (1983). Improvement of reliabilities of regulations using a hierarchical structure in a genetic network. *science*, *220*, 671-680.

Kou, S., Golden, B., & Poikonen, S. (2022). Optimal TSP tour length estimation using standard deviation as a predictor. *Computers & Operations Research*, *148*, 105993.

Kou, S., Golden, B., & Poikonen, S. (2023). Optimal TSP tour length estimation using Sammon maps. *Optimization Letters*, *17*(1), 89-105.

Kulović, M. (2004). Freight transport costs model based on truck fleet operational parameters. *PROMET-Traffic&Transportation*, *16*(6), 321-325.

Kweon, H. (2019). Comparisons of estimated circuity factor of forest roads with different vertical heights in mountainous areas, Republic of Korea. *Forests*, *10*(12), 1147.

Lee, J., & Choi, M. (1994). Optimization by multicanonical annealing and the traveling salesman problem. *Physical Review E, 50*(2), R651.

Levinson, D., & El-Geneidy, A. (2009). The minimum circuity frontier and the journey to work. *Regional science and urban economics, 39*(6), 732-738.

LKH Version 2.0.10, http://webhotel4.ruc.dk/~keld/research/LKH/

Merchán, D., & Winkenbach, M. (2019). An empirical validation and data‐driven extension of continuum approximation approaches for urban route distances. *Networks*, *73*(4), 418-433.

Miller, C. E., Tucker, A. W., & Zemlin, R. A. (1960). Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, *7*(4), 326-329.

Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. IEEE Transactions on computers, 100(5), 401-409.

Sanches, D., Whitley, D., & Tinós, R. (2017). Improving an exact solver for the traveling salesman problem using partition crossover. Proceedings of the Genetic and Evolutionary Computation Conference,

Sdoukopoulos, E., Boile, M., & Anagnostopoulou, A. (2015). Estimating truck operating costs for domestic trips–case studies from Greece. *Logistics, Supply Chain, Sustainability and Global Challenges*, *6*(1), 41-49.

Shayanfar, E., Schonfeld, P. M., & Wang, J. J. (2019). Prioritizing highway development projects based on market access in Appalachia. *Transportation Research Record*, *2673*(8), 333-342.

Stein, D. M. (1978a). An asymptotic, probabilistic analysis of a routing problem. *Mathematics of Operations Research*, *3*(2), 89-101.

Stein, D. M. (1978b). Scheduling dial-a-ride transportation systems. *Transportation Science*, *12*(3), 232-249.

Topal, E., & Ramazan, S. (2012). Mining truck scheduling with stochastic maintenance cost. *Journal of Coal Science and Engineering (China)*, *18*, 313-319.

TSPLIB, http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/

U.S. Bureau of Labor Statistics (2023), Average price data (in U.S. dollars), selected items, https://www.bls.gov/charts/consumer-price-index/consumer-price-index-average-price-data.htm Last Accessed Jul. 20, 2023

Vig, V., & Palekar, U. S. (2008). On estimating the distribution of optimal traveling salesman tour lengths using heuristics. *European journal of operational research*, *186*(1), 111-119.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272.

What is the Justice40 Initiative? Justice40 Initiative https://www.transportation.gov/equity-Justice40 Accessed May 5, 2023

Winebrake, J. J., Green, E. H., Comer, B., Li, C., Froman, S., & Shelby, M. (2015). Fuel price elasticities for single-unit truck operations in the United States. *Transportation Research Part D: Transport and Environment*, *38*, 178-187.

Wu, X., Liu, P., & Lu, X. (2021). Study on operating cost economy of battery-swapping heavy-duty truck in China. *World Electric Vehicle Journal*, *12*(3), 144.