Article

Congestion-Aware Bi-Modal Delivery Systems Utilizing Drones

Mark Beliaev [1,*], Negar Mehr [2] and Ramtin Pedarsani [1]

1   Electrical & Computer Engineering, University of California Santa Barbara, Santa Barbara, CA 93106, USA
2   Aerospace Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA
*   Correspondence: mbeliaev@ucsb.edu

Abstract: With e-commerce demand rising, logistic operators are investing in alternative delivery methods such as drones. Because of their aerial reach, drones can provide much needed utility in the last mile by taking the load off of vehicles delivering parcels to customers on the road. Our goal is to assess the potential drones have in mitigating traffic congestion. To do so, we develop a mathematical model for a bi-modal delivery system composed of parcels carrying trucks and drones, combining it with an optimization problem that can be solved for the socially optimal routing and allocation policy efficiently. Within this formulation, we include multiple stakeholder perspectives by modeling the objective function in terms of both traffic congestion and parcel latency. This allows our model to quantify the impact of drones on reducing traffic congestion, and simultaneously finds the path routing that minimizes the given objective. To account for the effects of stopping trucks on road latency, we simulate roads shared between trucks and cars by utilizing SUMO. We then use quadratic optimization techniques to test our proposed framework on a variety of real-world transportation networks. Our findings highlight the trade-off between reducing traffic congestion and increasing parcel latency—while routing trucks along less time-efficient paths may alleviate traffic congestion, this disproportionately increases the parcel latency. This suggests the need for a balanced approach that considers both factors when solving for the routing policy.

Keywords: traffic congestion; air transportation; logistics; computational modeling; optimization

1. Introduction

As the need for e-commerce continues to grow, last mile delivery becomes an ever more present bottleneck in the supply chain [1]. E-commerce sales have increased at a 15% annual rate over the past decade, with a notable 40% surge in 2020 [2]. Although this surge may be attributed to the pandemic, US e-commerce showed a 49% year-over-year increase in March 2021 [3], pointing to a paradigm shift in consumer trends. This particularly affects last mile logistics, as delivery in the last mile accounts for 28% of a product's value [4]. Increase in demand for e-commerce has placed additional pressure on retailers, especially when it comes to meeting customer's growing expectations in the last mile.

With consumers relying more on online orders, we observe a significant rise in delivery truck traffic, affecting travel times for both private and public vehicles inside densely populated urban areas [5,6]. Looking specifically at the last mile, this steady growth of urban freight flow has many other negative impacts [7]. Specifically, increase in demand for freight transport is associated with traffic accidents, property damage, delivery delay, increased emissions, parking shortages, and noise pollution [8–10]. One case study investigating logistic operations in London underlines a challenge facing the last mile: as light goods vehicle traffic grows, central London's traffic congestion is expected to increase 60% by 2031 [11]. To tackle these problems, researchers have been investigating bi-modal delivery systems that include drones [12].

Unmanned Aerial Vehicles, or drones, are advantageous for last mile logistics networks because of their aerial reach and innate agility [13]. Europe has made significant progress towards incorporating drones within their airspace system [14]. In the United States, the Federal Aviation Administration has approved the use of drones based on environmental assessments for Zipline, Wing Aviation, and UPS Flight Forward [15]. Additionally, large courier companies such as Amazon and DHL have launched programs that aim to deliver parcels using drones in the near future [16,17]. Although these examples highlight the popularity of this emerging technology, there is a lot to uncover regarding the sustainability, impact, and effectiveness of implementing drones within the last mile.

Given the growing demand of e-commerce, we need to pay additional attention to the impacts it has on last mile logistics in cities [18,19]. A study looking at 48 public sector initiatives for improving freight transport highlights the lack of scientific research that assesses the outcomes of such strategies and the trade-offs involved [20]. In fact, a recent review investigating the literature around last mile logistics found that receivers and shippers were the most discussed group of stakeholders, while the perspectives of residents were the least discussed topic [21]. Hence, to effectively implement drones in the last mile, we need to assess their impact from multiple viewpoints.

To achieve this goal, we set out to estimate how much of a difference drones could make in reducing road congestion brought on by package delivery services. Inspired by our prior work [22], we formulate a mathematical model for a bi-modal parcel delivery system made up of trucks and drones. Alongside this model, we present an optimization framework which solves for the socially optimal routing and allocation policies. In our optimization formulation, we want to include the societal cost of traffic congestion induced by the delivery system as well as the overall latency experienced by customers.

To accomplish this, we suggest a network model that consists of two distinct graphs for the road and aerial networks as displayed in Figure 1. Our model's capacity to account for the societal effects of various routing and allocation strategies on the road network as well as the resulting delivery time of packages is the key feature which makes it congestion aware. This approach allows us to formulate our objective function as a trade-off between societal cost and parcel latency, thus yielding an optimization problem that can be solved to determine the socially-optimal routing strategy.

In order to capture how stopping trucks affect latency on individual links, we simulate roads shared between trucks and cars by utilizing SUMO [23], an open source package for traffic simulation. We use the results from these simulation studies to construct a linear model for link latency, which results in our primary objective function having a quadratic relationship with respect to the flow of stopping and non-stopping vehicles. In this work, we expand on our prior analyses by more accurately considering how stopping trucks affect the road, simulating new road configurations, generalizing our latency model, and performing additional case studies on realistic transportation networks.

We list the paper's primary contributions:

- Using a bi-modal delivery system consisting of parcel carrying trucks and drones, as well as regular vehicles, we develop a quadratic optimization framework to solve for the socially-optimal routing and allocation scheme;
- We use SUMO simulations to study the impact stopping trucks have on road capacity and latency, deriving a mathematical latency function for roads shared by stopping trucks and regular vehicles;
- By performing numerical evaluations on the trade-off between societal and parcel latency, we quantify the potential drones have for alleviating traffic congestion.
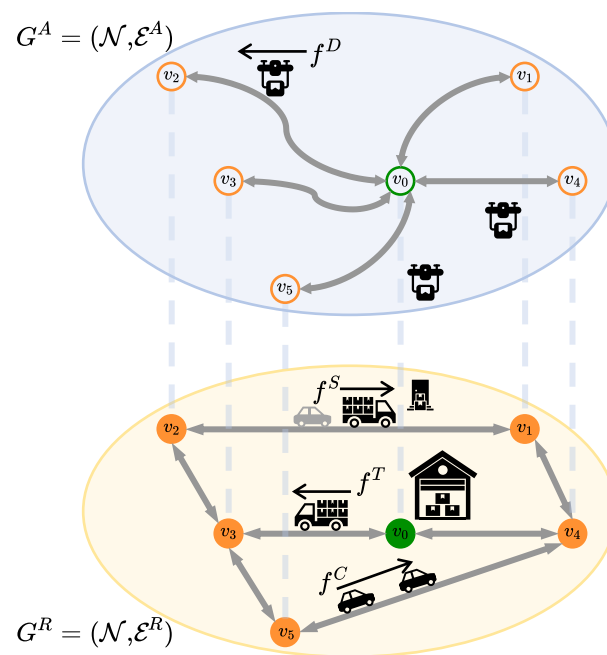
Figure 1. An illustration of a bi-modal delivery system that makes use of drones and trucks. A central depot depicted by node v_0 is responsible for dispatching parcels to the remaining nodes via trucks f^T and drones f^D. Trucks operate within the road network denoted by graph G^R, while drones operate in the separate aerial network denoted by graph G^A. We see stopping trucks f^S making deliveries and affecting passenger cars f^C that share the road. Alternatively, drones f^D take paths independent of the road network, and hence do not contribute negative effects to road congestion.

2. Related Work

In the following section, we briefly go through two areas of research that are connected to this work. We first go over works that consider delivery systems utilizing both trucks and drones. Afterwards, we look into research that studies the impact stopping vehicles have on road congestion.

2.1. Bi-Modal Drone Delivery Systems

The use of drones in delivery systems has drawn a lot of attention [24]. In the past decade, several works have provided theoretical contributions to the field by extending well-known combinatorial optimization (CO) problems, namely the Traveling Salesman Problem (TSP) and the more general Vehicle Routing Problem (VRP). The objective of a TSP is to find the shortest route for one traveler given a list of destinations and corresponding distances between them. The VRP extends this problem by considering a fleet of multiple travelers, a dilemma that is common for logistic operators concerned with delivering goods to many customers via trucks. One work formalized extensions of the TSP by considering scenarios where a drone works in collaboration with a delivery truck to distribute packages [25]. A subsequent work considered drones in the VRP setting [26], deriving multiple worst-case results that quantify potential savings obtained by incorporating drones. Following this, many works have dealt with variants of these formulations, where we point the reader to a thorough literature review [12]. In most of these works, the contribution arises from solving the CO problem, where one must typically rely on a combination of integer linear programming and heuristics.

Unlike the CO problems discussed thus far, which focus on minimizing parcel latency, our work additionally considers societal latency experienced by private drivers. Formulating a quadratic objective function posed as a trade-off between parcel latency and societal latency, our framework efficiently solves for a socially optimal routing scheme, allowing us to focus on the effect drones have on road congestion and overall parcel latency. Although the works mentioned thus far also deal with the routing of drones, their optimizations are concerned with the viewpoints of receivers and senders, individually optimizing for travel time, system cost, or total delivery volume. While our work deals with a simpler formulation compared to the CO problems mentioned, we consider the viewpoints of private drivers alongside receivers and senders, allowing us to study the impact drones have on mitigating traffic congestion from multiple stakeholder perspectives.

2.2. Impact of Stopping Vehicles on Road Congestion

As mentioned in the previous section, so far, there have been no works that consider the effect of replacing parcel carrying trucks with drones when it comes to traffic congestion experienced by private drivers. On the other hand, the study of urban freight transport and its negative impact is not a new research topic [27]. One point of view considers the effect of vehicle loading bays, where a recent work used a new cellular automation model to show that vehicles leaving and entering the loading bay have a negative impact on adjacent lane traffic [28]. In accordance, simulations performed based on real traffic data show that the effective capacity of the considered road is reduced when even a small amount of stopping buses are introduced [29]. When it comes to last mile logistics, loading bays are seen as an inefficient solution due to their high implementation cost, though they can provide societal benefits for residents [30]. As it is more common for delivery vehicles to stop curbside, we do not designate loading bays for our stopping vehicles.

In contrast to loading bays, allowing delivery vehicles to park directly at the curb requires no additional infrastructure. The effects of street parking have also been studied using a cellular automation model [31], where it was shown that traffic congestion increases as more vehicles are allowed to temporarily park curbside. We also note an interesting conclusion from an earlier simulation study that states that the influence of parking space occupancy on vehicle delay is more severe on narrow roads as compared to wide roads [32]. Unlike the cellular automation models used in both of these works, our studies rely on the open source traffic simulation package SUMO [23], where we directly simulate roads with different numbers of lanes, stopping vehicles, and regular vehicles. This allows us to construct a model for road latency, which is a key component to our overall optimization framework concerned with finding socially optimal routing policies within bi-modal delivery systems.

3. Materials and Methods

3.1. Network Model

The present study employs a dual network architecture to model the various delivery mechanisms, with no interconnections between the two networks. We let graph $G^R = (\mathcal{N}, \mathcal{E}^R)$ denote the road network and graph $G^A = (\mathcal{N}, \mathcal{E}^A)$ denote the aerial network, where $\mathcal{N}$ represents a shared set of nodes, while $\mathcal{E}^R$ and $\mathcal{E}^A$ represent two disjoint sets of edges. Each edge $e \in \mathcal{E}^R, \mathcal{E}^A$ connects two different nodes (v, w), indicating a directed edge from v towards w. It is assumed that both graphs are connected. We label node v_0 as the primary delivery hub, leaving all remaining nodes as potential destinations. Within the aerial network G^A, each edge $e \in \mathcal{E}^A$ is constrained to originate from node v_0, as the versatility of drones enables them to have unique routes between the main delivery hub and destinations. We illustrate the two networks in Figure 1 and provide Table 1 as a reference for notations defined.

Table 1. Notations.

		Network Model
$\mathcal{N}$	$\triangleq$	Set of nodes
$v \in \mathcal{N}$	$\triangleq$	Individual node
v_0	$\triangleq$	Delivery hub node
$\mathcal{E}^R, \mathcal{E}^A$	$\triangleq$	Sets of edges on the road and aerial network
$e \in \mathcal{E}^R, \mathcal{E}^A$	$\triangleq$	Individual edge
G^R, G^A	$\triangleq$	Road and aerial graphs, respectively
$\mathcal{P}$	$\triangleq$	All paths inside the road network
$\mathcal{P}_v$	$\triangleq$	Subset of paths incoming at node v
$p \in \mathcal{P}$	$\triangleq$	Individual path
e_p^-	$\triangleq$	Last edge along path p
$\mathcal{O}_v$	$\triangleq$	Set of outgoing edges at node v
$\mathcal{I}_v$	$\triangleq$	Set of incoming edges at node v
		Demand ($\frac{parcels}{hour}$)
d_v	$\triangleq$	Total parcel demand at node v
d_v^R	$\triangleq$	Parcel demand *satisfied* by trucks at node v
d_v^A	$\triangleq$	Parcel demand *satisfied* by drones at node v
		Flow ($\frac{vehicles}{hour}$)
$x_p \in \mathbf{x}$	$\triangleq$	Vector of truck flows along all paths $p \in \mathcal{P}$
f_e^T	$\triangleq$	Edge flows of trucks
f_e^S	$\triangleq$	Edge flows of stopping trucks
f_e^C	$\triangleq$	Edge flows of nominal vehicles
f_e^0	$\triangleq$	Capacity of edge e, used to compute edge latency
f_v^A	$\triangleq$	Flow of drones on aerial path incoming at v
		Latency (*hours*)
L	$\triangleq$	Average parcel latency
L^S	$\triangleq$	Average societal latency
ℓ_e^R	$\triangleq$	Latency of edge e in the road network
ℓ_e^C	$\triangleq$	Latency of edge e without trucks
ℓ_e^0	$\triangleq$	Free flow latency on edge e
ℓ_v^A	$\triangleq$	Latency of aerial path arriving at v
		Parameters
m	$\triangleq$	Number of parcels carried by each delivery truck
β	$\triangleq$	Total nominal flow
C	$\triangleq$	Operational cost for running the entire system
c^T, c^D	$\triangleq$	Cost per truck and drone, respectively
γ	$\triangleq$	Trade-off between parcel and societal latency
ω_e	$\triangleq$	Vector used to compute latency on edge e

We introduce the set of paths $\mathcal{P}$ which is specific to the road network, where $p \in \mathcal{P}$ denotes a path within G^R. The notation of paths for the aerial network is omitted as each path would only contain one edge, given that G^A is a star network. In the road network, each path $p \in \mathcal{P}$ connects the delivery hub v_0 to one of the other nodes using a collection of edges $e \in \mathcal{E}^R$, where repeated edges are not permitted. The subset of paths that terminate at node v is denoted as $\mathcal{P}_v \subset \mathcal{P}$, and the last edge of a path p is defined as e_p^-. Additionally, for each node $v \in \mathcal{N}$, we define the subsets $\mathcal{O}_v \subset \mathcal{E}^R$ and $\mathcal{I}_v \subset \mathcal{E}^R$ which represent the set of edges that originate and terminate at node v, respectively.

To setup the road network, for each $e \in \mathcal{E}^R$, we define the capacity f_e^0 and free flow latency ℓ_e^0 in units of cars per hour and hours, respectively. For the aerial network, we directly define the latency as ℓ_v^A for each node $v \in \mathcal{N}$, as there is only one aerial path for

each destination node. We will discuss these parameters further in the following sections, as they will be used to compute latency in our traffic model.

Finally, we define the decision variable used in our network model. Since we set out to optimize over a set of feasible routing strategies, we define our path flows as the decision variable $\mathbf{x}$, where the elements $x_p \in \mathbf{x}$ correspond to the flow of parcel-carrying trucks along a specific path $p \in \mathcal{P}$ in units of vehicles per hour. For the rest of this section, we describe how the parcel demand, vehicular flow, latency, and societal cost are derived in our model using the decision variable $\mathbf{x}$.

3.1.1. Demand

For each node $v \in \mathcal{N}$, we define the parcel demand d_v in units of parcels per hour. In reality, we use the set $\mathcal{N} \setminus \{v_0\}$ because the delivery hub does not have any demand, but to avoid cumbersome notation, we leave this distinction for the reader. To ensure the delivery of parcels at all destinations, for each node $v \in \mathcal{N}$, the following condition must be met:

$$d_v = d_v^R + d_v^A, \tag{1}$$

where d_v^R and d_v^A correspond to the components of node v's parcel demand satisfied by trucks within the road network and drones within the aerial network, respectively.

For the road network, given our decision variable $\mathbf{x}$, we can calculate the satisfied demand d_v^R at node v by summing the truck flows along all paths p incoming at node v:

$$d_v^R = m \sum_{p \in \mathcal{P}_v} x_p, \tag{2}$$

where m is the number of parcels each truck carries. Note that we are only considering paths that arrive at node v, which is not equivalent to the total truck flow incoming at node v. In addition, since our paths are predefined, we can compute (2) as a linear transformation over our decision variable $\mathbf{x}$.

For the aerial network, we set the parcel demand satisfied by drones d_v^A at all nodes $v \in \mathcal{N}$ equal to the leftover parcel demand required by our system after considering trucks:

$$d_v^A = d_v - d_v^R. \tag{3}$$

We want to emphasize that setting d_v^A as the *leftover* parcel demand is a key design choice as it decouples the two networks, allowing us to formulate a quadratic objective function to find the optimal routing scheme as we will show in Section 3.2.

3.1.2. Flow

Although the demands are given in units of parcels per hour, the corresponding flow rate for individual paths and edges is presented in units of vehicles per hour. Given the decision variable $\mathbf{x}$, we can define f_e^T as the flow of trucks along a given edge $e \in \mathcal{E}^R$ using:

$$f_e^T = \sum_{p \in \mathcal{P}:e \in p} x_p, \tag{4}$$

where we sum truck flows x_p over the paths $p \in \mathcal{P}$ which contain the corresponding edge e. Note that, since the road network G^R is predefined, we can compute (4) as a linear transformation over our decision variable $\mathbf{x}$.

In addition to having a designated truck flow f_e^T, each edge $e \in \mathcal{E}^R$ has a corresponding flow of stopping trucks making deliveries f_e^S. To calculate this flow, we consider that trucks along paths $p \in \mathcal{P}_v$ delivering to node v will stop along their final edge e_p^- as well as all

edges $e \in \mathcal{O}_v$ outgoing at node v. This way, given an edge $e = (v, w)$ connecting node v to w, we can calculate the flow of stopping trucks f_e^S:

$$f_e^S = \sum_{p \in \mathcal{P}_v} \frac{x_p}{1 + |\mathcal{O}_v|} + \sum_{\{p \in \mathcal{P}_w | e_p^- = e\}} \frac{x_p}{1 + |\mathcal{O}_w|}, \tag{5}$$

where $|\mathcal{O}_v|$ corresponds to the number of edges outgoing at node v, and the denominator states that the sum of paths arriving for delivery will divide their stops equally around the one stopping edge and $|\mathcal{O}_v|$ outgoing edges. Essentially, (5) sums the contribution of stopping trucks given by paths incoming at node v, and paths using edge e to deliver at node ω. This formulation lets us capture the notion that trucks will spread their deliveries around the destination node.

The last vehicular flow we consider on the road network is that of non-delivery vehicles or cars f_e^C for each edge $e \in \mathcal{E}^R$. We can think of f_e^C as the nominal vehicular flow along the edge, to outline that it is not affected by our decision variable **x**, which determines the flow of trucks along the road network.

In regard to the aerial network, we define f_v^A as the flow of drones on the aerial path linking the central delivery hub v_0 to the destination node $v \in \mathcal{N}$. Unlike trucks, we assume that each drone can deliver one package at a time. Given aerial path $e \in \mathcal{E}^A$ and its corresponding destination node $v \in \mathcal{N}$, we set the flow of drones f_v^A equivalent in magnitude to the parcel demand d_v^A assigned to drones: $f_v^A = d_v^A$. This way we can use (2) and (3) to derive the flow of drones f_v^A directly as the leftover parcel demand required after considering truck deliveries.

3.1.3. Parcel Latency

To determine the average latency encountered by parcels within the system, we sum over all contributions to parcel latency and divide by the overall demand:

$$L = \frac{1}{\sum_{v \in \mathcal{N}} d_v} \left[\sum_{p \in \mathcal{P}} m x_p \left(\sum_{e \in p} \ell_e^R \right) + \sum_{v \in \mathcal{N}} f_v^A \ell_v^A \right], \tag{6}$$

where the two terms added inside the brackets represent the contributions to latency from parcels delivered via the road network and the aerial network, respectively. By dividing the overall sum by the total demand, as specified in (6), we obtain the desired average parcel latency measured in units of hours. We proceed to explicate how these two components are derived.

To determine the latency experienced by parcels in the road network, we first compute the latency for each individual edge, denoted as ℓ_e^R for all edges $e \in \mathcal{E}^R$. Once these latencies have been calculated, we compute the latency along paths $p \in \mathcal{P}$ by summing the edge latencies along the path, i.e., $\sum_{e \in p} \ell_e^R$. These path latencies are then weighted by the flow of parcels $m x_p$ on that path. As will be demonstrated in Section 4.1, the edge latency ℓ_e^R is modeled as a linear function of the flow components on that edge, resulting in the overall expression for L being quadratic with respect to the decision variable **x**.

For the aerial network, the flow of drones f_v^A has no impact on the latency of aerial paths, as they will not cause any aerial congestion in the regimes of interest. As such, we model the latency of drones routed to the delivery node v as a constant value ℓ_v^A predetermined for each node $v \in \mathcal{N}$. Similar to the road network, these latencies are weighted by the flow of parcels along the same path f_v^A.

3.1.4. Societal Cost

Due to their frequent stops, delivery trucks have a negative impact on traffic congestion in dense urban areas [6,33–36]. To measure this impact, we look at the average latency

experienced by the nominal flow of traffic f^C, defining this as the societal cost L^S in units of hours:

$$L^S = \frac{1}{\beta} \sum_{e \in \mathcal{E}^R} f_e^C \ell_e^R, \tag{7}$$

where the value obtained is normalized using the parameter β, which can be interpreted as the total flow of nominal vehicles f^C within the road network. In order to give higher consideration to roads that have more affected drivers, we weight the edge latencies ℓ_e with their corresponding car flows f_e^C inside the summation of (7). When referring to L^S, we use the terminology "societal latency" and "societal cost" interchangeably.

3.1.5. Cost

Lastly, the cost of operating this delivery system is defined as C in units of dollars per hour, allowing us to directly restrict the usage of trucks and drones within our system. The cost consists of two components, one associated with the road network and the other associated with the aerial network. To compute these components, we define c^T as the cost per truck and c^D as the cost per drone. Using the total hourly demand of trucks and drones at node v, represented by $\frac{1}{m} d_v^R$ and d_v^A, respectively, we can derive the total cost:

$$C = \frac{c^T}{m} \sum_{v \in \mathcal{N}} d_v^R + c^D \sum_{v \in \mathcal{N}} d_v^A. \tag{8}$$

3.2. Optimization

Recall that we need to solve for the optimal allocation of flow in our network given a predefined set of parcel demands. To do this, we set our decision variable to the flows of trucks $x_p \in \mathbf{x}$ along the given paths $p \in \mathcal{P}$, where the flows of drones f_v^A are not controlled in the optimization as they are linear functions of our decision variable. We now go on to discuss the objective function and constraints used by our model.

3.2.1. Objective Function

Our goal is to find an allocation and routing policy for a delivery system that minimizes both parcel latency and societal costs. To achieve this, our objective function comprises two components: the average latency encountered by parcels L, and the average latency experienced by nominal vehicles L^S. This objective function can be expressed as:

$$J(f^T) = \gamma L + (1 - \gamma) L^S, \tag{9}$$

where $0 < \gamma < 1$ is used to weigh the relative importance between parcel latency and societal cost.

3.2.2. Constraints

The first constraint in our model is the supply cost defined in (8), which is represented as an equality constraint: $C(\mathbf{x}) \leq C_0$, where C_0 is a constant that represents the maximum cost permitted, measured in dollars per hour. Additionally, we must satisfy the system's parcel demand by adhering to (1)–(3).

3.2.3. Overall Optimization Problem

The overall optimization problem can be written as:

$$\min_{\mathbf{x}} \quad J(\mathbf{x}) = \gamma L + (1 - \gamma) L^S \tag{10}$$
$$\text{subject to} \quad x_p \geq 0 \qquad \forall p \in \mathcal{P},$$
$$C(\mathbf{x}) \leq C_0,$$
$$\text{and} \quad d_v^R(\mathbf{x}) \leq d_v, \qquad \forall v \in \mathcal{N}.$$

Note that the last constraint regarding parcel demand is expressed via an inequality since (3) guarantees the equality: $d_v^R(\mathbf{x}) + d_v^A(\mathbf{x}) = d_v$, i.e., drones will fulfill the leftover parcel demand. Hence, (10) makes sure trucks do not over satisfy the parcel demand, preventing negative values for drone demand d_v^A.

Using our expressions for average latency in (6), (7), and our derived latency model which accounts for the impact of stopping trucks (defined in the following Section 4.1), we represent this objective function using the standard quadratic form:

$$J(\mathbf{x}, \gamma) = \mathbf{x}^\mathsf{T} Q \mathbf{x} + \mathbf{a}^\mathsf{T} \mathbf{x}, \tag{11}$$

where $^\mathsf{T}$ is used for transpose operation, and the parameter γ has been incorporated into the expressions for matrix Q and vector $\mathbf{a}$. The corresponding set of inequalities from (10), along with the expressions for matrix Q and vector $\mathbf{a}$ are placed in Appendix C.

We note here that matrix Q is not positive semi-definite in general, making this Quadratic Program (QP) non-convex and difficult to scale. Considering large networks that contain thousands of nodes, using only one central delivery hub is impractical from a logistical perspective, and for such settings, we advise dividing the network into regions containing individual delivery hubs. Nonetheless, one may want to extend our method to larger networks without utilizing heuristics that simplify the problem.

To address this, we propose a convex version of our framework which reconsiders the definition of stopping trucks f_e^S given in (5). Instead of having trucks along paths $p \in \mathcal{P}_v$ delivering to node v stop along their final edge e_p^- and all edges $e \in \mathcal{O}_v$ outgoing at node v, we assume trucks will stop evenly across their path. This formulation makes the flow of stopping trucks f_e^S along an edge e proportional to the truck flow f_e^T on that edge, causing the quadratic matrix Q derived in Appendix C to become positive semi-definite. We will mainly focus on the non-convex formulation to derive our results, only using the convex version in Section 4.3 to compare the two in solution quality and scalability.

Finally, to solve our QP, we rely on Gurobi [37], a software for mathematical optimization. Since we can not say in general if matrix Q is positive semi-definite, we utilize Gurobi's methods for solving non-convex QPs. We note that, since version 9.0 of the Gurobi software, non-convex QPs can be solved globally using mixed-integer bilinear programming techniques. We provide details on the specific implementations used throughout the report in Appendix B, and plan to attach our code as Supplementary Material.

4. Results

4.1. Simulation Studies

As we expect trucks to have a negative impact on road congestion, our model should consider this when searching for optimal routing strategies. To do this, we derive a latency model for our optimization framework by relying on simulation studies. Using the open-source traffic simulation package SUMO [23], we formulate the latency ℓ_e^R as a function of the stopping and non-stopping flow components on a specific edge e. In this section, we will outline the simulation setup employed and subsequently present the results of our analysis.

4.1.1. Setup

Two vehicle types are used for all simulations—trucks that stop at their delivery locations, and cars that drive without stopping. We list the vehicle parameters used for our simulations in Table A1 of Appendix A. With these vehicle types, we simulated three variants of roads, namely, two-lane, three-lane, and four-lane roads. We placed equally spaced truck stops along all roads, where each truck has one stop. We then simulated over a set of two parameters for each of these roads: the ratio of stopping trucks to cars and the total flow of vehicles. Since we can expect the flow of stopping trucks f_e^S along a given edge e to be a small portion of the total flow on that edge, we consider ratios smaller than 10%. Hence, we varied the ratio from 0.001 to 0.1, and for each value we increased the total flow in increments of 10 cars per hour until reaching a prescribed maximum value. The

road parameters used throughout the simulations are listed in Table A2. We provide the rest of our implementation details in Appendix A. After collecting these data, we analyzed our results to see how both capacity and latency were affected by the input parameters.

4.1.2. Analysis

Although our end goal for these simulation studies is to derive a latency model, we begin our analysis by looking at how stopping trucks affect the capacity (maximum vehicular flow) on roads with a varying number of lanes. For a given ratio of stopping to total flow, we increase the total flow in incrementally and measure the flow out of all vehicles throughout the simulation. Since the flow in of traffic is constant during the first hour of simulation, we expect the road to be at a steady state in the middle of this timeframe. To measure this, we smooth the flow out using a 30 min sliding window, and record the steady state flow on the road as the midpoint of this averaged flow out.

We show this result in Figure 2, comparing the effect stopping trucks have on capacity for different roads by looking at the relationship between flow in and flow out both with and without trucks on the road. To do this, we first calculate the road's capacity f^0 without trucks on the road by measuring the maximum amount of flow in the road can handle before flow out starts to decrease. These plots are shown on the left side, where we can see that all roads can handle 100% of their capacity f^0 when the ratio of trucks is close to zero. As we increase the ratio of trucks to 10% of the flow in, we can see that the effective capacity of all three roads decreases as they are no longer able to handle 100% of their capacity f^0, where the two-lane road even struggles at 75%. The results show that the capacity of two-lane roads is more severely impacted by stopping trucks as compared to roads with more lanes.
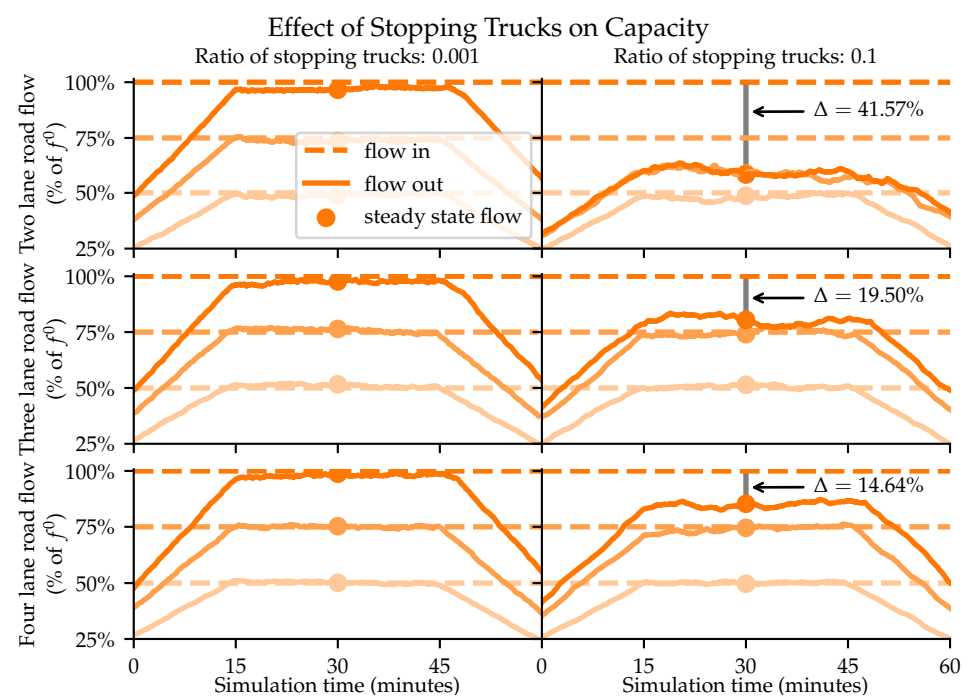
Figure 2. The three sets of plots show flow in (dashed line) vs. flow out (solid line) for each road, with (**right**) and without (**left**) stopping trucks. The *x*-axis shows simulation time, and the *y*-axis measures flow as a percentage of the road's respective capacity f^0. Each plot has three sets of data corresponding to flow in equaling 50%, 75%, and 100% of the capacity f^0, where we use matching color saturation to pair the data. By comparing the plots on the left and right, we can measure the effective decrease in capacity when 10% of the flow in becomes stopping trucks. We can see that stopping trucks have a greater impact on decreasing capacity on roads with fewer lanes, where we note that the two-lane road's capacity drops by 41.57%.

Now that we have shown how our simulation measures the effective capacity decrease for varying roads, we perform an identical analysis on the latency experienced by vehicles. For each simulation, we examined the travel times experienced by all vehicles, and then averaged these values to compute the steady state road latency given the current ratio and total flow, as previously described. We show this result in Figure 3, comparing all three roads by plotting the percentage increase over the given road's free flow latency ℓ^0 as we vary total flow in. The latency experienced by vehicles when there are no trucks is depicted by a dashed line, where we see it slightly increase as we approach the road's capacity. On the other hand, when we increase the ratio of stopping trucks to 10% of flow in, the solid lines show that latency increases greatly as flow in approaches the road's capacity. Moreover, we observe that this effect is greatest for two-lane roads shown in orange, while still present for three- and four- lane roads shown in green and purple, respectively.
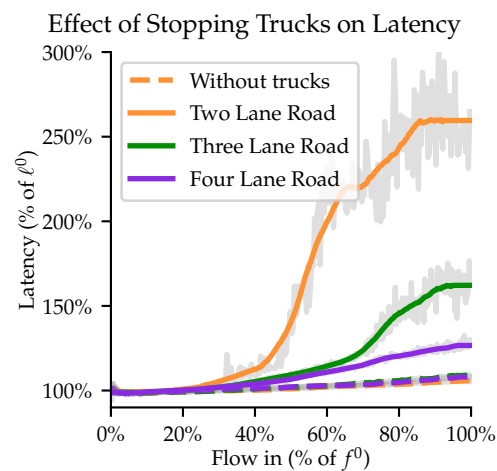
Figure 3. The plot above uses solid and dashed lines to represent the latency with and without trucks, respectively, while gray lines show the results without smoothing. When relevant, the flow of stopping trucks is 10% of the total flow. The *x*-axis and *y*-axis show total flow and latency, respectively, representing both as a percentage of the road's capacity f^0 and free flow latency ℓ_e^0. We use orange, green, and purple to represent two-, three-, and four-lane roads, respectively. We can see from this plot that stopping trucks have a greater impact on increasing latency on roads with less lanes.

To approximate the observed effect that stopping trucks have on latency, we model the relationship between latency and truck flow as linear. Specifically, we define our latencies ℓ_e^R for edge e with respect to all three components of flow on the road f_e^T, f_e^S, f_e^C, the road's capacity f_e^0, and its free flow latency ℓ_e^0:

$$\ell_e^R = \ell_e^0 \Big(1 + \langle \omega_e, [\frac{f_e^S}{f_e^0}, \frac{f_e^T + f_e^C}{f_e^0}] \rangle \Big), \tag{12}$$

where ω_e is the 2-dimensional vector of coefficients for edge $e \in \mathcal{E}^R$ and $\langle a, b \rangle$ represents the inner product between vectors a and b. As shown in Table 2, we can solve for ω by performing a linear regression over our simulation results, where Figure A1 in Appendix A provides a visual depiction of the planar fits found by the linear regression. We note that all models derived from our regressions showed a significant ($r^2 > 0.75$) correlation between the flows on the road and the corresponding latency. We also note that our derived weight vectors ω have strictly positive terms, since latency can only increase when additional flow is added to any of the components. Using the results of our simulations, we can now define affine transformations that map the network flow components to edge latencies.

Table 2. Linear regression results showing derived weights ω and squared correlation coefficient r^2.

No. of Lanes	$\omega[0]$ (*Stopping*)	$\omega[1]$ (*Total*)	r^2
2	15.76	0.02	0.80
3	4.26	0.06	0.76
4	1.92	0.06	0.82

4.2. Case Study: Sioux Falls

Before considering larger networks, we first perform a case study on the Sioux Falls transportation network shown in Figure 4, which is a popular choice in the literature for performing tests on traffic assignment [38]. Because drones are sent out from only one central hub, networks such as the Sioux Falls, which represent suburbs with modest populations, are a good fit for our experiments. The network contains 24 nodes and 76 directed edges, where we choose node v_{13} as our central delivery hub. The dataset also provides us with real locations of all nodes, as well as the capacity and free flow latency of all edges. To derive nominal flow on the road f_e^C, we use solutions provided by the study which correspond to the best routing strategy for $\beta = 360{,}600$ trips. Finally, we set the number of lanes for each edge by referring to a prior study [39].
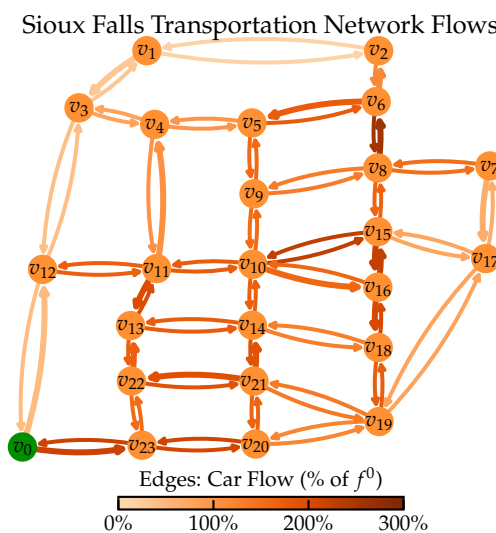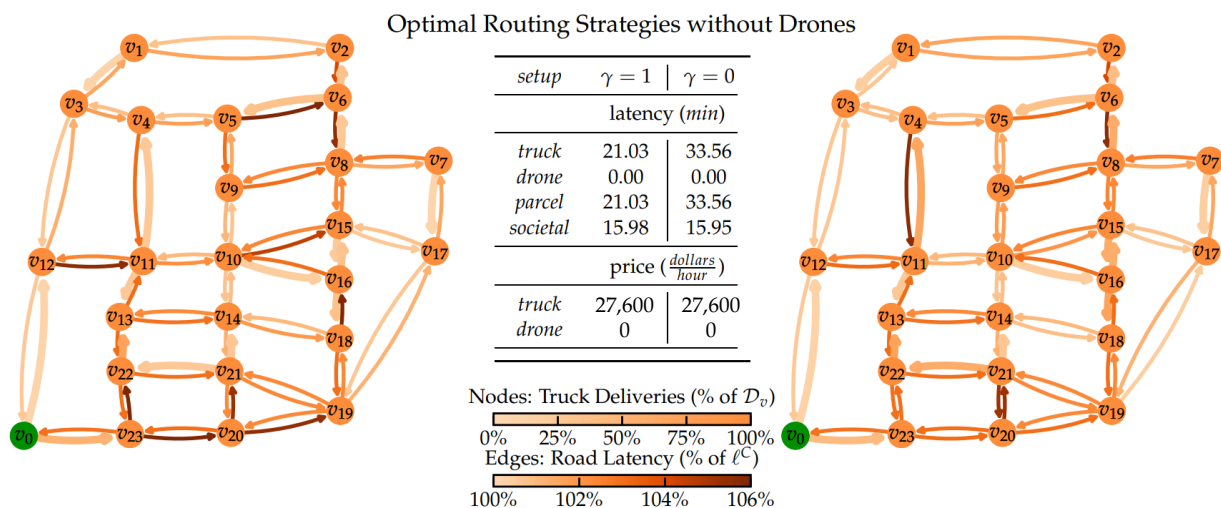

Figure 4. The figure on the left depicts the Sioux Falls transportation network. We color each edge $e \in \mathcal{E}$ to depict the nominal vehicular flow f_e^C as a percentage of the road's capacity f_e^0, both of which were directly taken from the referenced transportation study [38]. The total amount of vehicular flow was set to $\beta = 360{,}600$, which is the total number of trips to which the provided flows correspond. The central delivery hub v_0 is highlighted in green, and this location was chosen manually as it corresponds to real parcel delivery centers found in Sioux Falls. The edges have widths corresponding to the number of lanes on that road, which we directly took from a prior study [39]. We note that the downtown area of Sioux Falls is located around nodes v_{15} and v_{16}. The figure is not to scale, as the position of the nodes has been adjusted to allow for better visualization.

To setup our QP, we consider the 50 shortest paths for each origin–destination pair in our decision variable **x**. We measure the length of a path p by summing societal latency experienced on each edge $e \in p$ when there are no trucks. We define this value for all edges in the road network $e \in \mathcal{E}$ as ℓ_e^C, calculating it by using (12) and setting $f_e^T = 0$. We then optimize our objective function (11) using parcel demands of 5000 parcels per hour for each node, and allow a maximum supply cost of 40,000 dollars per hour. We set $c^T = 30$ dollars per truck and $c^D = 0.5$ dollars per drone, allowing each truck to carry a total of $m = 125$ parcels [40]. Lastly, we choose 25 kilometers per hour for the drone speed, as this allows for longer flight distances [41].

4.2.1. Routing without Drones

We first observe the optimal routing strategy we can obtain without drones. We accomplish this by setting our demand constraint given in (10) as an equality, forcing trucks to complete all deliveries. The result is portrayed in Figure 5, where we note that the supply cost of delivering all parcels by trucks is smaller than the maximum allowed cost C_0 by design, as introducing drones will require a larger budget. We can see from the results that there is a disproportionate trade-off between parcel latency and societal latency, as setting $\gamma = 0$ to minimize for societal latency only slightly changes the traffic conditions on the road, while increasing the parcel latency by over 50% as compared to when $\gamma = 1$. This occurs because **x** contains 50 paths for each origin–destination pair, meaning some trucks are rerouted on very time-inefficient paths. Logistic operators can take this into account by restricting **x** to contain paths they are okay with taking, setting an upper bound on the parcel latency. We do not do this in our study, and instead set the number of paths per origin–destination pair to 50 as larger numbers did not affect the result.
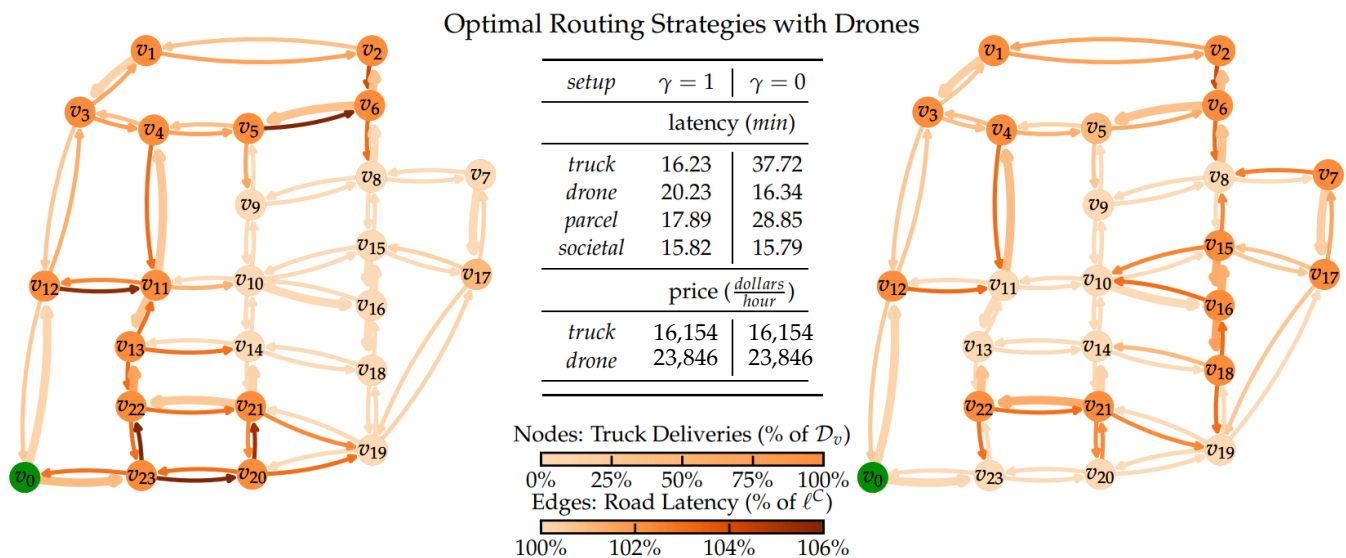

Optimal Routing Strategies without Drones

setup	$\gamma = 1$	$\gamma = 0$
	latency (*min*)	
truck	21.03	33.56
drone	0.00	0.00
parcel	21.03	33.56
societal	15.98	15.95
	price ($\frac{dollars}{hour}$)	
truck	27,600	27,600
drone	0	0

Nodes: Truck Deliveries (% of $\mathcal{D}_v$)

0%   25%   50%   75%   100%

Edges: Road Latency (% of ℓ^C)

100%   102%   104%   106%

Figure 5. We depict the optimal routing strategy derived for minimizing parcel latency (**left**) $\gamma = 1$, and societal latency (**right**) $\gamma = 0$. The table lists values reported by the solution, where we round the cost to the nearest dollar. The network diagrams show how the transportation network is affected by the routing strategy. We saturate the node colors to depict the portion of parcels delivered by trucks at the corresponding node. We saturate the edge colors to represent the change in societal latency experienced on the road once trucks are introduced. Specifically, we compare the solution's edge latency ℓ_e^R with the nominal latency ℓ_e^C derived in (12) when truck flow f_e^T is set to zero.

4.2.2. Routing with Drones

When we introduce drones into our delivery system, we can mitigate traffic congestion on the road and decrease the overall parcel latency. Drones delivering to node v experience a constant latency ℓ_v^A corresponding to the distance of the aerial path divided by the drone speed. We experiment with two setups, one where the distance is measured by a straight line between v_0 and v, and one where we scale this distance by $\sqrt{2}$ to create relatively longer paths, effectively making the drones deliver packages slower.

We first look at the results when drones can take direct paths to their destinations, which we depict in Figure 6. In this setup, we consider drones as being generally faster than trucks because, for most nodes $v \in \mathcal{N}$, the corresponding aerial path is more efficient than any truck path $p \in \mathcal{P}_v$. Since drones are more costly by design ($mc^D > c^T$), the strategy is to use as many drones as possible to route our packages and send the remainder by trucks. Which nodes we send drones to depends on if we care more about parcel latency or societal latency. Since the downtown nodes v_{15} and v_{16} are hard to get to via trucks, the solution for $\gamma = 1$ routes our drones to all surrounding nodes. When we instead optimize for societal latency by setting $\gamma = 0$, we re-route drones so that trucks can take the roads that least affect societal latency. This results in a decrease of congestion on roads surrounding nodes v_{20}, v_{21}, v_{22}, and v_{23}. Although the decrease in societal latency from 15.95 to 15.79 min when introducing drones may seem negligible, we see that drones allow us to alleviate congestion on some roads in the downtown area without increasing parcel latency. Note that, in both cases, the proportion of drone to truck deliveries stays the same as shown by the cost. This is expected since both solutions will utilize the maximum amount of drones that our system can afford. We see that, by adjusting γ, our framework can account for multiple stakeholder perspectives, routing the trucks and drones accordingly.

Optimal Routing Strategies with Drones

setup	$\gamma = 1$	$\gamma = 0$
	latency (*min*)	
truck	16.23	37.72
drone	20.23	16.34
parcel	17.89	28.85
societal	15.82	15.79
	price ($\frac{dollars}{hour}$)	
truck	16,154	16,154
drone	23,846	23,846

Nodes: Truck Deliveries (% of $\mathcal{D}_v$)

0%   25%   50%   75%   100%

Edges: Road Latency (% of ℓ^C)

100%   102%   104%   106%

Figure 6. Refer to the caption for Figure 5.

Finally, we look at the results when drones can not take direct paths to their destinations, which we depict in Figure 7. We refer to the drones in this setting as "slower" since scaling up the lengths of aerial paths by a constant is equivalent to scaling down the speed of drones by the reciprocal of this constant. In this setting, we no longer have an incentive to use drones when minimizing parcel latency with $\gamma = 1$, since trucks generally deliver parcels faster. We see this when comparing our solutions for $\gamma = 1$ between Figures 6 and 7: instead of routing drones to all the surrounding downtown nodes, we route them to much fewer nodes. This occurs because the set of nodes to which drones deliver faster than trucks decreases when we increase the distance of the aerial paths. On the other hand, when optimizing for societal latency with $\gamma = 0$, we do not expect drone latency to have an effect on our solution since drones do not contribute to the societal latency and hence are always chosen over trucks. This is confirmed by our results in Figures 6 and 7, where we can see that both optimal routing strategies for $\gamma = 0$ are identical.
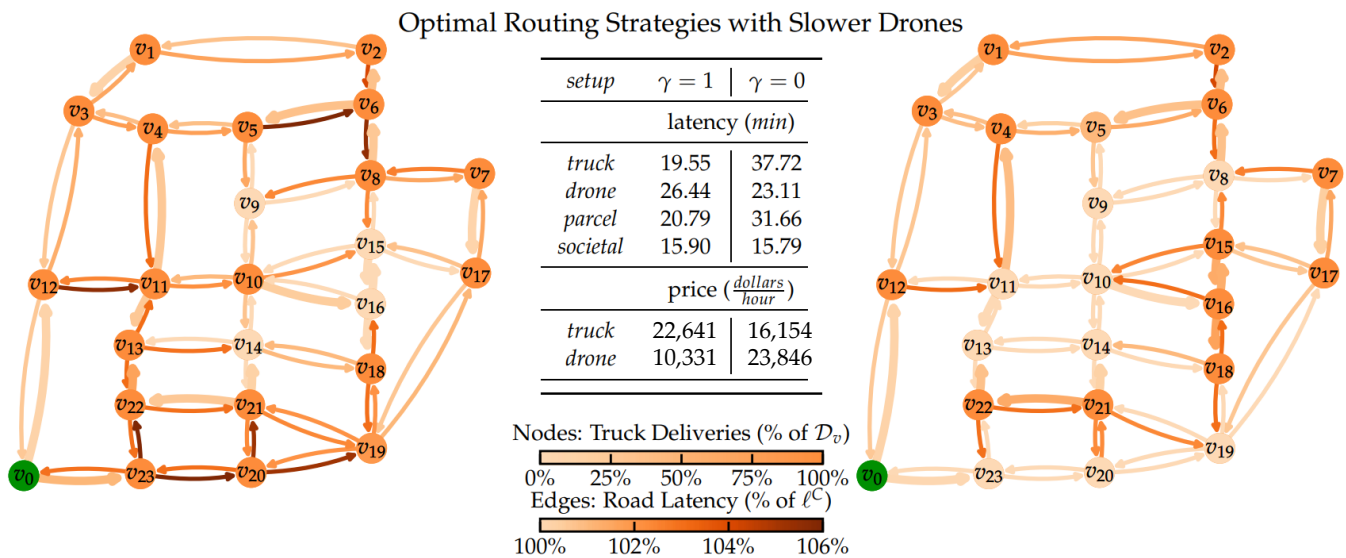
Figure 7. Refer to the caption for Figure 5.

We have shown that using drones can reduce both societal latency and parcel latency. In conclusion, our framework quantifies the potential of drones in mitigating traffic congestion, while solving for the allocation and routing that minimize a selected combination of societal latency and parcel latency. Before finalizing our work, we look at studies with larger networks to test the scalability of our framework.

4.3. Scaling to Larger Networks

As stated in Section 3.2, so far, we have utilized non-convex QP techniques to solve the optimization problem defined in (10). Here, we will test the scalability of our proposed method in the general non-convex case, and compare it to the simplified convex setting aforementioned. Our goal is to compare the optimal routing strategies derived by both formulations, where we calculate their corresponding parcel and societal latency using the true values for stopping truck flow f_e^S as defined by (5). This way we can see how the solution quality changes when we make the simplification to a convex QP.

We test our frameworks on the Sioux Falls, Anaheim, and Chicago transpiration networks, using the same dataset [38] for the network parameters, and scaling the demands and costs proportional with the network size. Due to missing lane data, we divide the edges equally into two-lane and three-lane roads based on the capacity. Table 3 lists the results, where the number of paths per origin–destination pair was first set to 5 so that the non-convex QP remained computationally feasible. We also show results for the convex QP when the number of paths per node was increased to 15, making the QP for Chicago's network have more than 10,000 decision variables as shown in the last row. We see from the results that both methods capture the trade-off between parcel and societal latency, the non-convex method is computationally efficient for smaller networks, while the convex method remains tractable even for larger networks. In addition, both methods provide optimal routing strategies with similar performance in terms of parcel latency, while the non-convex method is better at minimizing for societal latency as it uses the full definition for stopping trucks.

Table 3. Solutions for larger networks using both non-convex and convex formulations.

Setup			$\gamma = 1$		$\gamma = 0.5$		$\gamma = 0$		Runtime (sec.)
Network	# of Paths	Configuration	L	L^S	L	L^S	L	L^S	Average
Sioux Falls	115	non-convex	11.62	10.15	11.62	10.15	12.84	10.12	0.10
24 nodes	115	convex	11.62	10.15	11.62	10.15	11.94	10.16	<0.01
76 edges	345	convex	11.62	10.15	11.62	10.15	11.94	10.16	0.01
Anaheim	2067	non-convex	16.29	25.47	16.29	25.46	17.47	25.42	259.10
416 nodes	2067	convex	16.31	25.46	16.31	25.46	17.37	25.45	0.23
914 edges	6197	convex	16.31	25.46	16.31	25.46	18.20	25.45	2.16
Chicago	4656	non-convex	22.52	16.67	22.64	16.39	24.21	16.09	269.22
933 nodes	1200	convex	23.21	16.93	23.21	16.93	25.20	16.85	0.38
2950 edges	13,966	convex	23.20	16.94	23.21	16.94	25.39	16.85	10.32

5. Discussion

In this research, we have developed a network model for a congestion-aware bi-modal delivery system that comprises trucks and drones. To capture the effect our delivery system had on traffic congestion, we used simulation studies to derive a mathematical latency function for roads shared between regular vehicles and stopping trucks. With this, we were able to formulate our objective function as a trade-off between the parcel latency experienced by our system and the traffic congestion experienced by the road network. This allowed us to construct a QP to solve for the optimal routing of our delivery system incorporating multiple stakeholder viewpoints. We have demonstrated the effectiveness of our framework by applying it on the well-studied Sioux Falls transportation network, extending it to larger networks, and comparing it with a simplified convex formulation. To conclude, we examine the implications and shortcomings of our work, and discuss relevant directions for improvement.

One limitation of our framework is the planar fit latency model illustrated in Figure A1. To address this, one can utilize a piecewise linear fit. This addition would change our main objective function to piecewise quadratic, which would no longer allow us to rely on the same QP technique for our optimization, as we would have integer constraints. We point interested readers to a study of such piecewise quadratic solvers and leave this extension for future work [42].

In addition to the inaccuracies of our latency estimates, one may also consider the non-convexity of our general framework as a limitation. However, we would like to note that non-convex Quadratic Programming (QP) can be solved efficiently for problem sizes of relevance, e.g., where the central delivery hub is responsible for less than 1000 nodes. For larger metropolis areas, we expect there to be multiple delivery centers. This can be handled with our method by using heuristics that partition the nodes according to their delivery centers. Furthermore, restricting the number of paths per origin–destination pair allows us to keep the problem tractable without lowering solution quality with respect to parcel latency.

As shown by our case study in Section 4.2, although societal latency can improve by routing trucks along less time-efficient paths, this disproportionately increases parcel latency. Due to this, we suggest that receivers provide the allowed paths per origin–destination pair, while minimizing for societal latency in the framework. This will cause the QP to have an upper bound on the parcel latency while searching for the routing and allocation strategy that minimizes traffic congestion.

In conclusion, our proposed network model and optimization framework can assist delivery systems in determining routing policies for their trucks and drones, while evaluating the impact of different routing policies on overall parcel latency and traffic congestion.

Appendix A. SUMO

To run our simulation studies on different road configurations, we rely on the open source traffic simulation software SUMO [23]. Specifically, we used the Python package *sumolib-1.8.0* provided by SUMO. We detail the corresponding parameters used throughout our simulations in Tables A1 and A2.
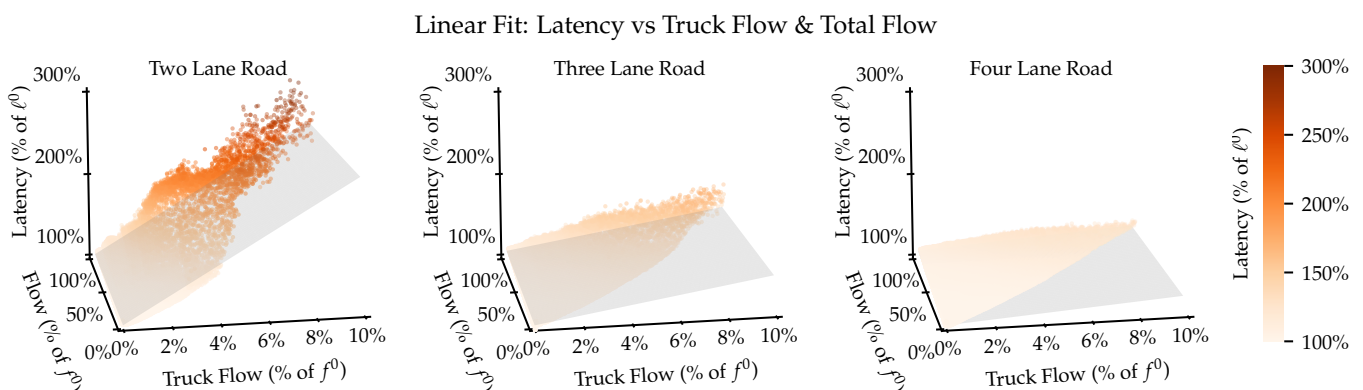
Figure A1. The three plots show the linear fit for each road, where we model latency as a function of total flow and stopping truck flow. The scattered data points represent the values of steady state latency for each configuration of stopping truck flow and total flow, where we use increasing saturation to help visualize the vertical axis corresponding to latency. The gray plane represents the solution of our linear regression for each road, where we set the intercept to be at 100% of free flow latency ℓ_e^0 since, when there is no flow on the road, we expect the latency to remain unchanged. Based on data visualization alone, we can see that stopping trucks have a greater impact on decreasing latency on roads with less lanes. We provide the specific values for the coefficients and r^2 corresponding to each linear fit in Table 2.

Table A1. Vehicle parameters used for cars and trucks throughout our simulation studies.

Parameter	Car	Truck
Length m [43]	4.00	7.82
Max. Speed $\frac{km}{h}$ [44]	160.93	96.56
Max. Acceleration $\frac{m}{s^2}$ [44]	2.87	1.00
Max. Deceleration $\frac{m}{s^2}$ [44]	4.33	0.88

Table A2. Road parameters used throughout our simulation studies.

Parameter	Two-Lane	Three-Lane	Four-Lane
Length m	500	2000	3000
Speed Limit $\frac{km}{h}$	48.28	80.47	80.47
Number of Stops	25	100	150
Duration minutes	60	60	60
Max. Flow $\frac{vehicles}{hour}$	3000	4500	6000

Appendix B. Gurobi Implementation

Throughout our work, we use the Gurobi optimization software [37] to program our quadratic problem. We received permission to use this software via a free academic license. The version used is: Gurobi Optimizer version 9.5.1 build v9.5.1rc2 (linux64), which we ran on Ubuntu 20.04. The computer we used had 8 physical cores and 16 logical processors, allowing Gurobi to use up to 16 threads. Because the quadratic term defined in our optimization is not necessarily positive semi-definite, we rely on non-convex optimization techniques provided by Gurobi. As stated by the documentation, the non-convex QP is converted to a mixed integer program by considering bi-linear techniques that rely on dual simplex methods. The solutions reported throughout the paper were all found as optimal under the tolerance of 10^{-5}.

Appendix C. Objective Function Derivation

Below, we detail how we constructed the optimization problem provided in (10) as a QP. First, we define the common linear transformations that we use throughout the rest of the derivation:

$$A\mathbf{x} = f^T, \tag{A1}$$

$$B\mathbf{x} = \frac{1}{m}d_v^R, \tag{A2}$$

$$E\mathbf{x} = f^S, \tag{A3}$$

where we note that the above three equations correspond to the linear transformations defined in Equations (2), (4) and (5). Note that we remove the subscript for edges $e \in \mathcal{E}$ to signify that we are dealing with a vector. These matrices can be pre-computed given a graph G^R along with the set of allowed paths $\mathcal{P}$. In addition, we collect the following constant terms to form two diagonal matrices:

$$D_0 = diag(\frac{\omega[0]\ell^0}{f^0}), \tag{A4}$$

$$D_1 = diag(\frac{\omega[1]\ell^0}{f^0}), \tag{A5}$$

where we use $diag()$ to represent the diagonal operation we are preforming over the vector composed by considering the above quantities over all edges $e \in \mathcal{E}$.

Using these defined constant matrices, we can express our overall optimization problem using the following QP:

$$\min_{\mathbf{x}} \quad \mathbf{x}^\mathsf{T} Q_0 \mathbf{x} + \mathbf{a}_0^\mathsf{T} \mathbf{x} + b_0 \tag{A6}$$

$$\text{subject to} \quad x \geq 0 \quad \forall x \in \mathbf{x}, \tag{A7}$$

$$Q_1 \mathbf{x} \leq \mathbf{b}_1, \tag{A8}$$

$$\text{and} \quad \mathbf{a}_1^\mathsf{T} \mathbf{x} \leq b_2. \tag{A9}$$

We note that, when $\gamma = 0$, the objective function no longer depends on parcel latency L, making the optimization problem linear. The expressions for the corresponding terms are given below:

$$Q_0 = \frac{\gamma m}{\sum_{v \in \mathcal{N}} d_v} \left(A^\mathsf{T} D_0 E + A^\mathsf{T} D_1 A \right), \tag{A10}$$

$$\mathbf{a}_0 = \frac{\gamma m}{\sum_{v \in \mathcal{N}} d_v} \left(A^\mathsf{T} \ell^0 + A^\mathsf{T} D_1 f^C - B^\mathsf{T} \ell^A \right) \tag{A11}$$

$$+ \frac{1 - \gamma}{\beta} \left((D_0^\mathsf{T} f^C)^\mathsf{T} E + (D_1^\mathsf{T} f^C)^\mathsf{T} A \right), \tag{A12}$$

$$b_0 = \frac{\gamma}{\sum_{v \in \mathcal{N}} d_v} \left(\langle d_v, \ell^A \rangle \right) \tag{A13}$$

$$+ \frac{1 - \gamma}{\beta} \left(\langle f^C, \ell^0 \rangle + (D_1^\mathsf{T} f^C)^\mathsf{T} f^C \right), \tag{A14}$$

$$Q_1 = mB \tag{A15}$$

$$\mathbf{b}_1 = d_v \tag{A16}$$

$$\mathbf{a}_1 = (c^T - c^D m)(B^\mathsf{T} \underline{1}), \tag{A17}$$

$$b_2 = C_0 - c^D \sum_{v \in \mathcal{N}} d_v \tag{A18}$$

References and Notes

1. Joerss, M.; Schröder, J.; Neuhaus, F.; Klink, C.; Mann, F. *Parcel Delivery: The Future of Last Mile*; McKinsey & Company: Chicago, IL, USA, 2016; pp. 1–32.
2. Ali, F. US Ecommerce Grows 44.0 in 2020. 2021. Available online: https://www.digitalcommerce360.com/article/us-ecommerce-sales/ (accessed on 9 May 2022).
3. Adobe. Adobe Digital Economy Index, Q1 2021. 2021. Available online: https://business.adobe.com/resources/reports/adobe-digital-economic-index-april-2021.html (accessed on 6 May 2022).
4. Joerss, M.; Neuhaus, F.; Schröder, J. *How Customer Demands Are Reshaping Last-Mile Delivery*; McKinsey & Company: Chicago, IL, USA, 2016; pp. 1–32.
5. Weiss, C.; Onnen-Weber, U. The challenge of sustainable last mile distribution of CEP services in small towns. *Transp. Res. Procedia* **2019**, *39*, 597–604. [CrossRef]
6. Holguín-Veras, J.; Amaya, J.; Sánchez-Díaz, I.; Browne, M.; Wojtowicz, J. State of the art and practice of urban freight management Part II: Financial approaches, logistics, and demand management. *Transp. Res. Part Policy Pract.* **2020**, *137*, 383–410. [CrossRef]
7. Bosona, T. Urban freight last mile logistics—challenges and opportunities to improve sustainability: A literature review. *Sustainability* **2020**, *12*, 8769. [CrossRef]
8. Ranieri, L.; Digiesi, S.; Silvestri, B.; Roccotelli, M. A review of last mile logistics innovations in an externalities cost reduction vision. *Sustainability* **2018**, *10*, 782. [CrossRef]
9. Cardenas, I.; Borbon-Galvez, Y.; Verlinden, T.; Van de Voorde, E.; Vanelslander, T.; Dewulf, W. City logistics, urban goods distribution and last mile delivery and collection. *Compet. Regul. Netw. Ind.* **2017**, *18*, 22–43. [CrossRef]
10. Digiesi, S.; Fanti, M.P.; Mummolo, G.; Silvestri, B. Externalities reduction strategies in last mile logistics: A review. In Proceedings of the 2017 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Bari, Italy, 18–20 September 2017; pp. 248–253.
11. Allen, J.; Piecyk, M.; Piotrowska, M.; McLeod, F.; Cherrett, T.; Ghali, K.; Nguyen, T.; Bektas, T.; Bates, O.; Friday, A.; et al. Understanding the impact of e-commerce on last-mile light goods vehicle activity in urban areas: The case of London. *Transp. Res. Part Transp. Environ.* **2018**, *61*, 325–338. [CrossRef]
12. Macrina, G.; Pugliese, L.D.P.; Guerriero, F.; Laporte, G. Drone-aided routing: A literature review. *Transp. Res. Part Emerg. Technol.* **2020**, *120*, 102762. [CrossRef]
13. Frachtenberg, E. Practical drone delivery. *Computer* **2019**, *52*, 53–57. [CrossRef]
14. SESARJU. Smart ATM U-Space. 2021. Available online: https://www.sesarju.eu/U-space (accessed on 9 May 2022).
15. Food and Agriculture Organization. National Environmental Policy Act and Drones. 2022. Available online: https://www.faa.gov/uas/advanced_operations/nepa_and_drones/ (accessed on 9 May 2022).
16. Amazon. Amazon Prime Air. Available online: https://www.faa.gov/uas/advanced_operations/nepa_and_drones/ (accessed on 9 May 2022).
17. DHL. DHL drone delivery and Parcelcopter Technology: Discover DHL. 2021. Available online: https://www.dhl.com/discover/en-my/business/business-ethics/parcelcopter-drone-technology (accessed on 9 May 2022).

18. Viu-Roig, M.; Alvarez-Palau, E.J. The impact of E-Commerce-related last-mile logistics on cities: A systematic literature review. *Sustainability* **2020**, *12*, 6492. [CrossRef]

19. Cleophas, C.; Cottrill, C.; Ehmke, J.F.; Tierney, K. Collaborative urban transportation: Recent advances in theory and practice. *Eur. J. Oper. Res.* **2019**, *273*, 801–816. [CrossRef]

20. Holguín-Veras, J.; Leal, J.A.; Sánchez-Diaz, I.; Browne, M.; Wojtowicz, J. State of the art and practice of urban freight management: Part I: Infrastructure, vehicle-related, and traffic operations. *Transp. Res. Part A Policy Pract.* **2020**, *137*, 360–382. [CrossRef]

21. Kiba-Janiak, M.; Marcinkowski, J.; Jagoda, A.; Skowrońska, A. Sustainable last mile delivery on e-commerce market in cities from the perspective of various stakeholders. Literature review. *Sustain. Cities Soc.* **2021**, *71*, 102984. [CrossRef]

22. Beliaev, M.; Mehr, N.; Pedarsani, R. Congestion-aware Bi-modal Delivery Systems Utilizing Drones. In Proceedings of the European Control Conference, London, UK, 11–14 July 2022

23. Lopez, P.A.; Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Flötteröd, Y.P.; Hilbrich, R.; Lücken, L.; Rummel, J.; Wagner, P.; Wießner, E. Microscopic Traffic Simulation using SUMO. In Proceedings of the 21st IEEE International Conference on Intelligent Transportation Systems, Maui, HI, USA, 4–7 November 2018

24. Merkert, R.; Bushell, J. Managing the drone revolution: A systematic literature review into the current use of airborne drones and future strategic directions for their effective control. *J. Air Transp. Manag.* **2020**, *89*, 101929. [CrossRef] [PubMed]

25. Murray, C.C.; Chu, A.G. The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. *Transp. Res. Part Emerg. Technol.* **2015**, *54*, 86–109. [CrossRef]

26. Wang, X.; Poikonen, S.; Golden, B. The vehicle routing problem with drones: Several worst-case results. *Optim. Lett.* **2017**, *11*, 679–697. [CrossRef]

27. Lindholm, M. *Urban Freight Transport from a Local Authority Perspective—A Literature Review*; ISTIEE, Institute for the Study of Transport within the European Economic Integration: Trieste, Italy, 2013.

28. Hu, X.; Liu, T.; Hao, X.; Su, Z.; Yang, Z. Research on the influence of bus bay on traffic flow in adjacent lane: Simulations in the framework of Kerner's three-phase traffic theory. *Phys. Stat. Mech. Its Appl.* **2021**, *563*, 125495. [CrossRef]

29. Raj, P.; Asaithambi, G.; Ravi Shankar, A. Effect of curbside bus stops on passenger car units and capacity in disordered traffic using simulation model. *Transp. Lett.* **2022**, *14*, 104–113. [CrossRef]

30. Trott, M.; Baur, N.F.; der Landwehr, M.A.; Rieck, J.; von Viebahn, C. Evaluating the role of commercial parking bays for urban stakeholders on last-mile deliveries—A consideration of various sustainability aspects. *J. Clean. Prod.* **2021**, *312*, 127462. [CrossRef]

31. Hu, X.; Hao, X.; Wang, H.; Su, Z.; Zhang, F. Research on on-street temporary parking effects based on cellular automaton model under the framework of Kerner's three-phase traffic theory. *Phys. A Stat. Mech. Its Appl.* **2020**, *545*, 123725. [CrossRef]

32. Chen, J.; Li, Z.; Jiang, H.; Zhu, S.; Wang, W. Simulating the impacts of on-street vehicle parking on traffic operations on urban streets using cellular automation. *Phys. A Stat. Mech. Its Appl.* **2017**, *468*, 880–891. [CrossRef]

33. Kong, D.; Guo, X.; Yang, B.; Wu, D. Analyzing the Impact of Trucks on Traffic Flow Based on an Improved Cellular Automaton Model. *Discret. Dyn. Nat. Soc.* **2016**, *2016*, 1236846. [CrossRef]

34. Al Eisaeia, M.; Moridpourb, S.; Tay, R. Heavy Vehicle Management: Restriction Strategies. *Transp. Res. Procedia* **2017**, *21*, 18–28.

35. Thompson, R.G. Vehicle Orientated Initiatives for Improving the Environmental Performance of Urban Freight Systems. In *Green Logistics and Transportation: A Sustainable Supply Chain Perspective*; Fahimnia, B., Bell, M.G., Hensher, D.A., Sarkis, J., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 119–129. [CrossRef]

36. Dablanc, L.; Rodrigue, J.P. The geography of urban freight. In *The Geography of Urban Transportation*; The Guilford Press: New York, NY, USA, 2014.

37. Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*; Gurobi Optimization, LLC: Houston, TX, USA, 2022.

38. Stabler, B. Transportation Networks for Research Core Team. Available online: https://github.com/bstabler/TransportationNetworks (accessed on 9 May 2022).

39. Ukkusuri, S.V.; Yushimito, W.F. A methodology to assess the criticality of highway transportation networks. *J. Transp. Secur.* **2009**, *2*, 29–46. [CrossRef]

40. Kim, M.; Matson, E.T. A Cost-Optimization Model in Multi-agent System Routing for Drone Delivery. In Proceedings of the Highlights of Practical Applications of Cyber-Physical Multi-Agent Systems; Bajo, J., Vale, Z., Hallenborg, K., Rocha, A.P., Mathieu, P., Pawlewski, P., Del Val, E., Novais, P., Lopes, F., Duque Méndez, N.D., et al., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 40–51.

41. DJI. DJI Mavic 2 Specifications Sheet.

42. Cui, Y.; Chang, T.H.; Hong, M.; Pang, J.S. A Study of Piecewise Linear-Quadratic Programs. *J. Optim. Theory Appl.* **2020**, *186*, 523–553.

43. Sharpe, B.; Rodriguez, F. Market analysis of heavy-duty commercial trailers in Europe. 2018.

44. Bokare, P.; Maurya, A. Acceleration-Deceleration Behaviour of Various Vehicle Types. *Transp. Res. Procedia* **2017**, *25*, 4733–4749.