Who Has Access to E-Commerce and When?
Time-Varying Service Regions in Same-Day Delivery

Dipayan Banerjee  Alan L. Erera  Alexander M. Stroh  Alejandro Toriello
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
dipayan.banerjee@gatech.edu, aerera@isye.gatech.edu, alexmstroh@gatech.edu, atoriello@isye.gatech.edu

February 4, 2022

Abstract
Motivated by access and equity issues in e-commerce, we study the design of same-day delivery (SDD) systems under the assumption that service regions are allowed to vary over the course of the day; equivalently, that customers in different locations may have access to SDD for different lengths of time over the service day or may have no access at all. This contrasts with the bulk of the literature, in which a service region is defined in advance and all customers in the service region can place SDD orders during the same time window. Leveraging continuous approximation techniques to capture average-case system behavior, we derive optimal service region areas and corresponding SDD order cutoff times to maximize the expected number of orders served per day. We quantify the benefit of allowing the service regions to vary, both theoretically and empirically, and discuss related equity issues in SDD systems. We illustrate and validate our results with a case study set in the Phoenix, Arizona metropolitan area.

1 Introduction
Driven by increased internet access, the e-commerce retail sector has been expanding steadily in recent years. Changes in consumer behavior due to the COVID-19 pandemic [7, 17] have accelerated this trend: total e-retail volume in the U.S. between April 2020 and March 2021 surpassed $817 billion, representing an increase of over 30% from the prior year [48, 49]. In an effort to capture a larger share of this market, e-retailers have improved their delivery time guarantees. Same-day delivery (SDD), which was once leveraged as a service offering differentiator, has now become expected by consumers at large. Amazon, which has offered SDD to select premium subscribers for over a decade [1], continues to expand its same-day supply chain network in order to serve customers faster and provide SDD options in more cities [8]. Large American retailers such as Walmart, Target and Costco have recently partnered with third party managers of their SDD systems [9]. Some smaller niche retailers, including Sephora (beauty products) and Michael’s (arts and crafts), have done the same in order to provide SDD to their customers [22, 52].
SDD allows e-commerce firms to compete more effectively with brick-and-mortar retail by providing the customer with near-instant demand fulfillment; however, delivering cost-effectively given a same-day deadline requires careful planning. The last-mile component of traditional parcel delivery often generates more than 50% of the total cost of delivery [18], and SDD systems face even greater potential cost inefficiencies since they may suffer from reduced opportunities for consolidation due to a high degree of dynamism.

An important design problem for modern e-commerce systems is the question of where and when (how late in the service day) to offer the SDD promise. A service region that is too small or an early order deadline will result in fewer SDD customers and market share, while a large region or a late SDD order cutoff time may result in costly operations or failed deliveries and loss of customer goodwill. Considerations of equity and access are also important. For example, Amazon faced criticism in recent years for perceived racial bias in SDD service region design [21], which the company later addressed [40, 41]. If designed well, however, e-commerce systems including those offering same-day delivery have the potential to help customers unable to travel to traditional brick-and-mortar retail stores, and thus may improve access to food and other important household goods. With this motivation in mind, our goal in this study is the selection of SDD service regions and order cutoff deadlines from the perspective of an e-retailer operating a single fulfillment center (i.e., depot) with a fixed delivery fleet.

Our objective is to choose a service region and deadline that maximize the expected order volume the retailer can feasibly serve each day. In particular, we seek an understanding of the potential system gains that result by allowing the service region to vary over the course of the service day, by offering different order cutoff times to different parts of the overall region. Our results indicate that the system may indeed benefit by allowing such variation. The intuition behind this result is clear: customers that are far away from the fulfillment center (e.g. in suburban areas) cannot be served as efficiently as nearby (e.g. in-town) customers, and thus the system may operate more efficiently if we only accept faraway SDD orders early in the day but allow nearby customers to continue placing orders until later.

Interestingly, this tiered approach may also allow the SDD system to increase its overall footprint, by offering SDD farther away from the fulfillment center than a system with a single common order cutoff. Our results thus contribute to the growing discussion in the literature regarding fairness, equity, and access in SDD and e-commerce more broadly. Some recent work in SDD [10, 12] assumes all customers in the service region should be treated roughly the same according to some metric of customer service, such as expected waiting time or order acceptance rate. Our models suggest that imposing such requirements implicitly
constrains the system to a reduced service area, thus potentially denying SDD to customers who live outside the smaller region. There is no single agreed-upon metric for what constitutes a more fair or equitable outcome [see e.g. 27], and thus different companies, managers and customers may perceive different system choices as more or less desirable; nevertheless, our results allow decision makers to quantify the impact of important SDD system variables, such as service region size and order cutoff times, in order to make informed decisions.

While there has been significant research attention devoted to SDD in the past decade, work has focused primarily on the operational management of SDD systems (decisions made over the course of a service day) rather than on system design at longer time scales (every few weeks or months). This research stream seeks to optimize day-to-day operations of SDD systems, including vehicle routing plans and order acceptance mechanisms [e.g., 11, 24, 25, 29, 50, 51]. While these studies are crucial to efficiently manage a defined system, they generally do not focus on designing elements of the system itself. In particular, operationally-focused SDD literature often assumes a fixed service region from which SDD demand realizes.

More recent work [4, 10, 42] has focused on longer time frames in SDD and on some system design variables, including the partitioning of a service region into vehicle zones and related fleet sizing questions. Our work shares some methodological features with this literature stream, particularly in the use of continuous approximation techniques to capture the average-case behavior of SDD systems. Nonetheless, as with the operational SDD literature, the models that have been developed to date all assume a given, fixed service region and ignore the question of choosing the SDD system’s overall footprint.

1.1 Contributions

We consider our main contributions to be the following:

1) Using continuous approximations of order arrivals and vehicle routing times, we propose a mathematical optimization model for maximizing order quantities served in a single-depot SDD system when the service region is allowed to vary between vehicle dispatches. The decision space for the model includes choosing the order accumulation time between successive dispatches as well as determining the size of time-varying service regions from which orders accrue.

2) We perform an in-depth theoretical analysis of this system design question for a few important SDD system variations. Specifically, we study a setting in which multiple vehicles each dispatch once per day to analyze the marginal benefits of increasing the fleet size. We also study a setting in which...
one vehicle dispatches multiple times per day to analyze the marginal benefits of re-using a particular vehicle. We leverage our theoretical results to design efficient solution procedures.

(3) We study the quantifiable effects of allowing time-varying service regions compared to traditional designs with a fixed service region. We describe the effects of such dynamics on both the service provider as well as the customers.

(4) We conduct an extensive computational study using the Phoenix, Arizona metro area road network, and use these results to motivate a discussion on issues of profitability, equity, and access in SDD systems.

Section 1 concludes with a review of the relevant literature. A formal definition of our general model is given in Section 2. In Section 3 we analyze a one-vehicle, one-dispatch variant of the model to motivate more complex settings. In Section 4 we study the setting in which multiple vehicles each dispatch once per day. In Section 5 we study the setting in which one vehicle dispatches multiple times per day. In Section 6 we perform computational validation and discuss managerial insights. Section 7 contains concluding remarks. Appendices contain proofs and other omitted material.

1.2 Literature Review

The majority of the SDD literature has focused on operational problems, in which system features are fixed and a system manager must determine an optimal policy to guide decision-making over a short horizon (typically a single service day). Such works typically focus on vehicle dispatching and routing as customer information is dynamically revealed. Proposed solutions are compared to offline heuristics or current best practices. Specific problems considered in the literature include the same-day delivery problem for online purchases [13, 51] and the dynamic dispatch waves problem [24, 25, 26]. Other works integrate autonomous vehicles [45], drones [11, 16, 46], and additional extensions [47, 53]. Operational SDD problems are closely related to the broad problem classes of stochastic VRPs [32, 33] and dynamic VRPs [35, 36].

Operational SDD problems are often modeled as mixed-integer linear programs (MILPs), Markov decision processes (MDPs), or a combination of such models. Because of their underlying stochasticity and extremely large decision spaces, these problems are generally solved without optimality guarantees; solution techniques include approximate dynamic programming [e.g., 24, 47], neighborhood search [13], and tailored heuristics [16]. Such models may be sufficient for day-to-day operational usage. However, it is
difficult to perform high-level SDD system design with detailed operational models since they often require significant computational effort to approximately solve even moderately-sized instances, without optimality guarantees, over a single set of design parameters. While simulation is an option for gaining managerial insights [39, 44], the lack of transparency and interpretability in simulation-based methods motivates a need for simpler analytical approaches to SDD system design problems.

While we are not aware of any literature directly studying service region sizing and design for SDD systems, a few papers examining operational problems have considered how service regions influence their modeling and results. Notably, [12] formulates an operational SDD model where the dispatcher of the system can choose whether or not to accept SDD orders, but is constrained to accept orders across different customer zones at the same rate. The authors note that the benefits of enforcing such fairness constraints come at the cost of lowering the total quantity of served orders. Another work [11] that also allows a dispatcher to accept or reject orders for SDD observes that as the service day progresses, the operator is less likely to accept orders from customers living farther away from the depot if the dispatcher is to maximize the number of orders served.

Seminal works in the area of continuous approximations for vehicle routing show that the expected length of vehicle tours can be functionally approximated by the number of stops in the tour, the region from which demand points originate, and the probability distribution governing the points’ locations. The foundational Beardwood-Halton-Hammersley (BHH) Theorem [5] states that the expected length of an optimal traveling salesperson problem (TSP) tour over \( n \) points in a region of area \( A \) approaches \( \beta \sqrt{An} \) as \( n \) grows, where \( \beta \) is a region-, distribution- and metric-dependent routing constant. Many studies analyze BHH-type approximations of vehicle tour lengths in various settings [14, 15, 30, 31]. Various works have focused on empirical estimation of BHH routing constants on stylized regions [3, 23] and real-world road networks [28]. Comprehensive surveys of the continuous approximation literature, from fundamental works to recent results and applications, are given by [2, 19].

Recent papers [4, 10, 42] use continuous approximation methods to design last-mile e-commerce systems with very short delivery deadlines. In [42], the authors assume that orders for SDD arrive from a defined, fixed service region until some defined cutoff time. The objective is to minimize the total routing time to serve all of the accrued SDD orders. In a similar setting, again with a predefined fixed service region and cutoff time, [4] minimize the total number of vehicles needed to serve SDD orders assuming the region is to be partitioned into single-vehicle zones. Similarly, [10] also use continuous approximations to partition
a fixed e-commerce service region into single-vehicle delivery zones, enforcing the additional requirement that the expected order-to-delivery time is equitable across all customers. This is in contrast to a basic SDD setting, in which customers simply share the same delivery deadline.

2 Model Formulation

We consider an SDD system with a single fulfillment center (or depot) from which a fleet of uncapacitated, homogeneous vehicles is dispatched. Customer orders arrive via a two-dimensional (random) point process beginning at the start of the service day. All orders are to be served (i.e., delivered) and all vehicles must return to the depot by the end of the service day. Our goal is to design this system by selecting the service region: the geographical area, potentially varying over time, from which customers are permitted to place SDD orders. The objective is to maximize the expected number of SDD orders served each day. We solve this design problem via a continuous approximation model of the system characterized as follows.

**Service Day:** The beginning of the service day is denoted as time $t=0$. The end of the service day, which represents both the order delivery and vehicle return deadlines, is denoted as time $t=T$. We assume without loss of generality that $T=1$ and all other parameters are appropriately scaled.

**Service Region:** At the start of the service day, and after each dispatch, we must determine the service region from which SDD orders accrue until the next vehicle dispatch. Hence, if vehicles dispatch a total of $k$ times, the service region changes at most $k-1$ times over the course of the day; no service region exists after the final vehicle dispatch since no orders can be placed after that time.

We assume, unless stated otherwise, that service regions grow concentrically from the depot, either in all directions or in a fixed direction (i.e., as a wedge). Specifically, regions are constructed so that the driving time from the depot to any point along the outer edge of a region is equal. In practice, the shape of a region depends on the road network topology and the depot’s location. In our computational case studies, we consider travel in a real-world road network. In this initial discussion, we illustrate our model with the simpler $\ell_1$ or $\ell_2$ metrics.

As a result, we can characterize a service region by its area $A$; a region can be equivalently characterized by its maximum driving time radius. To illustrate, consider the service regions in Figure 1a. We assume the travel time is given by the $\ell_2$ metric, so regions are circular. The initial region has area $A_1$, and after the first vehicle dispatch, the service region shrinks to an area $A_2 = A_1/2$; equivalently, the drive time radius
decreases by a factor of $\sqrt{2}$ from the first service region to the second. Note that the first service region includes the full area within the outer circle, including the inner circle. Figure 1b illustrates the same service region structure restricted to the depot’s northeast quadrant.

**Customer Orders:** SDD order requests accumulate continuously at a rate of $\lambda$ orders per unit time per unit area starting at $t = 0$. At any given time, SDD orders accumulate only within the current service region. All accumulated orders must be served (i.e., delivered) by $T$. We assume the order rate per time and area remains constant for any region we choose to serve. In practice, this may only apply to regions within a certain size; in Section 4.2, we discuss bounding the maximum service area.

**Vehicle Dispatches:** The fleet is comprised of $m$ homogeneous vehicles. Vehicles are not explicitly constrained by capacity nor are they restricted to carry an integer number of orders. However, each vehicle in the fleet is allowed at most $D$ dispatches in total over the service day, where $D$ is an integer. At each dispatch time, a vehicle leaves the depot with all of the accumulated orders since the previous dispatch, implying a first in, first out (FIFO) order processing approach. Equivalently, dispatches do not batch or differentiate orders based on geography, and therefore those orders are distributed uniformly across the service region associated with the dispatch.

**Routing Time Function:** The time it takes for a vehicle to dispatch from the depot, serve $n \in \mathbb{R}_{\geq 0}$ orders uniformly distributed over a region of area $A$, and return to the depot is given by a deterministic, continuous routing time function $f(A, n) = c_0 \sqrt{n}$, where $c_0$ is a known positive constant. For our analysis, we equivalently define the routing time function as $f(A, \tau) = cA \sqrt{\tau}$, where $\tau$ is the accumulation time since the previous dispatch and $c = c_0 \sqrt{\lambda}$. The structure of the routing function is derived from the BHH theorem [5] discussed earlier, which has been empirically shown to work well for relatively small $n$.

As a basic illustrative example, consider the system in Figure 2 with one vehicle ($m = 1$) that dispatches twice ($D = 2$) over the course of the day. At the beginning of the day, the service region is $A_1$, as depicted in
Over a duration of $\tau_1$, a total of $\lambda A_1 \tau_1$ SDD orders accumulate in this service region. At time $\tau_1$, the vehicle dispatches from the depot to serve these accumulated orders. Simultaneously, the service region shrinks to $A_2$, as depicted in Figure 1. Over a duration of $\tau_2$, a total of $\lambda A_2 \tau_2$ orders accumulate over this smaller service region. At time $\tau_1 + \tau_2$, the vehicle dispatches from the depot to serve these orders. Note that this example is feasible: all accumulated orders are served, the vehicle never dispatches before it returns to the depot from a prior trip, and the vehicle returns to the depot for the final time before $T = 1$.

The objective of this decision problem is to choose a set of feasible accumulation times and service regions in order to maximize the number of SDD orders served. We formally define the $d$-th time-ordered dispatch as a tuple $((\tau_d, A_d, i_d))$, where $\tau_d$ defines the order accumulation time (since the previous dispatch, or since $t = 0$ for the first dispatch) for vehicle $i_d$ serving all of the accumulated orders in a region of area $A_d$. A set of dispatches $\{(\tau_d, A_d, i_d)\}_{d=1}^{mD}$ defines a policy. A policy is feasible for our model if the following conditions are satisfied:

\[
\sum_{d=1}^{d} \tau_d + f(A_d, \tau_d) \leq 1 \quad \forall d \in [mD], \tag{1a}
\]

\[
\sum_{d=1}^{d} \tau_d + f(A_d, \tau_d) \leq \sum_{d=1}^{d'} \tau_d \quad \forall d \in [mD], \; d' \text{ s.t. } i_d = i_{d'}, \; d < d', \tag{1b}
\]

\[
i_d \in [m] \quad \forall d \in [mD], \tag{1c}
\]

\[
A_d, \tau_d \geq 0 \quad \forall d \in [mD]. \tag{1d}
\]

The goal is to maximize the total number of SDD orders served, $\sum_{d=1}^{mD} \lambda A_d \tau_d$, subject to the constraints (1a)-(1d). Constraint (1a) ensures that all vehicles will return to the depot by the end of the service day. Constraint (1b) ensures that each vehicle returns to the depot prior to any of its subsequent dispatches. Constraint (1c) assigns each dispatch to a vehicle in the fleet. Lastly, (1d) enforces non-negativity for the service area and accumulation time variables.

This is the most general statement of the problem; we next study specific variants motivated by practical
considerations. Concurrently, we use results derived for these specific variants to analyze features of the general model. Henceforth, we denote a setting with \( m \) vehicles and \( D \) dispatches per vehicle as \( \langle m, D \rangle \) for clarity and notational convenience. We let \( z_{m,D} \) denote the optimal objective value of an \( \langle m, D \rangle \) problem with the constraints and objective described above.

3 One vehicle, One Dispatch

We begin our analysis by studying the simplest case, with one vehicle that is permitted to dispatch once per day. Such a system may be of interest to a small retailer with limited resources and limited scope for online optimization during the service day. More importantly, studying such \( \langle 1, 1 \rangle \) systems can provide insights on how to approach the optimization of more complicated families of problem instances.

In this \( \langle 1, 1 \rangle \) setting, the system designer is responsible for two choices: determining the service area \( A_1 \) and the duration of time \( \tau_1 \) during which customers can place orders. Since \( m = D = 1 \), Problem (I) simplifies to the following:

\[
\begin{align*}
\max_{A_1, \tau_1 \geq 0} & \quad \lambda A_1 \tau_1 \\
\text{s.t.} & \quad \tau_1 + cA_1 \sqrt{\tau_1} \leq 1.
\end{align*}
\]

Intuitively, we face a tradeoff between the two decision variables. If the service area is too large, we can only accumulate SDD orders for a shorter duration to ensure that the vehicle has sufficient time to service all customers. Similarly, if we allow customers to place SDD orders for an excessive duration, we must concurrently shrink the service region to shorten the vehicle’s tour duration. Our goal is to balance these factors in such a way that the total number of SDD orders is maximized.

We first observe that it is inefficient for the vehicle to idle at the depot after completing its dispatch. If the vehicle returns to the depot prior to \( T \), the continuity of \( f \) implies that we can increase the service area, accumulation time, or both; this improves the objective while maintaining feasibility. As a result, constraint (2b) is tight at optimality. This observation, which will prove useful in analyses of more complicated systems, allows us to reduce the decision space to only the accumulation time variable. Specifically, given a fixed accumulation time \( \tau_1 \in (0, 1] \), the service area that maximizes the number of orders fulfilled in the \( \langle 1, 1 \rangle \) setting is given by \( A_1 = \frac{1-\tau_1}{c\sqrt{\tau_1}} \) via rearranging the constraint.
Figure 3: Objective value vs. accumulation time $\tau_1$ for $(1, 1)$ system, $\lambda = c = 1$.

We can now reformulate the problem solely over the variable $\tau_1$:

$$\max_{\tau_1 \in [0, 1]} \frac{\lambda}{c} (1 - \tau_1) \sqrt{\tau_1}.$$

This problem can be solved analytically via the first-order condition. The optimal solution is $\tau_1^* = \frac{1}{3}$, with optimal objective value $z_{1,1} = \frac{\lambda}{c} \frac{2}{3\sqrt{3}}$, and it follows that $A_1^* = \frac{\lambda}{c\sqrt{3}}$. Figure 3 depicts the plot of the objective value, $(1 - \tau_1) \sqrt{\tau_1}$, versus the accumulation time $\tau_1$ when $\lambda = c = 1$. Note the peak at $\tau_1 = \frac{1}{3}$, which is invariant to the values of $c$ and $\lambda$.

Consider the following example of a $(1, 1)$ system. For simplicity, suppose that the travel time between points is given by the $\ell_1$ metric to approximate a grid-like road network. Orders accumulate at a rate of 0.5 per hour per square mile within the chosen service region. The service day ranges from 9 AM to 6 PM, and the vehicle travels at 20 mph. Using the empirically estimated BHH constant of 1.0533 (with units of orders$^{-1/2}$) from [4], we arrive at parameter values of $\lambda = 4.5$ and $c = \frac{1.0533 \sqrt{4.5}}{20 \times 9} \approx 0.0124$. Via the results above, the optimal service area is approximately 93.02 sq. mi.; the vehicle dispatches at noon with approximately $93.02 \times 3 \times \lambda = 139.53$ orders and returns to the depot at 6 PM. The service region is a diamond centered at the depot with a driving radius of approximately 6.82 mi.

In practice, a scenario may arise where it is necessary to impose a restriction on the size of the service area. For example, the modeling of customer order arrivals and the routing time function may rely on a specific customer density which is bounded geographically, or the SDD retailer may only have regulatory authorization to operate within a certain area. To account for such a restriction, we can introduce the con-
straint $A_1 \leq B$ into the model. Proposition 1 extends the optimization results of the base $(1, 1)$ model under this constraint.

**Proposition 1.** The optimal dispatching policy for the $(1, 1)$ model where the service area is bounded by $A_1 \leq B$ is to serve an area of $A_1^* = \min\{\frac{2}{c\sqrt{3}}, B\}$ after accumulating orders for a duration $\tau_1^*$, where $\tau_1^*$ uniquely solves $\tau_1^* + cA_1^* \sqrt{\tau_1^*} = 1$.

**Proof.** See Appendix A.1

4 Multiple Vehicles, One Dispatch Each

Suppose the SDD system has a finite fleet of $m > 1$ vehicles, each dispatching once per day. The analysis of this $(m, 1)$ setting is more complex, but it admits more sophisticated managerial insights. Specifically, studying this setting allows us to answer the following fundamental question: can allowing service regions to vary over time improve the total order service rate of an SDD system?

We must now determine the service region $A_d$ and accumulation time $\tau_d$ for each dispatch $d \in [m]$ (or, equivalently, for each vehicle). Figure 4 depicts an example dispatching policy (not necessarily optimal) when $m = 2$; observe the difference in the service area associated with each vehicle. The formal optimization problem associated with the $(m, 1)$ model is as follows:

$$\max_{A, \tau \geq 0} \sum_{d=1}^{m} \lambda A_d \tau_d$$  \hspace{1cm} (3a)

s.t.  \hspace{1cm}$$\sum_{\delta=1}^{d} \tau_\delta + cA_d \sqrt{\tau_d} \leq 1 \hspace{1cm} \forall d \in [m].$$  \hspace{1cm} (3b)

Constraints (3b) in particular define a $2m$-dimensional non-convex feasible region, implying that the problem may be difficult to solve by conventional methods.

However, as in the single-vehicle case, it is inefficient for a vehicle to leave idle time after its return to
the depot. If a vehicle \( d \in [m] \) returns to the depot prior to \( T \), the continuity of \( f \) implies that we can slightly increase the service area associated with the vehicle’s dispatch. This increases \( \lambda A_d \tau_d \) (and thus the overall objective) while satisfying vehicle \( d \)'s feasibility condition. Additionally, increasing \( A_d \) does not affect the operation or service area of any other vehicle, so overall feasibility is maintained as well. Therefore, the constraints (3b) are all tight at optimality. By simple inspection, the dispatching policy illustrated in Figure 4 is thus suboptimal. Analogous to the \( (1, 1) \) setting, we can again reduce the decision space to only the accumulation time variables. Proposition 2 formalizes this result.

**Proposition 2.** Given a set of positive accumulation times, \( \{\tau_1, \tau_2, \ldots, \tau_m\} \), for the \( (m, 1) \) model, the service areas that maximize the total number of orders served are given by \( A_d = \frac{1 - \sum_{d=1}^{m} \tau_d}{c\sqrt{\tau_d}} \), for all \( d \in [m] \).

**Proof.** See Appendix A.2.

Here is another interpretation of the argument above: given a set of accumulation times, each vehicle is indifferent to the service regions associated with the other \( m - 1 \) vehicles. Hence, given a set of accumulation times, each vehicle operates within its own region with a truncated service day. Nevertheless, this does not imply that vehicles can be dispatched in a greedy fashion throughout the service day. The dispatcher must still determine the set of optimal accumulation times, which are linked since each accumulation time influences the departure times of later dispatches.

Applying Proposition 2, we arrive at the following optimization problem:

\[
\max_{\tau \geq 0} \quad \frac{\lambda}{c} \sum_{d=1}^{m} (1 - \sum_{\delta=1}^{d} \tau_{\delta}) \sqrt{\tau_d}
\]

\[
\text{s.t.} \quad \sum_{d=1}^{m} \tau_d \leq 1.
\]

Solving this problem to optimality may still be computationally inefficient due to the non-linear, non-convex objective. We therefore seek an efficient solution method.

### 4.1 Model analysis and structural properties

Consider the perspective of the system manager immediately after the first vehicle dispatches, i.e., at \( t = \tau_1 \). Once this occurs, the first vehicle has no further bearing on the service areas or accumulation times associated with the remaining \( m - 1 \) vehicles. Intuitively, the subsequent decisions are “memoryless” with respect to the first vehicle, equivalent to starting with \( m - 1 \) vehicles but with a reduced service day.
Algorithmically, we can use this property to derive a recursive solution procedure for the \( \langle m, 1 \rangle \) family of instances. Consider the specific case of \( m = 2 \). At the time of the first vehicle’s dispatch, we know exactly how to optimize the second vehicle’s dispatch over the service day’s remaining duration, \( 1 - \tau_1 \). This is true no matter the actual value of \( \tau_1 \): by our analysis in Section 3, the second vehicle should dispatch after accumulating orders for one third of the remaining day in order to optimize its total orders served. Given that \( \tau_2 = (1 - \tau_1) / 3 \) in an optimal solution, we can efficiently optimize solely over \( \tau_1 \) to optimize the overall problem. In general, given knowledge of the structure of an \( \langle m - 1 \rangle \)-vehicle optimal solution, we can efficiently solve for \( \tau_1 \) in the \( m \)-vehicle problem. Theorem 3 formalizes this idea into a solution approach.

**Theorem 3.** Given the optimal dispatch policy \( \{ (\tau_{m,d}^*, A_{m,d}^*) \}_{d=1}^m \) for the \( \langle m, 1 \rangle \) model with an objective value of \( z_{m,1} = \frac{\lambda}{c} \sum_{d=1}^m (1 - \sum_{\delta=1}^d \tau_{m,\delta}^*)^{1.5} \tau_{m,d}^* \), we can formulate the \( \langle m + 1, 1 \rangle \) optimization problem as

\[
\max_{0 \leq \tau_{m+1,1} \leq 1} \frac{\lambda}{c} (1 - \tau_{m+1,1}) \sqrt{\tau_{m+1,1}} + (1 - \tau_{m+1,1})^{1.5} z_{m,1}.
\]  

Furthermore, we can perform the following updates to the \( \langle m, 1 \rangle \) optimal policy to obtain the \( \langle m + 1, 1 \rangle \) optimal policy:

\[
\tau_{m+1,d}^* \leftarrow (1 - \tau_{m+1,1}^*) \tau_{m,d}^* -1 \quad \forall d \in \{2, \ldots, m+1\},
\]

\[
A_{m+1,d}^* \leftarrow A_{m,d-1}^* \sqrt{1 - \tau_{m+1,1}^*} \quad \forall d \in \{2, \ldots, m+1\},
\]

\[
A_{m+1,1}^* \leftarrow \frac{1 - \tau_{m+1,1}^*}{c} \sqrt{\tau_{m+1,1}^*}.
\]

**Proof.** See Appendix A.3

Beginning with the solution for \( m = 1 \), which we computed in Section 3, we can iteratively compute the optimal dispatching solution and objective for any \( m \) using the method above. We next seek to guarantee that the optimization problem in (5) is efficiently solvable.

Another property of optimal dispatching solutions proves useful to this end. Intuitively, we expect that the optimal total quantity of orders served increases as the number of vehicles \( m \) increases. Concurrently, the optimal first accumulation time shrinks towards zero as the number of vehicles increases. Theorem 4 formalizes this result.

**Theorem 4.** As the number of vehicles \( m \) in the \( \langle m, 1 \rangle \) model increases, the optimal accumulation time of
the first vehicle $\tau_{m,1}^*$ strictly decreases, and the total number of SDD orders $z_{m,1}$ served strictly increases. Furthermore, as $m \to \infty$, $\tau_{m,1}^* \to 0$ and $z_{m,1} \to \infty$; specifically, $z_{m,1} = \Theta(\sqrt{m})$.

Proof. See Appendix A.4.

It follows that the optimal solution to problem (5) lies within the interval $[0, \tau_{m,1}^*]$. Hence, we can instead solve the following problem within our solution procedure:

$$\max_{0 \leq \tau_{m+1,1} \leq \tau_{m,1}^*} \frac{\lambda}{c} (1 - \tau_{m+1,1}) \sqrt{\tau_{m+1,1}} + (1 - \tau_{m+1,1})^{1.5} z_{m,1}. \tag{6}$$

Additionally, the objective function in Problem 6 is concave over $[0, \tau_{m,1}]$ (see Lemma 19, Appendix A.8). As a result, optimizing Problem 6 is guaranteed to be efficient, as is our overall solution method. Observe that we began with a formulation over $2m$ decision variables in Problem 3, reduced the decision space in half via Proposition 2, and finally arrived at a recursive, one-variable concave maximization problem.

Recall our example from Section 3 with $\ell_1$ travel distances and times, $\lambda = 9$, and $c \approx 0.0124$. Using the approach described above, implemented in MATLAB 2019b using fminbnd to solve (6) to optimality, we calculated the optimal solutions for these parameters up to $m = 4$. The results are displayed in Table 1, and the solutions for $m = 1, 2, 3$ are illustrated to scale in Figure 5. Note that the relative scale of areas and quantities across different values of $m$ is invariant to $\lambda$ and $c$. Additionally, the accumulation and dispatch departure times are invariant to $\lambda$ and $c$; e.g., the second dispatch’s optimal departure time in the $\langle 3,1 \rangle$ model is always 11:34 AM when the service day is 9 AM to 6 PM regardless of the values of $\lambda$ and $c$.

These computed values suggest clear trends concerning the structure of optimal $\langle m,1 \rangle$ dispatching policies, which may provide important insights to system managers. First, we observe diminishing marginal returns in the total order quantity as more vehicles are added to the system. This potential trend suggests that, at some point, operating an additional vehicle may provide no practical benefit in the $\langle m,1 \rangle$ setting. Indeed, it can be shown that this is always the case; Proposition 5 formalizes this property.

**Proposition 5.** There is a strictly decreasing marginal gain in additional orders served in the $\langle m,1 \rangle$ model when adding an additional vehicle. That is, $(z_{m+2,1} - z_{m+1,1}) < (z_{m+1,1} - z_{m,1})$ for all $m \geq 1$.

Proof. See Appendix A.5.

We also observe that, over the course of the day, dispatch accumulation times seem to increase while service areas seem to decrease. The proof of Proposition 6 shows that these observations indeed hold for any
<table>
<thead>
<tr>
<th>Dispatch #</th>
<th>Accum. Time (hrs.)</th>
<th>Area (sq. mi.)</th>
<th>Radius (mi.)</th>
<th>Depart. Time</th>
<th>Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.00</td>
<td>93.02</td>
<td>6.82</td>
<td>12:00 PM</td>
<td>139.53</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>139.53</td>
</tr>
<tr>
<td>$m = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.66</td>
<td>153.16</td>
<td>8.75</td>
<td>10:39 AM</td>
<td>126.92</td>
</tr>
<tr>
<td>2</td>
<td>2.45</td>
<td>84.02</td>
<td>6.48</td>
<td>1:06 PM</td>
<td>102.82</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>229.74</td>
</tr>
<tr>
<td>$m = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.12</td>
<td>200.12</td>
<td>10.00</td>
<td>10:07 AM</td>
<td>111.91</td>
</tr>
<tr>
<td>2</td>
<td>1.45</td>
<td>143.33</td>
<td>8.47</td>
<td>11:34 AM</td>
<td>104.01</td>
</tr>
<tr>
<td>3</td>
<td>2.14</td>
<td>78.63</td>
<td>6.27</td>
<td>1:42 PM</td>
<td>84.27</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300.19</td>
</tr>
<tr>
<td>$m = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.84</td>
<td>239.71</td>
<td>10.95</td>
<td>9:50 AM</td>
<td>100.24</td>
</tr>
<tr>
<td>2</td>
<td>1.01</td>
<td>190.60</td>
<td>9.76</td>
<td>10:51 AM</td>
<td>96.68</td>
</tr>
<tr>
<td>3</td>
<td>1.31</td>
<td>136.51</td>
<td>8.26</td>
<td>12:10 AM</td>
<td>89.85</td>
</tr>
<tr>
<td>4</td>
<td>1.94</td>
<td>74.89</td>
<td>6.12</td>
<td>2:06 PM</td>
<td>78.80</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>359.57</td>
</tr>
</tbody>
</table>

**Table 1:** Example computed optimal dispatching policies for the $(m,1)$ model, up to $m = 4$.

**Figure 5:** Service regions and dispatching policies to scale for $m = 1,2,3$ vehicles, $(m,1)$ setting.
The fact that optimal dispatch areas are strictly decreasing is an important design implication since, from a customer’s perspective, it implies that SDD offerings will not “fluctuate” during the course of a service day.

**Proposition 6.** In the optimal dispatch policy for the \((m, 1)\) model, accumulation times are strictly increasing while service areas are strictly decreasing; that is, \(\tau_{m,1}^* < \tau_{m,2}^* < \cdots < \tau_{m,m}^*\) and \(A_{m,1}^* > A_{m,2}^* > \cdots > A_{m,m}^*\).

**Proof.** See Appendix A.6.

Furthermore, the result verifies the intuition that the system can operate more efficiently by limiting faraway customers to an earlier cutoff, and simultaneously offering nearby customers SDD until later in the day. As the fleet size increases, the system’s overall footprint (the largest area served) increases, but some customers are worse off. For instance, by increasing the fleet from one to two vehicles, we can significantly increase the area where we offer SDD and the total number of orders served; however, some customers will experience a reduced SDD order cutoff. In the example, customers outside of a 6.48-mile radius but within a 6.82-mile radius would only be able to place SDD orders until 10:39 AM instead of noon.

### 4.1.1 Bounding the multiple dispatch case

The \((m, 1)\) model is appealing from an operational perspective, as it is simple to implement; each vehicle is only dispatched once, with a planned return at the end of the day. A manager may therefore wonder how much the system gains by adding dispatches, which may complicate the depot’s operations.

Consider the general \((m, D)\) model with arbitrary \(m\) and \(D\), and recall that \(z_{m,1} = \Theta(\sqrt{m})\) by Theorem 4. A simple corollary of this result characterizes the objective’s growth for arbitrary \(m\) and \(D\).

**Corollary 7.** \(z_{m,D} = \Theta(\sqrt{mD})\), and, for any fixed \(D\), \(z_{m,D} = \Theta(\sqrt{m})\).

**Proof.** See Appendix A.7.

Corollary 7 generalizes the growth rate of the \((m, 1)\) model to the case of an arbitrary number of dispatches. In particular, it implies \(z_{m,D}/z_{m,1} = \Theta(\sqrt{D})\), which means that the system’s potential gains from allowing \(D\) dispatches per vehicle instead of one are limited. However, Corollary 7 does not rule out the unlimited growth of \(z_{m,D}\) as \(D \to \infty\) for a fixed fleet size \(m\). In Section 5, we strengthen this upper bound by leveraging results from the \((1, D)\) case.
4.2 Impact of constrained service regions

The preceding \( \langle m, 1 \rangle \) results assume the SDD service area is unbounded and can be chosen as large as necessary. In particular, as the fleet size \( m \) grows, the largest service area \( A_{m,1}^* \) tends to infinity. Nonetheless, in practical situations it is natural to expect that the service area must be limited, either explicitly, such as by regulations that determine the maximum area where a company can offer SDD, or implicitly because SDD demand decreases or disappears once we are too far from the depot. Motivated by these considerations, we now consider the \( \langle m, 1 \rangle \) model where service areas are bounded by a maximum area \( B > 0 \).

As before, the resulting optimization problem is non-linear and non-convex; therefore, we are interested in an efficient solution method. A natural idea is to compare \( A_{1}^* \) in the unconstrained solution to \( B \). By Proposition 6, as long as the first service region has an area smaller than than \( B \), the unconstrained solution is feasible, and therefore optimal for the constrained problem. If \( A_{1}^* > B \), we fix \( A_1 = B \) and choose \( \tau_1 \) so the first vehicle returns at \( T \). We then re-optimize with respect to the remaining \( m - 1 \) vehicles and the remaining service day, repeating the process as required. Theorem 8 states that this intuitive procedure, formalized in Algorithm 1, indeed produces an optimal dispatching policy.

**Theorem 8.** For the \( \langle m, 1 \rangle \) model with an upper bound \( B > 0 \) on the service areas, Algorithm 1 returns an optimal policy. Additionally, the optimal areas satisfy \( A_{m,1}^* \geq A_{m,2}^* \geq \cdots \geq A_{m,m}^* \).

**Proof.** See Appendix A.8.

4.3 Value of varying service regions

We now return our focus to the original question regarding the benefit of allowing service regions to vary over time. To this end, we study the same \( \langle m, 1 \rangle \) model with the additional requirement that the service area must stay constant over the course of the day until the final vehicle’s dispatch. Formally, we add the constraint \( A = A_1 = A_2 = \cdots = A_m \). The full optimization problem for the fixed-area \( \langle m, 1 \rangle \) model is as follows:

\[
\begin{align*}
\max_{A, \tau} & \quad \sum_{d=1}^{m} \lambda A \tau_d \\
\text{s.t.} & \quad \sum_{\delta=1}^{d} \tau_{\delta} + cA\sqrt{\tau_d} \leq 1 \quad \forall d \in [D], \\
& \quad A \geq 0,
\end{align*}
\]
Algorithm 1 Iterative solution procedure for the constrained \( (m, 1) \) model

1: given vehicles \( m \), area upper bound \( B \), parameters \( c, \lambda \) 
2: initialize remaining service day time \( T_{\text{var}} \leftarrow 1 \), remaining vehicles \( w \leftarrow m \) 
3: while \( w > 0 \) do 
4: calculate the optimal policy \( \{ (\tau_{w,d}^*, A_{w,d}^*) \}_{d=1}^m \) to the unconstrained \( (w, 1) \) model as given by (3) 
5: Let \( \tau_{w,d}^* \leftarrow \tau_{w,d}^* T_{\text{var}} \forall d \in [w] \) 
6: Let \( A_{w,d}^* \leftarrow A_{w,d}^* T_{\text{var}} \sqrt{T_{\text{var}}} \forall d \in [w] \) 
7: if \( A_{w,1}^* \leq B \) then 
8: \( A_{m,m-w+d}^* \leftarrow A_{w,d}^* \forall d \in [w] \) 
9: \( \tau_{m,m-w+d}^* \leftarrow \tau_{w,d}^* \forall d \in [w] \) 
10: \( w \leftarrow 0 \) 
11: else 
12: \( A_{m,m-w+1}^* \leftarrow B \) 
13: \( \tau_{m,m-w+1}^* \leftarrow T_{\text{var}} + cB \left( cB - \sqrt{(cB)^2 + 4T_{\text{var}}} \right) \) 
14: \( w \leftarrow w - 1 \) 
15: \( T_{\text{var}} \leftarrow T_{\text{var}} - \tau_{m,m-w+1}^* \) 
16: end if 
17: end while 
18: return optimal dispatching policy \( \{ (\tau_{m,d}^*, A_{m,d}^*) \}_{d=1}^m \) 

\[ \tau_d \geq 0 \quad \forall d \in [D]. \quad (7d) \]

In the variable-area \( (m, 1) \) optimization model (3), we simplified the optimization problem by noting that, for a given set of accumulation times, the service areas should be as large as possible in order to serve a maximal number of orders. In the fixed-area setting, this is generally not possible since all areas must be equal. However, we can show that it is still a dominant dispatching policy to have all of the vehicles return to the depot exactly at the end of the service day. This implies that, for any given fixed service area, all of the constraints (7b) are tight at optimality. Proposition 9 formalizes this result.

**Proposition 9.** Consider a variant of the \( (m, 1) \) model where each service region serves a fixed area of size \( A > 0 \). The set of accumulation times that maximize the total number of orders served are such that \( \sum_{d=1}^D \tau_d + cA \sqrt{\bar{\tau}_d} = 1 \) for all dispatches \( d \in [D] \).

**Proof.** See Appendix A.9

Unfortunately, while the resulting problem is more tractable than (7), we don’t have a method analogous to that described in Theorem 3 to optimize for the order-maximizing area. Therefore, we rely on general-purpose numerical optimization software to solve for the optimal dispatching policy. To facilitate global...
<table>
<thead>
<tr>
<th>Dispatch #</th>
<th>Accum. Time (hrs.)</th>
<th>Area (sq. mi.)</th>
<th>Radius (mi.)</th>
<th>Depart. Time</th>
<th>Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>93.02</td>
<td>6.82</td>
<td>12:00 PM</td>
<td>139.53</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>93.02</td>
<td>6.82</td>
<td></td>
<td>139.53</td>
</tr>
<tr>
<td>$m = 2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.21</td>
<td>122.71</td>
<td>7.83</td>
<td>11:12 AM</td>
<td>135.51</td>
</tr>
<tr>
<td>2</td>
<td>1.39</td>
<td>122.71</td>
<td>7.83</td>
<td>12:36 PM</td>
<td>85.57</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>221.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.76</td>
<td>146.66</td>
<td>8.56</td>
<td>10:45 AM</td>
<td>128.92</td>
</tr>
<tr>
<td>2</td>
<td>1.22</td>
<td>146.66</td>
<td>8.56</td>
<td>11:58 AM</td>
<td>89.24</td>
</tr>
<tr>
<td>3</td>
<td>0.89</td>
<td>146.66</td>
<td>8.56</td>
<td>12:51 PM</td>
<td>64.94</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>283.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.46</td>
<td>167.28</td>
<td>9.15</td>
<td>10:27 AM</td>
<td>122.23</td>
</tr>
<tr>
<td>2</td>
<td>1.08</td>
<td>167.28</td>
<td>9.15</td>
<td>11:32 AM</td>
<td>89.96</td>
</tr>
<tr>
<td>3</td>
<td>0.82</td>
<td>167.28</td>
<td>9.15</td>
<td>12:21 PM</td>
<td>68.65</td>
</tr>
<tr>
<td>4</td>
<td>0.64</td>
<td>167.28</td>
<td>9.15</td>
<td>1:00 PM</td>
<td>53.94</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>334.79</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Example computed optimal dispatching policies for the fixed-area $\langle m, 1 \rangle$ model, up to $m = 4$.

Optimality certification, such software may require bounds on all decision variables. Proposition 10 provides an efficiently computable upper bound on $A$ to the optimization routine.

**Proposition 10.** Let $\{(\tau_d^*, A_d^*)\}_{d=1}^m$ denote the optimal solution to the variable-area $\langle m, 1 \rangle$ problem. The optimal area $A_m^*$ associated with the fixed-area $\langle m, 1 \rangle$ problem satisfies $A_m^* \leq \frac{1-\tau^-}{c\sqrt{\tau^-}}$, where $\tau^- \in [0, \frac{1}{3}]$ uniquely solves

$$\frac{1}{c} \sqrt{\tau(1-\tau)} = \frac{1}{m} A_m^* \sum_{d=1}^m \tau_d^*.$$  (8)

**Proof.** See Appendix A.10.

We solve this optimization problem via BARON 21.1.13 [38, 43] within a numerical tolerance not exceeding $10^{-7}$. For comparison with the time-varying $\langle m, 1 \rangle$ model, Table 2 presents the computed solutions to the fixed-area $\langle m, 1 \rangle$ model with parameters identical to the problem studied in Table 1. Figure 6 compares the variable-area and fixed-area solutions to scale for $m = 2$.

As before, the relative scale of areas and quantities across different values of $m$ is invariant to $\lambda$ and...
Additionally, the accumulation and dispatch departure times are invariant to $\lambda$ and $c$. Therefore, the relative objective value gaps between the fixed-area and variable-area models are invariant to $\lambda$ and $c$. We can use Tables 1 and 2 to compare the objective values between the two models. When $m = 2$, 3.9% more SDD orders can be served by allowing the service regions to vary. This gap is 6.0% and 7.4% for $m = 3$ and $m = 4$, respectively. Empirical evidence for up to $m = 10$ suggests that both the relative and absolute gap in the optimal order fill rate between the fixed-area and variable-area models increase with $m$, albeit at a decreasing rate. Therefore, we can conclude that allowing service areas to vary over time leads to significant gains in the SDD order quantity served in the $(m, 1)$ setting for realistic values of $m$.

These results also highlight the equity trade-offs involved in SDD service region design. In the example results described in Tables 1 and 2 when $m = 2$, varying service regions between dispatches yields a system footprint (largest service region) equivalent to an 8.75-mile driving radius. By restricting the system to a single, unchanging service region, we reduce the footprint to a 7.83-mile radius, a decrease of over 30 square miles. In other words, by requiring all customers in the chosen service region to be treated equally, we implicitly deny SDD to other customers that could be served in a more flexible system.

5 One Vehicle, Multiple Dispatches

The results in the previous section illustrate the effects of a changing fleet size on the SDD system using the $(m, 1)$ model, which assumes each vehicle makes a single dispatch per day. In this section we study the potential benefit of allowing a vehicle to make additional dispatches over varying service areas, using the $(1, D)$ model. One potential use of the $(1, D)$ model, which we illustrate in Section 6 on a real-world road
network, is for partitioning schemes, where each vehicle is responsible for a wedge-shaped region emanating from the depot. Our analysis of the \langle 1, D \rangle model is also useful to compare the \langle m, 1 \rangle and \langle m, D \rangle models.

5.1 Model formulation

Formally, we wish to find an optimal dispatch policy \{ (\tau_d^*, A_d^*) \}_{d=1}^D for the following problem:

\[
\max_{A, \tau \geq 0} \sum_{d=1}^D \lambda A_d \tau_d \\
\text{s.t.} \quad \sum_{d=1}^D \tau_d + cA_D \sqrt{\tau_D} \leq 1, \quad (9a) \\
\quad cA_d \sqrt{\tau_d} \leq \tau_{d+1} \quad \forall d \in [D-1]. \quad (9b)
\]

Constraint (9b) requires the vehicle to return to the depot by the end of the service day after its final dispatch, while constraints (9c) require the vehicle to return to the depot prior to departing on its next dispatch. Figure 2 is an example of a feasible dispatching policy for \( D = 2 \).

Analogously to the previous models, it is inefficient to leave the vehicle idle between dispatches. If a vehicle waits at the depot after completing its \( d \)-th dispatch, the total number of orders served can be increased by slightly increasing \( A_d \). Proposition 11 formalizes this observation.

**Proposition 11.** Given a set of positive accumulation times \{ \tau_1, \tau_2, \ldots, \tau_D \} for the \langle 1, D \rangle model, the set of service areas which maximize the total number of served orders are given by \( A_d = \frac{\tau_{d+1}}{c\sqrt{\tau_d}} \) for all \( d < D \), and \( A_D = \frac{1 - \sum_{d=1}^D \tau_d}{c\sqrt{\tau_D}} \).

**Proof.** See Appendix A.11.

This result implies that the dispatching policy in Figure 2 is suboptimal. Having no idle vehicle time during the course of the day after the first dispatch is a property found in other SDD planning models [e.g. 4, 25, 42] with deterministic order arrivals. More generally, this result also suggests that minimizing vehicle idle time between dispatches may be beneficial at the operational level.

Knowing that we can choose service areas to maximize orders served given a set of accumulation times, we focus on choosing the best set of accumulation times for the system. This reduces (9) to

\[
\max_{\tau \geq 0} \frac{\lambda}{c} \left( \sum_{d=1}^{D-1} \tau_{d+1} \sqrt{\tau_d} + \left( 1 - \sum_{d=1}^D \tau_d \right) \sqrt{\tau_D} \right) \quad (10a)
\]
Table 3: Optimal dispatching policies for the \((1,D)\) model for up to \(D = 4\) dispatches.

<table>
<thead>
<tr>
<th>Dispatch #</th>
<th>Accum. Time (hrs.)</th>
<th>Area (sq. mi.)</th>
<th>Radius (mi.)</th>
<th>Depart. Time</th>
<th>Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m = 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>93.02</td>
<td>6.82</td>
<td>12:00 PM</td>
<td>139.53</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(m = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>107.41</td>
<td>7.33</td>
<td>10:00 AM</td>
<td>53.71</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>53.71</td>
<td>5.18</td>
<td>2:00 PM</td>
<td>107.41</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>161.12</td>
</tr>
<tr>
<td>(m = 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
<td>109.06</td>
<td>7.38</td>
<td>9:09 AM</td>
<td>8.20</td>
</tr>
<tr>
<td>2</td>
<td>1.58</td>
<td>88.23</td>
<td>6.64</td>
<td>10:43 AM</td>
<td>69.49</td>
</tr>
<tr>
<td>3</td>
<td>4.12</td>
<td>41.66</td>
<td>4.56</td>
<td>2:50 PM</td>
<td>85.90</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>163.60</td>
</tr>
<tr>
<td>(m = 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.0016</td>
<td>109.09</td>
<td>7.39</td>
<td>9:00 AM</td>
<td>0.09</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td>106.94</td>
<td>7.31</td>
<td>9:09 AM</td>
<td>8.69</td>
</tr>
<tr>
<td>3</td>
<td>1.61</td>
<td>87.44</td>
<td>6.61</td>
<td>10:46 AM</td>
<td>70.20</td>
</tr>
<tr>
<td>4</td>
<td>4.13</td>
<td>41.04</td>
<td>4.53</td>
<td>2:53 PM</td>
<td>84.66</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>163.64</td>
</tr>
</tbody>
</table>

\[
s.t. \quad \sum_{d=1}^{D} \tau_d \leq 1. \tag{10b}\n\]

We solve this simplified \(d\)-dimensional problem (with a non-concave objective) over a convex set via BARON within a tolerance not exceeding \(10^{-7}\). We compute optimal solutions for up to \(D = 4\) for the same setting considered in Tables 1 and 2. Summary results are presented in Table 3 and optimal policies for up to \(D = 3\) are illustrated in Figure 7. As in the previous settings, the relative scale of areas and quantities across different values of \(D\) is invariant to \(\lambda\) and \(c\), and the accumulation and dispatch departure times are invariant to \(\lambda\) and \(c\).

We observe an increase in total orders served of 15.5% when using two dispatches instead of one. However, the marginal improvement when adding dispatches shrinks rapidly: only 1.5% more orders are served when using three dispatches instead of two, and only 0.002% more orders are served when using four dispatches instead of three. A similar trend is evident when observing the first dispatch times and quantities as the number of total dispatches increases. When \(D = 3\), the first dispatch accumulates only a handful
of orders for less than the first ten minutes of the day before dispatching. When $D = 4$, the first dispatch is entirely insignificant: the vehicle dispatches less than a minute into the day to serve less than one-tenth of an order. As a practical design implication, this suggests a vehicle should not be dispatched more than twice in a service day in SDD settings similar to the one we describe, as the marginal gains from additional dispatches are negligible.

As in the multiple-vehicle case, these results also highlight equity trade-offs in terms of SDD access. Compared to a single dispatch, by allowing two dispatches that serve different regions, we increase the system’s overall footprint from 93 to 107 square miles. However, customers outside a 5.18-mile driving radius but within a 6.82-mile radius see their SDD order cutoff reduced from noon to 10 AM.

Recall the behavior of the variable-area $\langle m, 1 \rangle$ model as $m$ increases: despite decreasing marginal returns, the total number of orders served grows with $\sqrt{m}$, and the first (largest) service area grows to infinity. Naturally, we ask whether the same is true in the $\langle 1, D \rangle$ systems; Lemma 12 and Theorem 13 state that this is not the case. Specifically, we show that the maximum number of orders that can be served with any number of dispatches $D$ is no more than twice the number of orders served by the optimal $\langle 1, 1 \rangle$ solution. Our empirical calculations suggest that this factor is in fact tighter, approximately 1.18 times the optimal $\langle 1, 1 \rangle$ order quantity.
Lemma 12. In the optimal dispatch policy for a $\langle 1, D \rangle$ model, optimal service areas are bounded with respect to a function of the $D$-th optimal accumulation time. Specifically, $A_d^* \leq \frac{2}{\lambda} \sqrt{\tau_D}$ for all $d < D$.

Proof. See Appendix A.12

Theorem 13. For any $D$, $z_{1,D} \leq 2z_{1,1} = 4\lambda/c3\sqrt{3}$.

Proof. See Appendix A.13

5.1.1 Multiple vehicles, multiple dispatches

Theorem 13 allows us to more precisely analyze the $\langle m, D \rangle$ model and compare it to $\langle m, 1 \rangle$, where we only allow one dispatch per vehicle. In Section 4, we showed that, $z_{m,D} = \Theta(\sqrt{D})$ when $m$ is fixed. A direct application of Theorem 13 provides a stronger result: for a fixed $m$, the total quantity of orders served is bounded above, regardless of the number of dispatches per vehicle $D$.

Theorem 14. For a fixed $m$ and for any $D$, $z_{m,D} \leq 2mz_{1,1}$ and $z_{m,D} \leq 16z_{m,1} \sqrt{m/27} \approx 3.1z_{m,1}\sqrt{m}$.

Proof. See Appendix A.14

As a consequence of this result, $z_{m,D}/z_{m,1}$ is bounded above by a constant for any fixed $m$. In other words, there is limited benefit to considering additional dispatches per vehicle regardless of the fleet size.

5.2 Value of varying service regions

We return to our primary question of quantifying the benefit associated with allowing areas to vary between dispatches. As we did previously in the multi-vehicle setting, we now consider the fixed-area variant of the $\langle 1, D \rangle$ model with the added constraint $A = A_1 = A_2 = \cdots = A_m$. The resulting optimization problem is as follows:

$$\max_{A, \tau} \sum_{d=1}^{D} \lambda A \tau_d$$

s.t.

$$\sum_{\delta=1}^{D} \tau_\delta + cA \sqrt{\tau_D} \leq 1,$$  \hspace{1cm} (11b)

$$cA \sqrt{\tau_d} \leq \tau_{d+1} \hspace{2cm} \forall d \in [D-1],$$  \hspace{1cm} (11c)

$$A \geq 0,$$ \hspace{6cm} (11d)

$$\tau_d \geq 0 \hspace{3cm} \forall d \in [D].$$ \hspace{1cm} (11e)
As in previous models, at optimality a vehicle does not idle after a dispatch. This property allows us to reduce the search space for the optimization problem. Proposition 15 formalizes this property for the fixed-area \( (1, D) \) model, expressed in terms of the total accumulation time.

**Proposition 15 (\cite{4}, Theorems 2 and 3).** Consider the fixed-area \( (1, D) \) model. Given a fixed total accumulation time \( \sum_{\delta=1}^{D} \tau_{D} \in (0, 1) \), the area \( A \) and set of accumulation times \( \tau_{1}, \ldots, \tau_{D} \) that maximize the total number of orders served satisfy

\[
c A \sqrt{\tau_{d}} = \tau_{d+1}, \quad \forall d < D, \quad \sum_{\delta=1}^{D} \tau_{\delta} + c A \sqrt{\tau_{D}} = 1.
\]

In other words, after the first dispatch, the vehicle never idles at the depot, and it returns to the depot exactly at the end of the service day after the last dispatch.

Therefore, constraints (11b) and (11c) hold at equality at an optimal solution. Additionally, values from the optimal solutions to the variable-area \( (1, D) \) problem can be used to derive an upper bound virtually identical to Proposition 10 on the optimal area in the fixed-area problem (see Appendix A.10 for further details). We again calculate optimal solutions via BARON with tolerance not exceeding \( 10^{-7} \) for the same parameter settings. Table 4 summarizes results for up to \( D = 4 \) dispatches, and Figure 8 compares the variable-area and fixed-area solutions for \( D = 2 \).

**Figure 8:** Variable-area solution vs. fixed-area solution to scale, \( (1, 2) \) setting.

We observe some structural similarities between optimal policies for the variable-area and fixed-area settings, but one notable difference: as the number of dispatches increases, the area served by the fixed-area model actually decreases. In a fixed-area setting, when dispatches increase the system perceives gains from shrinking its footprint while offering SDD until later to its reduced customer base.
<table>
<thead>
<tr>
<th>Dispatch #</th>
<th>Accum. Time (hrs.)</th>
<th>Area (sq. mi.)</th>
<th>Radius (mi.)</th>
<th>Depart. Time</th>
<th>Orders</th>
</tr>
</thead>
<tbody>
<tr>
<td>m = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td>93.02</td>
<td>6.82</td>
<td>12:00 PM</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m = 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1.06</td>
<td>77.35</td>
<td>6.22</td>
<td>10:03 AM</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.97</td>
<td>77.35</td>
<td>6.22</td>
<td>1:02 PM</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.20</td>
<td>72.57</td>
<td>6.02</td>
<td>9:11 AM</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.20</td>
<td>72.57</td>
<td>6.02</td>
<td>10:23 AM</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.96</td>
<td>72.57</td>
<td>6.02</td>
<td>1:21 PM</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m = 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0057</td>
<td>72.15</td>
<td>6.01</td>
<td>9:00 AM</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.20</td>
<td>72.15</td>
<td>6.01</td>
<td>9:12 AM</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.21</td>
<td>72.15</td>
<td>6.01</td>
<td>10:25 AM</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.95</td>
<td>72.15</td>
<td>6.01</td>
<td>1:22 PM</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Optimal dispatching policies for the fixed-area \( \langle 1, D \rangle \) model for up to \( D = 4 \) dispatches.

The fixed-area solutions also exhibit minimal marginal gains in the total orders served beyond \( D = 2 \) dispatches. Additionally, the first dispatch rapidly shrinks to insignificance. Most importantly, we observe that allowing service areas to vary in a \( \langle 1, 2 \rangle \) model leads to an additional 3.2% orders served over the course of the service day. As in the \( \langle m, 1 \rangle \) case, this highlights equity trade-offs in SDD access that we discuss further in the following section, in the context of a real-world case study.

6 Computational Examples

In this section, we describe results of a case study designed using realistic data from a metropolitan area, including with dispatch time functions calibrated with drive times from its road network. For selected examples, we also validate our models against a more detailed operational setting in which we simulate a service day with order arrivals given by a Poisson point process, and dispatch durations calculated on the road network. We use these computational examples to motivate further discussion regarding the roles of equity and access considerations in the design of SDD systems.

The study is set in the Phoenix, Arizona metropolitan area, with the depot located in the major suburb
of Glendale, Arizona. Each service day begins at 9 AM and ends at 6 PM. We assume a homogeneous order rate of 0.2 orders per hour per square mile. For additional realism, we assume each delivery incurs a service time of one minute, which may include time taken to load the package onto the vehicle at the depot or time taken for the vehicle driver to drop off packages at residences. Recall that our model assumes a dispatch time function of the form \( c_0 \sqrt{An} \), while including a per-order service time would seemingly require the routing time function to include a linear component. We instead adhere to the original functional form, and demonstrate that the model provides reliable solutions even when a small per-order service time is present.

For each instance in the study, we choose a distinct “best-fit” value of the BHH routing constant \( c_0 \) via a method detailed in Appendix B. A distinct value of \( c_0 \) is required for each instance because the value of the constant exhibits dependence on various parameters (particularly area and orders served) when calibrated for real-world road networks with multiple road types. At a high level, the method for choosing a value of \( c_0 \) for a particular model proceeds as follows. First, the model is solved with an initial guess of the BHH constant. Then, using the values of \( A \) and \( n \) associated with each dispatch, a BHH constant is calculated for each dispatch. The largest of these constants (or, alternately, some type of weighted average of these constants) is set as the new overall BHH constant, which is used to re-solve the model. This process is repeated until the BHH constant converges. Our routing constant estimation process is necessitated by the fact that areas change between dispatches; we refer to [6, 28] for recent examples of routing constant estimation for static real-world regions.

For the order arrival process, customer locations are generated uniformly at random along the road network using VeRoViz [34]. Specifically, initial customer locations are generated uniformly at random within 30 meters of an existing road, and each customer is then automatically assigned to its closest location on the road network. Isochrones and actual driving times between customer locations are queried via Openroute-service [20]. We calculate optimal vehicle tours with a standard arc-based asymmetric TSP formulation implemented in Gurobi 9.1.1 via Python 3.7.3. We created all maps in Leaflet via VeRoViz.

6.1 Multiple vehicles, one dispatch each

We first study the multi-vehicle model; specifically, we consider the \((2,1)\) case with two vehicles, each dispatching once per day. If the system planner allows service areas to vary between each dispatch, the first and second vehicles serve areas of 186 square miles and 102 square miles, respectively. To construct the corresponding service region for each vehicle, we seek an isochrone (i.e., a zone for which all of its locations
can be reached from the depot within a certain driving time) with the given area. In this case, the region reachable from the depot in 22 min. 21 sec. of driving time has an area of 186 square miles; this isochrone corresponds to the first vehicle’s service region. Similarly, the second vehicle’s service region has a drive time radius of 17 min. 5 sec. around the depot. Figure 9 illustrates the service regions for each vehicle. The policy implied by the continuous approximation model is as follows: the first vehicle dispatches at 10:39 AM, serves 61.60 orders, and returns at the 6 PM deadline; the second vehicle dispatches at 1:06 PM, serves 49.91 orders, and returns at the 6 PM deadline.

As a point of comparison, we also examine a system design in which the service regions are fixed between dispatches. Under this design assumption, the continuous approximation model implies a service area of 151 square miles (corresponding to a driving time radius of 20 min. 26 sec.) for each dispatch. The first vehicle dispatches at 11:12 AM, serves 66.66 orders, and returns at the 6 PM deadline. The second vehicle dispatches at 12:36 PM, serves 42.09 orders, and returns at the 6 PM deadline.

In order to validate these recommendations, we also assess the performance of the system in an operational setting. We consider a simulated version of a service day in which SDD orders arrive according to a Poisson point process with the same rate (0.2 per hour per square mile), with locations chosen randomly as described above. Vehicle dispatches to customer locations include a per-order service time of one minute and driving time given by the solution of a TSP that uses actual driving times between locations.

We implement the following operational version of the multi-vehicle dispatching policy. For each dispatch, orders accumulate from the beginning of the service day. As orders arrive into the system, the dispatcher re-calculates an optimal TSP tour (including the one minute per-delivery service time) that serves all accumulated demand. The SDD dispatcher allows orders to accrue until the calculated dispatch time equals the remaining time in the service day, at which point the first dispatch occurs. If an order arrives that would cause the vehicle to finish after the deadline, the dispatch occurs immediately but that order is not included, ensuring the vehicle returns before the end of the service day. However, if this order originates within the

**Figure 9**: Service regions for variable-area (2, 1) solution.
second vehicle’s region, it is added to the second vehicle’s load. The dispatch procedure for the second
vehicle is analogous to the first. We simulate 120 service days for each system design and serve orders
according to the aforementioned operational policy. We report average quantities and dispatch durations for
the operational simulations in Table 5 along with 95% confidence intervals (in parentheses). The predicted
amounts are remarkably close to their simulated counterparts. In particular, predicted total orders served
nearly coincide with the simulated operational quantities in both the variable- and fixed-area models.

<table>
<thead>
<tr>
<th></th>
<th>Variable Areas</th>
<th></th>
<th>Fixed Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted</td>
<td>Simulated</td>
<td>Predicted</td>
</tr>
<tr>
<td>Dispatch 1 Quantity</td>
<td>61.60 (± 0.59)</td>
<td>64.51 (± 0.59)</td>
<td>66.66 (± 0.63)</td>
</tr>
<tr>
<td>Dispatch 1 Duration (min.)</td>
<td>440.59 (± 1.88)</td>
<td>426.06 (± 1.88)</td>
<td>407.48 (± 2.18)</td>
</tr>
<tr>
<td>Dispatch 2 Quantity</td>
<td>49.91 (± 0.55)</td>
<td>46.94 (± 0.55)</td>
<td>42.09 (± 0.61)</td>
</tr>
<tr>
<td>Dispatch 2 Duration (min.)</td>
<td>293.71 (± 2.40)</td>
<td>285.51 (± 2.40)</td>
<td>323.80 (± 2.47)</td>
</tr>
<tr>
<td>Total Quantity Served</td>
<td>111.50 (± 0.86)</td>
<td>111.45 (± 0.86)</td>
<td>108.75 (± 0.98)</td>
</tr>
<tr>
<td>Total Dispatch Duration (min.)</td>
<td>734.29 (± 3.50)</td>
<td>711.57 (± 3.50)</td>
<td>731.27 (± 4.17)</td>
</tr>
</tbody>
</table>

Table 5: Predicted and simulated (operational) results for \((2, 1)\) instance.

We now examine the perspective of an e-retailer choosing between these two system designs. We con-
sider three criteria: profitability, access, and equity. Generally, the most important of these is profitability,
since margins on last-mile delivery tend to be small. The variable cost of a design may be proportional to
the average total dispatch duration (i.e., total routing and service time); however, the empirical difference in
this quantity is negligible between the two designs (711.57 min. versus 710.22 min. for the variable-area and
fixed-area model, respectively). Thus, the differentiating factor is the revenue earned by serving customers:
in the operational simulations, the variable-area design serves 2.53% more customers than the fixed-area
model on average. Whether this represents sufficient reason to select the variable-area model likely depends
on factors whose monetary value is not directly measurable, which we discuss next.

Consider the number of customers who have access to the SDD system at any level of service. Assuming
customers are distributed uniformly, the variable-area design provides SDD access to approximately 23%
more customers than the fixed-area design (calculated by comparing the area served by the first dispatch in
the variable-area design and the service region of the fixed-area design). From the e-retailer’s viewpoint,
greater customer access to SDD via the variable-area model can aid in establishing a larger customer base
for future expansions. The benefits to customers located further from the depot are evident, especially since
an e-retailer with a small fleet may be offering a niche product unavailable via other means. However, expanding the number of customers who have access to the SDD system makes some other customers worse off. In this case, we predict that the customers located in the region depicted in Figure 10, which has an area of 49 square miles, can place orders until 12:36 PM in the fixed-area design but can only place orders until 10:39 AM in the variable-area design.

This phenomenon motivates an analysis of equity issues. In an ideal scenario (with respect to equity), every potential customer in a metropolitan area would receive access and a high level of service. While a large established e-retailer may have the resources to provide such offerings, as Amazon did in response to criticism in 2016 [40, 41], the small e-retailer in this example likely cannot do so while remaining profitable. One measure of equity in this setting is the variation in service level between customers in the system. By this criteria, the fixed-area model is perfectly equitable: every single potential customer in the service region faces the same SDD order cutoff time. On the other hand, in the variable-area model, approximately 45% of the potential customers in the system (i.e., within the boundaries of the first vehicle’s service region) face a cutoff time nearly 2.5 hours earlier than the other 55%. This bias against distant customers may motivate the SDD e-retailer to prefer the less-profitable fixed-area model, especially if customers located farther from the depot are disproportionately from a particular socioeconomic group. It should be noted, however, that there are many characterizations of equity within logistics systems [e.g., 27]. A Rawlsian [37] approach to equity — often referred to as the maximin criterion — seeks to maximize the utility of the least well-off. By this measure, the variable-area design is more equitable because it provides some level of access to customers outside the fixed-area system. Ultimately, the choice of system design depends on which of these considerations have more weight for the system manager.
6.1.1 Unconstrained vs. constrained service areas

We conclude our discussion of \( \langle m, 1 \rangle \) systems with an illustration of how constraining the service area impacts system design, using results from Section 4.2. Suppose the same e-retailer has a fleet of three vehicles, each dispatching once daily. The unconstrained \( \langle 3, 1 \rangle \) model implies concentric service regions with areas of 246, 176, and 97 square miles for the first, second, and third dispatches, respectively (Figure 11a). The three dispatches would occur at 10:07 AM, 11:34 AM, and 1:42 PM and serve a total of 147.65 orders.

To focus on a smaller base of customers, the e-retailer may instead seek to constrain the area of each service region to \( B = 150 \) square miles. In this scenario, Algorithm 1 implies that the service areas of the first two dispatches are identical 150 square miles, while the third dispatch serves an area of 88 square miles. Figure 11b illustrates the service regions; note that the outer ring represents the service region for both the first and second dispatches. The three dispatches occur at 11:13 AM, 12:37 PM, and 2:25 PM and serve a total of 140.35 orders. We highlight two observations. First, our model predicts a relatively small reduction in the total quantity served when constraining the service area (approximately 5%). Second, constraining the areas extends the order placement windows, albeit for a smaller group of customers.

6.2 One vehicle, multiple dispatches

We now study the single vehicle, multiple dispatch model. The depot is located at the same address; however, we assume that the overall system has been partitioned into four geographical quadrants, each served by a single vehicle dispatching twice daily. We focus specifically on the \( \langle 1, 2 \rangle \) subsystem in the southeastern quadrant. The method used to estimate the routing constants in this subsection is described in Appendix B.

As before, we first consider the setting in which the system planner allows service areas to vary between each dispatch. The first and second dispatches serve areas of approximately 123 square miles and 61 square miles, respectively. These areas correspond to driving time radii of 31 min. 57 sec. and 24 min. 37 sec., respectively. The service regions for each dispatch are illustrated in Figure 12a. Based on the optimal solution’s policy, the vehicle first dispatches at 10 AM, serves 24.57 orders, and returns to the depot at 2 PM. The vehicle departs immediately on its second dispatch at 2 PM, serves 49.14 orders, and returns to the depot at the 6 PM deadline.

For comparison, we again examine an alternative subsystem design in which the service regions are fixed between dispatches. Under this design assumption, the approximation model implies a service area of
approximately 89 square miles (corresponding to a driving time radius of 28 min. 24 sec.) for each dispatch, illustrated in Figure 12b. The vehicle first dispatches at 10:04 AM, serves 18.54 orders, and returns to the depot at 1:02 PM. The vehicle dispatches immediately on its second dispatch at 1:02 PM, serves 52.80 orders, and returns to the depot at the 6 PM deadline. Note that, as expected, the total quantity served in operational simulations is greater in the variable-area design (73.70) than in the fixed-area design (71.73), and thus the discussions on profitability, access, and equity in Section 6.1 remain relevant in the \langle 1, 2 \rangle case considered here.

7 Conclusions

We studied the design of SDD systems, and particularly investigated whether these systems can benefit by allowing their service regions to vary over the course of the service day. We perform structural analyses for two important settings. First, we examine the case of multiple vehicles each dispatching once daily, which allows us to assess the marginal benefit of adding vehicles to a delivery fleet. Second, we consider one vehicle dispatching multiple times daily, and show that allowing a second dispatch indeed increases order volume, but additional dispatches offer negligible benefit. For each of these settings, we derive theoretical
properties that allow us to efficiently optimize the model and calculate vehicle dispatching policies. In order to quantify the value of allowing service areas to vary, we also calculate solutions to the problem of maximizing orders served with fixed service areas for each setting.

Our case study set in the Phoenix metropolitan area verifies our model’s applicability to real-world settings and its predictions’ fidelity when compared to a detailed operational model. Our findings suggest that variable-area system designs earn more revenue and provide some level of SDD access to more customers, but fixed-area models entail a greater degree of equity for customers within the system. Future studies may consider differentiating other system parameters, such as varying prices over the course of the day, in order to increase system efficiency and profitability.

Acknowledgements

Dipayan Banerjee’s work was supported via a U.S. National Science Foundation Graduate Research Fellowship (DGE-1650044) and the U.S. Federal Highway Administration’s Eisenhower Transportation Research Fellowship. The authors would like to thank the University at Buffalo’s Optimator Lab for their development of the open-source VeRoViz package.

References


Appendix A  Omitted Proofs

A.1 Proof of Proposition 1

First consider the case where $B \geq \frac{2}{c\sqrt{3}}$. In this case, the optimal dispatch policy for the unconstrained-area problem of $\tau_1^* = \frac{1}{3}$, $A_1^* = \frac{2}{c\sqrt{3}}$ is feasible and therefore optimal for the constrained-area problem. Furthermore, it is also the dispatch policy prescribed by Proposition 1, which proves this case.

Now consider the case where $B < \frac{2}{c\sqrt{3}}$. Let us define time $\tau_B$ as the time which solves $\tau_B + cB\sqrt{\tau_B} = 1$ over $\tau_B \in (0, 1]$. As $\tau_B + cB\sqrt{\tau_B}$ is strictly increasing with respect to $\tau_B$ over this domain, it must be a unique solution to the equation. Furthermore, as $B < \frac{2}{c\sqrt{3}}$, $\tau_B \in \left(\frac{1}{3}, 1\right]$. For any given $\tau_1 \in [0, \tau_B]$ the optimal service area choice is to set $A_1 = B$, and for any given $\tau_1 \in [\tau_B, 1]$ the optimal service area choice is to set $A_1 = \frac{1 - \tau_1}{c\sqrt{3}}$. It follows that the maximal number of orders served is equal to $\lambda B\tau_1$ when $\tau_1 \in [0, \tau_B]$, which is maximized when $\tau_1 = \tau_B$. Additionally, the maximal number of orders served is equal to $\frac{2}{c\sqrt{3}}(1 - \tau_1)\sqrt{\tau_1}$ when $\tau_1 \in [\tau_B, 1]$, which is also maximized when $\tau_1 = \tau_B$. Thus in the case that $B < \frac{2}{c\sqrt{3}}$, we have that $\tau_1^* = \tau_B$ and $A_1^* = B$. As this is the dispatch policy prescribed by Proposition 1, this case is proven as well. □

A.2 Proof of Proposition 2

Given a set of fixed, positive, accumulation times, $\{\tau_1, \tau_2, \ldots, \tau_m\}$, consider the $d$-th vehicle to dispatch. Inequality (3b) constrains the service area by: $A_d \leq \frac{1 - \sum_{d=1}^m \tau_d}{c\sqrt{d}}$. As the objective value (3a) increases linearly with each $A_d$, we have that $A_d = \frac{1 - \sum_{d=1}^m \tau_d}{c\sqrt{d}}$ maximizes the number of orders fulfilled for the $d$-th vehicle. As this is true for all vehicles $d$, we are done. □

A.3 Proof of Theorem 3

Consider the optimization problem for the $(m+1, 1)$ model:

$$\max_{\tau} \frac{\lambda}{c} \left( \sum_{d=1}^{m+1} \left( 1 - \sum_{\delta=1}^{d} \tau_{\delta} \right) \sqrt{\tau_d} \right) \tag{A1a}$$

s.t. $\sum_{d=1}^{m+1} \tau_d \leq 1,$ \hspace{1cm} (A1b)

$\tau_d \geq 0$ \hspace{1cm} \forall d \in [D]. \tag{A1c}$

We can re-formulate the problem as follows:

$$\max_{\tau} \frac{\lambda}{c} \left( (1 - \tau_1)\sqrt{\tau_1} + \sum_{d=2}^{m+1} \left( (1 - \tau_1) - \sum_{\delta=2}^{d} \tau_{\delta} \right) \sqrt{\tau_d} \right) \tag{A2a}$$

s.t. $\sum_{d=2}^{m+1} \tau_d \leq (1 - \tau_1), \tag{A2b}$

$\tau_d \geq 0$ \hspace{1cm} \forall d \in [D]. \tag{A2c}$

Note that, given a value of $\tau_1 \in [0, 1)$, choosing the optimal values for $\tau_2, \ldots, \tau_{m+1}$ equates to solving the $(m, 1)$ problem over a reduced service day. With this in mind, define $\tau_d'$ such that $\tau_d' = \frac{\tau_d}{(1 - \tau_1)}$ for all $d \geq 2$ in order to equate the remaining accumulation times as proportions of the remaining service day. Thus, we
can again re-formulate the problem as:

$$
\max_{\tau} \frac{\lambda}{c} \left( (1 - \tau_1) \sqrt{\tau_1} + (1 - \tau_1)^{1.5} \sum_{d=2}^{m+1} \left( 1 - \frac{d}{\sum_{\delta=2}^{d} \delta} \right) \sqrt{\tau_d} \right)
$$

(A3a)

\[ \text{s.t. } \sum_{d=2}^{m+1} \tau_d' \leq 1, \quad \tau_1 \leq 1, \quad \tau_d \geq 0 \quad \forall d \in [D]. \] (A3b)

There are no constraints in this optimization problem linking the $\tau_d'$ decision variables to the $\tau_1$ decision variable. Thus we can independently optimize for the $\tau_d'$ decision variables; this equates to solving the $\langle m, 1 \rangle$ model to optimality. By the presumptions of the Theorem, we have that $\tau_d' = \tau_{m,d-1}$ for all $d \geq 2$. Additionally, by Proposition 2 we have that $A_d' = A_{m,d-1}'$ for all $d \geq 2$. What remains in the $\langle m + 1, 1 \rangle$ model is to optimize:

$$
\max_{0 \leq \tau_1 \leq 1} \frac{\lambda}{c} (1 - \tau_1) \sqrt{\tau_1} + (1 - \tau_1)^{1.5} z_{m,1},
$$

which proves the first claim given in Theorem 3. Once this function is optimized for $\tau_1$, we can use Proposition 2 to determine that $A_1' = \frac{1}{c} (1 - \tau_1)' (\tau_1)'^{0.5}$. Furthermore, we can translate the optimal $(\tau_d, A_d')$ decision variables back to the $(\tau_d, A_d)$ decision space by performing the updates:

$$
\tau_d' \leftarrow (1 - \tau_1) \tau_d' = (1 - \tau_1) \tau_{m,d} \quad \forall d \geq 2
$$

and

$$
A_d \leftarrow \frac{1 - \sum_{\delta=2}^{d} \tau_\delta'}{\sqrt{\tau_d}} = \frac{(1 - \tau_1)^{0.5} (1 - \sum_{\delta=2}^{d} \tau_\delta')}{\sqrt{\tau_d}} = (1 - \tau_1)^{0.5} A_d' = (1 - \tau_1)^{0.5} A_{m,d} \quad \forall d \geq 2,
$$

which completes the proof.

\[ \square \]

### A.4 Proof of Theorem 4

Fix a number of vehicles $m$. Consider the objective function of the one variable optimization problem for $m + 1$ vehicles as given in Theorem 3

$$
\frac{\lambda}{c} (1 - \tau_{m+1,1}) \sqrt{\tau_{m+1,1}} + (1 - \tau_{m+1,1})^{1.5} z_{m,1}.
$$

When solving for the optimal value of $\tau_{m+1,1} \in [0, 1]$, first order conditions imply that $\tau_{m+1,1}^*$ is the unique value of $\tau_{m+1,1} \in (0, \frac{1}{2}]$ which satisfies the equation

$$
\frac{1}{3} = \tau_{m+1,1} + \frac{c}{\lambda} z_{m,1} \sqrt{(1 - \tau_{m+1,1}) (1 - \tau_{m+1,1})}.
$$

By the uniqueness of $\tau_{m+1,1}^*$ and the fact that $\tau_{m+1,1}^* \neq 0$, we are able to claim that $z_{m+1,1} > z_{m,1}$. From this, it directly follows that $\tau_{m+2,1} < \tau_{m+1,1}^*$ as $\tau_{m+2,1}^*$ is the unique value of $\tau_{m+2,1} \in (0, \frac{1}{3}]$ which satisfies the
equation
\[ \frac{1}{3} = \tau_{m+2,1} + \frac{c}{\lambda} z_{m+1,1} \sqrt{(\tau_{m+2,1})(1 - \tau_{m+2,1})}. \]

Thus, the first part of Theorem 4 is proven.

Now consider an \( m \) vehicle, one dispatch each policy where each vehicle (feasibly) accumulates orders for \( \frac{1}{m+1} \) units of time, that is, \( \tau_d = \frac{1}{m+1} \) for all \( d \). By Proposition 2, we would like these vehicles to each serve a maximal area of \( A_d = \frac{1 - \tau_d}{c \sqrt{\tau_d}} \). It follows that the total number of orders served by this policy is equal to
\[ \frac{\lambda}{c} \sum_{d=1}^{m} \left( 1 - \sum_{\delta=1}^{d} \tau_\delta \right) \sqrt{\tau_d} = \frac{\lambda}{c} \sum_{d=1}^{m} \left( 1 - \frac{d}{m+1} \right) \sqrt{\frac{1}{m+1}} = \frac{\lambda m}{2c \sqrt{m+1}} > \frac{\lambda}{4c} \sqrt{m}, \]
which tends to infinity as \( m \to \infty \). As the optimal \( m \) vehicle policy serves at least as many orders as this policy, it follows that \( z_{m,1} \to \infty \); specifically, \( z_{m,1} = \Omega(\sqrt{m}) \).

We next show that \( z_{m,1} = \Theta(\sqrt{m}) \) by constructing an upper bound. We relax the problem by removing the vehicle return deadline, instead limiting the duration of each dispatch to not exceed 1. The relaxed problem is as follows:

\[
\begin{align*}
\max_{A, \tau \geq 0} & \quad \frac{\lambda}{c} \sum_{d=1}^{m} A_d \tau_d \\
\text{s.t.} & \quad c A_d \sqrt{\tau_d} \leq 1 \quad \forall d \in [m], \\
& \quad \sum_{d=1}^{m} \tau_d \leq 1.
\end{align*}
\]

Without loss of optimality, we may assume the constraints (A4b) hold at equality. This implies that, for all \( d \in [m], A_d = \frac{1}{c \sqrt{\tau_d}} \). As such, we may rewrite the problem without the \( A_d \) variables:

\[
\begin{align*}
\max_{\tau \geq 0} & \quad \frac{\lambda}{c} \sum_{d=1}^{m} \sqrt{\tau_d} \\
\text{s.t.} & \quad \sum_{d=1}^{m} \tau_d \leq 1.
\end{align*}
\]

The optimal solution to this problem is \( \tau_1 = \tau_2 = \cdots = \tau_m = \frac{1}{m} \). The corresponding objective value is
\[ \frac{\lambda}{c} m \sqrt{\frac{1}{m}} = \frac{\lambda}{c} \sqrt{m}. \] Thus, for all \( m, z_{m,1} \leq \frac{\lambda}{c} \sqrt{m} \), implying \( z_{m,1} = \Theta(m) \). We conclude that \( z_{m,1} = \Theta(m) \).

What remains to be seen is that as \( m \to \infty \), \( z_{m,1} \to 0 \). By the construction of \( z_{m+1,1}^* \) via the first order conditions described above, we have that
\[
\begin{align*}
z_{m+1,1} &= \frac{\lambda}{c} \left( 1 - z_{m+1,1}^* \right) \sqrt{z_{m+1,1}^* + \left( 1 - z_{m+1,1}^* \right)^1 z_{m,1}} \\
&= \frac{\lambda}{c} \frac{1 - z_{m+1,1}^*}{\sqrt{z_{m+1,1}^*}} \left( z_{m+1,1}^* + \frac{c}{\lambda} z_{m,1} \sqrt{z_{m+1,1}^* (1 - z_{m+1,1}^*)} \right).
\end{align*}
\]
As we know that \( z_{m,1} \to \infty \) as \( m \to \infty \), we can equivalently state that as \( m \to \infty \), \( \frac{1 - \tau_{m+1,1}^*}{3\sqrt{\tau_{m+1,1}}} \to \infty \). This implies that \( \tau_{m,1}^* \to 0 \), which completes the proof.

\section*{A.5 Proof of Theorem 5}

Fix a number of vehicles \( m \). Consider the objective function of the one variable optimization problem given in Theorem 3. It can be shown via first order conditions that \( \tau_{m+1,1}^* \) is the unique value of \( \tau_{m+1,1} \in (0, \frac{1}{3}] \) which satisfies the equation \( 1 = 3\tau_{m+1,1} + \frac{3c}{\lambda} z_{m,1} \sqrt{\tau_{m+1,1}}(1 - \tau_{m+1,1}) \) and that \( z_{m+1,1} = \frac{\lambda}{c} \frac{1 - \tau_{m+1,1}^*}{3\sqrt{\tau_{m+1,1}}} \) (see the proof of Theorem 4). This leads to the relation

\[
\frac{z_{m+1,1} - z_{m,1}}{\lambda} = \frac{1 - \tau_{m+1,1}^*}{c \sqrt{\tau_{m+1,1}}} - \frac{1 - 3\tau_{m+1,1}^*}{3 \sqrt{\tau_{m+1,1}}(1 - \tau_{m+1,1})},
\]

which decreases as \( \tau_{m+1,1}^* \) decreases and tends to 0 as \( \tau_{m+1,1}^* \) tends to 0. Since \( \tau_{m+1,1}^* \) is decreasing as \( m \) increases by Theorem 4 we have that \( z_{m+1,1} - z_{m,1} \) also decreases as \( m \) increases. Therefore, we can conclude that \( (z_{m+1,1} - z_{m,1}) < (z_{m+1,1} - z_{m,1}) \). Additionally, since \( \tau_{m+1,1}^* \to 0 \) as \( m \to \infty \) by Theorem 4 we have that \( (z_{m+1,1} - z_{m,1}) \to 0 \) as \( m \to \infty \).

\section*{A.6 Proof of Proposition 6}

For the sake of induction, assume for a given \( m \) that \( \tau_{m,1}^* < \tau_{m,2}^* < \cdots < \tau_{m,m}^* \). We now show that it must be true that \( \tau_{m+1,1}^* < \tau_{m+1,2}^* < \cdots < \tau_{m,m+1}^* \). By Theorem 3 we have that \( \tau_{m+1,d}^* = (1 - \tau_{m+1,1}^*) \tau_{m,d-1}^* \) for all \( d \geq 2 \). Thus, we can infer that \( \tau_{m+1,2}^* < \tau_{m+1,3}^* < \cdots < \tau_{m+1,m+1}^* \) by induction. What remains to be seen is if \( \tau_{m+1,1}^* < \tau_{m+1,2}^* \).

Consider the objective value given in (A5), \( \frac{\lambda}{c} \frac{1 - \tau_{m+1,1}^*}{\sqrt{\tau_{m+1,1}}} + \frac{\lambda}{c} (1 - \tau_{m+1,1}^* - \tau_{m+1,2}^*) \sqrt{\tau_{m+1,2}^*} \). The only term of this summation that depends on either \( \tau_{m+1,1} \) or \( \tau_{m+1,2} \), but not their sum, is given by

\[
\frac{\lambda}{c} (1 - \tau_{m+1,1}^*) \sqrt{\tau_{m+1,1}} + \frac{\lambda}{c} (1 - \tau_{m+1,1}^* - \tau_{m+1,2}^*) \sqrt{\tau_{m+1,2}^*}.
\]  \hfill (A6)

We claim that (A6) can never be maximized when \( \tau_{m+1,1} \geq \tau_{m+1,2} \). Consider a fixed \( \theta = \tau_{m+1,1} + \tau_{m+1,2} \), and note that \( 0 < \theta \leq 1 \). We can rewrite (A6), without the multiplicative constant, as a function \( h_\theta : [0, \theta] \to \mathbb{R} \) of \( \tau_{m+1,1} \) that we wish to maximize in the interval \( \tau_{m+1,1} \in [0, \theta] \):

\[
h_\theta(\tau_{m+1,1}) = (1 - \tau_{m+1,1}) \sqrt{\tau_{m+1,1}} + (1 - \theta) \sqrt{\theta - \tau_{m+1,1}}.
\]

Differentiating once and twice gives

\[
h'_\theta(\tau_{m+1,1}) = \frac{1 - 3\tau_{m+1,1}}{2\sqrt{\tau_{m+1,1}}} - \frac{1 - \theta}{2\sqrt{\theta - \tau_{m+1,1}}},
\]
\[ h''_\theta (\tau_{m+1,1}) = -\frac{1}{4\tau_{m+1,1}^{3/2}} - \frac{3}{4\sqrt{\tau_{m+1,1}}} - \frac{1 - \theta}{4(\theta - \tau_{m+1,1})^{3/2}}. \]

Observe that \( h'_{\theta} (\theta/4) > 0, h'_{\theta} (\theta/2) < 0, \) and \( h''_{\theta} (\tau_{m+1,1}) < 0 \) for all \( \tau_{m+1,1} \in (0, \theta) \). It follows that \( h_{\theta} \)'s unique maximizer is located in the interval \((\theta/4, \theta/2)\), implying that \( A6 \) can never be maximized when \( \tau_{m+1,1} \geq \tau_{m+1,2} \) (i.e., when \( \tau_{m+1,1} \geq \theta/2 \)). Thus, we have that \( \tau^*_m < \tau_{m+1,1} < \tau_{m+1,2} \).

To finish our proof by induction, what remains to be seen is if a base case value of \( m \) yields \( \tau^*_m < \tau^*_{m+1} < \cdots < \tau^*_m \). From Table 1 we see that this is indeed true for \( m = 2 \). Thus, we have shown that accumulation times are strictly increasing throughout the service day. From this fact, and Proposition 2, it directly follows that the optimal service areas are strictly decreasing throughout the service day.

**A.7 Proof of Theorem 7**

Suppose we relax the \( (m,D) \) problem by removing constraint \( (1b) \), which requires that each vehicle’s dispatches are non-overlapping; this results in the \( (m,D,1) \) problem. Therefore, \( z_{m,1} \leq z_{m,D} \leq z_{mD,1} \). Since \( z_{mD,1} = \Theta(\sqrt{mD}) \) by Theorem 4, it follows that \( z_{m,D} = \Theta(\sqrt{mD}) \). Additionally, since \( z_{m,1} = \Theta(\sqrt{m}) \) by Theorem 4, for any fixed \( D \) we also have that \( z_{m,D} = \Theta(\sqrt{m}) \).

**A.8 Proof of Theorem 8**

We prove a series of results which together imply the correctness of Algorithm 1.

**Lemma 16.** Let \( z_{m,1} \) denote the optimal objective value of the unconstrained \( (m,1) \) problem, and let \( \hat{z}_{m,1} = z_{m,1} \). For notational convenience, let \( \theta_m = \tau^*_m,1 \) denote the corresponding optimal first dispatch time. Then,

\[
\hat{z}_{m,1} \sqrt{\frac{1 - \theta_m}{1 - 2\theta_m} - \frac{1 - \theta_m}{\theta_m}} \leq 0 \tag{A7}
\]

for all \( m \).

**Proof.** By the proof of Theorem 4 we have that

\[
\hat{z}_{m,1} = \frac{1 - \theta_m}{3\sqrt{\theta_m}} \tag{A8}
\]

By the results of the \( (1,1) \) model and Theorem 4 we know that \( \theta_m \leq \frac{1}{3} \). Additionally, observe that for all \( \theta_m \in (0, \frac{1}{3}] \),

\[
\frac{1 - \theta_m}{3\sqrt{\theta_m}} \leq \sqrt{\frac{1 - 2\theta_m}{\theta_m}}. \tag{A9}
\]

Hence,

\[
\hat{z}_{m,1} \leq \sqrt{\frac{1 - 2\theta_m}{\theta_m}}, \tag{A10}
\]

which implies

\[
\hat{z}_{m,1} \sqrt{\frac{1}{1 - 2\theta_m}} \leq \sqrt{\frac{1}{\theta_m}}. \tag{A11}
\]
which in turn implies
\[ \hat{z}_{m,1} \sqrt{\frac{1 - \theta_m}{1 - 2\theta_m}} \leq \sqrt{\frac{1 - \theta_m}{\theta_m}}, \] (A12)
as desired.

Henceforth, let \( A_1^* \) denote the optimal first dispatch area in the unconstrained problem, and define \( \tau_1^* \) such that \( \tau_1^* + cA_1^* \sqrt{\tau_1^*} = 1 \). The next lemma proves the correctness of the algorithm when \( m = 2 \) and also serves as the base case for the inductive proof of the general result.

**Lemma 17.** Let \( m = 2 \) and \( B < A_1^* \). In an optimal solution to the \( B \)-constrained problem, it must hold that \( B = A_1^* \geq A_2 \).

**Proof.** Suppose we are given an optimal solution \((\tau_1, A_1), (\tau_2, A_2)\) such that \( B > A_1 \). We know that
\[ \tau_1 + cA_1 \sqrt{\tau_1} = 1 \] (A13)
and
\[ \tau_2 + cA_2 \sqrt{\tau_2} = 1 - \tau_1. \] (A14)
As a preliminary note, if we are to solve the unconstrained \((1,1)\) problem on a truncated service day of length \( 1 - \tau_1 \) by re-scaling the service day to have unit length, then we must use the following parameters instead of \( \hat{\lambda}, \hat{c}_0, \) and \( \hat{c} \):
\[ \hat{\lambda} = \lambda(1 - \tau_1), \]
\[ \hat{c}_0 = \frac{c_0}{1 - \tau_1}, \]
\[ \hat{c} = \hat{c}_0 \sqrt{\hat{\lambda}} = \frac{c}{\sqrt{1 - \tau_1}}. \]
First, suppose that \( \tau_1 > \frac{1}{2} \). Recall from the analysis of the unconstrained \((1,1)\) problem that the total quantity is maximized when the accumulation time is \( \tau = \frac{1}{2} \). Additionally, the derivative of the total quantity as a function of the accumulation time is negative for all \( \tau \in (\frac{1}{2}, 1) \) in that problem. Therefore, we can decrease \( \tau_1 \) by a sufficiently small amount (and increase \( A_1 \) by a corresponding amount such that (A13) still holds) to increase the quantity served by the first dispatch without decreasing the quantity served by the second dispatch. By contradiction, the solution \((\tau_1, A_1), (\tau_2, A_2)\) cannot be optimal, so it must hold that \( \tau_1 \leq \frac{1}{2} \) in the optimal solution to the constrained problem.

Now, let us consider the case when \( A_1 \leq A_2 \). This implies
\[ \frac{1 - \tau_1}{c \sqrt{\tau_1}} \leq \frac{2}{\hat{c} \sqrt{3}} \] (A15)
or, equivalently,
\[ \frac{1 - \tau_1}{c \sqrt{\tau_1}} \leq \frac{2 \sqrt{1 - \tau_1}}{c \sqrt{3}}. \] (A16)
Rearranging gives $\tau_1 \geq \frac{3}{7} > \frac{1}{3}$, a contradiction to our previous result. Thus, it must hold that $A_1 > A_2$.

Since $B > A_1 > A_2$, we can express the quantity served by each of the first and second dispatches in terms of $\tau_1$:

\[ q_1(\tau_1) = \frac{\lambda}{c} (1 - \tau_1) \sqrt{\tau_1}, \quad (A17) \]
\[ q_2(\tau_1) = \frac{\lambda}{c} \cdot \frac{2}{3\sqrt{3}} = \frac{2\lambda}{3c\sqrt{3}} (1 - \tau_1)^{3/2}. \quad (A18) \]

We then take the derivative of the total quantity with respect to $\tau_1$:

\[ q'(\tau_1) = \frac{\lambda}{c} \left( \frac{1}{2} \tau_1^{-1/2} - \frac{3}{2} \tau_1^{1/2} - \frac{1}{\sqrt{3}} (1 - \tau_1)^{1/2} \right). \quad (A19) \]

It can be verified that this expression is negative for all $\tau_1$ for which the corresponding $A_1 < A_1^*$ (i.e., for all $\tau_1 \in (\tau_1^*, \frac{1}{3})$). Thus, we can decrease $\tau_1$ by a sufficiently small quantity, increase $A_1$ accordingly, and re-optimize $q_2$ accordingly such that the total quantity served increases. This contradicts the optimality of the given solution. Therefore, any solution with $A_1 < B$ cannot be optimal. \hfill \Box

The following two results prove the correctness of the algorithm when $m \geq 3$.

**Lemma 18.** Let $m \geq 3$ and $B < A_1^*$. Then, any solution with $A_1 < B$ and $A_1 \leq A_2$ cannot be optimal for the $B$-constrained problem.

**Proof.** We will prove this claim by contradiction. Suppose we are given a candidate optimal solution to the constrained $<m,1>$ problem $(\langle \tau_1, A_1 \rangle, \langle \tau_2, A_2 \rangle, \ldots, \langle \tau_m, A_m \rangle)$ with $m \geq 3$, $B > A_1$, and $A_2 \geq A_1$. By our previous discussion, we know that $\tau_1 \in (\tau_1^*, \frac{1}{3}]$.

Let $t_2 = \tau_1 + \tau_2$. We will show that we can slightly simultaneously perturb $\tau_1$ and $\tau_2$ (while leaving their sum $t_2$ unchanged) such that the total quantity served increases. Specifically, we wish to perform the following operations: decrease $\tau_1$ by some small $\varepsilon > 0$, increase $\tau_2$ by the same $\varepsilon$, increase $A_1$ by some $\delta_1$ such that (A13) is maintained, and decrease $A_2$ by some $\delta_2$ such that (A14) is maintained.

It remains to be seen whether the rate of increase of $q_2$ outpaces the rate of decrease of $q_1$ when we perform the above operations. With a slight abuse of notation, let us represent $q_1$ and $q_2$ as functions of $\tau_1$ under the assumption that $t_2$ is fixed:

\[ q_1(\tau_1) = \lambda A_1 \tau_1 = \lambda \left( \frac{1 - \tau_1}{c \sqrt{\tau_1}} \right) \tau_1 = \frac{\lambda}{c} (1 - \tau_1) \sqrt{\tau_1}, \quad (A20) \]
\[ q_2(\tau_1) = \lambda A_2 \tau_2 = \lambda A_2 (t_2 - \tau_1) = \lambda \left( \frac{1 - t_2}{c \sqrt{t_2 - \tau_1}} \right) (t_2 - \tau_1) = \frac{\lambda}{c} (1 - t_2) \sqrt{t_2 - \tau_1}. \quad (A21) \]

Omitting the constant factors $\frac{\lambda}{c}$, the derivatives of both quantities with respect to $\tau_1$ are

\[ q'_1(\tau_1) = \frac{1}{2} \tau_1^{-1/2} - \frac{3}{2} \tau_1^{1/2}, \quad (A22) \]
\[ q'_2(\tau_1) = \frac{t_2 - 1}{2\sqrt{t_2 - \tau_1}} \]  

(A23)

To show that the perturbation procedure increases the total quantity served by the first and second dispatches, we must prove that the sum of these derivatives evaluated at \( \tau_1 \) is negative, i.e., that

\[ h(\tau_1) = \frac{3}{2} \tau_1^{1/2} - \frac{1}{2} \tau_1^{-1/2} + \frac{1-t_2}{2\sqrt{t_2 - \tau_1}} \geq 0. \]  

(A24)

Since \( A_2 \geq A_1 \), by (A13) and (A14), it must hold that \( \tau_2 \leq \tau_1 \). This implies \( t_2 \leq 2\tau_1 \), which further implies

\[ \frac{1-t_2}{2\sqrt{t_2 - \tau_1}} \geq \frac{1-2\tau_1}{2\sqrt{2\tau_1 - \tau_1}}. \]  

(A25)

Therefore,

\[ h(\tau_1) \geq \frac{3}{2} \tau_1^{1/2} - \frac{1}{2} \tau_1^{-1/2} + \frac{1-2\tau_1}{2\sqrt{2\tau_1 - \tau_1}}. \]

\[ \geq \frac{3}{2} \tau_1^{1/2} - \frac{1}{2} \tau_1^{-1/2} + \frac{1-2\tau_1}{2\sqrt{2\tau_1 - \tau_1}}. \]

\[ = \frac{3}{2} \tau_1^{1/2} - \frac{1}{2} \tau_1^{-1/2} + \frac{1}{2} \tau_1^{-1/2} - \tau_1^{1/2} \]

\[ = \frac{1}{2} \tau_1^{1/2} \]

\[ > 0. \]

Thus, if \( A_2 \geq A_1 \), we can find sufficiently small \( \epsilon, \delta_1, \delta_2 > 0 \) such that \( (\tau_1, A_1 + \delta_1), (\tau_2, A_2 - \delta_2), \ldots, (\tau_m, A_m) \) is an improved feasible solution. Hence, by contradiction, if \( B < A_1^* \), then the optimal solution to the constrained problem must have either \( A_1 = B \) or \( A_1 > A_2 \).

Applying induction implies that the \( B \)-constrained optimal solution must have \( B \geq A_1 \geq A_2 \geq \cdots \geq A_m \).

**Lemma 19.** Let \( m \geq 3 \) and \( B < A_1^* \). Then, in an optimal solution to the constrained \( (m, 1) \) problem, \( A_1 = B \).

**Proof.** We proceed by induction with the result in Lemma [17] as the base case. Assume that the claim is true for \( m - 1 \) vehicles. For the purposes of contradiction, suppose we are given a candidate optimal solution to the constrained \( m \)-vehicle problem \( (\tau_1, A_1), (\tau_2, A_2), \ldots, (\tau_m, A_m) \) with \( m \geq 3 \) and \( B > A_1 \). Observe first that \( \tau_1 > \tau_1^* \). By Lemma [18] we may assume that \( A_1 > A_2 \).

Because the full solution is assumed optimal, the final \( m - 1 \) dispatches are also optimized over the truncated service day induced by \( \tau_1 \). By the induction hypothesis, if the final \( m - 1 \) vehicles were optimized with respect to the constrained problem but not the unconstrained problem, it would hold that \( A_2 = B \). However, because \( A_2 < B \), we know that the the final \( m - 1 \) vehicles must be optimized with respect to the unconstrained \( (m - 1) \)-vehicle problem as well. Therefore, by Theorem [3] \( \tau_2 = \theta_{m-1}(1 - \tau_1) \), where \( \theta_{m-1} \)
is the optimal first dispatch time in the unconstrained $\langle m + 1, 1 \rangle$ problem. Consequently, $A_1 > A_2$ implies

$$\frac{1 - \tau_1}{c\sqrt{\tau_1}} > \frac{1 - \tau_1 - \theta_{m-1}(1 - \tau_1)}{c\sqrt{\theta_{m-1}(1 - \tau_1)}}$$  \hspace{1cm} (A26)$$

which in turn implies

$$\tau_1 < \frac{1}{\frac{1}{\theta_{m-1}} - 1 + \frac{1}{\theta_{m-1}}} < \frac{\theta_{m-1}}{1 - \theta_{m-1}}.$$  \hspace{1cm} (A27)$$

Let $z_{m-1}$ represent the optimal total quantity served in the unconstrained $\langle m - 1, 1 \rangle$ problem, and let $\frac{\hat{c}}{c} z_{m,1} = z_{m,1}$. Theorem 3 implies that the total quantity as a function of $\tau_1$ is

$$q(\tau_1) = \frac{\lambda}{c}(1 - \tau_1)\sqrt{\tau_1} + \frac{\lambda}{c} z_{m-1,1}(1 - \tau_1)^{3/2}$$  \hspace{1cm} (A28)$$

when $A_1 > A_2$. Our goal is to show that $q'(\tau_1) < 0$ for all $\tau_1 \in (\tau^*_1, \frac{\theta_{m-1}}{1 - \theta_{m-1}})$ so that we can slightly reduce $\tau_1$ (equivalently, slightly increase $A_1$) and improve the total quantity served. As such, we henceforth ignore the scaling factor $\frac{\hat{c}}{c}$.

Differentiating gives

$$q'(\tau_1) = -\frac{3}{2} \tau_1^{-1/2} + \frac{1}{2} \tau_1^{-1/2} - \frac{3}{2} \hat{z}_{m-1,1}(1 - \tau_1)^{1/2}.$$  \hspace{1cm} (A29)$$

Note that, by definition, $q'(\tau_1^*) = 0$. Therefore, it suffices to show that $q''(\tau_1) < 0$ for all $\tau_1 \in (0, \frac{\theta_{m-1}}{1 - \theta_{m-1}}]$.

Differentiating twice gives

$$q''(\tau_1) = -\frac{3}{4} \tau_1^{-3/2} - \frac{1}{4} \tau_1^{-3/2} + \frac{3}{4} \hat{z}_{m-1,1}(1 - \tau_1)^{-1/2},$$  \hspace{1cm} (A30)$$

and differentiating thrice gives

$$q'''(\tau_1) = \frac{3}{8} \tau_1^{-3/2} + \frac{3}{8} \tau_1^{-5/2} + \frac{3}{8} \hat{z}_{m-1,1}(1 - \tau_1)^{-3/2} > 0.$$  \hspace{1cm} (A31)$$

Observe that $\lim_{\tau_1 \downarrow 0} q''(\tau_1) = -\infty$. Additionally,

$$q''\left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right) = -\frac{3}{4} \left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-1/2} - \frac{1}{4} \left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-3/2} + \frac{3}{4} \hat{z}_{m-1,1} \left(1 - \frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-1/2}$$

$$= -\frac{3}{4} \left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-1/2} - \frac{1}{4} \left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-3/2} + \frac{3}{4} \hat{z}_{m-1,1} \left(\frac{1 - 2\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-1/2}$$

$$= \frac{3}{4} \left(\hat{z}_{m-1,1} \sqrt{\frac{1 - \theta_{m-1}}{1 - 2\theta_{m-1}}} - \sqrt{\frac{1 - \theta_{m-1}}{\theta_{m-1}}} - \frac{1}{4} \left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-3/2}ight)$$

$$
\leq -\frac{1}{4} \left(\frac{\theta_{m-1}}{1 - \theta_{m-1}}\right)^{-3/2}$$
where the penultimate step is implied by Lemma 16. It follows that \( q''(\tau_1) < 0 \) for all \( \tau_1 \in \left( 0, \frac{\theta_{m-1}}{1-\theta_{m-1}} \right) \). Because \( q'(\tau^*_1) = 0 \), it follows that \( q'(\tau_1) < 0 \) for all \( \tau_1 \in \left( \tau^*_1, \frac{\theta_{m-1}}{1-\theta_{m-1}} \right) \). Thus, we can decrease \( \tau_1 \) by a sufficiently small quantity, increase \( A_1 \) accordingly, and re-optimize the remaining dispatches accordingly such that the total quantity served increases. This contradicts the optimality of the given solution. Therefore, any solution with \( A_1 < B \) cannot be optimal. \[ \square \]

Note that \( \theta_{m-1} < \frac{\theta_{m-1}}{1-\theta_{m-1}} \), so by the proof above we have that \( q''(\tau_1) < 0 \) for all \( \tau_1 \in \left( 0, \theta_{m-1} \right) \). This implies that the function in Problem 6 is concave over the interval, as noted earlier.

### A.9 Proof of Proposition 9

Given a fixed value for \( A > 0 \), let us assume for the sake of contradiction that there exists such optimal dispatch policy \( \{(\tau^*_d, A)\}_{d=1}^m \) such that at least one of the inequalities defined by constraint (7b) hold strictly. Fix such a policy and consider the first dispatch, \( d' \), such that, \( \sum_{\delta=1}^{d'} \tau_\delta + cA \sqrt{\tau_{d'}} < 1 \). There must exist some \( \epsilon \in (0, 1) \) such that \( \sum_{\delta=1}^{d'} \tau_\delta + \epsilon + cA \sqrt{\tau_{d'}} = 1 \). We can then feasibly replace the \( d' \)-th dispatch with \( (\tau^*_{d'} + \epsilon, A) \) by removing the next \( \epsilon \) orders from the subsequent dispatch(es). Any remaining dispatches will remain feasible since they are either: completely removed from the dispatch policy (i.e., their quantity is set to zero), set to depart at the same time of day with strictly less orders to serve, or set to depart at the same time of day with the exact same order amount to serve as before. This process of shifting orders to earlier dispatches can be repeated until (7b) holds at equality for each of the first \( m-1 \) dispatches. Eventually, it will be the case that (7b) holds strictly for \( d = m \). Then, this last dispatch could feasibly serve some \( \delta > 0 \) additional orders, which contradicts the assumed optimality of the given initial dispatch policy. Thus, constraints (7b) must hold at equality for all \( d \) in an optimal solution. \[ \square \]

### A.10 Proof of Proposition 10

Let \( \{(\tau^*_d, A^*_d)\}_{d=1}^m \) denote the optimal solution to the variable-area \( (m, 1) \) problem. Since \( A^*_1 > A^*_2 > \cdots > A^*_m \) by Proposition 6, \( \{(\tau^*_d, A^*_m)\}_{d=1}^m \) is a feasible solution to the fixed-area problem. The objective value associated with this feasible solution, \( \lambda A_m \sum_{d=1}^m \tau_d \), is therefore a lower bound on the optimal objective value associated with the fixed-area problem. Let \( \tau^*_{m-1} \) and \( A^*_{m-1} \) denote the service area and first dispatch time, respectively, associated with the optimal solution to the fixed-area problem. In any feasible solution to the fixed-area problem, all dispatches are associated with the same area (by definition) and all dispatches return at the same time \( T \) (by Proposition 9). Therefore, in the optimal solution to the fixed-area problem, a vehicle that departs later in the day serves a smaller quantity. This further implies that, in the optimal solution to the fixed-area problem, a vehicle that departs later in the day is associated with a shorter accumulation time. Hence, in the optimal solution, the total quantity served by each dispatch is decreasing over the course of the day.

Therefore,

\[
\lambda A^*_{m-1} \tau^*_{m-1} \geq \frac{\lambda}{m} \sum_{d=1}^m A_m \tau_d \\
\implies A^*_{m-1} \tau^*_{m-1} \geq \frac{1}{m} \sum_{d=1}^m A_m \tau_d.
\]
For notational purposes, let $b_m$ denote the right-hand side of the previous inequality. Since $A^*_1 = \frac{1 - \tau_1^*}{c\sqrt{\tau_1^*}}$, it follows that $\frac{1}{c}\sqrt{\tau_{e-1}^*}(1 - \tau_{e-1}^*) \geq b_m$. This implies that $\tau_{e-1}^* \geq \tau^*$, where $\tau^*$ is the solution to the optimization problem

$$\min \tau \quad \text{s.t.} \quad \frac{1}{c}\sqrt{\tau}(1 - \tau) \geq b_m.$$ 

By our work on the $\langle 1, 1 \rangle$ problem, we know that there exists exactly one $\tau \in [0, \frac{1}{3}]$ satisfying $\frac{1}{c}\sqrt{\tau}(1 - \tau) = b_m$; additionally, because $\frac{1}{c}\sqrt{\tau}(1 - \tau)$ is increasing in the interval $[0, \frac{1}{3}]$, this $\tau$ is necessarily the desired solution $\tau^*$. Since a lower bound on the first accumulation time corresponds to an upper bound on the service area, it follows that $A^*_1 = \frac{1 - \tau^*}{c\sqrt{\tau^*}}$. Note that the bound above remains valid if, for a different fixed-area model variant, the largest dispatch is not the first or if the vehicle is required to return to the depot at some time before the end of the service day. Hence, an analogous procedure can be used to derive an essentially identical upper bound on the optimal area for the fixed-area $\langle 1, D \rangle$ model discussed later. The equivalent equation to (8) in that case is

$$\frac{1}{c}\sqrt{\tau}(1 - \tau) = \frac{1}{m} \times \min_{d \in [D]} \{A^*_d\} \times \sum_{d=1}^D \tau^*_d,$$

where $\{(\tau^*_d, A^*_d)\}_{d=1}^D$ is the optimal solution to the variable-area $\langle 1, D \rangle$ problem with the same parameters.

A.11 Proof of Proposition 11

Given a set of positive accumulation times, $\{\tau_1, \tau_2, \ldots, \tau_D\}$, consider the $d$-th dispatch, where $d < D$. Inequality (9c) constrains the service area by $A_d \leq \frac{\tau_d^*}{c\sqrt{\tau_d^*}}$. As the objective value (9a) increases linearly with $A_d$, choosing the service area such that this inequality holds at equality will maximize the quantity served on the $d$-th dispatch for all $d < D$. Now, consider the last dispatch. Inequality (9b) constrains the service area by: $A_D \leq \frac{1 - \tau_D^*}{c\sqrt{\tau_D^*}}$. As the objective value (9a) scales linearly with $A_D$, we choose this service area such that this inequality holds at equality in order to serve the maximal number of orders served by the $D$-th dispatch, which completes the proof. 

A.12 Proof of Lemma 12

For a set of accumulation times to be a part of an optimal dispatching solution to the $\langle 1, D \rangle$ model (10), by first order conditions it must be true that both

$$\tau_1^* = \frac{(\tau_2^*)^2}{4\tau_D^*}$$

and

$$\tau_d^* = \frac{(\tau_{d+1}^*)^2}{4(\sqrt{\tau_D^*} - \sqrt{\tau_{d-1}^*})^2}$$

for all $d \in \{2, \ldots, D - 1\}$. By Proposition 11, these equations imply that $A_1^* = \frac{2}{c}\sqrt{\tau^*_D}$, and

$$A_d^* = \frac{2}{c}(\sqrt{\tau^*_D} - \sqrt{\tau^*_{d-1}}) \leq \frac{2}{c}\sqrt{\tau^*_D}$$
for all $d \in \{2, \ldots, D-1\}$.

### A.13 Proof of Theorem 13

By Lemma 12, we have that $A_d^* \leq \frac{2}{c} \sqrt{\tau_d^*}$ for all $d < D$. Thus, we can consider a relaxation of the original problem defined by (9), where each of the first $D - 1$ dispatches serves an area of $A_d = \frac{2}{c} \sqrt{\tau_d}$ without any regard for returning to the depot in time for the next dispatch. That is, consider the relaxation:

$$
\max_{A_D, \tau \geq 0} \quad \lambda A_D \tau_D + \sum_{d=1}^{D-1} \frac{2 \sqrt{\tau_d}}{c} \tau_d
$$

$$\text{s.t.} \quad \sum_{d=1}^{D} \tau_d + cA_D \sqrt{\tau_D} \leq 1. \quad (A32a)$$

In this relaxed system, it is a strictly dominant strategy for the final dispatch to serve an area large enough such that the vehicle will arrive back to the depot exactly at the end of the service day, implying that $A_D = \frac{1 - \tau_1 - \tau_D}{c \sqrt{\tau_D}}$. Without loss of optimality, as each of the first $D - 1$ dispatches serve the same area, and themselves have no explicit concerns of arriving back before a future dispatch, everything can be served on the first dispatch while the remaining $D - 2$ dispatches serve nothing. Thus, this relaxed problem can be re-formulated as:

$$
\max_{\tau_1, \tau_D \geq 0} \quad \lambda \left( 1 - \tau_1 - \tau_D \right) \tau_D + \lambda \left( \frac{2 \sqrt{\tau_D}}{c} \right) \tau_1
$$

$$\text{s.t.} \quad \tau_1 + \tau_D \leq 1. \quad (A33a)$$

Since this objective function is equivalent to $\frac{\lambda}{c} \left( 1 + \tau_1 - \tau_D \right) \sqrt{\tau_D}$, we see that $\tau_1 + \tau_D \leq 1$ will hold at equality for an optimal dispatching policy, so after a substitution of $\tau_1 = 1 - \tau_D$ we arrive at the problem:

$$
\max_{\tau_D \in [0,1]} \quad \frac{2\lambda}{c} \left( 1 - \tau_D \right) \sqrt{\tau_D}. \quad (A34a)
$$

This problem is identical to the optimization problem presented in Section 3 for the $\langle 1, 1 \rangle$ model with the caveat that the objective value is exactly twice is large. Therefore, this relaxation of the $\langle 1, D \rangle$ model has an optimal objective value of $\frac{\lambda}{c} \frac{1}{3 \sqrt{3}}$, as desired.

### A.14 Proof of Theorem 14

Assume $\lambda = c = 1$ without loss of generality. By an intermediate result within the proof of Theorem 4 (Section A.4), we know that $z_{m,D} \geq \frac{1}{4} \sqrt{m}$. Since $z_{1,1} = \frac{2}{3 \sqrt{3}}$, we have that

$$
\frac{3 \sqrt{3}}{8} \cdot \frac{z_{1,1}}{\sqrt{m}} \leq z_{m,1} \quad \Rightarrow \quad z_{1,1} \leq \frac{8}{3 \sqrt{3} m} \cdot z_{m,1}
$$

by rearrangement.

Now, consider the $\langle m, D \rangle$ problem. Suppose we relax the problem by allowing accumulation times of different vehicles to overlap (accumulation times associated with the same vehicle must still be non-overlapping). Note that this relaxation results in $m$ independent “copies” of the $\langle 1, D \rangle$ problem. Therefore,
we have that \( z_{m,D} \leq m \cdot z_{1,D} \). Applying the inequality above and Theorem \[13\] gives

\[
 z_{m,D} \leq m \cdot z_{1,D} \leq 2m \cdot z_{1,1} \leq \frac{16}{3} \sqrt{\frac{m}{3}} \cdot z_{m,1},
\]
as desired.

\[\square\]

**Appendix B  Empirical routing constant estimation**

For several combinations of \( n \) and service area \( A \) (where the corresponding region is given by a driving time isochrone), we simulate 100 TSP tours through the depot and \( n \) randomly generated locations in the region. For every tour with length \( \ell \) minutes (which includes the one minute per-customer service time), the corresponding ratio is calculated as \( \ell/\sqrt{An} \). Table 6 displays the average ratio, in units of

\[
\text{minutes} \quad \frac{\text{miles}}{\text{customers}^{1/2}}
\]

for combinations of \( n \) and \( A \) for the first and second computational studies (in which the service regions extend in all directions). Table 7 similarly displays the average ratio for the \langle 1,2 \rangle computational study (in which the service regions extend only in the southeast direction from the depot).

Observe that, in both tables, the ratios are decreasing in \( A \) for each fixed \( n \). This behavior is due to the

<table>
<thead>
<tr>
<th>Area (square miles)</th>
<th>50</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 15 )</td>
<td>3.9843</td>
<td>3.6079</td>
<td>3.3794</td>
<td>3.2168</td>
<td>3.1422</td>
</tr>
<tr>
<td>20</td>
<td>4.1277</td>
<td>3.7227</td>
<td>3.4420</td>
<td>3.3359</td>
<td>3.2638</td>
</tr>
<tr>
<td>30</td>
<td>4.3613</td>
<td>3.8719</td>
<td>3.6152</td>
<td>3.4748</td>
<td>3.3804</td>
</tr>
<tr>
<td>35</td>
<td>4.4324</td>
<td>3.9334</td>
<td>3.6820</td>
<td>3.5255</td>
<td>3.4646</td>
</tr>
<tr>
<td>40</td>
<td>4.5305</td>
<td>4.0033</td>
<td>3.7583</td>
<td>3.5878</td>
<td>3.5192</td>
</tr>
<tr>
<td>45</td>
<td>4.6508</td>
<td>4.0865</td>
<td>3.7945</td>
<td>3.6859</td>
<td>3.5702</td>
</tr>
<tr>
<td>50</td>
<td>4.7073</td>
<td>4.1285</td>
<td>3.8710</td>
<td>3.7285</td>
<td>3.6164</td>
</tr>
<tr>
<td>55</td>
<td>4.7457</td>
<td>4.1722</td>
<td>3.9372</td>
<td>3.7906</td>
<td>3.6980</td>
</tr>
<tr>
<td>60</td>
<td>4.8385</td>
<td>4.2638</td>
<td>4.0008</td>
<td>3.8204</td>
<td>3.7591</td>
</tr>
<tr>
<td>65</td>
<td>4.9181</td>
<td>4.3541</td>
<td>4.0510</td>
<td>3.8970</td>
<td>3.8073</td>
</tr>
<tr>
<td>70</td>
<td>4.9878</td>
<td>4.3775</td>
<td>4.0946</td>
<td>3.9454</td>
<td>3.8265</td>
</tr>
<tr>
<td>75</td>
<td>5.0602</td>
<td>4.4432</td>
<td>4.1207</td>
<td>3.9606</td>
<td>3.8691</td>
</tr>
<tr>
<td>80</td>
<td>5.1412</td>
<td>4.4760</td>
<td>4.1852</td>
<td>4.0170</td>
<td>3.9145</td>
</tr>
</tbody>
</table>

**Table 6:** Empirically estimated routing constants for multi-vehicle computational studies
Table 7: Empirically estimated routing constants for single-vehicle computational studies

<table>
<thead>
<tr>
<th>Area (square miles)</th>
<th>40</th>
<th>80</th>
<th>120</th>
<th>160</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 15</td>
<td>4.3503</td>
<td>3.9787</td>
<td>3.7695</td>
<td>3.5506</td>
</tr>
<tr>
<td>25</td>
<td>4.5813</td>
<td>4.1344</td>
<td>3.9172</td>
<td>3.6664</td>
</tr>
<tr>
<td>35</td>
<td>4.7489</td>
<td>4.3157</td>
<td>3.9742</td>
<td>3.7747</td>
</tr>
<tr>
<td>45</td>
<td>4.9635</td>
<td>4.4301</td>
<td>4.1197</td>
<td>3.8773</td>
</tr>
<tr>
<td>55</td>
<td>5.1088</td>
<td>4.5194</td>
<td>4.2320</td>
<td>3.9693</td>
</tr>
<tr>
<td>65</td>
<td>5.2501</td>
<td>4.6383</td>
<td>4.3370</td>
<td>4.0813</td>
</tr>
<tr>
<td>75</td>
<td>5.3761</td>
<td>4.8075</td>
<td>4.4258</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>5.5541</td>
<td>4.8831</td>
<td>4.5097</td>
<td></td>
</tr>
</tbody>
</table>

average inter-node distance increasing as \( A \) increases, which subsequently increases the likelihood of travel on faster major roads and highways (which are relatively sparse compared to residential streets) between customers. On the other hand, the ratios are increasing in \( n \) for each fixed \( A \) for a similar reason. As the density of points in a given area increases, it becomes more difficult to leverage faster roads when traveling between customers. This effect is slightly exacerbated by the linearly accumulating per-order service times. For comparison, if point-to-point travel times were given by the Euclidean or Manhattan metric and per-customer service times were ignored, we would expect the average ratios to be asymptotically decreasing in \( n \) and invariant to \( A \).

To estimate the corresponding ratio for any given \( A \) and \( n \), we define a function \( \beta(A, n) \) by linearly interpolating between the respective tabular values for each corresponding computational study. Similarly, the value of \( \beta(\cdot, \cdot) \) is linearly extrapolated for arguments \( A, n \) which may fall just outside the domain given in the corresponding table. Scaling the value \( \beta(A, n) \) appropriately produces the routing constant \( c_0 \) used in the planning models. Suppose that we solve a two-dispatch planning model with an estimated routing constant. Our tabular results indicate that each dispatch’s calculated values of \( A \) and \( n \) will likely subsequently correspond to two different routing constants \( \beta(A_1, n_1) \) and \( \beta(A_2, n_2) \) which may be significantly different from our initial estimate. This requires us to iteratively re-estimate the routing constant and re-solve the planning model until the routing constant converges to a final value. The following pseudocode formalizes this procedure. An analogous method is used when the model requires three total dispatches.

In our studies, convergence occurs rapidly (within ten iterations for a tolerance of \( \varepsilon = 10^{-5} \)) given a reasonable starting estimate. We use the following resulting routing constants, scaled as appropriate, to determine service regions in the computational studies where the regions extend in all directions: 4.1176 for the \( \langle 2, 1 \rangle \) case with variable areas, 4.0627 for the \( \langle 2, 1 \rangle \) case with constant areas, 4.0630 for the unconstrained \( \langle 3, 1 \rangle \) case, and 4.0657 for the constrained \( \langle 3, 1 \rangle \) case.

The above constant estimation procedure is not the only valid method; rather, it easily admits modifications. To illustrate, our computed examples for the \( \langle 1, 2 \rangle \) setting use routing constants calculated with
Algorithm 2 Convergence of BHH routing constant

1: given interpolant function $\beta(\cdot, \cdot)$, initial estimate $\beta_{\text{init}}$, tolerance $\epsilon$
2: initialize $\beta_{\text{old}} \leftarrow \beta_{\text{init}}$, $\beta_{\text{new}} \leftarrow \infty$
3: while $|\beta_{\text{old}} - \beta_{\text{new}}| > \epsilon$ do
4: set $\beta_{\text{old}} \leftarrow \beta_{\text{new}}$
5: use $\beta_{\text{old}}$ to calculate optimal policy areas and quantities $A_1, A_2, n_1, n_2$
6: set $\beta_{\text{new}} \leftarrow \max\{\beta(n_1, A_1), \beta(n_2, A_2)\}$
7: end while
8: return $\beta_{\text{new}}$

a slightly different procedure that updates the constant via a weighted average (instead of by selecting the most conservative options). Specifically, line 6 of Algorithm 2 is replaced with

$$\text{set } \omega \leftarrow \frac{n_1}{n_1 + n_2}$$
$$\text{set } \beta_{\text{new}} \leftarrow \beta(\omega n_1 + (1 - \omega) n_2, \omega A_1 + (1 - \omega) A_2).$$

Using this method and the estimated values from Table 7, the resulting routing constants are 4.3684 for the variable-area $\langle 1, 2 \rangle$ setting and 4.3477 for the fixed-area $\langle 1, 2 \rangle$ setting.