Approximations for Planar Covering Routes: an Analysis and Application to Public School Transportation

Dipayan Banerjee
Loyola University Chicago

Karen Smilowitz
Northwestern University

Abstract. Public school bus routes can change from year to year as students and their home locations change. However, school administrators benefit from the ability to predict future transportation needs on multi-year time scales. With this motivation in mind, this paper develops planning models for school bus routing when student locations are not known with certainty. Continuous approximation methods have been shown to be valuable in other settings where future spatial realizations of transportation demand are unknown. We first show how applying standard continuous approximation techniques to school bus routing entails unique methodological challenges not found in other routing problems. Specifically, we argue that the typical square-root functional form derived from the classical Beardwood-Halton-Hammersley Theorem does not necessarily produce appropriate route length approximations for a broad class of planar covering route problems, in which vehicles visit facilities (e.g., school bus stops, parcel lockers) that collectively 'cover' demand points (e.g., students' homes, delivery recipients) within a certain radius. We provide both empirical evidence and analytical results in support of this argument. Next, we propose a continuous approximation framework for robust estimation of a school's minimum required bus fleet size when student locations are distributed independently and identically at random. We focus on the non-trivial case in which the maximum allowable route duration (and not the physical bus capacity) is the limiting factor. Prior experience with school administrators suggests that the ability to quickly evaluate 'what-if' scenarios in real time is critical during the planning process; therefore, our approach requires minimal computation after one-time setup. Finally, we conduct computational experiments to validate our approach for a stylized planar region and actual school attendance regions in the United States.

Keywords. school bus routing, continuous approximation, public sector operations, covering routes

---

1. Introduction

Widespread funding challenges and school bus driver shortages have impacted the transportation capabilities of public school districts in the United States. These issues have necessitated policy changes to reduce the number of buses and drivers used. For example, Baltimore County Public Schools no longer offers transportation to high school students located less 1.7 miles from their school, an increase from the previous radius of 1.5 miles (Tooten 2023). Other strategies considered by policymakers include staggering school bell times to facilitate reusing buses on multiple routes (Lambeck 2025) or eliminating busing for certain grade levels entirely (Morris 2024). Unfortunately, due to a combination of declining public school enrollment and policies regulating the maximum duration of a school bus route, existing buses are often not filled

to their physical capacity.

Although public school policy decisions and transportation contracts generally have a multi-year scope, annual student turnover induced by natural grade progression and demographic shifts makes long-term public school transportation planning inherently difficult. For example, a school's attendance region boundaries will likely remain unchanged for at least five years; however, the residential locations of students attending the school in five years will differ from the locations of the students attending the school when the boundaries were drawn. Therefore, policymakers would benefit from the ability to estimate a school's expected future transportation needs — specifically, its required number of bus routes (i.e., its necessary *fleet size*) — without knowing future students' exact locations. It would be especially useful to understand how operational parameters, such as bus capacity and the maximum allowable route duration, impact these fleet size calculations. However, computationally intensive optimization methods are often not viable for this purpose: prior experience with public school administrators suggests that rapid real-time evaluation of 'what-if' scenarios is critical during the planning process, and community stakeholders may disapprove of solutions produced by complex algorithms despite their promised improvements (e.g., Scharfenberg 2018).

The need for these fleet size calculations is motivated by decision support for medium- and long-term planning and policymaking. At the tactical level, school district administrators require estimates of each school's transportation resource needs to plan future bus leases and driver contracts. At the strategic level, administrators may seek to estimate the transportation-related impacts of significant changes to district schedules, policies, and attendance boundaries. For example, as in Bertsimas et al. (2019), some districts stagger school start times into tiers so that each bus may be used for multiple schools' routes. If so, the temporal gap between each tier effectively limits the maximum duration of each bus route. Understanding how the length of this gap impacts each school's required bus fleet size would allow policymakers to gauge the transportation-related impacts of modifying school bell schedules. Note that, for such applications, it is not necessary to know the exact routes themselves.

Continuous approximation (CA) methodologies are often applied to similar planning problems in transportation and logistics when future demand locations are not known with certainty. CA is particularly effective when interpretability, transparency, and computational efficiency are of high priority. At a high level, CA methods replace stochastic discrete realizations (i.e., customer locations or optimal vehicle routes) with deterministic continuous planning approximations (density distributions or expected route lengths) to facilitate understanding of 'average-case' system behavior. Such routing approximations generally rely on

the classical Beardwood-Halton-Hammersley (BHH) Theorem, which implies that the expected duration of a traveling salesman problem (TSP) tour over n random points independently and identically distributed (IID) in a planar region grows proportionally to $\sqrt{n}$ (Beardwood et al. 1959). Practically, this square-root functional form allows for straightforward closed-form approximations of TSP and vehicle routing problem (VRP) route lengths after empirical estimation of an appropriate proportionality constant.

It may therefore seem natural to directly apply existing BHH-based CA methods to the aforementioned school transportation planning problem, especially because of the constraint on bus route durations. However, the structure of a school bus route is fundamentally different from that of a TSP. In a typical school bus route, students' residences are not visited directly. Rather, buses visit designated bus stops; these stops are located such that each student must walk no more than a pre-specified maximum distance to their assigned stop. The chosen bus stops — of which there are significantly fewer than the number of students — collectively cover all of the student locations. A school bus route is an example of a *covering route*, in which a vehicle visits a set of facilities (i.e., bus stops) such that at least one is within a given radius of each demand location (i.e., student residence); we use the term broadly to encompass both covering tours (Gendreau et al. 1997) and covering paths (Zeng et al. 2019a,b). Other recent applications of covering routes include diagnostic reconnaissance (Di Placido et al. 2022), collaborative truck-drone delivery (Zhou et al. 2023), and urban last-mile delivery with parking considerations (Le Colleter et al. 2023).

Our contributions in this work are twofold. First, we provide empirical evidence and derive formal mathematical results to argue that BHH-based CA methods are not necessarily appropriate for planar covering routes of realistic size. Specifically, we present data suggesting that covering routes may be significantly longer than analogous routes over facilities located IID at random. To understand this discrepancy, we then prove that covering facility locations induced by various multiobjective optimization criteria violate the necessary IID assumption of the BHH Theorem.

Second, we develop a CA framework for school bus fleet sizing in planar regions. As is common in the CA literature, we assume for planning purposes that a school's attendance region will be partitioned into zones, each serviced by a distinct bus route. To minimize the number of bus routes (equivalently, zones) required, we model and solve the optimization problem of maximizing the area of an arbitrary zone subject to constraints on physical bus capacity and route duration. We then leverage the solution of this problem to derive a fleet size estimate that depends on student density, bus speed, and several other features. Based on our previous empirical and mathematical results, our proposed approach makes no assumption on the

functional form of covering route lengths, instead relying on linear interpolation between empirical point estimates together with numerical integration. We validate our CA approach via computational experimentation, and we demonstrate how our approach can be used to support real-world public school transportation planning via case studies involving actual schools and street networks in the United States.

The remainder of this paper proceeds as follows. This section concludes with a brief review of the relevant literature. Section 2 formally describes the structure of school bus routes and covering routes. Section 3 conducts empirical and mathematical analyses of planar covering route lengths. Section 4 develops the CA school bus fleet sizing model. Section 5 conducts computational experiments on planar regions and real-world road networks. Section 6 provides concluding remarks and discusses avenues for further study. The appendices contain proofs, technical details, and data omitted from the main body.

1.1 Related Work

We provide an overview of prior work on closely related topics here. Additional relevant references are discussed in their appropriate contexts later in the paper.

1.1.1 Planning Approximations of Vehicle Route Lengths

Various tactical and strategic planning problems in transportation and logistics require approximations of vehicle route lengths when future demand locations are not known with certainty. The most common of these approximations involves the aforementioned square-root functional form based on the BHH Theorem. More specifically, the expected TSP route length is typically modeled as a continuous function of the form $\beta\sqrt{An}$, where n represents the number of IID demand locations and A represents the area of the distribution's bounded support in the plane. The routing constant β, which depends on the underlying probability distribution and distance metric, is usually taken from a previously published estimate (e.g., Applegate et al. 2011) or re-estimated via Monte Carlo simulation by solving many randomly generated TSP instances and finding the square-root curve of best fit. Constant and linear terms may be included alongside the square-root term to capture setup, linehaul travel, and/or per-stop service times, as in Stroh et al. (2022).

While relatively less common, similar planning approximations exist for other routing problems. Functional forms of varying complexity have been proposed for VRP route lengths by Daganzo (1984a), Figliozzi (2008), and others using heuristic constructions and empirical analysis. Recent years have seen the derivation of asymptotic results analogous to the BHH Theorem for other problems, such as the traveling repairman problem (Blanchard et al. 2024) and a minimum-cost covering path problem (Zeng et al. 2019a,b). While

the BHH Theorem is directly useful for approximating real-world TSP lengths is convenient, it is unclear whether these other problems' asymptotic results can be used to produce accurate planning approximations of comparable accuracy for instances of realistic size. In this work, we argue that (i) planar covering routes are structurally different from planar TSPs and thus require different planning approximations, and (ii) adherence to any particular functional form is not necessarily required to develop a practically useful CA model for school transportation policymaking. We defer to Franceschetti et al. (2017) and Ansari et al. (2018) for comprehensive reviews of CA methods and applications.

1.1.2 School Bus and Covering Route Optimization

In the operations research literature, school bus routing problems (SBRPs) fundamentally involve optimizing the locations of bus stops, optimally assigning students to bus stops that are sufficiently close to their residences, and optimally routing a fleet of one or more buses to pick up students from these bus stops and transport them to school (for morning operations; afternoon routes operate in the opposite manner). Richer models include other decisions such as adjusting school bell times (Vercraene et al. 2023) and scheduling the start or end time of each route (Shafahi et al. 2018). Park and Kim (2010), Ellegood et al. (2020), and Wang et al. (2025a) provide detailed reviews of the literature on school bus routing optimization, illustrating both the variety of objectives considered in the literature (discussed further in Section 2) and the variety of heuristic procedures developed for solving these problems. In addition, the authors' personal experience suggests that many school districts in the United States use route planning software licensed from private companies such as Tyler Technologies and HopSkipDrive. Because of their proprietary nature, it is unclear what types of algorithms are implemented in such software. Manual adjustments by transportation planners based on heuristics or intuition are also not uncommon.

To our knowledge, Ellegood et al. (2015) is the only previous work to leverage CA methods for public school transportation planning. The authors conduct an analysis of mixed-load busing strategies (where a single route may serve students attending different schools) versus more common non-mixed-load busing strategies (where each route serves students who all attend the same school) in a planar setting, finding that the former strategy may be beneficial in some cases. Their analysis uses a closed-form route length approximation that includes a term proportional to the square root of the number of bus stops on the route; their analytical expressions are supported by routing constants previously estimated by Daganzo (1984a,b) and Campbell (1993) in the context of standard TSP and VRP tours. The analysis assumes a particular

density of bus stops per square mile, but it does not detail how (or whether) these stops are optimized. Taken together, these features seemingly suggest that stop locations and routes behave in line with the BHH Theorem's underlying assumptions. We present an alternate perspective in this work wherein the number and locations of bus stops are also optimized, resulting in fundamentally different planning approximations.

Because students are picked up from nearby bus stops instead of their residences, school bus routing optimization is part of a class of routing problems in which vehicles visit only a subset of points that, in turn, cover other nearby points. Comprehensive reviews of such problems are presented by Glock and Meyer (2023), Moradi et al. (2024), and Dursunoglu et al. (2025); none of the relevant papers referenced in these three reviews develop or leverage CA methods. With respect to the classification scheme proposed by Glock and Meyer (2023), the planar school bus routes studied in this work have the following characteristics: student locations are passive nodes, potential bus stop locations are active nodes that provide complete coverage, and the underlying topological space is modeled as continuous. We later consider discrete underlying topological spaces in Sections 5.2–5.4 when we apply our planar models to real-world street networks.

2. Motivation: School Transportation Planning

Consider a single neighborhood that is assigned to a nearby school; the school and neighborhood are stylized as a point and compact planar subregion in $\mathbb{R}^2$, respectively (e.g., Figure 1). There are a total of n students in this subregion requiring transportation to the school, all of whom are assigned to the same dedicated bus route. However, the exact locations of these students are unknown in advance and presumed to be IID uniformly at random for planning purposes. To facilitate long-term decision-making, school transportation planners seek an approximate answer to the following fundamental question: what is the expected length of this bus route, measured as the total distance traveled between the first visited bus stop and the school?

Answering this question requires understanding the basic operational structure of a school bus route. After student locations are known, certain locations in the subregion are designated as bus stops; because student locations change as students progress grade levels and move residences between one school year to the next, bus stop locations may also change each year. Each student eligible for transportation is assigned to a bus stop located no more than c miles (henceforth referred to as the *coverage radius*) from their home. In practice, the

Figure 1: Example school and planar subregion (neighborhood)

coverage radius may reflect a codified legal requirement, such as the half-mile distance specified in the Ohio Administrative Code (2020), or simply a planning guideline used internally by transportation planners. The school bus visits each bus stop to pick up all of the students, then completes the route by traveling directly to the school. Figure 2 depicts such a bus route assuming that travel is modeled by the rectilinear (i.e., Manhattan) metric. Student locations in the subregion are marked in light blue, while bus stops and their corresponding coverage radii are marked in red.

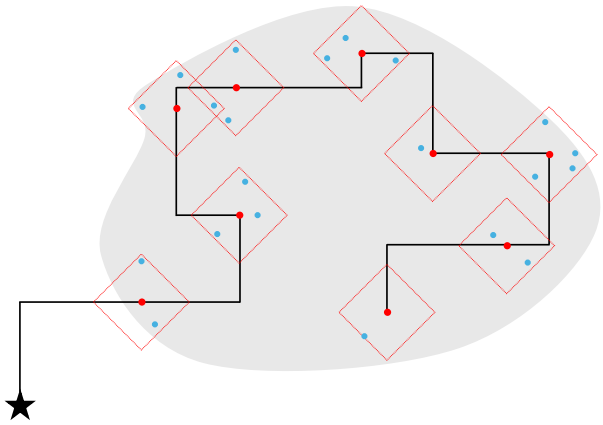


Figure 2: Example school bus route

Constructing the 'best' school bus route for a given set of students is a challenging and often ill-defined multiobjective optimization problem. Common decision criteria include (i) minimizing the number of bus stops, as additional bus stops incur marginal costs and increase operational complexity; (ii) minimizing travel distance or route duration, to conserve fuel and facilitate reusing buses for other schools; and (iii) minimizing student-to-stop walking distance. The relative prioritization or weighting of these competing objectives, among others, differs on a case-by-case basis. Regardless of how the bus stops and their sequence are chosen, a school bus route is an example of a planar *covering route*: a vehicle route in which the demand locations (e.g., the students' homes) are not themselves visited by the vehicle directly, but a smaller set of covering facilities (e.g., the bus stops) are visited instead. School bus routes can be specifically categorized as *open* covering routes because bus routes' starting and ending locations differ, whereas *closed* covering routes appear in contexts where vehicles begin and end their routes at a dedicated depot.

Generalizing away from the specific context of public school transportation, how might a system planner model the expected length of a covering route in a planar region of area A? Given an estimate m for the number of covering facilities to visit, a natural approach would be to leverage the classical square-root functional form of the BHH Theorem. One could then use an existing empirical estimate for the routing constant β (e.g., Applegate et al. 2011, Stroh et al. 2022, Vinel and Silva 2018), allowing for easy and computationally inexpensive modeling of the expected route length as $\beta\sqrt{m}$. Unfortunately, as we show in the following section, this straightforward approach may lead to noticeable prediction errors because of covering routes' unique structural properties.

3. Covering Route Length Estimation

In this section, we argue that CA functional forms, routing constants, and estimation procedures developed for non-covering routing problems — that is, standard TSPs — are not entirely appropriate for routing problems that involve planar covering of demand locations by facilities.

Our argument is comprised of two related claims. We claim that the spatial distribution of covering facilities in such routing problems is not IID uniform. This may be generally accepted knowledge in the facility location literature (e.g., Drezner et al. 2024); however, we specifically demonstrate that this behavior is significant enough to impact CA predictions. Furthermore, we claim that rigorous empirical estimation of covering route lengths cannot simply rely on random IID sampling of facility locations from any distribution — uniform or not. Rather, proper estimation requires randomly generating demand locations, then determining the resulting facility locations via optimization.

We provide both empirical and analytical support for our argument. First, we generate and optimize covering facility locations for a large number of random instances. We compare the average route lengths over these covering facilities to the analogous average route lengths over facility locations IID uniformly at random. We find that route lengths over covering facility locations are 10–20% longer on average. Intuitively, this phenomenon occurs because covering facilities tend to 'spread out' to ensure that all demand is covered, whereas IID uniform facility locations do not necessarily represent a feasible solution to the coverage problem. Second, we use this idea to formally prove that covering facility locations are not IID (with respect to the uniform or any non-uniform distribution) for a variety of optimization criteria, which violates a core underlying assumption of standard CA methods.

This section complements Carlsson and Jia (2015) and Zeng et al. (2019a,b), who devise asymptotically near-optimal configurations for certain types of planar covering routes. In practice, many covering routes are in decidedly non-asymptotic regimes; for example, a viable school bus route likely cannot visit more than 10–20 bus stops. Our work supplements the aforementioned papers by providing practical insights for covering routes which visit as few as four facilities. As such, this section is also closely related to the work of Vinel and Silva (2018) and Choi and Schonfeld (2022) on TSP length estimation for small instances.

3.1 Empirical Analysis

As noted in Section 1.1.2, there exist many approaches to modeling and solving SBRPs and covering route

problems. In this empirical analysis, for illustrative purposes, we consider one particular covering route optimization problem that lexicographically prioritizes the number of facilities required, the distance between demand locations and facilities, and the route length. As such, the problem is decomposed into sequential facility location and routing subproblems. The formal mathematical analysis in Section 3.2 extends our empirical insights to a broader array of covering path approaches, including the simultaneous optimization of facility locations and routes as in Zeng et al. (2019a,b).

Consider a 3×3 mile square region whose southwest corner is located at the origin. A total of n demand locations are distributed in this region independently and uniformly at random. Upon realization of the demand locations, covering facilities are located as follows. First, a minimum set cover problem is solved to determine the fewest number of facilities k required to ensure that every demand location is within 0.5 miles of at least one facility. Second, a constrained k-median problem is solved to minimize the average distance from each demand location to its closest facility (while still adhering to the half-mile coverage constraint). The set of candidate facility locations is designated as the 0.05-mile lattice to approximate the computationally intractable problem of locating these facilities in continuous space. This procedure is repeated 1000 times for each $n \in \{30, 50, 70\}$. Across these 1000 instances, an average of 10.19, 12.35, and 13.74 facilities are required for $n = 30$, 50, and 70, respectively.

For each instance, we optimize two routes: an open route that visits every covering facility and ends at the origin, and a closed route that visits every covering facility while beginning and ending at the origin. Figure 3 plots the average open route length for each n as a function of the number of facilities visited, and Figure 4 plots the corresponding closed route lengths. For comparison, for each $k \in \{7, 10, \ldots, 16\}$, we additionally solve 1000 open and closed TSP instances over the origin and k facility locations IID uniformly at random. Figures 3 and 4 also plot the average route lengths for these instances. Tables B1 and B2 in Appendix B contain detailed numerical results.

It is clear that, for these realistic values of n, routes visiting facilities that cover IID uniform demand locations are significantly longer than routes visiting IID uniform facilities. For example, open routes that visit 12 facilities located uniformly at random are over 16% shorter on average than open routes that visit 12 optimally located covering facilities! This gap seemingly increases in the number of facilities. Regardless, this data demonstrates that using standard TSP planning approximations for covering routes may lead to noticeable prediction errors and invalid optimization models.

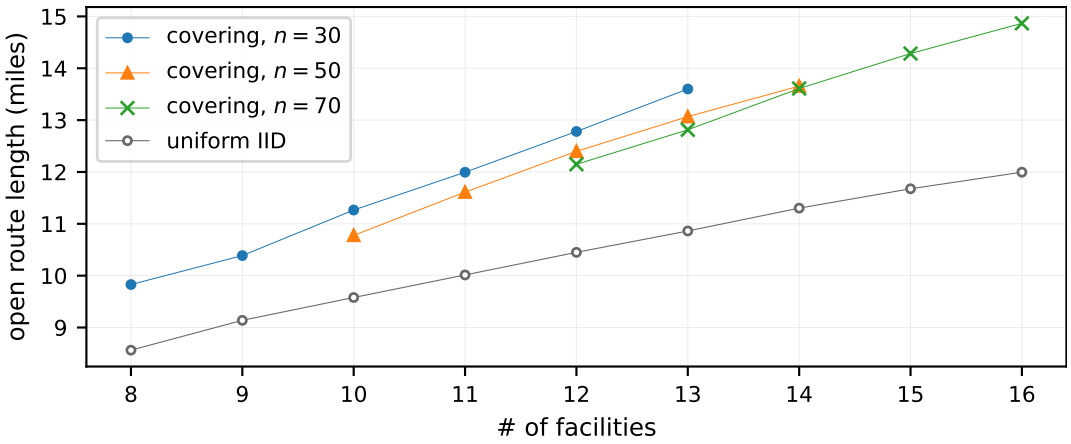


Figure 3: Comparison of empirically estimated open route lengths

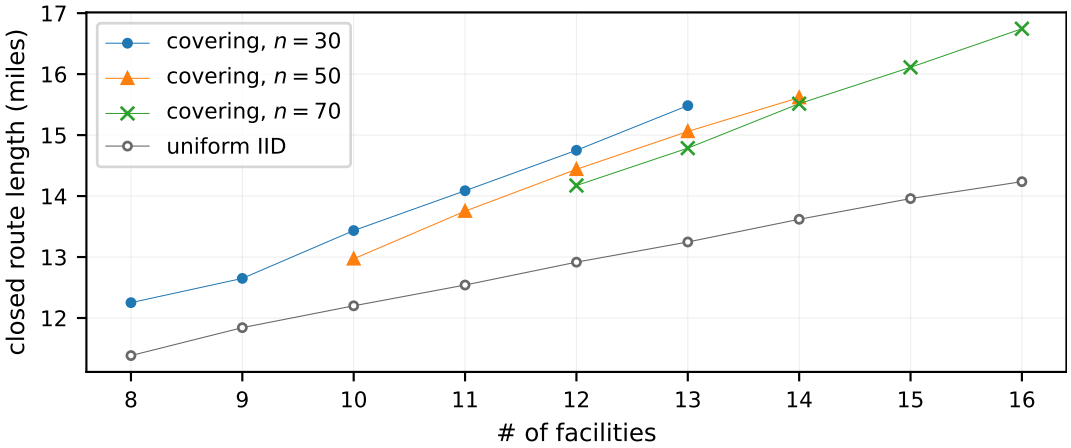
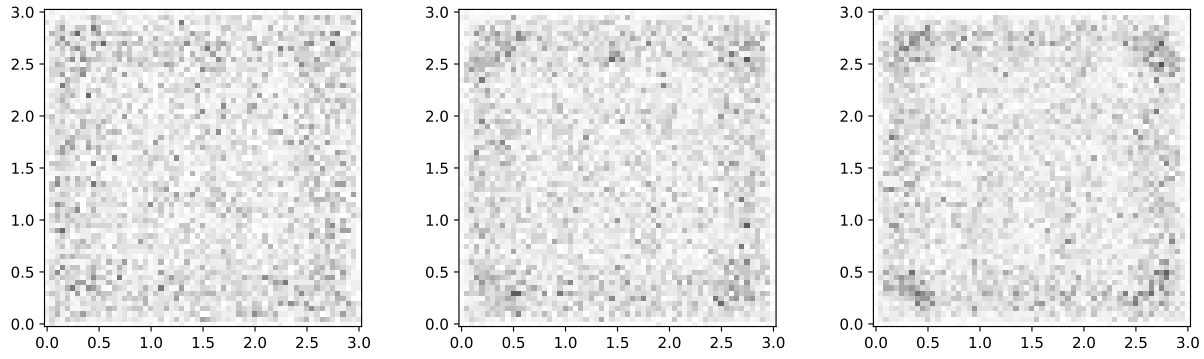


Figure 4: Comparison of empirically estimated closed route lengths

This behavior can be partially explained by the intuitive notion that facilities located optimally with respect to IID uniform demand points are themselves not uniformly distributed. Figure 5 displays relative frequency heatmaps of optimal facility locations across all trials. A non-uniform pattern emerges for $n = 30$ and becomes more evident as n increases.

Then, for a single realization of $n = 70$ demand points, is it the case that the optimal facility locations will be independently and identically distributed according to a spatial distribution similar to that visualized in Figure 5c? If so, it would significantly ease the process of estimating a routing approximation via Monte Carlo simulation. For an appropriate number of trials, one could simply sample a desired number of facilities IID from this distribution, then solve the corresponding routing problems to produce an empirical estimate. Such an approach is convenient because it removes the additional computational burden of randomly generating demand locations and then optimizing facility locations for each trial.



(a) $n = 30$ **(b)** $n = 50$ **(c)** $n = 70$

Figure 5: Heatmaps of facility locations (axes in miles)

This is unfortunately not the case. For example, if we randomly select 15 facility locations IID from the empirical distribution in Figure 5c, the average observed closed route length (over 1000 trials) is 13.87 miles. This is nearly identical to the corresponding average route length of 13.96 miles when facilities are sampled IID uniform, and therefore a considerable underestimate of the corresponding route length over optimized facilities of 16.11 miles (see Table B1). Similar results are obtained for the other parameters considered in Figures 3 and 4. As such, it is seemingly not the non-uniformity of optimal covering facility locations that leads to increased route lengths, but rather the apparent non-independence of these locations.

3.2 Probabilistic Analysis

Regardless of the specific optimization criteria used, it is often suboptimal for multiple facilities to be in very close proximity. We use this intuitive 'anti-clustering' tendency to prove the results stated in this subsection, which concern the non-independence of optimal locations in three different facility location problems encountered in practical school transportation contexts and other covering route applications. Full proof details for Propositions 1–3 are included in Appendix A.

In practice, every additional stop on a school bus route increases operational risk and complexity. As such, a planner may seek routing solutions that minimize the number of designated bus stops, then choose among these solutions based on other criteria (e.g., walking distance and/or route length). For example, Glenview School District 34 in Illinois "makes every effort to minimize the number of bus stops by locating them central to where students live." We can represent this as the classical planar set cover problem. Formally, let U denote the unit square and 2^U its power set. Recall that c represents the coverage radius of a facility. Define the function $H^{\text{SC}}_{n,c} : U^n \to 2^U$ to take as input n demand locations in U and return

a unique optimal solution to the set covering problem, where 'ties' between multiple potential optimal solutions are broken based on some well-defined tiebreaking rule. That is, $H^{\text{SC}}_{n,c}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = S$ is the unique optimal solution to

$$\underset{S \in 2^U}{\text{lexmin}} \quad |S|,\ t\big(\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}, S\big) \qquad \text{s.t.} \quad \min_{\mathbf{y} \in S}\{d(\mathbf{x}_i, \mathbf{y})\} \leq c \quad \forall\, i \in [n], \tag{1}$$

where $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is the set of demand locations, d is the rectilinear distance metric, S represents the chosen set of facility locations, and $t(\cdot)$ encodes an arbitrary tiebreaking rule. Suppose the n demand locations are chosen IID uniformly at random from U. Proposition 1 states that the resulting facilities S are themselves not IID with respect to <u>any</u> continuous probability distribution on U.

Proposition 1. *Let $X_1, X_2, \ldots$ be points chosen IID uniformly at random from U and let $Z_1, Z_2, \ldots$ be points chosen IID from some continuous probability distribution having support U. For a given $n \geq 5$, define $S = \{Y_1, \ldots, Y_{|S|}\} = H^{SC}_{n,c}(X_1, \ldots, X_n)$. If $|S| \geq 5$, then $(Y_1, \ldots, Y_{|S|}) \nsim (Z_1, \ldots, Z_{|S|})$.*

An alternative route planning approach is to fix the number of bus stops — potentially more than the bare minimum required — and locate these bus stops to minimize the maximum home-to-stop walking distance for any student. This is the well-known planar k-median problem. Define the function $H^k_n : U^n \to U^k$ to take as input n demand locations in U and return a unique optimal solution to the k-median problem, again with a well-defined tiebreaking rule. That is, $H^k_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) = S$ is the unique optimal solution to

$$\underset{\mathbf{y}_1, \ldots, \mathbf{y}_k \in U}{\text{lexmin}} \quad \left(\max_{i \in [n]} \left\{ \min_{j \in [k]} \big\{ d(\mathbf{x}_i, \mathbf{y}_j) \big\} \right\}\right),\ t\big(\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}, \{\mathbf{y}_1, \ldots, \mathbf{y}_k\}\big). \tag{2}$$

Observe that k is a valid number of stops in practice if the primary objective value does not exceed c. Proposition 2 is the result analogous to Proposition 1 for this problem.

Proposition 2. *Let $X_1, X_2, \ldots$ be points chosen IID uniformly at random from U and let $Z_1, Z_2, \ldots$ be points chosen IID from some continuous probability distribution having support U. Let $n > k \geq 4$ be positive integers, and define $(Y_1, \ldots, Y_k) = H^k_n(X_1, \ldots, X_n)$. Then, $(Y_1, \ldots, Y_k) \nsim (Z_1, \ldots, Z_k)$.*

When making a stop to pick up students, a school bus incurs extra time slowing down to a complete halt, then activating safety lights and signage. Each additional stop thus increases the duration of the route. A third planning approach is to prioritize minimizing the total duration of the route (while ensuring coverage of student locations) including both travel and stopping time. The resulting solution may visit more bus

stops than the bare minimum required, and students may need to walk longer distances on average. We refer to this as the planar *minimum-duration covering route problem* (MDCRP); its open variant is studied by Zeng et al. (2019a,b), while its closed variant is related to the problem in Carlsson and Jia (2015).

Let $a > 0$ denote the additional route duration incurred for each stop, not including travel time. Assume without loss of generality that the vehicle travels at unit speed. Define the function $H^{\text{MD}}_{n,c,a} : U^n \to 2^U$ to take as input n demand locations in U and return a unique optimal solution to the desired variant of the MDCRP (with a given depot location), including a well-defined tiebreaking rule as before. Proposition 3 is the analogous non-IID result for this setting.

Proposition 3. *Let $X_1, X_2, \ldots$ be points chosen IID uniformly at random from U and let $Z_1, Z_2, \ldots$ be points chosen IID from some continuous probability distribution having support U. For a given $n \geq 5$, define $S = \{Y_1, \ldots, Y_{|S|}\} = H^{MD}_{n,c,a}(X_1, \ldots, X_n)$. If $|S| \geq 5$, then $(Y_1, \ldots, Y_{|S|}) \nsim (Z_1, \ldots, Z_{|S|})$.*

3.3 Discussion

As discussed in Section 1.1.1, CA-based methods often explicitly or implicitly rely on the BHH Theorem. The theorem assumes the vehicle's stop locations are IID; Arlotto and Steele (2016) show that it is not sufficient for stop locations to have identical marginal distributions. The mathematical results in this section prove non-independence of stop locations for a variety of covering route settings, and the empirical data suggests that this non-independence may significantly affect route length estimation. We therefore caution against assuming that covering routes behave like TSPs when using CA methods in planning models.

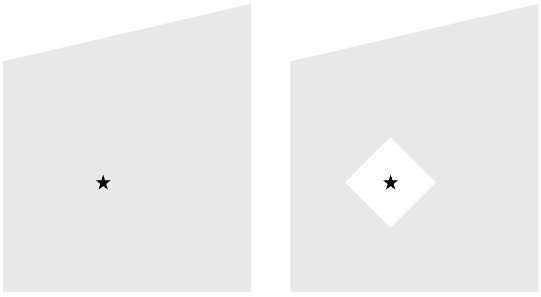
Our results entail two further questions. First, is there a known, parametrized stochastic process that can adequately mimic spatial facility distributions for covering route problems? There may be a connection to the Strauss process (Strauss 1975), in which points are less likely to be located near each other, but even sampling from the Strauss process is not trivial (Huber 2022). Second, which functional form(s) with what parameter(s) provide the best closed-form approximations for covering route problems of practical size?

We leave both of these questions for future work. In the remainder of this paper, we instead demonstrate how a planning problem involving planar covering routes (specifically, school bus fleet size estimation) can be modeled and solved without relying on a particular functional form. Our proposed CA model combines empirical point estimates with simple linear interpolation. While this approach does not allow for the closed-form mathematical analysis often afforded by the square-root BHH routing approximation (see e.g., Stroh et al. 2022), it retains the computational tractability and transparency for which CA methods are valued.

4. Application: School Bus Fleet Sizing

Section 2 presented the motivation of a single bus route serving students located in a single subregion/ neighborhood. However, a school that serves many students dispersed across a broad *attendance region* may require multiple buses in its fleet; a single route is often insufficient due to physical capacity or time constraints. In this section, we analyze the following strategic planning question: how many buses are required to transport a school's students whose exact locations are yet unknown?

Formally, consider a planar attendance region containing a school located at point $\mathbf{o}$. Reflecting common practice, students located within a w-mile radius of the school are not provided bus transportation. We refer to the resulting region in which students are provided transportation to the school as the *service region*, denoted $\mathcal{R}$. Figures 6a and 6b illustrate how the service region is defined via subtraction of a school-centered rectilinear ball from the attendance region. We assume that student locations are IID at random in $\mathcal{R}$ according to a slow-varying density function $\delta : \mathcal{R} \to \mathbb{R}_{>0}$, implying a total of $\int_{\mathcal{R}} \delta(\mathbf{r})\mathrm{d}\mathbf{r}$ students requiring transportation. The slow-varying density assumption, common in CA applications (e.g., Ellegood et al. 2015, Tsao et al. 2012), is revisited in Section 5.



(a) Attendance region **(b)** Service region

Figure 6: Example planar regions

Each bus route is constrained in two ways. First, each bus has a physical capacity of $q_{\max}$ students. Second, no route can exceed $d_{\max}$ minutes in duration, where this time limit is a planning guideline chosen by system managers. We are primarily interested in this latter constraint for methodological and practical reasons: the minimum required fleet size is trivially $\left\lceil q_{\max}^{-1} \int_{\mathcal{R}} \delta(\mathbf{r})\mathrm{d}\mathbf{r} \right\rceil$ when the density is sufficiently high, but the impact of $d_{\max}$ is not immediately clear while also critically important for policymaking purposes.

As before, the coverage radius c denotes the maximum allowable walking distance (in miles) between a student and their designated bus stop. Each additional stop on a route adds a minutes to the route duration, and each additional student picked up adds b minutes to the route duration. Thus, $a + bz$ of total stoppage time is required to pick up a group of z students at a bus stop; this incorporates the extra time required for the bus to come to a complete halt, activate safety lights and signage, and then load the waiting students. Finally, s denotes the rectilinear travel speed of the bus in miles per minute.

Prior experience indicates that, in practice, (i) school districts need to balance route durations with other operational priorities, and (ii) school bus routes are developed via a combination of straightforward heuristics, proprietary commercial software, and manual adjustments. As such, we seek to estimate a school's future busing needs given its particular approach to bus stop selection and route construction. The CA model developed in the following subsections is flexible in this regard.

4.1 Continuous Approximation Model

Following a common approach in the routing literature, we assume for planning purposes that $\mathcal{R}$ will be partitioned into distinct compact *zones* such that each bus in the fleet will be responsible for the students located in a single zone. Making each zone as large as possible will minimize the number of zones required. Therefore, determining a school's fleet size is, in a sense, equivalent to determining the maximum possible area of each zone. We thus begin by analyzing this area maximization problem. Intuitively, a zone's maximum area should be inversely related to its student density due to the capacity constraint $q_{\max}$. Similarly, a zone's maximum area should be inversely related to its distance from the school, as more of a distant zone's $d_{\max}$ duration 'budget' is spent on linehaul travel. Characterizing the latter relationship requires approximating the expected duration of the open covering route serving a given zone.

Consider a hypothetical square zone of area A containing n students distributed uniformly at random. For the bus route servicing this zone, let the function $L(n, A)$ denote the length of the section of the route inside the zone itself, and let $T(n, A)$ denote the number of bus stops on the route (not including the school). Note that both quantities are actually random variables that depend on the spatial realization of the n students; their expected values are $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$. Next, for such a zone centered at $\mathbf{r} \in \mathcal{R}$, we approximate its student population as $n \approx A\delta(\mathbf{r})$ due to the slow-varying density assumption. Then, the expected duration of this zone's bus route can be modeled as

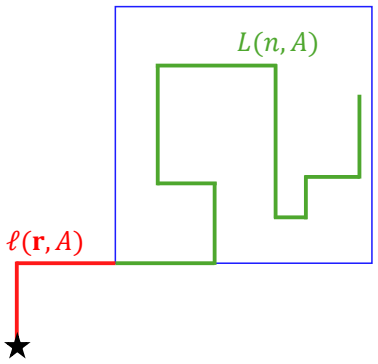


Figure 7: Components of travel distance

$$f(\mathbf{r}, A) = a\,\mathbb{E}\big[T\big(A\delta(\mathbf{r}), A\big)\big] \;+\; bA\delta(\mathbf{r}) \;+\; s^{-1}\,\mathbb{E}\big[L\big(A\delta(\mathbf{r}), A\big)\big] \;+\; s^{-1}\ell(\mathbf{r}, A). \tag{3}$$

The former two terms represent the time spent stopped on the route, while the latter two terms represent the time spent traveling. Specifically, the third term is the time spent traveling inside the zone, while the fourth

term is the linehaul travel duration, where $\ell(\mathbf{r}, A)$ is an estimate of the expected linehaul distance. Figure 7 illustrates the components of the route's travel distance (not necessarily to scale) with $L(n, A)$ in green, $\ell(\mathbf{r}, A)$ in red, and the zone boundary in blue.

Because bus routes tend to travel towards the school, $\ell(\mathbf{r}, A)$ should approximate the distance between the school and the nearest part of the zone. For zones close to the school, this distance is simply the walking radius w. For zones farther away, this distance is approximated as

$$\max\left\{|r_1 - o_1| - \tfrac{1}{2}\sqrt{A}\ ,\ 0\right\} + \max\left\{|r_2 - o_2| - \tfrac{1}{2}\sqrt{A}\ ,\ 0\right\}, \tag{4}$$

where $\mathbf{r} = (r_1, r_2)$ is the center of the hypothetical zone and $\mathbf{o} = (o_1, o_2)$ is the school location. The first term of (4) represents the horizontal component of rectilinear distance; it equals 0 when the zone is directly north or south of the school, and it equals $|r_1 - o_1| - \frac{1}{2}\sqrt{A}$ otherwise. Similarly, the second term of (4) represents the vertical component of rectilinear distance; it equals 0 when the zone is directly east or west of the school, and it equals $|r_2 - o_2| - \frac{1}{2}\sqrt{A}$ otherwise. Figure 8 visualizes the calculation of (4). Combining all of these ideas, we define the linehaul distance $\ell(\mathbf{r}, A)$ as the maximum of w and the expression in (4). It remains to characterize $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$ as functions of n and A, which we discuss next.

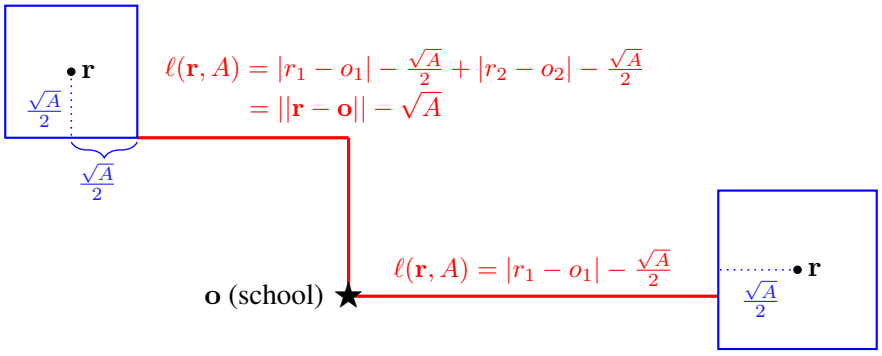


Figure 8: Linehaul distance calculation examples for two hypothetical square zones

4.2 Estimating $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$

As discussed earlier, school districts construct their bus routes based on their specific operational priorities. Thus, the functions $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$ should be empirically estimated in a manner that reflects the route construction procedure employed by the specific school district under consideration.

In this work, we consider a setting in which the primary objective is to minimize the number of bus stops required while respecting the c-mile coverage requirement. Then, we seek routes that perform well with

respect to both route duration and average student home-to-stop walking distance. This approach, which decomposes the full problem into sequential facility location and routing optimization problems, reflects a particular set of operational priorities. Any given district may plan routes in a different manner that reflects their own unique priorities, perhaps by decomposing the SBRP optimization differently or by heuristically solving a single MDCRP-type problem. We emphasize that the overall CA framework is agnostic to the particular SBRP objective(s) and heuristic(s) used; we only require that the procedure used to estimate $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$ should resemble the procedure used to develop operational routes in practice.

The steps of our estimation procedure, detailed with explicitly stated optimization models in Appendix A.4, are briefly summarized as follows.

1. We randomly generate an instance of n IID uniform student locations in a square of area A.
2. We determine the minimum number of bus stops required by solving a planar set covering problem with a coverage radius of c.
3. Generally, bus routes that have a larger geographical 'footprint' tend to be longer. With this in mind, we choose among the optimal solutions from Step 2 by minimizing the sum of the stops' maximum vertical and horizontal displacements.
4. Then, to finalize the bus stop locations, we choose among the optimal solutions from Step 3 by minimizing the total distance between each student and their assigned stop.
5. Finally, we determine the length of the shortest open route that visits all of these bus stop locations and ends at the southwest corner of the square.

The results of steps 2 and 5 correspond to the values of $T(n, A)$ and $L(n, A)$ for this instance. For a particular choice of n and A, repeating these steps several times and taking the average over all instances produces empirical point estimates for $\mathbb{E}\big[T(n, A)\big]$ and $\mathbb{E}\big[L(n, A)\big]$. Upon producing point estimates for several choices of n and A, we linearly interpolate to generate continuous estimator functions $\bar{L}(\cdot)$ and $\bar{T}(\cdot)$ that approximate the true values of the functions $\mathbb{E}\big[T(\cdot)\big]$ and $\mathbb{E}\big[L(\cdot)\big]$. Figure B1 in Appendix B provides examples of these interpolated approximations for values of n and A relevant to our computational studies.

4.3 Fleet Sizing

We can now replace the expectations with their corresponding empirical interpolated approximations:

$$\bar{f}(\mathbf{r}, A) = s^{-1}\bar{L}\big(A\delta(\mathbf{r}), A\big) + a\bar{T}\big(A\delta(\mathbf{r}), A\big) + bA\delta(\mathbf{r}) + s^{-1}\ell(\mathbf{r}, A). \tag{5}$$

We henceforth refer to $\bar{f}$ as the *route duration function*.

Practically, a routing zone cannot be arbitrarily large for two reasons. First, a bus cannot hold more students than its capacity $q_{\max}$, but $\lim_{A\to\infty} A\delta(\mathbf{r}) = \infty$ for any $\mathbf{r}$. Second, a route's duration cannot exceed the duration constraint $d_{\max}$, but $\lim_{A\to\infty} \bar{f}(\mathbf{r}, A) = \infty$ for any $\mathbf{r}$. We define $M(\mathbf{r})$ as the maximum area of a square bus routing zone centered at $\mathbf{r}$; formally,

$$M(\mathbf{r}) \coloneqq \max\ A \qquad \text{s.t. } \bar{f}(\mathbf{r}, A) \le d_{\max},\ A\delta(\mathbf{r}) \le q_{\max} \tag{6}$$

for all $\mathbf{r} \in \mathcal{R}$. Both $\bar{f}(\mathbf{r}, A)$ and $A\delta(\mathbf{r})$ are continuously increasing in A. Therefore, $M(\mathbf{r})$ is either equal to $q_{\max}/\delta(\mathbf{r})$, or the value of A that solves $\bar{f}(\mathbf{r}, A) = d_{\max}$, or the area of $\mathcal{R}$ — whichever of the three values is smallest. Any general-purpose univariate root finding method can be employed to solve $\bar{f}(\mathbf{r}, A) = d_{\max}$ with little computational effort.

Our ultimate goal is to determine the minimum number of bus routes required. When $d_{\max}$ is sufficiently high such that the maximum route duration constraint is never binding for any $\mathbf{r} \in \mathcal{R}$, we require

$$\frac{1}{q_{\max}} \int_{\mathcal{R}} \delta(\mathbf{r}) \mathrm{d}\mathbf{r} \tag{7}$$

buses, rounded up. If so, we can interpret $\delta(\mathbf{r})$ as the 'load' at point $\mathbf{r}$. However, given our initial motivations, we are less interested in this case. How do we define the load of a point when the route duration might be the binding constraint? As in Banerjee et al. (2022), we use the value $1/M(\mathbf{r})$, which we interpret as an approximation of the number of buses required to service a zone of unit area centered at $\mathbf{r}$. Finally, following Erera (2000) and Banerjee et al. (2022), we integrate over the region to estimate the minimum number of bus routes required:

$$V_{\text{CA}} = \int_{\mathcal{R}} \frac{1}{M(\mathbf{r})} \mathrm{d}\mathbf{r}\ . \tag{8}$$

This integral satisfies a basic boundary condition: when the density of students is sufficiently high, V_{CA} is equal to the total number of students in the region divided by the capacity of each bus. That is, the values of (7) and (8) are identical when student density is the limiting factor. More generally, regardless of density,

V_{CA} is an estimate for the minimum number of bus routes required such that each route contains at most $q_{\max}$ students and is at most $d_{\max}$ minutes long.

4.4 Interpreting V_{CA}

Note that the minimum fleet size estimate V_{CA} is usually fractional, suggesting that rounding up this value is required to satisfy the constraints. However, both 4.2 and 4.8 both round up to 5 buses — how can we glean further insight from the exact decimal value of V_{CA}? In short, V_{CA} should not be interpreted as the expected minimum number of buses required across all potential realizations of student locations. Rather, $\lceil V_{\text{CA}} \rceil$ should be treated as an estimate of the minimum number of buses required in practice, and $\text{frac}(V_{\text{CA}}) = V_{\text{CA}} - \lfloor V_{\text{CA}} \rfloor$ should be used to gauge the likelihood and direction of the estimate's error.

To make this idea concrete, suppose that $V_{\text{CA}} \approx j + 0.5$ for some positive integer j. Then, across all possible actual realizations of student locations, we will require a minimum of exactly $\lceil V_{\text{CA}} \rceil = j + 1$ buses with high probability; we may very rarely encounter realizations that require exactly j or $j + 2$ buses to satisfy the $d_{\max}$ constraint. Similarly, suppose that $j \leq V_{\text{CA}} \ll j + 1$ (e.g., $V_{\text{CA}} \approx 4.1$). Then, in practice, we will often require $j + 1$ buses, but we will occasionally require only j buses when the realized students are somewhat conveniently located. Finally, suppose that $j \ll V_{\text{CA}} \leq j + 1$ (e.g., $V_{\text{CA}} \approx 4.9$). Then, in practice, we will often require $j + 1$ buses, but we will occasionally require $j + 2$ buses when the realized students are somewhat inconveniently located.

These interpretations of V_{CA} are valid to the extent that the routing approximations are good estimates of operational route durations. Therefore, we can judge the accuracy of our approach by randomly generating many realizations of student locations, constructing actual bus routes for these instances, and noting the minimum number of buses required to satisfy a hard $d_{\max}$ constraint in each instance. If V_{CA} is indeed a valid guideline, we should observe that a minimum of exactly $\lceil V_{\text{CA}} \rceil$ buses are typically required for values of $d_{\max}$ corresponding to $\text{frac}(V_{\text{CA}}) \approx 0.5$. Frequently requiring $j + 2$ buses in practice when $j \leq V_{\text{CA}} \ll j + 1$ would imply that our approach is noticeably underestimating actual route durations. Similarly, frequently requiring only j buses in practice when $j \ll V_{\text{CA}} \leq j + 1$ would imply that our approach is noticeably overestimating actual route durations.

Evaluating (8) is computationally inexpensive for a given combination of parameters once $\bar{T}(\cdot)$ and $\bar{L}(\cdot)$ have been estimated. Additionally, if V_{CA} is viewed as a function of the input parameters, it is clear that V_{CA} is non-increasing in $d_{\max}$. Taking an inverse optimization approach to our strategic planning problem,

a transportation policymaker can thus use univariate root finding to easily determine what value of $d_{\max}$ produces a desired V_{CA} output. For example, a policymaker who seeks robustness may wish to know what values of $d_{\max}$ entail frac$(V_{\text{CA}}) = 0.5$. This idea can be applied to other strategic policy levers, such as the walking radius w and bus capacity $q_{\max}$.

The same concept can also be used to judge the accuracy of the CA approach by fixing the number of buses and treating $d_{\max}$ as a soft constraint to be targeted in expectation. For practical and reasonable integer values of j, root finding can be used to find the value of $d_{\max}$ such that $V_{\text{CA}} = j$. We can again generate many random demand realizations and construct j routes for each realization. If we track the longest of the j routes in each realization, the resulting distribution will necessarily show some variation due to the inherent underlying randomness. However, if V_{CA} is a valid guideline, this distribution of the longest routes should be roughly centered at the aforementioned value of $d_{\max}$. We demonstrate both validation approaches in the following section's computational studies.

5. Computational Examples

In this section, we implement the fleet sizing approach from Section 4 and assess its performance. The methodological and managerial insights of our computational studies are summarized as follows.

(i) Operational simulations provide evidence that our CA approach produces accurate estimates in planar settings, both with uniform and slow-varying student densities (Section 5.1). For comparison purposes, we find that a 'traditional' route length approximation procedure — i.e., assuming that stop locations are IID, as in the BHH Theorem — produces less accurate fleet size estimates.

(ii) Our CA approach, while developed for metric planar regions, maintains its accuracy when applied to a case study set on the grid-like road network of a real-world suburban school (Section 5.2). Additionally, our approach remains fairly accurate even if the slow-varying density assumption is violated.

(iii) Our proposed approach's accuracy noticeably decreases when the road network structure is sparser and highly irregular, as we demonstrate in a follow-up case study (Section 5.3). Methodological implications and practical consequences are are discussed in Section 5.4.

Attendance region boundaries for the two real-world case studies are sourced from the 2015–2016 United States School Attendance Boundary Survey (National Center for Education Statistics 2016) and manually adjusted to reflect current conditions. To ensure privacy, no actual student or route data were used in the

case studies, and the schools themselves remain anonymous. We use Gurobi 12.0 via Python 3.7 for all optimization tasks in this section unless stated otherwise.

5.1 Planar Attendance Regions

5.1.1 Uniform Density

We begin by analyzing a four-by-five mile planar region with a centrally located school and a walking radius of one mile (Figure 9). A total of 250 students are distributed uniformly at random within the region, corresponding to a density of $\delta(\mathbf{r}) \approx 13.9$ students per square mile for all $\mathbf{r} \in \mathcal{R}$. Travel is rectilinear at a speed of 20 miles per hour or, equivalently, one-third of a mile per minute. The 'Planar #1' row of Table 1 lists the relevant operational parameters for this setting.

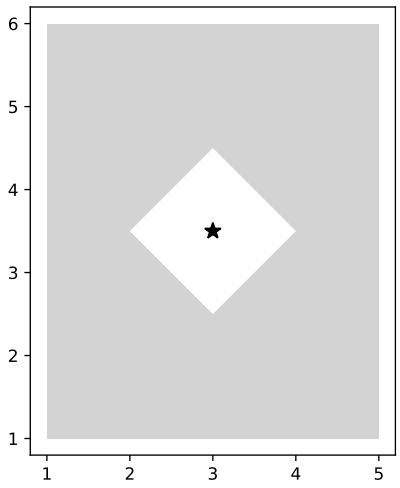


Figure 9: Service region for first planar study (axes in miles)

We construct $\bar{L}(n, A)$ and $\bar{T}(n, A)$ for this setting as follows. We consider squares with side lengths $\{1, 1.5, 2, 2.5, 3\}$ and $n \in \{5, 10, 15, 20, 25, 35, 50, 65, 80\}$. For each square and choice of n, we randomly generate 100 sets of student locations uniformly at random. Next, we follow the procedure in Section 4.2 to produce empirical point estimates for $\mathbb{E}\big[T(n, A)\big]$ and $\mathbb{E}\big[L(n, A)\big]$; we assume that bus stops may be located on a lattice with 0.05-mile spacings. Finally, we linearly interpolate to produce the continuous estimator functions $\bar{L}(n, A)$ and $\bar{T}(n, A)$. As an aside, we originally intended to conduct this estimation procedure for evenly spaced values of $n \in \{5, 20, 35, 50, 65, 80\}$. However, we found that the number of stops required and route lengths increase most rapidly for smaller n. We therefore conducted additional simulations for $n \in \{10, 15, 25\}$. Observe the rapidly increasing followed by plateauing behavior of $\bar{T}(n, 4)$ and $\bar{L}(n, 4)$ in Figure 10. Figure B1 in Appendix B plots the complete empirical data.

Table 1: Summary of computational studies' parameters

	Bus capacity (students)	Walking radius (mi.)	Stop coverage radius (mi.)	Per-stop time (min.)	Per-student time (min.)	Bus speed (mi./min.)
Planar #1	$q_{\max} = 70$	$w = 1$	$c = 0.5$	$a = 1$	$b = 1/6$	$s = 1/3$
Planar #2	70	1	0.5	1.25	1/6	1/4
S-1	81	0.5	0.25	1	1/6	Varies
S-2	81	1.5	0.5	1	1/6	Varies

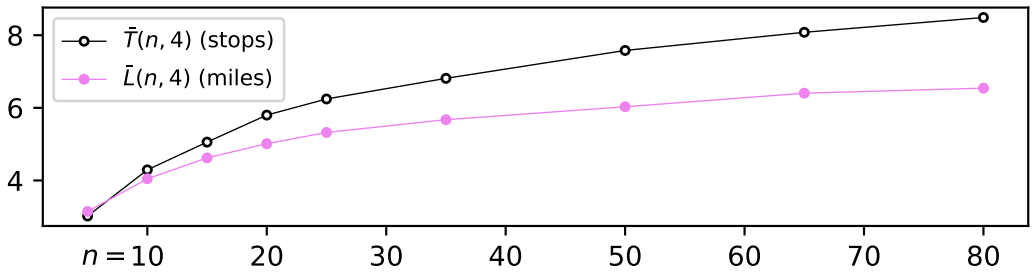


Figure 10: Example $\bar{T}(\cdot)$ and $\bar{L}(\cdot)$ functions

We can now evaluate the integral (8) to estimate V_{CA} for a given $d_{\max}$. Here and in the remainder of this paper, we numerically evaluate integrals using the open-source legacy version of QuadPy (Schlömer 2023) for quadrature together with the open-source TriPy package (Bolgert 2019) for triangulation. Figure 11 displays computed values of V_{CA} as a function of $d_{\max}$. Green, yellow, and red markers indicate $\lceil V_{\text{CA}} \rceil = 4$, 5, and 6 buses, respectively. The apparent convex relationship suggests that greater amounts of transportation resources are required to achieve similar reductions in route durations as $d_{\max}$ decreases.

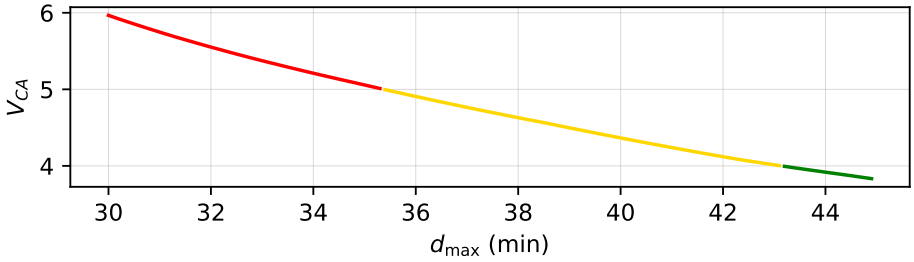


Figure 11: V_{CA} as a function of maximum route duration

The values in this plot can be used to easily assess the impact of lowering the maximum route duration. For example, $d_{\max} = 40$ minutes implies $V_{\text{CA}} = 4.37$, while $d_{\max} = 36$ minutes implies $V_{\text{CA}} = 4.91$. From a planning perspective, this suggests that five buses are required in practice with high likelihood to ensure that each route does not exceed 40 minutes in duration. However, there is a significant likelihood of requiring six buses in practice to ensure that each route does not exceed 36 minutes in duration. A conservative policymaker seeking robustness may target $V_{\text{CA}} \approx 4.5$ or 5.5; these correspond to $d_{\max} \approx 38.98$ and 32.28 minutes, respectively. We emphasize that such insights are near-instantaneous to derive after the one-time computational cost associated with estimating $\bar{T}(\cdot)$ and $\bar{L}(\cdot)$.

Of course, it is necessary to validate these insights before practical implementation. We do so by generating 100 random sets of 250 students uniformly at random within the service region. For each of these 100 operational instances, we heuristically solve a SBRP in four steps. First, with bus stop locations chosen

from a 0.05×0.05 mile lattice, we determine the minimum number of bus stops k required in the region by solving a set covering problem. Second, among all k-stop solutions that cover the students, we determine the sum of the maximum vertical and horizontal ranges of the bus stop locations. Third, with coverage and this sum as constraints, we finalize the k bus stop locations by solving a k-center problem minimizing the average distance traveled by each student to their assigned bus stop. Fourth, we seek bus routes that together visit all of these stops and minimize the length of the longest route. We metaheuristically solve this final min-max VRP with the Google OR-Tools implementation of guided local search with a time limit of five minutes; initial routes are constructed via the classical savings algorithm (Clarke and Wright 1964).

For each instance, we solve the SBRP with four, five, and six buses and record the maximum route duration for each. For a range of $d_{\max}$ corresponding to certain values of V_{CA}, we then check to determine the minimum number of buses required to ensure that no route exceeds $d_{\max}$ minutes. For each value of V_{CA}, Table 2 displays the corresponding value of $d_{\max}$ and the number of instances (out of 100) that require four, five, six, or seven-plus buses to satisfy the hard $d_{\max}$ constraint for all routes.

Table 2: Operational results, first planar study

		# of instances requiring...			
V_{CA}	$d_{\max}$ (min)	4 buses	5 buses	6 buses	7+ buses
4	43.14	42	58	0	0
4.25	40.91	1	99	0	0
4.5	38.98	0	100	0	0
4.75	37.11	0	99	1	0
5	35.37	0	47	53	0
5.25	33.75	0	3	97	0
5.5	32.28	0	0	100	0
5.75	30.99	0	0	86	14
6	29.86	0	0	33	67

These results align with the interpretations discussed in Section 4.4. In particular, we highlight that exactly five buses are always required when $V_{\text{CA}} = 4.5$, and exactly six buses are always required when $V_{\text{CA}} = 5.5$. We never observe any extreme prediction errors in this setting; i.e., six buses are never required when $V_{\text{CA}} \approx 4$, and four or seven buses are never required when $V_{\text{CA}} \approx 5$. Figure 12 displays the longest route duration for each of the 300 solutions (100 operational instances each solved with four, five, and six buses). The median of these observed longest route durations is 43.29 minutes when using four buses, 35.50 minutes when using five buses, and 30.25 minutes when using six buses. These are all within only 1.3% of their corresponding $d_{\max}$ values in Table 2.

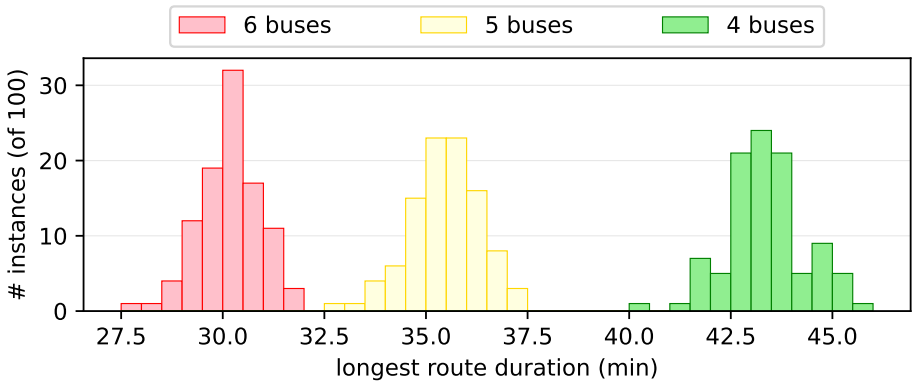


Figure 12: Histogram of observed longest route durations across 100 random instances, first planar study

These results provide evidence in support of our CA approach's accuracy in planar settings. However, the analysis in Section 3.1 implies a further question: is this accuracy maintained if the underlying route lengths were to be estimated in the 'traditional' manner (i.e., by generating IID uniform bus stop locations)? To test this, we repeat the initial estimation procedure, but we replace the optimized covering bus stop locations with an equal number of IID uniform bus stop locations in each trial. Formally, this modifies the $\bar{L}(\cdot)$ element of the expression in (5).

We find that this procedure leads to lower (and therefore less accurate) fleet size guidelines across the spectrum of realistic $d_{\max}$ values. As suggested by Section 3.1, this is because IID bus stop locations entail shorter route lengths; these imply larger $M(\mathbf{r})$ values, which in turn decrease the value of the integral (8). To illustrate, using the IID estimation procedure, $V_{\text{CA}} = 4$ is achieved at $d_{\max} = 41.26$ minutes, $V_{\text{CA}} = 5$ is achieved at $d_{\max} = 34.32$ minutes, and $V_{\text{CA}} = 6$ is achieved at $d_{\max} = 29.19$ minutes. These are 2.2–4.3% lower than the corresponding $d_{\max}$ values in Table 2.

Of particular managerial interest, this can lead to tangible practical differences for certain parameter combinations. For example, $d_{\max} = 31$ minutes implies $V_{\text{CA}} = 5.57$ when route lengths are estimated with IID bus stop locations, suggesting six buses will be required with near certainty. However, our more accurate proposed approach gives $V_{\text{CA}} = 5.75$, indicating some non-trivial likelihood of requiring seven buses. Indeed, seven buses are required to satisfy the hard constraint of $d_{\max} = 31$ minutes in 14 of the 100 operational simulations. When aggregated over a district with many schools, this unidirectional bias may lead to an underallocation of transportation resources.

5.1.2 Non-Uniform Density

Our CA framework is not limited to uniform density distributions, as $\delta(\mathbf{r})$ may vary slowly over the region. We next consider a similar planar region in which 250 students are now randomly distributed according to an inhomogeneous density function that slowly increases from $\delta(\mathbf{r}) \approx 11$ to $\delta(\mathbf{r}) \approx 17$ students per square mile in the southward direction (Figure 13). For additional variety in testing, we also move the school off-center, decrease the travel speed, and increase the per-stop time (see 'Planar #2' in Table 1).

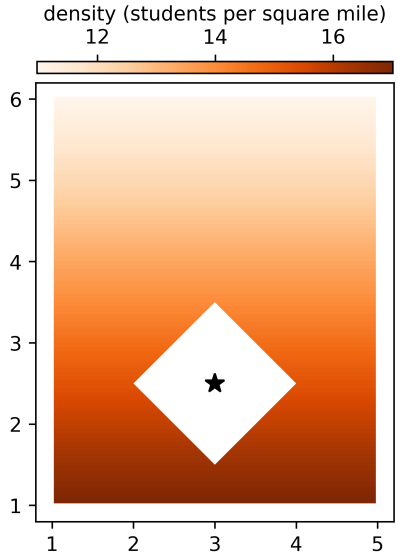


Figure 13: Service region for second planar study (axes in miles)

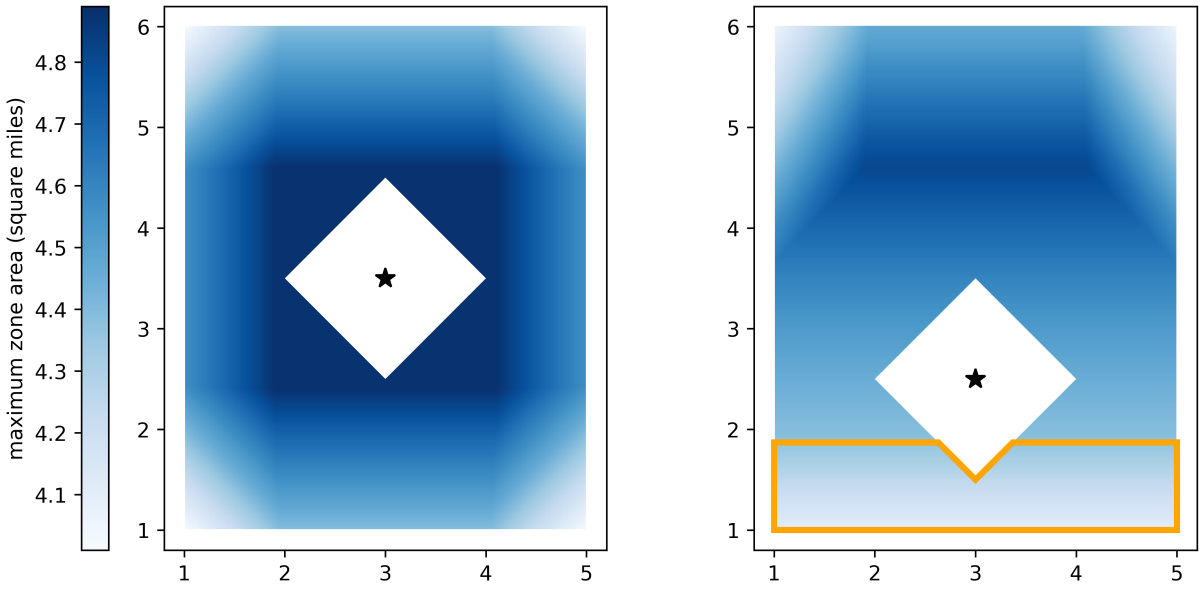
A key benefit of our CA approach is that $\bar{T}(\cdot)$ and $\bar{L}(\cdot)$ do not need to be re-estimated when these parameters change. Utilizing the previously constructed estimator functions, $d_{\max} = 50$ minutes implies $V_{\text{CA}} = 4.39$ and $d_{\max} = 40$ implies $V_{\text{CA}} = 5.66$; the increased busing needs are due to the slower travel speed s and higher per-stop time a. Operational tests again validate our results in this setting. In particular, we highlight that the median longest route duration across all instances is 53.85 minutes when using four buses, 44.23 minutes when using five buses, and 37.81 minutes when using six buses. These are all within 0.9% of the $d_{\max}$ values corresponding to $V_{\text{CA}} = 4$, 5, and 6. Table 3 contains detailed results.

Table 3: Operational results, second planar study

		# of instances requiring...			
V_{CA}	$d_{\max}$ (min)	4 buses	5 buses	6 buses	7+ buses
4	54.26	63	37	0	0
4.25	51.36	4	96	0	0
4.5	48.95	0	100	0	0
4.75	46.67	0	99	1	0
5	44.56	0	61	39	0
5.25	42.68	0	16	84	0
5.5	41.01	0	1	99	0
5.75	39.49	0	0	97	3
6	38.12	0	0	64	36

In the first planar study, all points in the region were 'duration-constrained' in the sense that, even when using the fewest possible number of buses (four), $\bar{f}\big(\mathbf{r}, M(\mathbf{r})\big) = d_{\max}$ for all $\mathbf{r} \in \mathcal{R}$. The second planar computational study was developed, in part, to demonstrate that this need not be the case. Figure 14 plots

$M(\mathbf{r})$ across each region for the corresponding values of $d_{\max}$ that entail $V_{\text{CA}} = 4$. For all points within the outlined section of Figure 14b, $M(\mathbf{r})$ is constrained by the bus capacity instead of the maximum route duration; i.e., $M(\mathbf{r})\delta(\mathbf{r}) = q_{\max}$ and $\bar{f}\big(\mathbf{r}, M(\mathbf{r})\big) < d_{\max}$.



(a) $M(\mathbf{r})$ for $d_{\max} = 43.14$ minutes, first planar study **(b)** $M(\mathbf{r})$ for $d_{\max} = 54.26$ minutes, second planar study

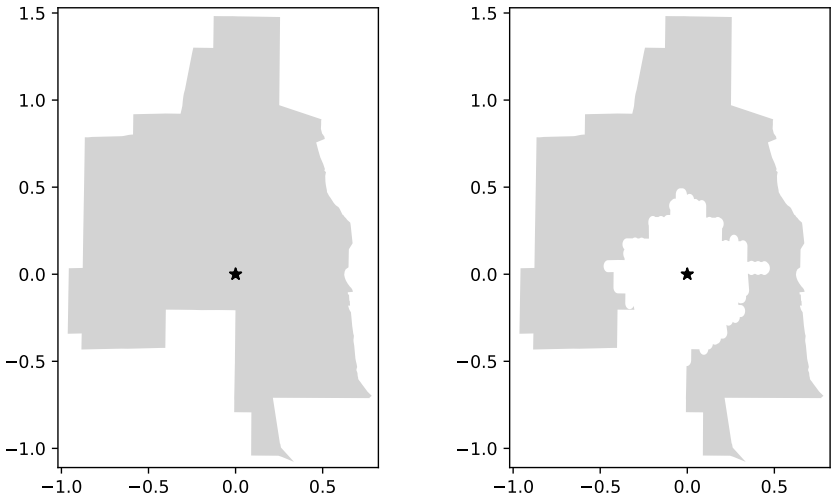
Figure 14: Maximum area functions for planar studies (axes in miles)

5.2 Case Study #1

Until this point, our work has focused on stylized planar attendance regions with rectilinear travel. In the two following case studies, we return to our original real-world motivation: estimating bus transportation needs for schools located on actual road networks. For realism, we assume that the buses used in both case studies have a capacity of $q_{\max} = 81$ students; this reflects the popular Thomas Saf-T-Liner C2 model known to be part of the school district's fleet. We also assume that buses in the case studies travel along road networks at variable speeds corresponding to two-thirds of each street's posted speed limit.

The subject of the first case study is a school in the midwestern United States, henceforth referred to as S-1. Figure 15a shows the attendance region of S-1, covering approximately 2.4 square miles Each bus stop has a coverage radius of 0.25 miles, and students within a walking radius of 0.5 miles of the school are not provided transportation. Figure 15b shows S-1's resulting service region. The 'S-1' row of Table 1 lists other relevant parameters.

Although the CA setup generally follows the steps of the planar studies, additional preliminary data



(a) S-1 attendance region **(b)** S-1 service region

Figure 15: Regions for first case study (axes in miles)

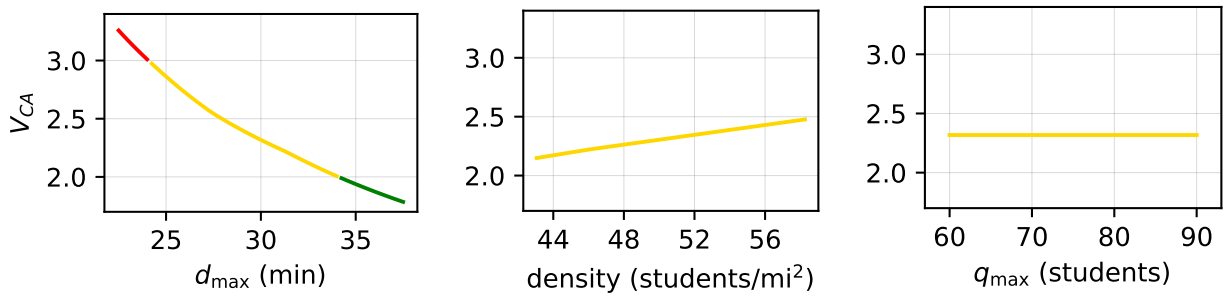
preparation is required to properly adapt our planar CA framework to a specific region's road network. We use OSMnx (Boeing 2025) to query road networks, distances, and travel times. In summary, potential bus stop locations are first selected from the actual road network with inter-stop spacings of approximately 0.05 miles in both the vertical and horizontal directions. Next, for each potential bus stop, OSMnx is also used to construct a quarter-mile *isopleth* — a polygon within which all students are located at most 0.25 miles from that bus stop when traveling along the actual road network. Similarly, a 0.5-mile isopleth is constructed for the school. Because of the intricacies and asymmetries of real-world road networks, these isopleths are not simple diamonds; observe the irregular shape of the school's isopleth removed from the attendance region to create the service region in Figure 15. Then, functions analogous to $\bar{T}(\cdot)$ and $\bar{L}(\cdot)$ are empirically estimated in squares with side lengths $\{0.55, 0.75, 0.95, 1.15, 1.35\}$ and $n \in \{5, 10, 15, 25, 35, 50, 65, 80, 95\}$. Full details are provided in Appendix B.

The schools in S-1's district are scheduled to start each morning and end each afternoon at staggered times. This allows each bus to be reused for multiple schools' routes, as in Bertsimas et al. (2019). Suppose, for example, that a transportation planner seeks to ensure that S-1's route durations are under 30 minutes to facilitate reusing buses. If so, how many buses are required?

5.2.1 Uniform Density

The service region in this case study contains 100 students. As before, we first assume that the students are IID uniformly at random, corresponding to a density of $\delta(\mathbf{r}) \approx 50.7$ students per square mile for all $\mathbf{r} \in \mathcal{R}$.

A maximum route duration of $d_{\max} = 30$ minutes corresponds to $V_{\text{CA}} = 2.32$, suggesting that three buses will be required in practice to achieve the planner's target. This insight is confirmed by our operational validation, as exactly three buses are required in all 100 of the randomly generated instances.



(a) Sensitivity to route duration **(b)** Sensitivity to density **(c)** Sensitivity to bus capacity

Figure 16: Sensitivity analyses, first case study (uniform distribution)

Before discussing validation further, we highlight some practical applications of the CA approach. Figure 16 displays the results of three sensitivity analyses. Green, yellow, and red markers indicate $\lceil V_{\text{CA}} \rceil = 2$, 3, and 4, respectively. Similar to Figure 11, Figure 16a plots V_{CA} as a function of $d_{\max}$. Assuming $d_{\max}$ is fixed at 30 minutes, Figure 16b plots V_{CA} as a function of the service region's density as it varies from 85% to 115% of the current density, and Figure 16c plots V_{CA} as a function of $q_{\max}$ as it varies from 60 to 90 students. Figure 16b may be useful to assess transportation needs if aggregate enrollment changes on long time scales that are typically considered in strategic planning. Figure 16c can be used to inform decisions about new bus leases or purchases; the bus capacity's lack of impact on V_{CA} implies that all of the routes are duration-constrained.

Returning to the original parameters, we use the procedure in Section 5.1 to construct actual routes for two and three buses across 100 randomly generated instances. Table 4 displays the results, which are in line with the planar studies. Most importantly, exactly three buses are required in every instance to satisfy the $d_{\max}$ value associated with $V_{\text{CA}} = 2.5$. The median longest observed maximum route duration when using two buses is 34.98 minutes, and the median longest observed maximum route duration when using three buses is 24.17 minutes. These are deviations of only 2.6% and 0.4%, respectively, from the corresponding $d_{\max}$ values of 34.10 and 24.07 minutes.

Table 4: Operational results, first case study (uniform density)

		# of instances requiring...		
V_{CA}	$d_{\max}$ (min)	2 buses	3 buses	4+ buses
2	34.10	27	73	0
2.25	30.88	0	100	0
2.5	27.94	0	100	0
2.75	25.82	0	95	5
3	24.07	0	44	56

5.2.2 Non-Uniform Density

Recall that the density function is assumed to be continuous and to vary slowly, if at all, over the region. How does the model perform if the density function contains discontinuities at which its value changes very abruptly? To test this, we maintain S-1's total bus-eligible population of 100 students but segment the service region into three distinct residential areas (shown in Figure 17) with noticeably different densities based on recent enrollment data. This new density function has $\delta(\mathbf{r}) \approx 31.5$, 89.8, and 64.6 students per square mile within S-1A, S-1B, and S-1C, respectively. With this new density function, the fleet size guideline decreases slightly to $V_{\text{CA}} = 2.25$ for $d_{\max} = 30$ minutes.

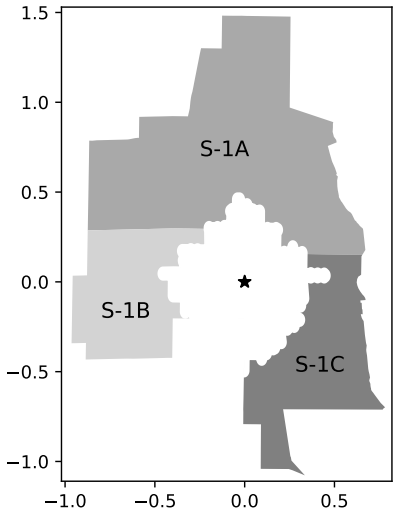


Figure 17: S-1A, S-1B, and S-1C

Table 5 displays the operational results. Observe that exactly three buses are required in 99 of the 100 random instances to satisfy the $d_{\max}$ constraint corresponding to $V_{\text{CA}} = 2.5$. The median longest observed maximum route duration when using two buses is 34.72 minutes, and the median longest observed maximum route duration when using three buses is 23.83 minutes. These are deviations of less than 4.7% and 0.6%,

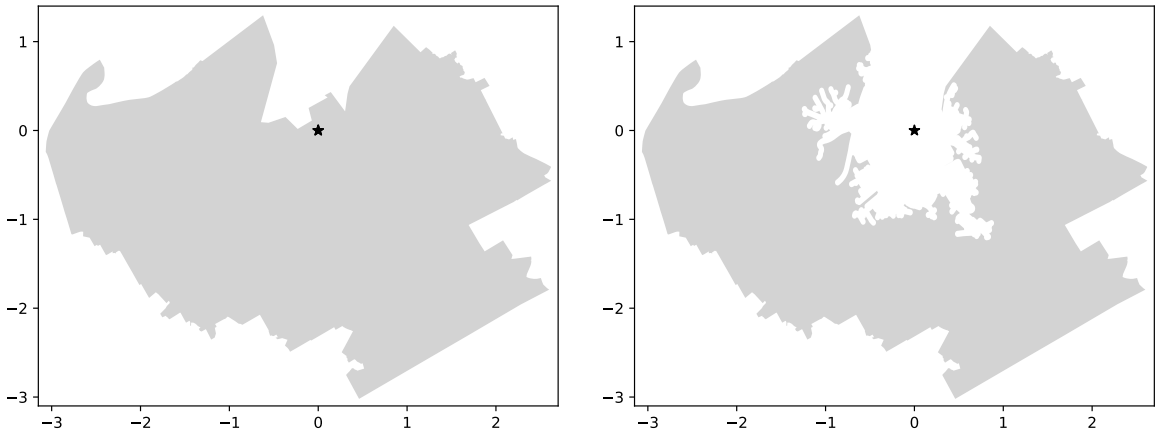
Table 5: Operational results, first case study (non-uniform density)

		# of instances requiring...		
V_{CA}	$d_{\max}$ (min.)	2 buses	3 buses	4+ buses
2	33.17	12	88	0
2.25	30.00	0	100	0
2.5	27.47	0	99	1
2.75	25.47	0	89	11
3	23.70	0	47	53

respectively, from the corresponding $d_{\max}$ values of 33.17 and 23.70 minutes. In summary, the model still performs well despite violating the slow-varying assumption, with only a slight reduction in operational accuracy when compared to the uniform density setting.

5.3 Case Study #2

The subject of the second case study is a school in the southeastern United States, henceforth referred to as S-2. Figure 18a shows the attendance region of S-2, covering approximately 14.3 square miles. Given S-2's larger geographical footprint, each bus stop now has a coverage radius of 0.5 miles, and students within a walking radius of 1.5 miles are not provided transportation. Figure 18b shows S-2's resulting service region. Table 1 lists other relevant parameters.



(a) S-2 attendance region **(b)** S-2 service region

Figure 18: Regions for second case study (axes in miles)

S-1's road network is fairly homogeneous, exhibiting a grid-like structure across most of its attendance region. It is therefore unsurprising that the CA approach, developed with rectilinear distances in mind, performs fairly well in the first case study. S-2, however, was chosen for testing in this paper specifically because its sprawling road network is relatively sparse and highly irregular: its streets are not geographically oriented in any predictable pattern, road types range from atypically narrow residential streets to segments of a major United States numbered highway, and dead-end streets are common. Planning for S-2 is further complicated by the presence of multiple large lakes and a street-level railroad with infrequent crossing locations. These differences are apparent when comparing Figures B2a and B6a in Appendix B, or when comparing the shapes of the walking radius isopleths in Figures 15b and 18b. For completeness, we seek to understand how the aforementioned inconvenient features affect the accuracy of our CA model.

We assume that a total of 250 students are IID uniformly at random in S-2's service region, corresponding to a density of $\delta(\mathbf{r}) \approx 17.5$ students per square mile for all $\mathbf{r} \in \mathcal{R}$, and our target maximum route duration is 45 minutes. Upon re-estimating the necessary empirical functions (as described in Appendix B), our CA approach produces an estimate of $V_{\text{CA}} = 5.53$ for $d_{\max} = 45$ minutes. Therefore, we should expect that exactly six buses are almost always required in practice. Instead, only 73 of the 100 operational instances require six buses, while the other 27 instances require only five buses, suggesting that the CA model is overestimating actual route durations and busing needs. Table 6 contains data for a wider range of V_{CA} and $d_{\max}$ values.

Table 6: Operational results, second case study (uniform density)

		# of instances requiring...			
V_{CA}	$d_{\max}$ (min.)	4 buses	5 buses	6 buses	7+ buses
4	60.15	85	15	0	0
4.25	56.83	40	60	0	0
4.5	53.91	8	92	0	0
4.75	51.61	0	99	1	0
5	49.56	0	93	7	0
5.25	47.75	0	73	27	0
5.5	46.10	0	44	56	0
5.75	44.61	0	23	77	0
6	43.26	0	8	92	0

This data shows a pattern of slight but noticeable overestimation by the CA model, particularly when compared to the accurate results in Tables 2–5. The median longest observed maximum route duration is 57.85 minutes when using four buses, 46.41 minutes when using five buses, and 39.32 minutes when using six buses. These values are 3.8%, 6.6%, and 9.1% lower, respectively, than the corresponding values of $d_{\max}$ in Table 6. This represents a reduction in accuracy from the previous computational examples. As we discuss further in Section 5.4, these results demonstrate the limitations of directly applying CA methods developed for planar settings to sparse, irregular road networks.

5.4 Discussion

We do not assume that $\bar{T}(n, A)$ and $\bar{L}(n, A)$ adhere to any particular structure. However, we observe these functions to be well-behaved in planar settings (see Figures 10 and B1), facilitating interpolation with limited data. The analogous functions are similarly well-behaved for S-1's region. However, this is not the case for S-2's decidedly non-planar region. As an example, Figure 19 plots the directly estimated and interpolated

values of $s^{-1}\bar{L}(20, A)$, $s^{-1}\bar{L}(50, A)$, and $s^{-1}\bar{L}(80, A)$ for both case studies. For each n, the values for S-1 are increasing in A in a fairly consistent manner. This is clearly not the case for S-2, for which we even observe a slight decrease from $s^{-1}\bar{L}(50, 1)$ to $s^{-1}\bar{L}(50, 1.96)$! The relative irregularity and sparsity of S-2's road network lead to noisy, unreliable estimator functions, which in turn produce lower-quality fleet size guidelines, suggesting that CA may not be the most appropriate planning tool in this case.

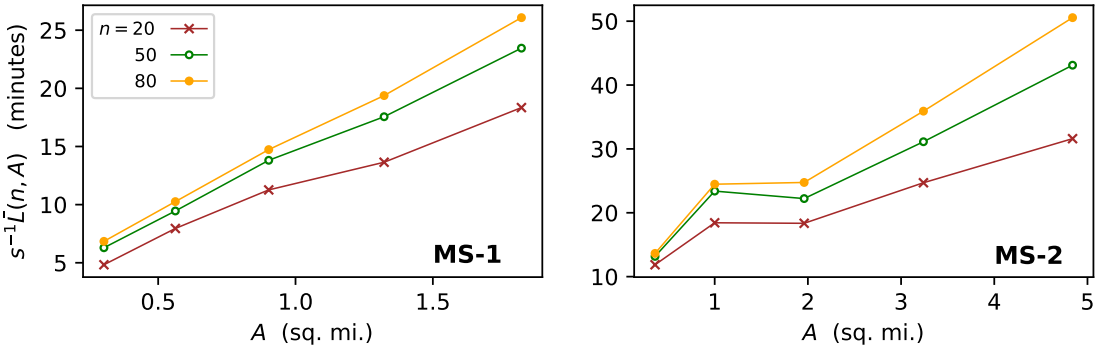


Figure 19: Comparison of selected empirical estimator functions for S-1 and S-2

These results highlight the challenges of developing parsimonious planning models when planar regions with metric distances are not a good approximation of reality. Covering routes seem to pose specific difficulties arising from the effects of road network topologies on the coverage 'footprint' of a particular stop location. Based on our experiences conducting these computational studies, we provide two practical recommendations for future research involving approximations for covering routes.

(i) It should be verified whether CA methods developed under planarity assumptions are appropriate for a given street network. This is especially important when dealing with covering routes, because road networks induce isopleths that are smaller than analogous rectilinear balls. Among S-1's quarter-mile walking isopleths, the median isopleth has 85% of the area of the corresponding rectilinear ball. For S-2's more irregular setting, the same value is significantly lower at only 64%. Thus, rectilinear balls may severely overestimate the actual coverage area of each bus stop.

In the same vein, a common approach in the routing literature is to use Euclidean or rectilinear distances scaled by a circuitry factor as a proxy for actual road network distances. We strongly caution against this practice when covering routes are involved, because irregularly shaped road network isopleths are not necessarily well-approximated by the diamond shape of rectilinear balls. Using such a scaling procedure without verification may result in inaccurate coverage of demand locations, with

potentially serious consequences if schoolchildren are required to walk longer distances or spend more time on a bus than expected.

(ii) When the sizes of routing zones are subject to change (i.e., when A is not fixed in a CA formula), the presence of significant geographical barriers may lead to unreliable and ill-behaved routing approximations. Care should be taken to identify such barriers prior to practical implementation. Whereas, for example, Banerjee et al. (2023) specifically avoid including a particular impassable mountain in their hypothetical same-day delivery service regions of varying area, the lakes and railroad in S-2's actual attendance region cannot be ignored by a school district legally mandated to serve students within its boundaries. Richer predictive models may be required, perhaps by extending the machine learning methods of Merchán and Winkenbach (2019) to non-urban settings and covering routes.

6. Concluding Remarks

Motivated by long-term policymaking for public school transportation systems, this work studied planning approximations for the length of planar covering routes. We provided empirical evidence to suggest that approximation methods developed for traditional TSP routes may lead to significant prediction errors when directly applied to covering routes. We explained this discrepancy by formally showing that, for several different objectives, covering routes' inherent structural properties violate a core assumption of the BHH Theorem. Leveraging these results, we then developed a CA approach for estimating a school's necessary bus fleet size given constraints on the duration of each route and the physical capacity of each bus. To facilitate its use in policymaking and acceptance by non-technical stakeholders, care was taken to ensure the approach's transparency and computational efficiency (after one-time offline setup). We conducted extensive computational experiments to demonstrate the applicability of our approach for school bus fleet sizing on stylized planar regions and grid-like real-world street networks. We also conducted an experiment on a highly irregular suburban street network to highlight the unique challenges of using parsimonious CA methods for covering routes in such settings.

A notable feature of our CA fleet sizing approach is that no functional form was assumed regarding the length of the bus route within each routing zone; simple linear interpolations were used instead. One avenue for future work is to develop closed-form expressions for planar covering route lengths in a similar fashion to the VRP routing functions proposed by Figliozzi (2008). As discussed at the end of Section 5.4, a related task is to develop learning-based length approximations for covering routes on ill-behaved street networks

for which planarity is not a reasonable simplifying assumption.

This work focused on transportation decision-making at the individual school level. Student density, attendance region boundaries, and the school's location were all assumed to be fixed. However, the proposed CA fleet sizing framework can potentially be used to support broader decision-making at the district level. For example, geographical student density is strongly influenced by the specific mechanism used to determine which students attend which schools; our CA approach could be used to estimate the future transportation needs associated with implementing a complex district-wide school choice policy, such as that of Ashlagi and Shi (2014). Similarly, our CA approach could be used to assess the transportation-related impacts of adjusting school boundaries, eliminating grade levels from certain schools, or closing schools entirely — all of which are being implemented in large school districts across the United States despite significant community pushback (e.g., Asmar 2024, Mallicoat 2024). With the methodological foundations laid in this paper, we intend to explore such policy issues in follow-up work.

Finally, our results are relevant to future applications of CA methods to other transportation and logistics problems involving covering routes. To illustrate, consider the recent work of Rojas et al. (2025) and Wang et al. (2025b), who both leverage the BHH Theorem to estimate route lengths for last-mile delivery vehicles that visit both customer residences and parcel lockers on each route. However, in both cases, each route tends to visit many IID customer locations and relatively few lockers; thus, unlike in this work, standard TSP-derived routing functions remain useful planning approximations. The structural insights from our work may be useful for similar settings in which delivery routes stop only at optimally located parcel lockers.

References

Ansari S, Başdere M, Li X, Ouyang Y, and Smilowitz K (2018) Advancements in continuous approximation models for logistics and transportation systems: 1996–2016. *Transportation Research Part B*, 107:229–252.

Applegate DL, Bixby RE, Chvátal V, and Cook WJ (2011) *The Traveling Salesman Problem: A Computational Study*. Princeton University Press.

Arlotto A and Steele JM (2016) Beardwood-Halton-Hammersley Theorem for stationary ergodic sequences: a counterexample. *Annals of Applied Probability*, 26(4):2141–2168.

Ashlagi I and Shi P (2014) Improving community cohesion in school choice via correlated-lottery implementation. *Operations Research*, 62(6):1247–1264.

Asmar M (2024) The week before the vote on Denver's proposed school closures was marked by protests and pushback. *Chalkbeat Colorado*, URL https://www.chalkbeat.org/colorado/2024/11/16/denver-school-closure-proposal-sparks-pushback-and-protests/.

Banerjee D, Erera A, Stroh A, and Toriello A (2023) Who has access to e-commerce and when? Time-varying service regions in same-day delivery. *Transportation Research Part B*, 170:148–168.

Banerjee D, Erera A, and Toriello A (2022) Fleet sizing and service region partitioning for same-day delivery systems. *Transportation Science*, 56(5):1327–1347.

Beardwood J, Halton JH, and Hammersley JM (1959) The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, 55(4):299–327.

Bertsimas D, Delarue A, and Martin S (2019) Optimizing schools' start time and bus routes. *Proceedings of the National Academy of Sciences*, 116(13):5943–5948.

Blanchard M, Jacquillat A, and Jaillet P (2024) Probabilistic bounds on the k-traveling salesman problem and the traveling repairman problem. *Mathematics of Operations Research*, 49(2):1169–1191.

Boeing G (2025) Modeling and analyzing urban networks and amenities with OSMnx. *Geographical Analysis*, 57(4):567–577.

Bolgert S (2019) TriPy 1.0.0. URL https://pypi.org/project/tripy/.

Campbell JF (1993) One-to-many distribution with transshipments: an analytic model. *Transportation Science*, 27(4):330–340.

Carlsson JG and Jia F (2015) Continuous facility location with backbone network costs. *Transportation Science*, 49(3):433–451.

Choi Y and Schonfeld PM (2022) Review of length approximations for tours with few stops. *Transportation Research Record*, 2676(3):201–213.

Clarke G and Wright JW (1964) Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12(4):568–581.

Daganzo C (1984a) The distance traveled to visit N points with a maximum of C stops per vehicle: an analytic model and an application. *Transportation Science*, 18(4):331–350.

Daganzo C (1984b) The length of tours in zones of different shapes. *Transportation Research Part B*, 188(2):135–145.

Di Placido A, Archetti C, and Cerrone C (2022) A genetic algorithm for the close-enough traveling salesman problem with application to solar panels diagnostic reconnaissance. *Computers & Operations Research*, 145:105831.

Drezner Z, Brimberg J, and Schöbel A (2024) Dispersed starting solutions in facility location: the case of the planar p-median problem. *Computers & Operations Research*, 169:106726.

Dursunoglu CF, Arslan O, Demir SM, Kara BY, and Laporte G (2025) A unifying framework for selective routing problems. *European Journal of Operational Research*, 320(1):1–19.

Ellegood WA, Campbell JF, and North J (2015) Continuous approximation models for mixed load school bus routing. *Transportation Research Part B*, 77:182–198.

Ellegood WA, Solomon S, North J, and Campbell JF (2020) School bus routing problem: contemporary trends and research directions. *Omega*, 95:102056.

Erera A (2000) *Design of Large-Scale Logistics Systems for Uncertain Environments*. Ph.D. thesis, University of California, Berkeley.

Figliozzi MA (2008) Planning approximations to the average length of vehicle routing problems with varying customer demands and routing constraints. *Transportation Research Record*, 2089(1):1–8.

Franceschetti A, Jabali O, and Laporte G (2017) Continuous approximation models in freight distribution management. *TOP*, 25(3):413–433.

Gendreau M, Laporte G, and Semet F (1997) The covering tour problem. *Operations Research*, 45(4):568–576.

Glenview School District 34 (2025) Bus Stop Procedures and Routing Criteria. URL https://resources.finalsite.net/images/v1578479531/glenview34org/dj4bftgdhr9cxw7khxik/Bus_Stop_Procedures_and_Routing.pdf.

Glock K and Meyer A (2023) Spatial coverage in routing and path planning problems. *European Journal of Operational Research*, 305(1):1–20.

Huber M (2022) Generating from the Strauss process using stitching. In *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing 2020*, pp. 241–251.

Lambeck L (2025) Cuts to school bus service mulled as 1,600 students are MIA. *Westport Journal*, URL https://westportjournal.com/education/cuts-to-school-bus-service-mulled-as-1600-students-are-mia/.

Le Colleter T, Dumez D, Lehuédé F, and Péton O (2023) Small and large neighborhood search for the park-and-loop

routing problem with parking selection. *European Journal of Operational Research*, 308(3):1233–1248.

Mallicoat M (2024) Three Duval elementary schools will close soon; three more after a year. *Jacksonville Today*, URL https://jaxtoday.org/2024/11/05/3-duval-elementary-schools-will-close-soon-3-more-after-a-year/.

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

Moradi N, Mafakheri F, and Wang C (2024) Set covering routing problems: a review and classification scheme. *Computers & Industrial Engineering*, 198:110730.

Morris C (2024) Youngstown schools will stop providing school buses for high schoolers. *Ideastream Public Media*, URL https://www.ideastream.org/education/2024-06-26/youngstown-schools-will-stop-providing-school-buses-for-high-schoolers.

National Center for Education Statistics (2016) School Attendance Boundary Survey 2015–2016. URL https://nces.ed.gov/programs/edge/sabs.

Ohio Administrative Code (2020) Rule 3301-83-13 — School bus routes and stops. URL https://codes.ohio.gov/ohio-administrative-code/rule-3301-83-13.

Park J and Kim BI (2010) The school bus routing problem: a review. *European Journal of Operational Research*, 202(2):311–319.

Rojas B, Larrain H, and Klapp M (2025) Strategic design of collection and delivery point networks for urban parcel distribution. Preprint, URL https://optimization-online.org/?p=29509.

Scharfenberg D (2018) Computers can solve your problem, you may not like the answer: what happened when Boston Public Schools tried for equity with an algorithm. *The Boston Globe*, URL https://apps.bostonglobe.com/ideas/graphics/2018/09/equity-machine/.

Schlömer N (2023) QuadPy 0.16.10. URL https://pypi.org/project/legacy-quadpy/.

Shafahi A, Wang Z, and Haghani A (2018) SpeedRoute: fast, efficient solutions for school bus routing problems. *Transportation Research Part B*, 117A:473–493.

Strauss DJ (1975) A model for clustering. *Biometrika*, 62(2):467–475.

Stroh A, Erera A, and Toriello A (2022) Tactical design of same-day delivery systems. *Management Science*, 68(5):3444–3463.

Tooten T (2023) Bus changes mean more Baltimore County high school students have to walk to campus. *WBAL-TV*, URL https://www.wbaltv.com/article/school-bus-changes-baltimore-county-walk-to-school/44892414.

Tsao YC, Mangotra D, Lu JC, and Dong M (2012) A continuous approximation approach for the integrated facility-inventory allocation problem. *European Journal of Operational Research*, 222(2):216–228.

Vercraene S, Lehuédé F, Monteiro T, and Péton O (2023) The dial-a-ride problem with school bell time adjustment. *Transportation Science*, 57(1):156–173.

Vinel A and Silva DF (2018) Probability distribution of the length of the shortest tour between a few random points: a simulation study. In *Proceedings of the 2018 Winter Simulation Conference*, pp. 3156–3167.

Wang K, Yang Y, Liao M, and Chen A (2025a) The school bus routing problem: a comprehensive review. *European Transport Studies*, 2:100043.

Wang Q, Lyu G, He L, and Teo CP (2025b) Does the Locker Alliance Network improve last mile delivery efficiency? *Management Science*, forthcoming.

Zeng L, Chopra S, and Smilowitz K (2019a) The covering path problem on a grid. *Transportation Science*, 53(6):1656–1672.

Zeng L, Smilowitz K, and Chopra S (2019b) Generalizing the covering path problem on a grid. Preprint, URL https://doi.org/10.48550/arXiv.1904.12258.

Zhou L, Silva DF, and Smith AE (2023) Locating drone stations for a truck-drone delivery system in continuous space. *IEEE Transactions on Evolutionary Computation*, 29(1):158–171.

Appendix A Omitted Technical Details

A.1 Proof of Proposition 1

Proof. Assume $c < 1$ to avoid triviality. Let U' denote the closed rectilinear ball centered at $(0.5, 0.5)$ with radius c. Let U'' denote the closed rectilinear ball centered at $(0.5, 0.5)$ with radius $2c$.

Let $S = \{\mathbf{y}_1, \ldots, \mathbf{y}_{|S|}\}$ be an optimal solution to the problem with $|S| \geq 5$. Suppose for the purposes of contradiction that at least five elements of S lie within U'; without loss of generality, assume $\mathbf{y}_1, \ldots, \mathbf{y}_5 \in U'$. Define $I \subseteq [n]$ such that $i \in I$ if and only if the nearest facility to $\mathbf{x}_i$ is among $\{\mathbf{y}_1, \ldots, \mathbf{y}_5\}$. Then, it must hold that $\mathbf{x}_i \in U''$ for all $i \in I$. As illustrated in Figure A1, let $\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4$ denote the 'corners' of U':

$$\hat{\mathbf{y}}_1 = \big(0.5, 0.5 + \min\{0.5, c\}\big),$$

$$\hat{\mathbf{y}}_2 = \big(0.5 + \min\{0.5, c\}, 0.5\big),$$

$$\hat{\mathbf{y}}_3 = \big(0.5, 0.5 - \min\{0.5, c\}\big),$$

$$\hat{\mathbf{y}}_4 = \big(0.5 - \min\{0.5, c\}, 0.5\big).$$

Observe that each $\mathbf{x}_i$ for $i \in I$ is at most c distance from at least one point in $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4\}$. As such, $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4, \mathbf{y}_6, \ldots, \mathbf{y}_{|S|}\}$ is an improved feasible solution to the set cover problem with only $|S| - 1$ selected facilities, contradicting the optimality of $\{\mathbf{y}_1, \ldots, \mathbf{y}_{|S|}\}$.

It follows that $\mathbb{P}(Y_1, \ldots, Y_5 \in U') = 0$, but $\mathbb{P}(Z_1, \ldots, Z_5 \in U') = \big[\mathbb{P}(Z_1 \in U')\big]^5 > 0$. Therefore, the joint distributions of $(Y_1, \ldots, Y_{|S|})$ and $(Z_1, \ldots, Z_{|S|})$ are different. □

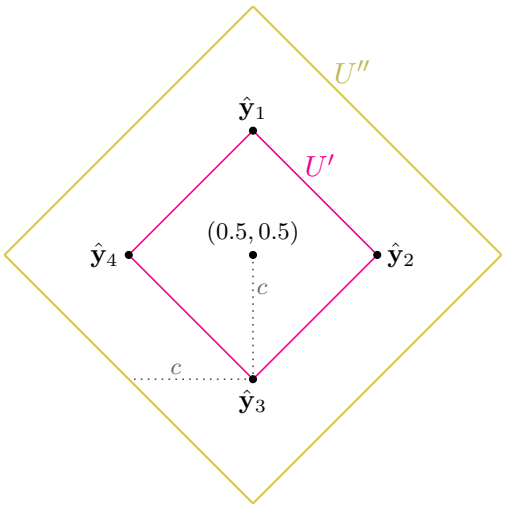


Figure A1: Illustration for proof of Proposition 1 (assuming $c \leq 0.5$) with boundaries of U' and U'' marked

A.2 Proof of Proposition 2

We begin by proving Proposition 2 for the case in which $Z_1, Z_2, \ldots$ are chosen IID with respect to the uniform distribution on U.

Proof. Let U' denote the closed rectilinear ball centered at $(0.5, 0.5)$ with area $\alpha = 0.25^{2n}$ and radius $\rho = \sqrt{\alpha/2}$. Let U'' denote the closed rectilinear ball centered at $(0.5, 0.5)$ with radius 2ρ and area $4\alpha = 0.25^{2n-1}$.

We first claim that $H_n^k(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\} \subset U'$ implies $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset U''$. Suppose for the purposes of contradiction that this is not the case. Then, there exists an instance with $\mathbf{y}_1, \ldots, \mathbf{y}_k \in U'$ and a nonempty set $I \subseteq [n]$ such that $\mathbf{x}_i \notin U''$ for all $i \in I$. By construction, $\min_{j \in [k]} \big\{d(\mathbf{x}_i, \mathbf{y}_j)\big\} > \rho$ for all $i \in I$. Define $\eta = \max_{i \in I} \big\{\min_{j \in [p]} \big\{d(\mathbf{x}_i, \mathbf{y}_j)\big\}\big\}$. Let $\varepsilon > 0$ satisfy $\varepsilon < \frac{1}{2}(\eta - \rho)$ and

$$\varepsilon < \frac{1}{2}\left[\min_{i \in I,\, j \in [4]} \big\{d(\mathbf{x}_i, \mathbf{w}_j)\big\}\right], \tag{A1}$$

where

$$\mathbf{w}_1 = (0.5 + \rho, 0.5 + \rho),$$
$$\mathbf{w}_2 = (0.5 + \rho, 0.5 - \rho),$$
$$\mathbf{w}_3 = (0.5 - \rho, 0.5 - \rho),$$
$$\mathbf{w}_4 = (0.5 - \rho, 0.5 + \rho).$$

Then, consider an alternate arrangement of facilities $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4, \mathbf{y}_6, \ldots, \mathbf{y}_k\}$ in which

$$\hat{\mathbf{y}}_1 = (0.5, 0.5 + \rho + \varepsilon),$$
$$\hat{\mathbf{y}}_2 = (0.5 + \rho + \varepsilon, 0.5),$$
$$\hat{\mathbf{y}}_3 = (0.5, 0.5 - \rho - \varepsilon),$$
$$\hat{\mathbf{y}}_4 = (0.5 - \rho - \varepsilon, 0.5).$$

It holds that $\min_{j \in [4]} \big\{d(\mathbf{x}_i, \hat{\mathbf{y}}_j)\big\} \leq \rho + \varepsilon < \eta$ for all $i \notin I$ (i.e., for all $\mathbf{x}_i \in U''$). Next, select an $\mathbf{x}_{i^*}$ with $i^* \in I$. Without loss of generality, assume that $\mathbf{x}_{i^*}$ is located in the upper quadrant formed by the diagonals of U. Observe that

$$\min_{\mathbf{u} \in U'} \big\{d(\mathbf{x}_{i^*}, \mathbf{u})\big\} = d\big(\mathbf{x}_{i^*}, (0.5, 0.5 + \rho)\big). \tag{A2}$$

By (A1), the vertical coordinate of $\mathbf{x}_{i^*}$ is at least $0.5 + \rho + 2\varepsilon$. Then,

$$d(\mathbf{x}_{i^*}, \hat{\mathbf{y}}_1) = d\big(\mathbf{x}_{i^*}, (0.5, 0.5 + \rho + \varepsilon)\big) = d\big(\mathbf{x}_{i^*}, (0.5, 0.5 + \rho)\big) - \varepsilon \leq \eta - \varepsilon < \eta. \tag{A3}$$

Generalizing for all $i^* \in I$, we can conclude that $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4, \mathbf{y}_6, \ldots, \mathbf{y}_k\}$ is an improved solution to the problem. Because $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \hat{\mathbf{y}}_3, \hat{\mathbf{y}}_4 \notin U'$, this contradicts the optimality of $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ and proves the claim. Figure A2 illustrates the constructions in this perturbation argument.

It follows that

$$\begin{aligned}
\mathbb{P}\left(Y_1, \ldots, Y_k \in U'\right) &\leq \mathbb{P}\left(X_1, \ldots, X_n \in U''\right) && \text{(A4a)} \\
&= \left(0.25^{2n-1}\right)^n && \text{(A4b)} \\
&= \left(0.25^{2n}\right)^{n-\frac{1}{2}} && \text{(A4c)} \\
&< \left(0.25^{2n}\right)^k && \text{(A4d)} \\
&= \mathbb{P}\left(Z_1, \ldots, Z_k \in U'\right). && \text{(A4e)}
\end{aligned}$$

Therefore, the random vectors $(Y_1, \ldots, Y_k)$ and $(Z_1, \ldots, Z_k)$ have different joint distributions. □

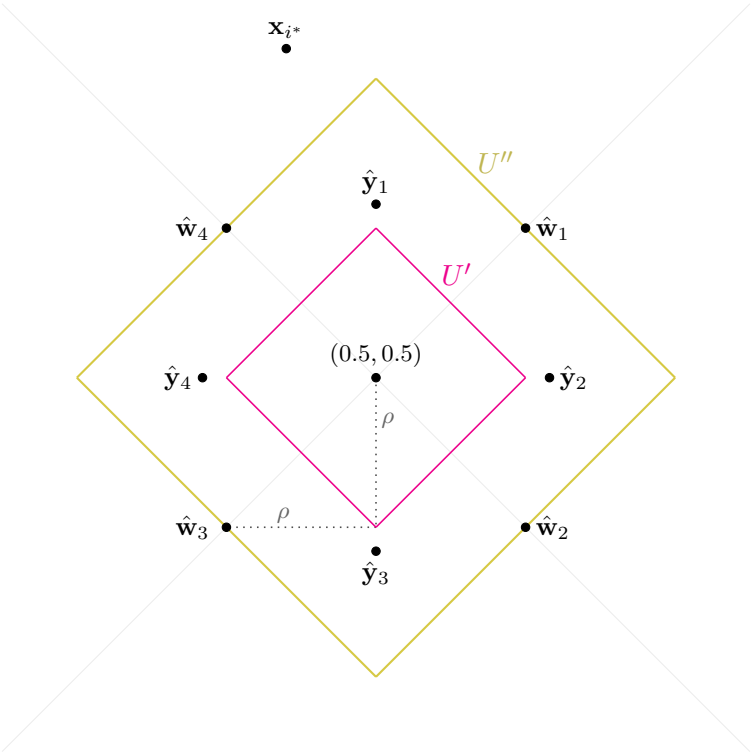


Figure A2: Illustration for proof of Proposition 2 (assuming IID uniform $Z_1, Z_2, \ldots$) with boundaries of U' and U'' marked

We can then use the same ideas to prove Proposition 2 for the remaining case in which $Z_1, Z_2, \ldots$ are chosen IID with respect to a continuous non-uniform distribution having support U.

Proof. Denote the non-uniform density function of Z_1 as $g(\cdot)$. By the non-uniformity and continuity of $g(\cdot)$, there exists some point v in the interior of U such that $g(v) > 1$. By continuity, there must exist a sufficiently large integer $\sigma \geq 2$ such that the closed rectilinear ball centered at v with area $\alpha = (2^{-2\sigma})^n$, denoted U', satisfies:

(i) $g(u) \geq 1$ for all $u \in U'$ and

(ii) $U'' \subset U$, where U'' is the closed rectilinear ball centered at v with radius twice that of U'.

An analogous perturbation argument as in the previous proof can be used to show that $H_n^k(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\} \subset U'$ implies $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset U''$.

Note that U'' has area $4\alpha = (2^{-2\sigma})^{n-\frac{1}{\sigma}}$. It follows that

$$\begin{aligned}
\mathbb{P}\left(Y_1, \ldots, Y_k \in U'\right) &\leq \mathbb{P}\left(X_1, \ldots, X_n \in U''\right) && \text{(A5a)} \\
&= \left((2^{-2\sigma})^{n-\frac{1}{\sigma}}\right)^n && \text{(A5b)} \\
&= \left((2^{-2\sigma})^n\right)^{n-\frac{1}{\sigma}} && \text{(A5c)} \\
&< \left((2^{-2\sigma})^n\right)^k && \text{(A5d)} \\
&\leq \mathbb{P}\left(Z_1, \ldots, Z_k \in U'\right). && \text{(A5e)}
\end{aligned}$$

Therefore, the random vectors $(Y_1, \ldots, Y_k)$ and $(Z_1, \ldots, Z_k)$ have different joint distributions. □

A.3 Proof of Proposition 3

Proof. Let U' denote the closed rectilinear ball centered at $(0.5, 0.5)$ with radius $\rho = \min\left\{\frac{a}{17}, c, \frac{1}{2}\right\}$. Let U'' denote the closed rectilinear ball centered at $(0.5, 0.5)$ with radius 2ρ.

Let $S = \{\mathbf{y}_1, \ldots, \mathbf{y}_{|S|}\}$ be an optimal solution to the problem with $|S| \geq 5$. Suppose for the purposes of contradiction that at least five elements of S lie within U'; without loss of generality, assume $\mathbf{y}_1, \ldots, \mathbf{y}_5 \in U'$. Define $I \subseteq [n]$ such that $i \in I$ if and only if the nearest facility to $\mathbf{x}_i$ is among $\{\mathbf{y}_1, \ldots, \mathbf{y}_5\}$. Then, it must hold that $\mathbf{x}_i \in U''$ for all $i \in I$. As shown in Figure A3, define:

$$\hat{\mathbf{y}}_1 = (0.5, 0.5 + \rho),$$

$$\hat{\mathbf{y}}_2 = (0.5 + \rho, 0.5),$$

$$\hat{\mathbf{y}}_3 = (0.5, 0.5 - \rho),$$

$$\hat{\mathbf{y}}_4 = (0.5 - \rho, 0.5).$$

Observe that each $\mathbf{x}_i$ for $i \in I$ is at most c distance from at least one point in $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4\}$. As such, $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4, \mathbf{y}_6, \ldots, \mathbf{y}_{|S|}\}$ is a feasible solution to the MDCRP.

Additionally, $d(\hat{\mathbf{y}}_i, \mathbf{y}_i) \leq \frac{2a}{17}$ for each $i \in [4]$. Thus, the total travel duration associated with $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4, \mathbf{y}_6, \ldots, \mathbf{y}_{|S|}\}$ is at most $2 \times 4 \times \frac{2a}{17} = \frac{16}{17}a$ greater than that of the original solution, while its total stopping time is a less than the original solution (due to the removal of the fifth stop). Therefore, the total route duration associated with $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_4, \mathbf{y}_6, \ldots, \mathbf{y}_{|S|}\}$ is at least $\frac{a}{17} > 0$ lower than that of $\{\mathbf{y}_1, \ldots, \mathbf{y}_{|S|}\}$, contradicting the optimality of $\{\mathbf{y}_1, \ldots, \mathbf{y}_{|S|}\}$.

It follows that, $\mathbb{P}(Y_1, \ldots, Y_5 \in U') = 0$, but $\mathbb{P}(Z_1, \ldots, Z_5 \in U') = \big[\mathbb{P}(Z_1 \in U')\big]^5 > 0$. Therefore, the joint distributions of $(Y_1, \ldots, Y_{|S|})$ and $(Z_1, \ldots, Z_{|S|})$ are different. □

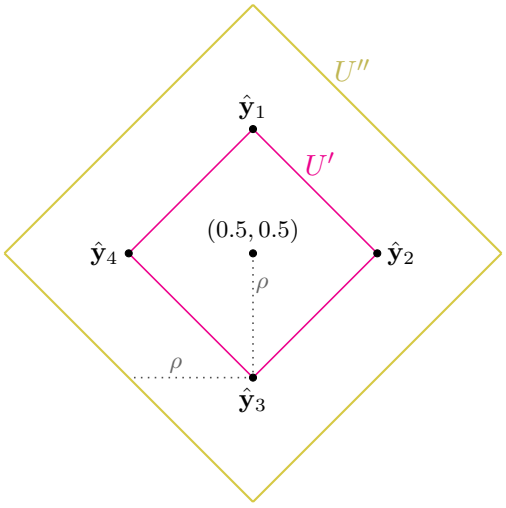


Figure A3: Illustration for proof of Proposition 3 with boundaries of U' and U'' marked

A.4 Estimating $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$

This appendix section provides additional technical details regarding the SBRP heuristic used to estimate $\mathbb{E}\big[L(n, A)\big]$ and $\mathbb{E}\big[T(n, A)\big]$ for a given combination of n and A. The following five steps correspond to those briefly summarized in Section 4.2.

Step 1. We randomly generate n IID uniform student locations in the square with vertices $(0, 0)$, $(0, \sqrt{A})$, $(\sqrt{A}, \sqrt{A})$, and $(\sqrt{A}, 0)$. Denote these student locations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$.

Step 2. We determine the minimum number of bus stops required to cover the student locations. Formally,

$$T(n, A) \coloneqq \min \quad |S| \tag{A6a}$$

$$\text{s.t.} \quad \min_{\mathbf{y} \in S} \{d\{\mathbf{x}_i, \mathbf{y})\} \leq c \qquad \forall i \in [n]. \tag{A6b}$$

Let $k = T(n, A)$ for notational simplicity.

Step 3. We minimize the sum of the k stops' maximum vertical and horizontal displacements relative to the origin. Let Δ represent the optimal value of this sum. That is, we solve

$$\Delta \coloneqq \min_{\mathbf{y}_1, \ldots, \mathbf{y}_k} \quad \left[\max_{j \in [k]} \left\{ y_1^j \right\} + \max_{j \in [k]} \left\{ y_2^j \right\} \right] \tag{A7a}$$

$$\text{s.t.} \quad \min_{j \in [k]} \{d(\mathbf{x}_i, \mathbf{y}_j)\} \leq c \qquad \forall i \in [n], \tag{A7b}$$

where each $\mathbf{y}_j = (y_1^j, y_2^j)$ represents a bus stop location in the square.

Step 4. We finalize the bus stop locations by minimizing the total distance between each student and their assigned stop:

$$\min_{\mathbf{y}_1, \ldots, \mathbf{y}_k} \quad \sum_{i=1}^{n} \min_{j \in [k]} \{d(\mathbf{x}_i, \mathbf{y}_j)\} \tag{A8a}$$

$$\text{s.t.} \quad \min_{j \in [k]} \{d(\mathbf{x}_i, \mathbf{y}_j)\} \leq c \qquad \forall i \in [n] \tag{A8b}$$

$$\max_{j \in [k]} \left\{ y_1^j \right\} + \max_{j \in [k]} \left\{ y_2^j \right\} \leq \Delta. \tag{A8c}$$

Step 5. Finally, we use a standard integer programming TSP formulation to determine the length $L(n, A)$ of the shortest open route that visits all of these finalized bus stop locations and ends at $(0, 0)$.

Steps 1–4 are also used to determine bus stop locations in the operational simulations, albeit in the entire service region instead of within a square.

Appendix B Computational Results and Data Preparation

B.1 Additional Data: Empirical Analysis

Table B1 provides the raw data associated with Figure 3 in Section 3.1. Two columns are included for each $n \in \{30, 50, 70\}$. The first column lists how many of the 1000 instances with n IID uniform demand

locations require $k \in \{7, 8, \ldots, 16\}$ covering facilities. The second column lists the average length of an open route over the origin and the k optimized facility locations for these instances. Note that the average route lengths are unreliable estimates for combinations of n and k with very few instances (e.g., $n = 50$ and $k = 15$); when five or fewer instances require k facilities for a particular n, the corresponding average route length is omitted from Figure 3. For comparison, the final column lists the average length (over 1000 trials) of an open route over the origin and $k \in \{7, 8, \ldots, 16\}$ facilities sampled IID from the uniform distribution. Table B2 provides the corresponding raw data associated with Figure 4.

Table B1: Comparison of open route lengths (in miles), covering facilities vs. IID uniform facilities

	Covering routes						IID uniform facilities
	$n = 30$		$n = 50$		$n = 70$		
# facilities	# instances	Avg. length	# instances	Avg. length	# instances	Avg. length	Avg. length
7	2	8.55	-	-	-	-	8.05
8	33	9.83	-	-	-	-	8.56
9	200	10.39	-	-	-	-	9.14
10	387	11.27	17	10.78	-	-	9.58
11	298	11.99	130	11.61	3	12.00	10.01
12	73	12.78	412	12.40	41	12.15	10.45
13	6	13.60	372	13.07	322	12.81	10.86
14	1	13.30	67	13.66	488	13.61	11.30
15	-	-	2	14.07	137	14.28	11.68
16	-	-	-	-	9	14.87	12.00

Table B2: Comparison of closed route lengths (in miles), covering facilities vs. IID uniform facilities

	Covering routes						IID uniform facilities
	$n = 30$		$n = 50$		$n = 70$		
# facilities	# instances	Avg. length	# instances	Avg. length	# instances	Avg. length	Avg. length
7	2	10.95	-	-	-	-	10.99
8	33	12.25	-	-	-	-	11.38
9	200	12.65	-	-	-	-	11.84
10	387	13.43	17	12.97	-	-	12.20
11	298	14.09	130	13.75	3	14.30	12.54
12	73	14.75	412	14.44	41	14.17	12.92
13	6	15.48	372	15.06	322	14.79	13.25
14	1	15.10	67	15.61	488	15.51	13.62
15	-	-	2	15.95	137	16.11	13.96
16	-	-	-	-	9	16.74	14.24

B.2 Additional Data: Planar Studies

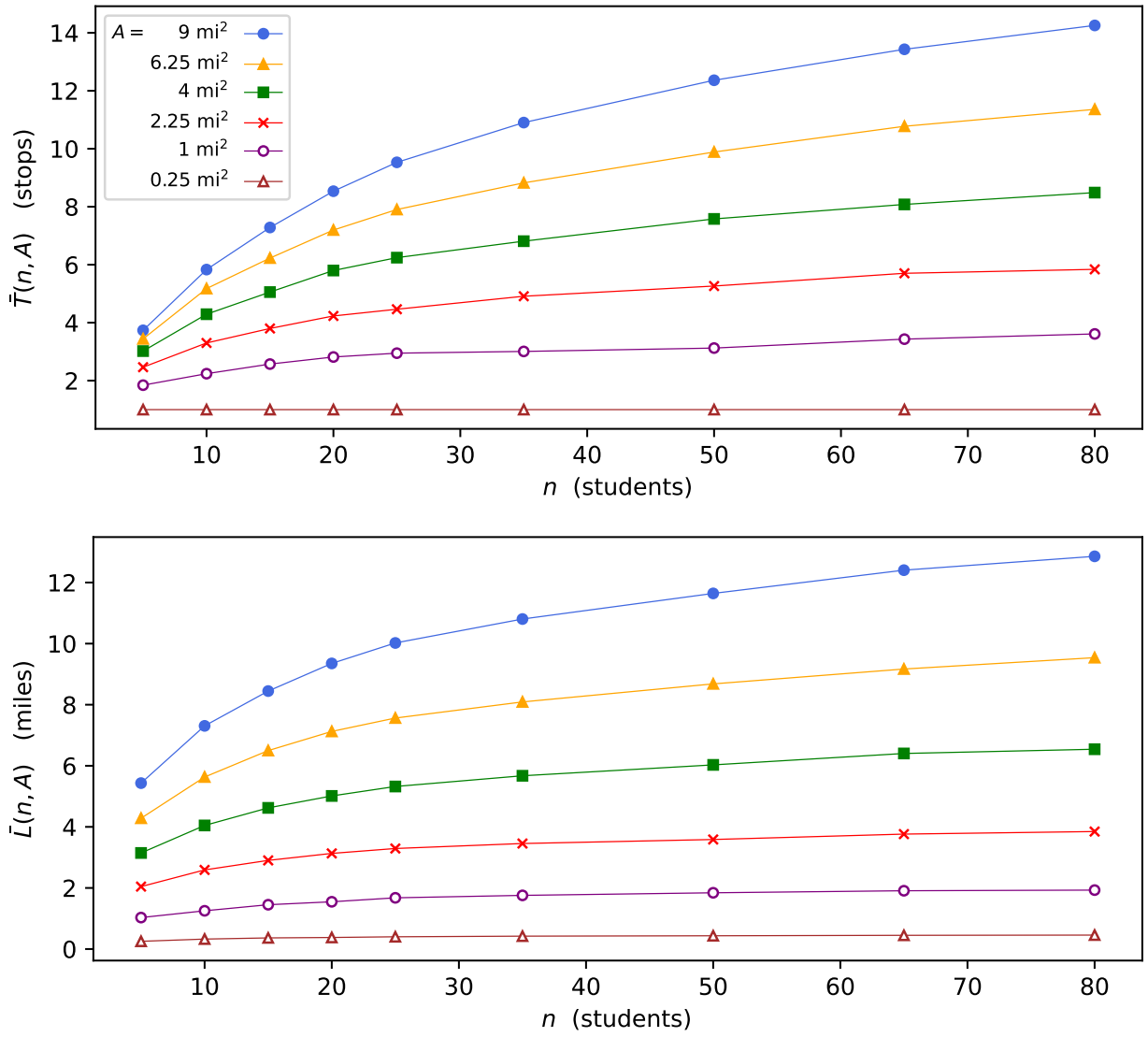
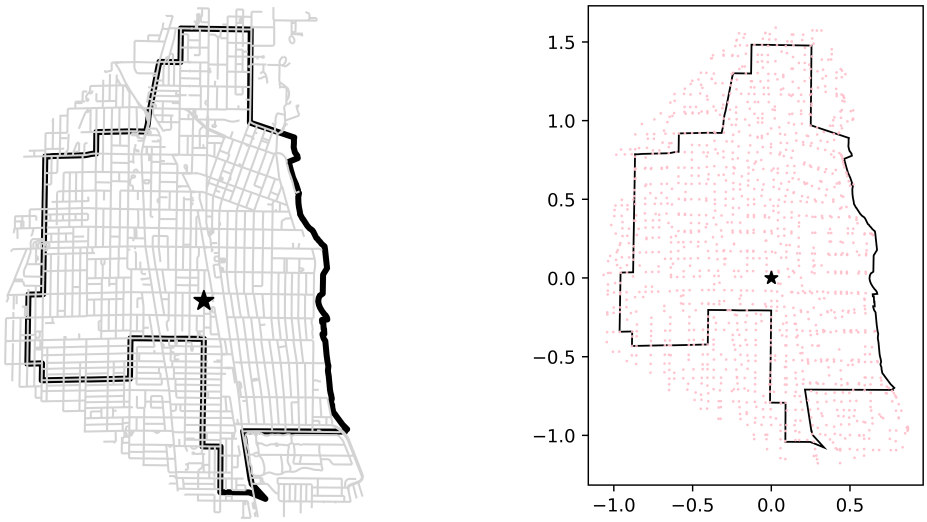


Figure B1: Empirically estimated $\bar{T}(\cdot)$ and $\bar{L}(\cdot)$ functions for planar studies

B.3 Data Preparation: Case Study #1

This section details the steps taken to prepare the data for the first case study. We attempted to mimic the planar CA approach as closely as possible throughout, but some adaptations were necessary to reflect practical aspects of school bus routing on real-world road networks.

We began by querying the full road network of the attendance region and a surrounding buffer using OSMnx. Figure B2a depicts this road network with the attendance region boundary and school location



(a) Road network **(b)** Potential stop locations (axes in miles)

Figure B2: S-1 road network structure and potential stop locations

marked in black. The buffer does not extend to the region's east due to an impassable geographical barrier. The network's graph representation contains approximately 5400 nodes, including both intersections and locations at which streets bend. We designated a subset of these nodes as potential bus stop locations by sampling a node every 0.05 miles in both the horizontal and vertical directions. The resulting set of potential stop locations, depicted in Figure B2b, includes both intersections and non-intersections.

Next, we used OSMnx to construct isopleth polygons for each potential bus stop location. The primary isopleths correspond to the quarter-mile coverage radius. These were used to quickly determine whether a randomly generated student location is within walking distance of a particular bus stop, as checking whether a given point is within a given polygon is computationally efficient. We then constructed smaller isopleths for each potential stop location with radii of 0.05, 0.1, 0.15, and 0.2 miles. These were used to quickly estimate distances between a student and a covering stop location. For example, if a student is inside a particular stop's 0.2-mile isopleth but not inside its 0.15-mile isopleth, we would conservatively estimate the student-to-stop distance as 0.2 miles. Figure B3 illustrates this idea. The entire shaded area represents the 0.2-mile isopleth of the bus stop location marked in black, while the yellow subset represents its 0.15-mile isopleth. Under our estimation scheme, the hypothetical student location marked in red (or

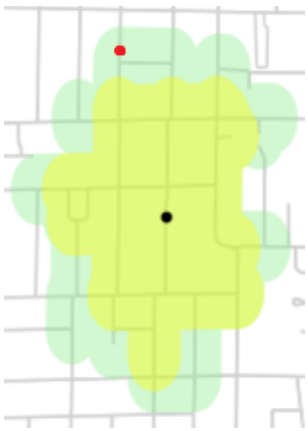


Figure B3: Distance approximation via isopleth polygons

any other student in the green area) is designated as being 0.2 miles from the bus stop. Finally, we also constructed a half-mile isopleth centered at the school and removed it from the attendance region (Figure 15a) to create S-1's service region (Figure 15b).

We next constructed the estimator functions for the CA model. In the planar case, we estimated bus stop needs and route lengths by generating random locations in arbitrary squares. This posed no issue due to the homogeneity of metric travel in planar regions. Real road networks, of course, do not necessarily entail homogeneity or metric distances. To mimic the planar estimation process, we selected a square with sides of length 1.35 miles from the buffered road network. The location of this square, shown in Figure B4, was chosen (i) for centrality and (ii) to allow for estimates with A as large as possible while remaining within the buffered network's boundary.

Within this larger square, we also considered smaller concentric squares with side lengths 0.55, 0.75, 0.95, and 1.15 miles. For each combination of side length and $n \in \{5, 10, 15, 20, 25, 35, 50, 65, 80, 95\}$, we generated 100 instances of n student locations uniformly at random in the square, then heuristically solved each instance by the procedure described in Section 4.2. A key difference from the planar setting was that routing optimization was conducted with respect to travel times, which are not necessarily linearly proportional to distances due to potentially varying speed limits on the road network. Therefore, the resulting empirical estimates can be viewed as analogous to $\bar{T}(n, A)$ and $s^{-1}\bar{L}(n, A)$. Figure B5 plots these empirical estimates.

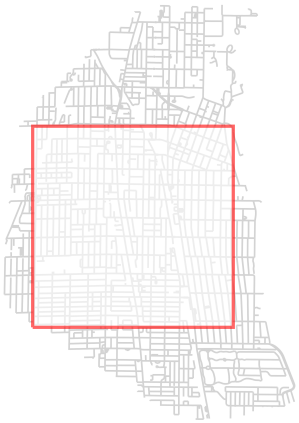


Figure B4: Location of 1.35 $\times$ 1.35 mile square used for routing parameter estimation

Lastly, a generalization of the planar linehaul travel time calculation was required. When a zone's center is sufficiently far from the school, estimating the linehaul travel time proceeded as in Figure 8 with road network travel times replacing rectilinear travel times. Otherwise, for nearer zones, the linehaul travel time was estimated as 1.8 minutes; this corresponds to the time taken to travel $w = 0.5$ miles at two-thirds of the region's legally defined maximum speed limit for residential streets (25 miles per hour). Thus, for an arbitrary zone, the linehaul travel time is the maximum of these two values.

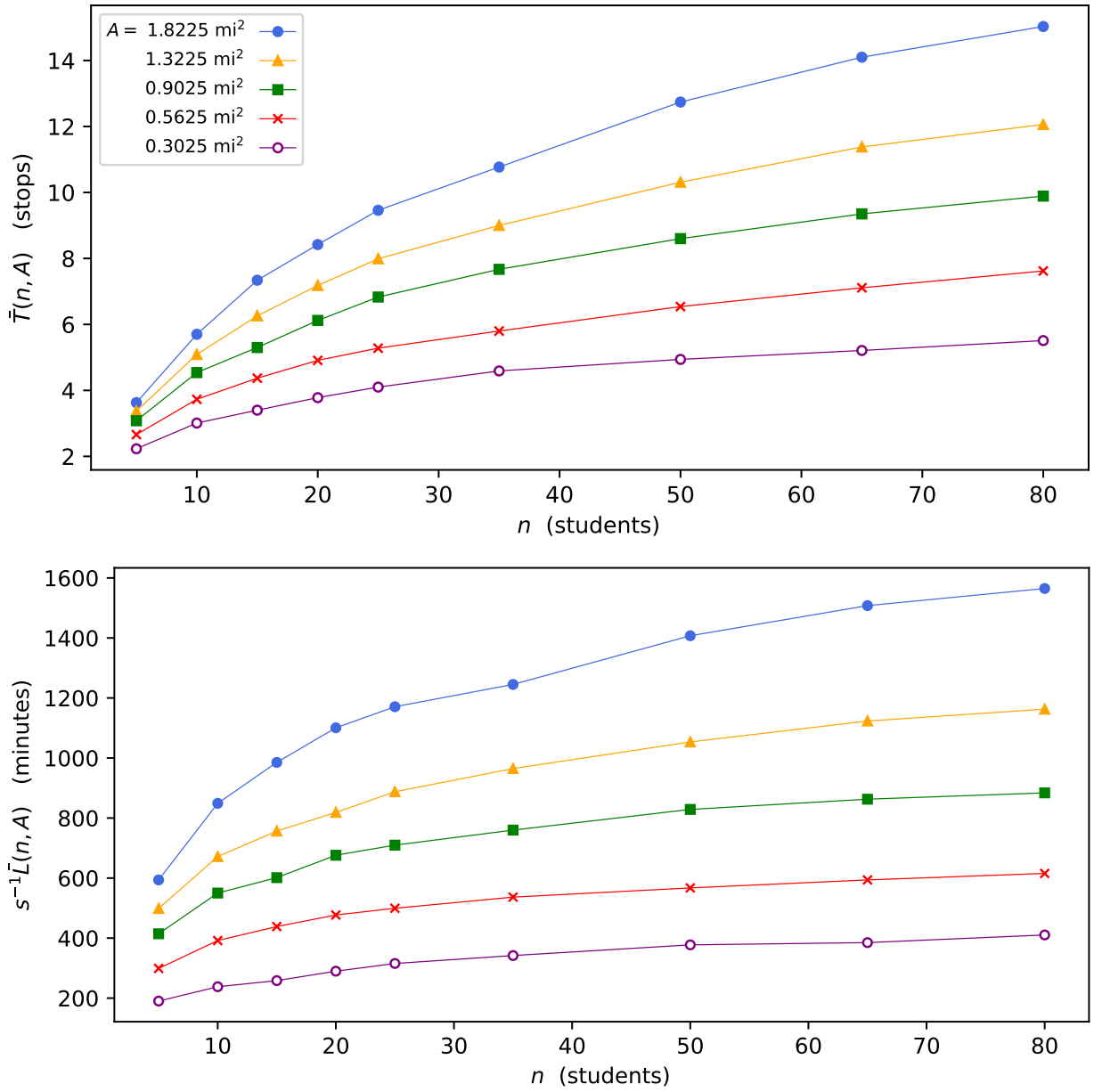
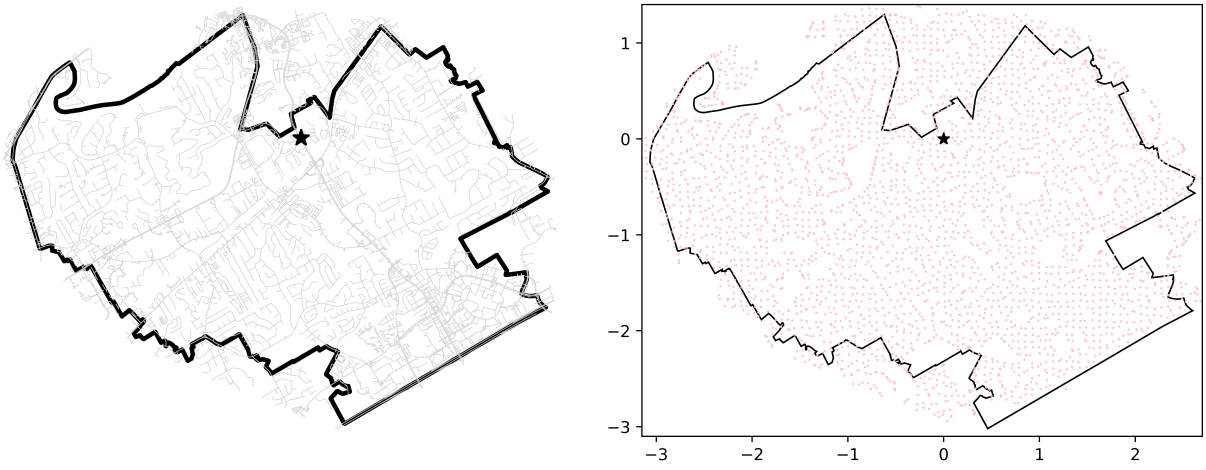


Figure B5: Empirical estimator functions for first case study

B.4 Data Preparation: Case Study #2

The same steps were followed with some minor differences to prepare data for the second case study. The queried road network for S-2 contains approximately 26600 nodes, which led to significantly increased memory and computational requirements for travel time matrix computation. To mitigate this, potential bus stop locations were selected approximately every 0.07 miles in both the horizontal and vertical directions; this entailed about half as many locations per square mile compared to the first case study. Figure B6 depicts

the road network and potential bus stop locations. Note the irregularity of the network and potential bus stop locations relative to those of S-1.



(a) Road network **(b)** Potential stop locations (axes in miles)

Figure B6: S-2 road network structure and potential stop locations

Given the size and shape of S-2's service region, we were able to designate a sufficiently large square within the service region for estimating $\bar{T}(n, A)$ and $s^{-1}\bar{L}(n, A)$. This square's location is depicted in Figure B7; smaller concentric squares with side lengths 0.6, 1, 1.4, and 1.8 miles were used to produce estimates for smaller values of A. Figure B8 plots the empirical estimator functions. Observe the eccentric behavior of the plots corresponding to $A = 1$ and $A = 1.96$ square miles; see Section 5.4 for further discussion of this issue.

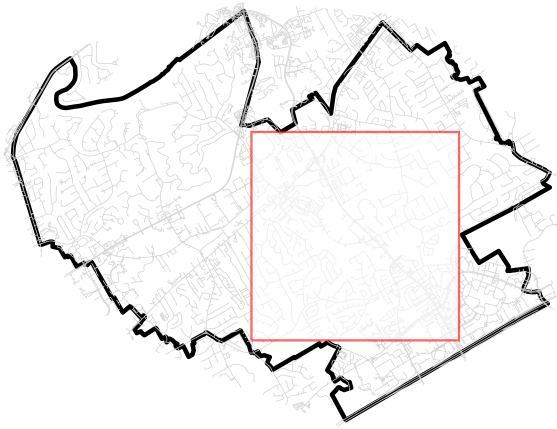


Figure B7: Location of 2.2 $\times$ 2.2 mile square used for routing parameter estimation

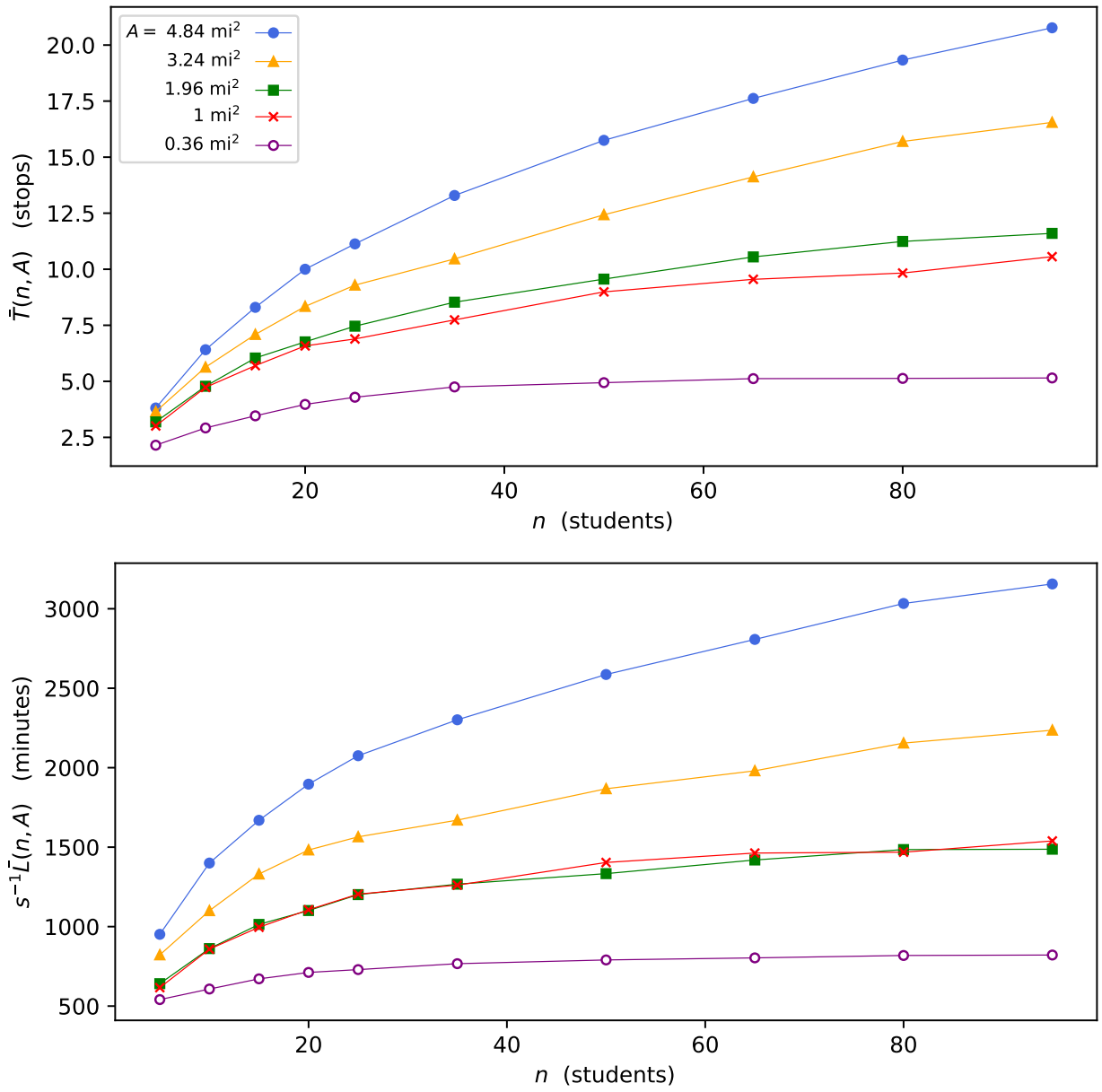


Figure B8: Empirical estimator functions for second case study

When a zone's center is sufficiently far from the school, estimating the linehaul travel time again proceeds as in Figure 8 with road network travel times replacing rectilinear travel times. Otherwise, for nearer zones, the linehaul travel time is estimated as 4.5 minutes; this corresponds to the time taken to travel $w = 1.5$ miles at two-thirds of the region's legally defined maximum speed limit for residential streets (30 miles per hour). Thus, for an arbitrary zone, the linehaul travel time is the maximum of these two values.