7-28-2022

# Planning and Real-Time Resource Allocation in Freight Logistics Systems Utilizing Emerging Transportation Technologies

Darshan Rajesh Chauhan
*Portland State University*

# Planning and Real-Time Resource Allocation in Freight Logistics Systems Utilizing Emerging Transportation Technologies

by

Darshan Rajesh Chauhan

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Civil and Environmental Engineering

Dissertation Committee:
Avinash Unnikrishnan, Chair
Stephen D. Boyles
Miguel A. Figliozzi
Thomas Schumacher

Portland State University
2022

Abstract

Transportation is a key driver of any national economy. In the United States, the transportation sector contributes $1.3 trillion to the economy, of which freight transportation represents more than 50%. Trucks alone account for more than 70% freight movements in the United States. In addition to worsening stress at ports of entry and traffic congestion in the system, freight also accounts for nearly one-third of the greenhouse gas emissions in the United States. Emerging transportation technologies like electric unmanned aerial vehicles (or drones) and electric vehicles can provide a more sustainable alternative to combat greenhouse gas emissions and reduce the congestion in the transportation network.

This dissertation extends the frontier in planning and real-time resource allocation in logistics systems that utilize emerging transportation technologies to move freight. A common theme throughout the dissertation is uncertainty. In network planning problems, uncertainty stems from inherent variation in problem parameters or the potential unavailability of data. In real-time operations, uncertainty arises due to the dynamic nature of the problem as the information is gradually revealed over time. The dissertation considers four application problems spanning both public sector and corporate applications. These problems involve a network planning component or a real-time operations component or both. The real-time operations are modeled as online resource allocation problems and multi-armed bandits-based reinforcement learning methodologies are proposed. The contributions are made by

developing novel problem formulations for each problem and proposing two new multi-armed bandit problems. A performance regret bound is also obtained for one of the proposed multi-armed bandit problems.

The four application problems are now very briefly described. The first application considers a network planning problem for locating electric drones equipped with automatic external defibrillators (AED) in an effort to combat out-of-hospital cardiac arrests in a service region. The second application considers a facility location and dynamic resource allocation problem applicable to a logistics company expanding to offer instant delivery using electric drones. The third application also considers a facility location and dynamic resource allocation problem but in the context of relief prepositioning and their equitable distribution post-disaster. Finally, the fourth application considers a dynamic truckload pickup and delivery problem in a service area using a fleet of electric trucks.

*Dedicated to my earliest role model and inspiration,*

*my dada (grandfather), Late Nanjibhai Chauhan,*

*and*

*my source of strength and blessings, my lucky charm,*

*my baa (grandmother), Kantaben Chauhan.*

**Acknowledgments**

This dissertation marks the culmination of my fantastic five years of graduate studies at Portland State University. Over this time I have met a lot of incredible people and cultivated lifelong friendships. I'll try my best to acknowledge everyone who has been a part of my memorable journey.

Foremost, I would like to sincerely thank Prof. Avinash Unnikrishnan, whom I dearly address as Dr. Avi. Through my time at graduate school, I have grown a lot personally, academically, and professionally, and Dr. Avi gets most of the credit for this. He has always given me the space to explore and was always there to pick me up when I failed (which is much more often than I would care to admit). I had heard a lot of great things about him even before I joined Portland State, and he has been all that and even more since I met him. I want to especially thank him for his patience, advice, continuous motivation, friendship, mentoring, and our countless discussions about this dissertation, research, and life in general.

I would like to thank Prof. Stephen Boyles for bringing in his expertise and for being a great collaborator to this research. His constructive feedback has greatly enriched the clarity and focus of this work.

I convey my many thanks to Prof. Miguel Figliozzi for bringing his expertise to a part of this dissertation. I also thank him for being a great teacher throughout my course of study at Portland State, especially his Freight and Logistics course which provided me with the overall picture of the freight logistics systems.

I thank Prof. Thomas Schumacher for agreeing to serve on my dissertation committee and for adding value to this work.

I am very grateful to the National Science Foundation (NSF) for enabling this work through much-needed financial support. This research is a product of NSF grants awarded to Prof. Unnikrishnan and Prof. Boyles (CMMI-1562109/1562291: Collaborative Research: Non-Additive Network Routing and Assignment Models, and CMMI-1826320/1826337: Collaborative Research: Real-Time Stochastic Matching Models for Freight Electronic Marketplace). I am also thankful for the support I received through the Maseeh Fellowship (by Portland State University), Laurels Graduate Fellowships (by Portland State University), Dean's Graduate Fellowship (by Portland State University Graduate School), the Walter H. Kramer Fellowships (by the Transportation Research and Education Center at Portland State University), and the NITC Scholarships (by the National Institute of Transportation and Communities (NITC) University Transportation Center).

My many thanks to Dr. Kalyana Rama J. S. for introducing me to research as an undergraduate junior. I immensely cherish the time I was working with him on my undergraduate thesis and my first research paper. I cannot thank Dr. Arkamitra Kar

enough for pushing me over the ledge for pursuing graduate studies, for everything he has taught me, and for being my go-to person when I'm too stressed with grad life. Continued research collaborations with him and our shared love of food, music, TV shows, and movies have been a bonus for me. Most of all, I appreciate his friendship and mentorship.

My time at the CEE department was enriched by a lot of people. Many thanks to Kiley, Sam, and Sarah for always physically being present for well-being, support, or just a casual chat through the ghost-town that was the COVID-19 pandemic! Bonus points to them for helping me jump through all the administrative hoops and making it a lot easier. The other thing that I am proud of is Pizza Club (Katherine gets the naming credits)!! Nothing beats a conversation with friends over pizza. Cheers to all: Rohan, Katherine, Santiago, Jaclyn, Melissa, Frank, Kayla, Tricia, Sam, Kiley, Sarah, Ayush, Arash, Amir, Chelle, Majid, Joe, Jael, and Travis. Many thanks to Tricia for taking over the mantle of Pizza Club "President"!

I'm glad to have met Dr. Jason Anderson here. Though not directly incorporated in this dissertation, I'm grateful for his course on Econometric Data Analysis which opened up a whole new world of exciting research for me. What I'll miss most are our frequent hour-long research meetings which were usually taken over by discussions on life, culture, sports, and travel after like 10 minutes.

Thanks to my roommates over the years, Ajinkya, Abhishek, Bikesh (weekends only :P), and Arjun, for making the mundane joyous. Ajinkya and my shared love for

food and cooking have been great, especially through the COVID-19 lockdown. It is because of the consistent efforts of Abhishek and Bikesh that I am now mediocre at bowling and not pathetic. It has been fun exploring Portland with them and enjoying the little things. Thanks to Arjun for always being up for a road trip. Our trips to California, Glacier-Yellowstone-Grand Tetons, and the Enchantments will be etched in my memory forever. Thanks also for helping me believe in myself during the times when I thought it was not possible.

Visits from friends to Portland also made my time here special. Monish's innumerable trips here were always fun. Doing the Christmas 2020 vacation hike to Mirror Lake with Raghavi, Anvita, and absolutely no winter gear whatsoever will always be something I'll laugh remembering! Aditya and Monish's visit in Summer 2021 was a blast back from our undergrad days. Dr. Kar's visit just after my dissertation defense in May 2022 made it real for me that I was almost at the finish line. It was a truly emotional and remarkable moment for me. Thank you for that.

Thanks to all my long-distance friends who were always a phone call away. They provided much-needed laughter, a space to decompress, and are the most awesome support system: Meena *fai*, Swati, Krisha, Shrineh, Mitul, Nilay, Anvita, Raghavi, Siddhi, Monish, Aditya, Amod, Jaideep, and Shashank.

As I prepare for the next stage of my career, I am indebted to Dr. Avi, Varun, and Mehrdad for their mentorship and guidance. Thanks to Monish, Arjun, Anvita, Raghavi, and Sai Sri for their valuable advice, help, and timely motivation.

Finally, a big shout out to my family: *baa*, mom, papa, *mota* mom, *mota* papa, Parth, Vaibhav, Bhumika, Chintan, and Piyu, who always believed in me, supported me and made sure I was doing fine all through my journey here over my regular video calls to India. Their cheery faces instantly lit my life on dull days. They are my greatest blessing and I am grateful every day to have them in my life. My deepest love and thanks to them!

**Table of Contents**

**List of Tables**

## List of Figures

# 1 Introduction

## 1.1 Background

Transportation is a key driver of a nation's economy. The demand for transportation, including all expenditures related to transportation-related goods and services, contributed $1.9 trillion to the US economy in 2019, representing an 8.9% share of the GDP (U.S. Department of Transportation, Bureau of Transportation Statistics, 2021). Just accounting for transportation services, the transportation sector contributed $1.3 trillion in which freight transportation represents slightly over a 50% share ($651.2 billion). As many sectors are dependent on freight transportation, the national GDP is highly correlated with the trends in freight volumes (see Figure 1.1). The trends in freight volumes are captured using the Freight Transportation Service Index (TSI) developed by the Bureau of Transportation Statistics (BTS) which is calculated monthly. According to BTS research, the changes in Freight TSI occur before changes in the economy, making it an important economic indicator.

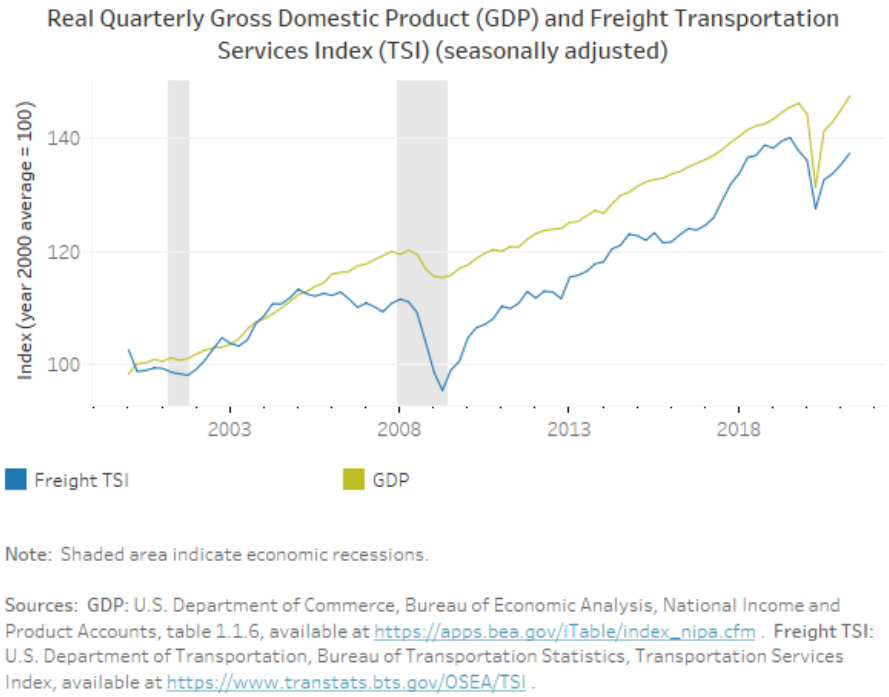



Figure 1.1: Quarterly trends of the US GDP and Freight Transportation Service Index from year 2000 onward

With a significant contribution to the GDP, the ability to move freight efficiently in, around, and out of a region has a direct implication on the regional economy. However historically, freight planning has received insufficient consideration in the transportation planning process (U.S. Department of Transportation, 2020). Combined with the growth in foreign trade spurred by the globalization of supply chains and containerization, the amount of freight moving through the nation has exploded resulting in stresses to not only the points of entry but the transportation network in general (U.S. Department of Transportation, Federal Highway Administration, 2012).

At $368 billion, trucking is the highest contributor to the GDP in the $1.3 tril-

lion transportation services sector (U.S. Department of Transportation, Bureau of Transportation Statistics, 2021). Trucks transported 11.84 billion tons of freight which represents 72.5% share of domestic tonnage shipped (American Trucking Association, 2019). Trucking is also the favored mode of freight transportation, with a mode share of greater than 50% when travel distances are less than 2000 miles (U.S. Department of Transportation, 2020). However, this dominance is not reflected in its use of data analytics to improve the efficiency and performance of its business operations (Wang et al., 2018). Improvements in information and communication technologies provide opportunities for using real-time information for improving the efficiency and optimizing the performance of freight systems, which is the central theme of this dissertation. Moreover, the advancements in truck-related freight logistics can also be applicable, under certain contexts, to the closely related fields of emergency services management, humanitarian logistics, ride-hailing/ride-sharing services, and package/food delivery services. To this end, the current work models freight movement in several application contexts.

## 1.2 Research Contributions

This work broadly contributes to planning and the real-time resource allocations in freight logistics systems. The freight systems employ emerging transportation technologies like battery-operated unmanned aerial vehicles (UAVs) or drones and electric freight vehicles. The application of these freight systems are considered in

the context of both for-profit corporate operations like instant delivery logistics and truckload delivery markets, and in public sector and/or non-profit based operations like emergency services management and humanitarian logistics. The major research contributions resulting from this dissertation are:

- **Development of a two-stage methodology**: As a part of this methodology, in the first stage, we locate facilities and/or allocate resources to them. In the second stage, requests arrive in an online manner and the goal is to allocate request to the located facilities. The fulfillment of requests lead to consumption of resources at the facility. The goal during the second stage is maximization of cumulative rewards obtained through successful allocation adhering to the resource budget constraints at the facilities. We use multi-armed bandits for designing the request allocation policy in the second stage. The development of this methodology contributes to limited research in two stage location-allocation models which use adaptive online policy for dynamic decision-making in the second stage. To the best of authors' knowledge, this is the first application of multi-armed bandits for designing the second stage online adaptive policy for transport-related applications.

  This model is employed two studies: first, facility location and dynamic resource allocation for incorporating instant delivery logistics in an existing company's last-mile supply-chain; and second, it is utilized for setting distribution centers and prepositioning scarce relief supply and their equitable distribution in post-

disaster operations.

- **Equitable linear contextual bandits with knapsacks**: A new multi-armed bandit (MAB) framework is proposed to learn an allocation policy in an online manner for while ensuring equity and considering globalized resource consumption budget (or, knapsack) constraints. We call this problem equitable linear contextual bandits with knapsack (E-linCBwK) problem. This framework is applied to obtain equitable distribution of scarcely available relief packages in post-disaster scenarios in an dynamic-stochastic setting. The fair or equitable distribution is rooted in the concept of envy Foley (1966), which is commonly used in public welfare applications. We propose a solution algorithm for E-linCBwK and obtain a sub-linear regret bound for the algorithm by formulating the E-linCBwK problem as an existing MAB problem proposed by Agrawal and Devanur (2016).

- **Linear contextual blocking bandits with context-dependent delays**: We formulate a new multi-armed bandit problem for the allocation of vehicles for real-time pickups and drop-offs. Examples of such problems include truck-load operations at a carrier company, ride-hailing services, and online food or freight delivery platforms. Here, we wish to learn the allocation of vehicles to each request. Once the vehicle is assigned, it cannot service other requests, and hence, it is "blocked". In addition, this "blocked" time is dependent on request characteristics. We propose a new methodology to solve this problem

based on blocking bandits (Basu et al., 2019, 2021). Specifically, we formulate linear contextual blocking bandits with context-dependent delays. We consider that both the revenue and the "blocked" times are linearly dependent on the context. The context would include information on route characteristics, traffic conditions, driver history, etc. We also propose an upper confidence bound algorithm to solve this problem.

The minor contributions of this dissertation are:

- **Robust Multi-Period Maximum Coverage Facility Location Considering Coverage Reliability**: We propose a new facility location problem with an objective of maximizing coverage considering coverage reliability constraints. The coverage reliability constraint is a chance-constraint that limits the probability of failure, and is commonly adopted by emergency medical services and fire departments as a performance measure. We use robust optimization and multiple periods to disentangle the uncertainty into two parts: estimation uncertainty, and temporal variation. A case study in the Portland, OR metropolitan area is analyzed for employing unmanned aerial vehicles (UAVs) or drones to deliver defibrillators in the region to combat out-of-hospital cardiac arrests. In our context, multiple periods represent periods with different wind speed and direction distributions. Further, we propose a novel Monte-Carlo simulation scheme to evaluate the value of adding robustness and multiple time periods.

- **Stratified knapsacks in equitable linear contextual bandits with knapsacks (E-linCBwK)**: The bandit with knapsacks (Agrawal and Devanur, 2016; Badanidiyuru et al., 2013) literature pose a limitation that the bandit episode ends when the first knapsack constraint is violated, independent of current resource consumption in other knapsack constraints. While this assumption is true when dealing with "globalized" knapsack constraints (i.e., exhaustion of any resource constraint would mean successful allocation can no longer happen. Alternatively, resource consumption in such constraints is almost always positive and non-zero, irrespective of the arm chosen). The assumption does not hold for, "stratified" constraints, i.e., exhaustion of a resource constraint does not necessarily imply that successful allocation can no longer happen. Or alternatively, resource consumption may be positive or zero and is dependent upon the choice of the arm. Additionally, the behavior of such constraints is already known based on the application. For example, consider 3 arms representing 3 facilities, and each facility has one type of resource which is used. This bandit problem will have 3 knapsack constraints, each one measuring the resource consumption at a single facility. Here, the exhaustion of resource at a facility does not mean that no further demand requests can be satisfied.

  We tackle this issue by introducing changing the context distribution after each set of stratified knapsacks is exhausted. The bandit algorithm has a "no-play" arm which can be used realize 0 reward and $\mathbf{0}$ resource consumption. Once a

set of stratified knapsacks is exhausted, we change the context distribution such that the corresponding arms would become unavailable for the rest of the time horizon. Essentially, this would mean that we are denying a demand request from our above illustrative example. Time-varying contexts would allow that context generation from a time-dependent unknown distribution. Therefore, we can modify the context such that the facility with an exhausted resource can no longer serve, and that the "no-play" is available only when no facility can serve a demand request.

- **Tackling budget loss during $Z$ estimation in equitable linear contextual bandits with knapsacks (E-linCBwK)**: A Lagrangian-like penalty parameter needs to be estimated in the linear contextual bandits with knapsacks (linCBwK) framework (Agrawal and Devanur, 2016) to relax the budget constraints and obtain upper confidence bounds for reward as in standard multi-armed bandit problems. For linCBwK, Agrawal and Devanur (2016) propose to a small multi-armed bandit instance with $T_0$ rounds, and having a uniform budget of $T_0$ (essentially, an unconstrained setting). Once, $Z$ is estimated, the learning happens over the remaining $(T - T_0)$ rounds with each knapsack constraint having a budget of $(B - T_0)$. However, in most instances, the budget $T_0$ allocated for $Z$ computation is rarely ever completely used. Wasting these resources would especially be detrimental in a scenario when resources are expensive or complete resource utilization is critical (example, medical sce-

narios, post-disaster conditions). Therefore, for our E-linCBwK framework, we consider that the remaining resources at the end of $Z$ computation for bandit learning after round $T_0$.

- **Electric truckload pickup and delivery problem**: We study a novel problem by considering fleet electrification in the conventional dynamic truckload pick-up and delivery problem. The aim of the problem is to tackle the common challenges which are related to fleet electrification like longer recharging times and limited charging infrastructure, which are more profound when considering electric freight vehicles. The important feature here is the consideration of recharging decisions during route construction for each dynamically-arriving request. For the computational analysis, we consider a service area of $200 \times 200$ miles which has five uncapacitated charging stations available. In the nominal scenario, we consider that the fleet consists of 10 electric freight vehicles and the request allocation decisions are made by a central command while incurring distance-based, time-based, recharging, and delay costs.

## 1.3 Organization

This work proposes contributions to planning and the real-time day-to-day operations of freight logistics systems. The contributions are also made broadly to the fields of emergency services management and humanitarian logistics, with our specific applications having a freight component associated with them. The chapters

having a planning component only are presented first, followed by chapters with both planning and real-time operations components, and finally, chapters with only real-time operations components are present. Finally, the dissertation concludes with a summary of the research and a discussion on avenues for future work.

Chapter 2 presents a robust multi-period maximum coverage facility location problem considering coverage reliability constraints. Coverage reliability constraints are a result of strict service standards adopted by emergency services like fire departments and ambulance services. We apply this model to combat cardiac events through the supply of automatic external defibrillators (AEDs) using AED-enabled unmanned aerial vehicles (UAVs). The UAVs are strategically deployed throughout the service region such that the coverage of the service is maximized subject to the coverage reliability constraints.

Chapter 3 presents a facility location and dynamic resource allocation problem for instant delivery logistics. With recent advancements in drone delivery infrastructure and the industry push towards instant delivery logistics, this chapter presents facility location and resource (product and battery capacity for UAVs) allocation at facilities. During the real-time operations, the allocated resources are utilized to meet demands arriving in a stochastic manner. We propose a multi-armed bandit formulation for tackling the real-time resource allocation problem. The goal is to maximize revenue generation while considering resource consumption constraints.

Chapter 4 presents a facility location and dynamic resource allocation problem

for humanitarian logistics. The chapter proposes an equitable relief prepositioning and distribution problem in post-disaster scenarios by building upon the framework developed in Chapter 3. A new multi-armed bandit framework explicitly considering an equity-based objective and resource budget constraints is proposed to solve the second stage dynamic resource allocation problem. We obtain a sub-linear performance regret bound of the proposed algorithm by showing that our problem can be modeled as an existing multi-armed bandit problem. Computational analyses are conducted on a case study based in the Portland Urban Metro area affected by the Cascadia Subduction Zone earthquake.

Chapter 5 proposes a new multi-armed bandit problem: linear contextual blocking bandits with context-dependent delays (linCBB). The problem is applied to the truckload pickup and delivery problem. A carrier operates a fleet of trucks in a prespecified region. Full truckload pickup and delivery jobs arrive in a stochastic manner. Every job allocation to a truck results in a reward and a "blocked" time when the truck cannot be used. We operate under the assumption that the reward and "blocked" times are linearly dependent on the context which consists of job and fleet characteristics.

# 2 Robust Multi-Period Maximum Coverage Drone Facility Location Considering Coverage Reliability

Cite the current chapter as:

Darshan R. Chauhan, Avinash Unnikrishnan, Miguel A. Figliozzi, Stephen D. Boyles, (2022), "*Robust Multi-Period Maximum Coverage Drone Facility Location Problem Considering Coverage Reliability*", Transportation Research Record: Journal of the Transportation Research Board, DOI: 10.1177/03611981221087240.

Author contributions:

The authors confirm contribution to the paper as follows: study conception and design: DRC, AU, MAF, SDB; mathematical formulation: DRC, AU; solution algorithms: DRC; data collection and coding: DRC; analysis and interpretation of results: DRC, AU, MAF, SDB; draft manuscript preparation: DRC; finalizing manuscript and revisions: DRC, AU, MAF, SDB. All authors reviewed the results and approved the final version of the manuscript.

## 2.1 Introduction

Public service agencies like hospitals, fire, rescue, and police departments are required to maintain high levels of service. For example, fire-related incidents require

90% reliability for a 4-minute response time (NFPA, 2020). Similarly, in the case of emergency medical services, the US Emergency Medical Services Act of 1997 requires a 95% response rate within 10 minutes (Lutter et al., 2017). In the United Kingdom, the National Health Service aims at serving 75% and 95% of demands in 8 and 14 minutes, respectively (Budge et al., 2010). As transportation systems are dynamic and stochastic an inherent uncertainty in travel time is present. This uncertainty in travel time leads to uncertainty in facility or demand coverage.

Drone or unmanned aerial vehicle (UAV) deliveries are being explored as a quicker, more cost-effective, and more reliable alternative for time-sensitive medical deliveries, emergency scenarios, humanitarian logistics, and other agricultural, security, and military applications (Ayamga et al., 2021; Nyaaba and Ayamga, 2021). Large corporations such as Amazon have secured operational licenses and begun field trials (Leonard, 2020). In addition, there is support from federal programs, such as the Federal Aviation Authority's UAS-BEYOND program (FAA, 2021), to test medical applications including delivery of automatic external defibrillators (AEDs), medical prescriptions, and medical emergency response. These medical applications are being field tested in the states of Nevada, North Carolina, and North Dakota, respectively.

Drones have some advantages when compared to traditional ground transportation modes. They can arrive faster by taking more direct paths and avoiding ground-based obstructions or congestion. For ground vehicles, congestion and associated

delays are key sources of travel time uncertainty. But for drone deliveries uncertainties arise because of weather conditions, mainly from uncertainty about wind speed and direction (Glick et al., 2022).

The effect of stochasticity in environmental factors on the performance of emergency departments is hard to quantify exactly, in addition to being data-intensive. However, reliable estimates for expected values (like, mean and variance) and extrema (like, minimum and maximum) are much easier to obtain. This is much more true for strategic decisions like facility location when the planning periods are longer. Robust optimization (RO) is a distribution-free approach that allows for incorporating stochasticity with limited information using uncertainty sets. The splitting of a planning period into multiple smaller periods would disaggregate uncertainties and possibly aid RO in tackling them.

This paper considers a robust multi-period maximum coverage facility location problem considering coverage reliability (MP-R) to improve decision-making. The coverage reliability constraints are captured using the chance constraints which provide probabilistic constraint satisfaction guarantee. The final model is developed by integrating the chance constrained approach with robust optimization, similar to Lutter et al. (2017), and expanded to multiple time periods. The contributions of this paper are:

- Developing a compact mixed-integer linear programming formulation for MP-R using polyhedral uncertainty sets (Bertsimas and Sim, 2004).

- Developing a case study in the Portland, OR metropolitan area to locate drone-launch sites to deliver defibrillators, considering uncertainty in travel times arising due to variation in wind speeds and directions.

- Analyzing the value of adding robustness and multiple time periods using a novel Monte-Carlo simulation scheme.

A brief literature review is presented in the next section, followed by the development of the mathematical model. The case study is developed and the computational analyses are discussed. Finally, the paper ends with brief conclusions and recommendations for future research.

## 2.2 Literature Review

A plethora of research has already been conducted in the field of emergency medical response. A vast majority of research has been focused around using ground vehicles (i.e., traditional ambulances) for optimizing coverage (Azizan et al., 2012; Enayati et al., 2018; Erdoğan et al., 2010; Schmid and Doerner, 2010), survival rates (Erkut et al., 2008; Zaffar et al., 2016), amount of relocation (Enayati et al., 2018; Naoum-Sawaya and Elhedhli, 2013; Schmid and Doerner, 2010), and crew shifts Erdoğan et al. (2010). Detailed literature reviews on ambulance location are presented in Aringhieri et al. (2017), Başar et al. (2012), and Li et al. (2011). Recently, there has been increasing interest around the usage of air-based vehicles for emergency med-

ical operations: AED-enabled drones for out-of-hospital cardiac arrests (Pulver and Wei, 2018; Pulver et al., 2016), drones supplying emergency relief packages (Chauhan et al., 2019, 2021), helicopters (Garner and van den Berg, 2017), and air ambulances (Røislien et al., 2017). This study focuses on locating AED-enabled drones for tackling out-of-hospital cardiac events in a planning region using a multi-period facility location formulation incorporating reliability in coverage.

Multi-period variants of traditional facility location problems have been studied for various contexts since the seminal work of (Ballou, 1968). Nickel and Saldanha-da Gama (2019) provides a review of multi-period facility location problems (MPFLP), and Vatsa and Jayaswal (2021) provides a brief review of studies considering uncertainties in MPFLP literature. Vatsa and Jayaswal (2021) note that while demand and cost uncertainties are widely tackled in the MPFLP literature, research tackling supply-side uncertainties (example, coverage capabilities) is relatively scarce. Kim et al. (2019) propose a MPFLP with drones considering uncertainty in flight distances. The study assumes that the probability of drone's successful return to the launch station is not time-period-dependent and that the time-periods are long enough that all drone trips complete in a time-period. Ghelichi et al. (2021) proposes a multi-stop drone location and scheduling problem for medical supply delivery. The study assumes deterministic travel speed for drones (i.e., ignoring weather conditions) in multiple periods, and time-periods are short and a drone-trip is assumed to last over multiple time periods. Our study assumes that the probability of timely arrival

at a demand location from a launch site is dependent on the time period, and that the time-periods are long enough that drones trips can be completed in a time period.

Erdoğan et al. (2010) state that appropriately defining coverage and incorporating uncertainty in travel times are the most important considerations in ambulance location. This study defines coverage based on the importance of covering the demand point. Therefore, the coverage importance metric can be a function of various population parameters like size and demographics, and other characteristics like history of emergency requests and equity considerations. Additionally, in most regions, the emergency response systems are required to maintain adequate service standards. We model the service standard reliability constraint as a chance constraint on probability of timely arrival for each demand point. Therefore, a demand point is considered covered only if the service standard reliability requirements are met for all time periods of the planning period.

The probability of timely arrival at a demand point is linked to the uncertainty in drone travel times which stems from variations in wind speed and directions. Due to dependency on environmental factors, the estimated values of probabilities of timely arrival are not deterministic, rather uncertain. Tackling parameter uncertainty has been a focus of the mathematical programming community for a long time. Two major approaches exist for tackling uncertainty: stochastic optimization (SO) and robust optimization (RO). SO assumes that a probability distribution of the uncertainty is available, whereas RO assumes no underlying distribution of the uncertainty

and considers it to be deterministic and set-based (Ben-Tal et al., 2009; Bertsimas et al., 2011). A set-based uncertainty structure of RO leads to better computational tractability than SO (Bertsimas et al., 2011). RO immunizes the solution from any manifestation of uncertainty in the described uncertainty set. In general, the larger the size of the uncertainty set, the lower is the objective value (considering maximization objective) and the lower is the probability of constraint violation (Bertsimas et al., 2011). This trade-off between expected objective values and constraint violation can be controlled by varying the size of the uncertainty set. Here, we use RO using polyhedral uncertainty sets (Bertsimas and Sim, 2004) to tackle uncertainty while maintaining computational tractability. This approach ensures that the robust counterpart of our linear optimization problem is also linear. We refer the interested reader to (Ben-Tal et al., 2009; Bertsimas et al., 2011; Delage and Ye, 2010; Gabrel et al., 2014; Goh and Sim, 2010; Gorissen et al., 2015) for a more comprehensive picture of RO.

## 2.3 Problem Description

This section first describes the modeling of the coverage reliability constraint and its assumptions. Later, we formulate a deterministic multi-period maximum coverage facility location problem with coverage reliability (abbreviated as MP-D). Finally, we provide a robust formulation of MP-D (abbreviated as MP-R) which accounts for uncertainty in the values of coverage failure probabilities.

Consider a set of demand points (represented as $I$) each with coverage importance $c_i$, a set of facilities (represented as $J$), and a set of all time periods (represented as $T$). Let $A$ be a $|I| \times |J|$ 1-0 accessibility matrix describing if the demand point $i$ can be covered by a facility $j$. We use $a_{ij}^t$ to represent the probabilistic nature of the $(i, j)$ element of $A$ in time period $t \in T$, while, $A_{ij}$ is used for the deterministic initial state of $(i, j)$ element of the matrix $A$. More specifically, if $A_{ij} = 1$, then, $a_{ij}^t = 1$ with probability $(1 - p_{ij}^t)$, and $a_{ij}^t = 0$ with probability $p_{ij}^t$. If $A_{ij} = 0$, then, $a_{ij}^t = 0$ always. Let, $\bar{p}_{ij}^t$ be our estimate of $p_{ij}^t$. Now, the service reliability requirement of achieving a service standard $\alpha$ can then be stated as

$$Pr\left[\sum_{j \in S_i} a_{ij}^t \geq 1\right] \geq \alpha\,, \tag{2.3.1}$$

where $S_i = \{j \in J \,|\, A_{ij} = 1\}$. The above equation potentially considers all the facilities that can access demand point $i \in I$. As a consequence, we assume that all the accessible facilities respond to the demand at location $i$. Under the assumption of independence among the values in $A$, equation (2.3.1) can modified as

$$Pr\left[\sum_{j \in S_i} a_{ij}^t \geq 1\right] = 1 - \prod_{j \in S_i} p_{ij}^t \equiv 1 - \prod_{j \in S_i} \bar{p}_{ij}^t \geq \alpha \tag{2.3.2}$$

For the above discussion, we have assumed that $\bar{p}_{ij}^t$ completely describe the distribution of variables $a_{ij}^t$. However, there are errors endemic to sampling (environmental factors) and measurement while estimating the value of $p_{ij}^t$. Therefore, the values of

$p_{ij}^t$ may not be known with complete certainty. We tackle this issue while formulating the MP-R model.

For MP-D and MP-R, the decision-making agency wishes to locate a maximum of $q$ facilities in each time-period to maximize the cumulative coverage importance achieved subject to coverage requirements described. Additionally, opened facility locations can be shifted between time periods subject to a facility relocation cost budget constraint.

### 2.3.1 Nomenclature

Sets and Indices

| | |
|---|---|
| $I$ | Set of all demand points ($i \in I$) |
| $J$ | Set of all candidate facility locations ($j, k \in J$) |
| $T$ | Set of all time periods ($t \in T := \{1, 2, \ldots, \lvert T \rvert\}$) |

Parameters

| | |
|---|---|
| $c_i$ | Coverage importance of demand point $i \in I$; $c_i \geq 0$ |
| $A_{ij}$ | 1, if the demand point $i \in I$ can be covered by facility $j \in J$, and 0, otherwise |
| $S_i$ | Set of facilities $j \in J$ that can cover the demand point $i \in I$; $S_i = \{j \in J \mid A_{ij} = 1\} \; \forall \; i \in I$ |
| $\bar{p}_{ij}^t$ | Nominal probability of failure of covering demand point $i \in I$ by facility $j \in J$ in time period $t \in T$; $0 < p_{ij}^t \leq 1$ |
| $\hat{p}_{ij}^t$ | Maximum deviation from nominal probability of failure of covering demand point $i \in I$ by facility $j \in J$ in time period $t \in T$; $0 \leq \hat{p}_{ij}^t < p_{ij}^t + \hat{p}_{ij}^t \leq 1$ |
| $q$ | Maximum number of facilities that can be located; $q \in \mathbb{Z}^+ \cup \{0\}$ |
| $\alpha$ | Required coverage threshold; $0 \leq \alpha \leq 1$ |
| $\Gamma_i^t$ | Maximum number of delivery paths to demand point $i$ that can achieve worst-case probability of failure simultaneously in time period $t \in T$; $\Gamma_i^t \in \mathbb{Z}^+ \cup \{0\}$ |
| $f_{jk}^t$ | Cost associated with shifting the facility from location $j \in J$ to location $k \in J$ at the beginning of time period $t \in T$ |
| $B$ | Facility shifting cost budget |

Decision Variables

$x_i$ 1, if demand location $i \in I$ is covered with given coverage threshold; 0, otherwise

$y_j^t$ 1, if candidate facility location $j \in J$ is open during time period $t \in T$ ; and 0, otherwise

$z_{jk}^t$ 1, if a facility is moved from location $j \in J$ to location $k \in J$ at the beginning of time period $t \in T\backslash\{1\}$; and 0, otherwise

### 2.3.2 Deterministic Formulation

Objective:

$$\max_{x,y,z} \quad \sum_{i \in I} c_i x_i \tag{2.3.3}$$

Subject to:

$$\prod_{j \in S_i} (\bar{p}_{ij}^t)^{y_j^t} \leq (1-\alpha)^{x_i} \qquad \forall \ i \in I, t \in T \tag{2.3.4}$$

$$\sum_{j \in J} y_j^t \leq q \qquad \forall \ t \in T \tag{2.3.5}$$

$$\sum_{t \in T\backslash\{1\}} \sum_{j \in J} \sum_{k \in J} f_{jk}^t z_{jk}^t \leq B \tag{2.3.6}$$

$$\sum_{k \in J} z_{jk}^t = y_j^{t-1} \qquad \forall \ j \in J, t \in T\backslash\{1\} \tag{2.3.7}$$

$$\sum_{j \in J} z_{jk}^t = y_k^t \qquad \forall\ k \in J, t \in T \backslash \{1\} \tag{2.3.8}$$

$$x_i \in \{0,1\} \qquad \forall\ i \in I \tag{2.3.9}$$

$$y_j^t \in \{0,1\} \qquad \forall\ j \in J, t \in T \tag{2.3.10}$$

$$z_{jk}^t \in \{0,1\} \qquad \forall\ j,k \in J, t \in T \backslash \{1\} \tag{2.3.11}$$

For the deterministic formulation, we assume that $p_{ij}^t = \bar{p}_{ij}^t$. Equation (2.3.3) represents maximizing coverage importance. In equation (2.3.4), the demand point $i \in I$ is covered only if the probability of failure to cover it is less than $(1 - \alpha)$ for all time periods $t \in T$. Note that all accessible open facilities respond to meet the demand at point $i \in I$. Equation (2.3.5) enforces that no more than $q$ facilities can be opened.

Equation (2.3.6) is a generalized cost constraint relating to the shifting facility locations. Note that using $f_{jj}^t = 0$ for all $j \in J, t \in T \backslash \{1\}$, and 1, otherwise, would limit the total number of facility location shifts to $B$. Equations (2.3.7) and (2.3.8) are transportation allocation constraints. Equations (2.3.9)–(2.3.11) are variable definitions. However, the formulation is not linear due to equation (2.3.4). Applying logarithm function on both sides of equation (2.3.4) yields:

$$\sum_{j \in S_i} w_{ij}^t y_j^t \leq \beta x_i \qquad \forall\ i \in I, t \in T \tag{2.3.12}$$

where $w_{ij}^t$ and $\beta$ represent $\log(\bar{p}_{ij}^t)$ and $\log(1-\alpha)$, respectively. The above formulation (equations 2.3.3 and 2.3.5–2.3.12) is referred to as the deterministic multi-period facility location problem considering coverage reliability, abbreviated as MP-D. MP-D is an integer linear program and can be solved using standard MIP solvers.

### 2.3.3 Robust Formulation

The parameter $p_{ij}^t$ represents the probability that the facility $j \in J$, in time period $t \in T$, will fail to cover the demand point $i \in I$ in a given service time threshold $\tau$. However, due to sampling errors stemming from environmental factors like variations in travel times throughout the day, the estimated values of parameters $p_{ij}^t$ are uncertain. Later, in the presented case study of delivering AED-enabled drones, this variation occurs primarily due to changing wind speeds and directions. As the complete probability distribution of $p_{ij}^t$ is arduous to obtain in comparison to the bounds of its variation, we use a robust optimization using polyhedral uncertainty sets (Bertsimas and Sim, 2004) to incorporate this uncertainty. Let, $\hat{p}_{ij}^t$ be the maximum deviation of $\bar{p}_{ij}^t$. For our robust model, we assume that $p_{ij}^t \in [\bar{p}_{ij}^t - \hat{p}_{ij}^t, \bar{p}_{ij}^t + \hat{p}_{ij}^t]$. Of all facilities servicing demand point $i$, up to $\Gamma_i^t$ facilities observe worst-case failure probabilities (i.e., $p_{ij}^t = \bar{p}_{ij}^t + \hat{p}_{ij}^t$), whereas the rest observe nominal failure probabilities (i.e., $p_{ij}^t = \bar{p}_{ij}^t$). This allocation happens in such a way that the probability of failing to serve demand point $i$ is maximized.

Objective:

$$\max_{x,y,z} \quad \sum_{i \in I} c_i x_i \tag{2.3.13}$$

Subject to:

$$\max_{\{U \subseteq S_i, |U| \leq \Gamma_i\}} \left[ \prod_{j \in U} (\bar{p}_{ij}^t + \hat{p}_{ij}^t)^{y_j^t} \prod_{j \in S_i \setminus U} (\bar{p}_{ij}^t)^{y_j^t} \right] \leq (1-\alpha)^{x_i} \quad \forall \ i \in I, t \in T \tag{2.3.14}$$

$$\sum_{j \in J} y_j^t \leq q \qquad \forall \ t \in T \tag{2.3.15}$$

$$\sum_{t \in T \setminus \{1\}} \sum_{j \in J} \sum_{k \in J} f_{jk}^t z_{jk}^t \leq B \tag{2.3.16}$$

$$\sum_{k \in J} z_{jk}^t = y_j^{t-1} \qquad \forall \ j \in J, t \in T \setminus \{1\} \tag{2.3.17}$$

$$\sum_{j \in J} z_{jk}^t = y_k^t \qquad \forall \ k \in J, t \in T \setminus \{1\} \tag{2.3.18}$$

$$x_i \in \{0,1\} \qquad \forall \ i \in I \tag{2.3.19}$$

$$y_j^t \in \{0,1\} \qquad \forall \ j \in J, t \in T \tag{2.3.20}$$

$$z_{jk}^t \in \{0,1\} \qquad \forall \ j,k \in J, t \in T \setminus \{1\} \tag{2.3.21}$$

Equation (2.3.13) represents the maximization of coverage importance. Incorporating uncertainty in the failure probabilities in equation (2.3.4) yields (2.3.14). The left hand side (lhs) of equation (2.3.14) seeks to find the absolute worst-case probability of failure such that at most $\Gamma_i$ facilities servicing the demand point $i \in I$

can individually observe worst-case failure probability. Demand point $i \in I$ is considered covered only if the left-hand side of equation (2.3.14) is less than $(1-\alpha)$. Generally, incorporating robustness into a problem imparts conservatism by realizing worst-case objective value subject to certain criteria (Ben-Tal and Nemirovski, 1999; Bertsimas and Sim, 2004). This leads to the robustness sub-problem being in conflict with the overall objective. Here, worst-case realizations of failure probability in equation (2.3.14) reduce the chance of the demand point $i$ being covered, and while the overall objective (2.3.13) want to increase the chances of demand point $i$ being covered. In other words, the current formulation is a bilevel optimization problem which cannot be solved directly using MIP solvers. Dualizing the robustness sub-problem would overcome this issue and align both objectives correctly, and yield a single level mixed-integer linear problem. Equations (2.3.15)-(2.3.21) have the same meaning as equations (2.3.5)-(2.3.11). Taking the logarithm of (2.3.14) yields:

$$\max_{\{U \subseteq S_i, |U| \leq \Gamma_i\}} \left[ \sum_{j \in U} \log(\bar{p}_{ij}^t + \hat{p}_{ij}^t) \cdot y_j^t + \sum_{j \in S_i \setminus U} \log(\bar{p}_{ij}^t) \cdot y_j^t \right] \leq \log(1-\alpha) \cdot x_i, \quad \forall\ i \in I, t \in T \tag{2.3.22}$$

Let, $\hat{w}_{ij}^t$, $w_{ij}^t$, and $\beta$ represent $\log(\bar{p}_{ij}^t + \hat{p}_{ij}^t)$, $\log(\bar{p}_{ij}^t)$, and $\log(1-\alpha)$, respectively. Note that $\hat{w}_{ij}^t \geq w_{ij}^t$. Rewriting $\hat{w}_{ij}^t$ as $w_{ij}^t + (\hat{w}_{ij}^t - w_{ij}^t)$, we re-write equation (2.3.22)

as:

$$\sum_{j \in S_i} w_{ij}^t y_j^t + \max_{\{U \subseteq S_i, |U| \leq \Gamma_i\}} \left[ \sum_{j \in U} (\hat{w}_{ij}^t - w_{ij}^t) \cdot y_j^t \right] \leq \beta x_i \quad \forall \ i \in I, t \in T \tag{2.3.23}$$

The optimization problem described on the lhs of equation (2.3.23) can be written as:

For each $i \in I$, $t \in T$:

$$SP_i^t : \quad \max_{\gamma} \quad \sum_{j \in S_i} w_{ij}^t y_j^t + \sum_{j \in S_i} (\hat{w}_{ij}^t - w_{ij}^t) y_j^t \gamma_{ij}^t \tag{2.3.24}$$

$$\sum_{j \in S_i} \gamma_{ij}^t \leq \Gamma_i^t \tag{2.3.25}$$

$$\gamma_{ij}^t \in \{0, 1\} \qquad \forall \ j \in S_i \tag{2.3.26}$$

The constraint coefficient matrix of the above sub-problem is totally unimodular, and $\Gamma_i^t$ are non-negative integer values. Therefore, $\gamma_{ij}^t$ can be linearized to the interval [0,1] without loss of optimality. Let, $\theta_i^t$ and $\sigma_{ij}^t$ be the dual variables associated with equations (2.3.25) and the upper bound of equation (2.3.26), respectively. Taking the dual of the formulation represented by equations (2.3.24)-(2.3.26), yields:

For each $i \in I$, $t \in T$:

$$SPD_i^t : \quad \min_{\sigma, \theta} \quad \sum_{j \in S_i} w_{ij}^t y_j^t + \sum_{j \in S_i} \sigma_{ij}^t + \Gamma_i^t \theta_i^t \tag{2.3.27}$$

$$\sigma_{ij}^t + \theta_i^t \geq (\hat{w}_{ij}^t - w_{ij}^t) y_j \qquad \forall\ j \in S_i \tag{2.3.28}$$

$$\sigma_{ij}^t \geq 0 \qquad \forall\ j \in S_i \tag{2.3.29}$$

$$\theta_i^t \geq 0 \tag{2.3.30}$$

Strong duality, along with the totally unimodular property, ensures that problems $SPD_i^t$ (equations (2.3.27)-(2.3.30)) and $SP_i^t$ (equations (2.3.24)-(2.3.26)), and consequently also the lhs of equation (2.3.14), are equivalent. Incorporating $SPD_i^t$ in the equation (2.3.23), updates the robust formulation (equations (2.3.13), (2.3.15)-(2.3.21), (2.3.23)) to:

Objective:

$$\max_{x,y,z,\sigma,\theta} \quad \sum_{i \in I} c_i x_i \tag{2.3.31}$$

Subject to:

$$\sum_{j \in S_i} w_{ij}^t y_j^t + \sum_{j \in S_i} \sigma_{ij}^t + \Gamma_i^t \theta_i^t \leq \beta x_i \qquad \forall\ i \in I, t \in T \tag{2.3.32}$$

$$\sigma_{ij}^t + \theta_i^t \geq (\hat{w}_{ij}^t - w_{ij}^t) y_j^t \qquad \forall\ j \in S_i,\ i \in I, t \in T \tag{2.3.33}$$

$$\sum_{j \in J} y_j^t \leq q \qquad \forall\ t \in T \tag{2.3.34}$$

$$\sum_{t \in T \setminus \{1\}} \sum_{j \in J} \sum_{k \in J} f_{jk}^t z_{jk}^t \leq B \tag{2.3.35}$$

$$\sum_{k \in J} z_{jk}^{t} = y_{j}^{t-1} \qquad \forall\ j \in J, t \in T \backslash \{1\} \qquad (2.3.36)$$

$$\sum_{j \in J} z_{jk}^{t} = y_{k}^{t} \qquad \forall\ k \in J, t \in T \backslash \{1\} \qquad (2.3.37)$$

$$x_i \in \{0,1\} \qquad \forall\ i \in I \qquad (2.3.38)$$

$$y_j^t \in \{0,1\} \qquad \forall\ j \in J, t \in T \qquad (2.3.39)$$

$$z_{jk}^t \in \{0,1\} \qquad \forall\ j,k \in J, t \in T \backslash \{1\} \qquad (2.3.40)$$

$$\sigma_{ij}^t \geq 0 \qquad \forall\ i \in I, j \in J, t \in T \qquad (2.3.41)$$

$$\theta_i^t \geq 0 \qquad \forall\ i \in I, t \in T \qquad (2.3.42)$$

The above formulation is referred to as the robust maximum coverage facility location problem considering coverage reliability, abbreviated as MP-R. MP-R is a mixed-integer linear program and can be solved using open-source or commercially-available MIP solvers. For cases when $|T|$ is large, the computational times using a MIP solver could be prohibitively large. The authors recommend decomposition-based methodologies for such cases. For example, applying Lagrangian relaxation to equations (2.3.32) and (2.3.33) decomposes MP-R into four sub-problems, of which three can be trivially solved. The development of computationally-efficient heuristics is left as a future research endeavor.

## 2.4 Computational Analysis

This section first describes the experimental setting of the case study conducted in Portland, OR metropolitan area. Later, three types of analysis are conducted: computational performance, the evaluating the value of considering robustness and multiple periods using a Monte Carlo simulation scheme, and finally, incorporating equity in decision-making.

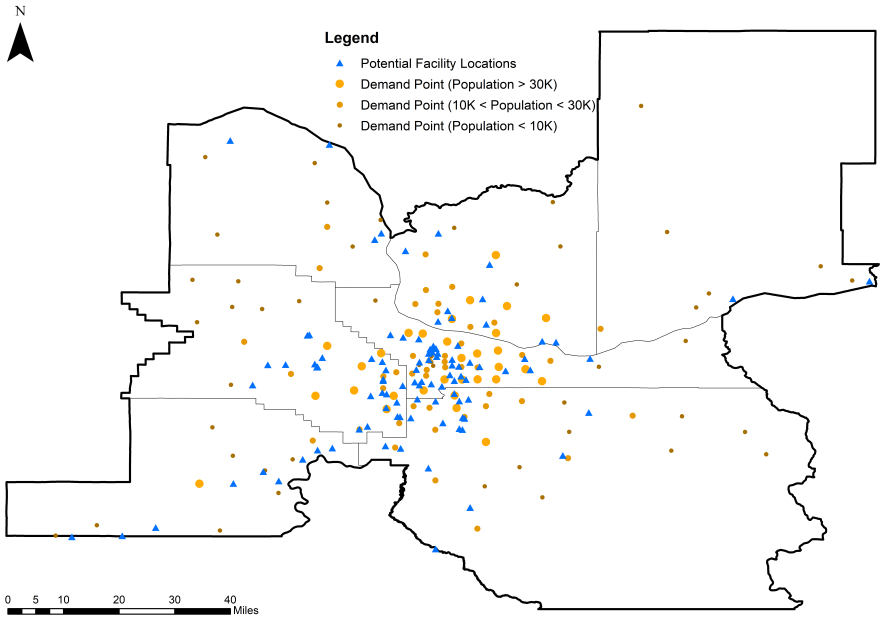



Figure 2.1: Locations of demand points and facility locations in Portland Metro Area

The feasibility of using UAVs or drones for delivering defibrillators to demand points in the Portland, OR metropolitan area is evaluated here. The Portland Metro service area consists of 122 ZIP Code Tabulation Areas (ZCTA) which act as demand points, and 104 community centers which act as potential launch sites as detailed in

Chauhan et al. (2019) and shown in Figure 2.1. We evaluate drones against two service standards: the National Fire Protection Association's emergency response standard of providing coverage reliability of 90% within in a response time of 4 minutes (NFPA, 2020), abbreviated as $SS1$; and, the 1997 US Emergency Medical Services Act service response standard of providing coverage reliability of 95% within a response time of 10 minutes (Lutter et al., 2017), abbreviated as $SS2$. Two service standards are selected to evaluate the effect of increasing response time on system performance and the value of data disaggregation using multiple time periods. All drones are equipped with an AED which weighs 1.5 kg each (Philips, 2022). A major factor leading to uncertainty in drone response times is wind speed and direction. The calculation of bounds of probability of failure (lower bound: $p\_best$; upper bound: $p\_worst$) for delivering from a launch site to a demand point is carried out using procedure described in Algorithm 2.1, similar to Glick et al. (2022), with sample size $n = 10,000$. The upper bound of probability of failure ($p\_worst$) is considered as worst-case probability ($\bar{p} + \hat{p}$). The nominal probability of failure ($\bar{p}$) is an average of bounds of variation weighted according to the distribution of wind directions.

A demand point is considered accessible by a launch site if the following two conditions are met. First, the amount of battery expended to go to the demand point and come back is less than the total available battery in the nominal scenario (calculated using the formula provided in Figliozzi (2017)). The total battery capacity of the drone is divided in two parts: total available battery and battery safety

**Algorithm 2.1** Calculating bounds of probability of failure

**Input:** Input sample size $n$, wind speed and direction distributions for each time period $t \in T$, maximum possible wind speed ($v\_wind\_max$), probability distribution of wind directions, response time ($\tau$), drone travel speed ($v\_drone$), and distance ($dist\_act$) and delivery angle from facility $j$ to demand point $i$.

**Output:** $p\_nominal, p\_worst$

1: Calculate $wt[i, j, t]$ which is the probability that the wind direction is not aligned with the delivery direction (i.e. difference is greater than 90°) from facility $j$ to demand point $i$ in time period $t$ using the input information. $t \in T$

2: Generate $windspeeds[t]$, an array of size $n$, following a lognormal distribution with given input parameters and a maximum value of $v\_wind\_max$.

3: $dist\_best = (v\_drone + windspeeds[t]) \cdot \tau$

4: $dist\_worst = (v\_drone - windspeeds[t]) \cdot \tau \; i \in I \; j \in J$

5: $p\_best[i, j, t] = \max\{len(where(dist\_best < dist\_act[i, j])),\ 1\}/n$

6: $p\_worst[i, j, t] = \max\{len(where(dist\_worst < dist\_act[i, j])),\ 1\}/n$

7: $p\_nominal[i, j, t] = (1 - wt[i, j, t]) \cdot p\_best[i, j, t] + wt[i, j, t] \cdot p\_worst[i, j, t]$

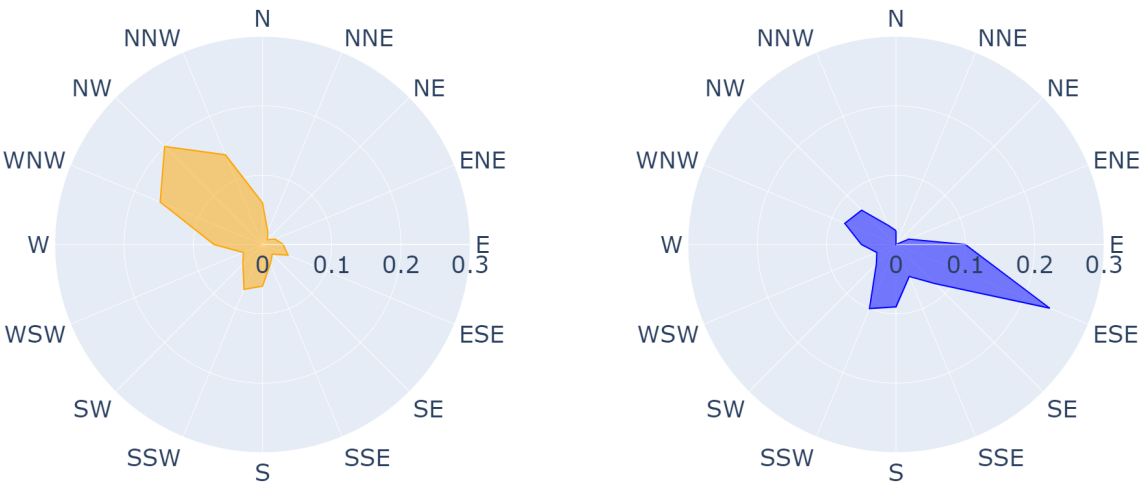
factor. As in Chauhan et al. (2019), we assume that drones ignore obstacles in urban landscape and travel over Euclidean distances, and that the energy consumed in VTOL operations are accommodated in battery safety factor. Second, the time required to reach the demand point in the most favorable wind direction and speed is less than the provided response time. The coverage importance metric is dependent on the normalized population of the demand point. The ZCTA population estimates for the demand points were adopted from 2017 American Community Survey 5-year estimates (Census-Bureau and American-FactFinder, 2018).

The summary of parameter specifications is provided below (Figliozzi, 2017; Glick et al., 2022).

- Maximum available battery: 777 Wh
- Battery Safety Factor: 20% of maximum available battery (Total available bat-

tery = maximum available battery − battery safety factor = 621.6 Wh)

- Sum of drone tare and battery mass: 10.1 kg
- Lift-to-drag ratio: 2.8445
- Total power transfer efficiency: 0.66
- Nominal travel speed of drone: 20 meters per second (mps)
- Maximum number of drones serving demand point $i$ in time period $t$ that can achieve worst-case probability of failure ($\Gamma_i^t$): 1
- $c_i$ = Normalized population of demand point $i$
  $= \left\lceil \dfrac{100 \times \text{population of demand point } i}{\text{maximum population of demand points}} \right\rceil$, where $\lceil u \rceil$ represents the ceiling function, i.e. the least integer greater than or equal to $u$.
- Wind speed and direction distributions (see Figure 2.2) are available openly at https://github.com/drc1807/RMP-MCFLP-CR
- Maximum possible wind speed: 68 miles per hour (30.3987 mps)
- The planning period is one year. In Portland, the wind direction is primarily in the NW direction in the summer months (April through September). Whereas, in winter months (October through March), the wind primarily flows in the ESE direction (see Figure 2.2). We investigate the value of using a multi-period formulation with $T = \{\text{Summer, Winter}\}$ over a single-period formulation with $T = \{\text{Whole Year}\}$.




(a) Summer (b) Winter

Figure 2.2: Wind direction distribution in Portland, OR

The model coverage (in %) is given as:

$$\text{Model Coverage} = \frac{\text{Objective Value of Model}}{\sum_{i \in I} c_i} \times 100 \tag{2.4.1}$$

The cost of shifting a facility located at $j \in J$ in time period $(t-1)$ to a location $k \in J, k \neq j$ in time period $t$ is considered to be 1, and 0, if the location does not change. Alternatively put, $f_{jj}^t = 0 \; \forall \; j \in J, t \in T \backslash \{1\}$, and 1, otherwise. This limits the total number of facility location shifts to the facility shifting cost budget $B$. The default value of $B$ used here is $\lfloor 0.35q \rfloor$, where $q$ is the maximum number of drone launch sites that can be opened.

The experiments are performed on four models: MP-R, MP-D, MP-R with $T = \{\text{Whole Year}\}$ (abbreviated as SP-R), and MP-D with $T = \{\text{Whole Year}\}$ (ab-

breviated as SP-D) considering a planning period of a whole year. Models are solved using Gurobi Gurobi Optimization (2020) in Python interface on a Windows 10 desktop with Intel i7-7700K processor and CPU specifications of 3.6 GHz, 4 cores, 8 logical processors, and 32 GB of RAM. Experiments to evaluate the computational efficiency with an increasing number of drone launch sites ($q$) are conducted, followed by the evaluation of the value added by robustness and granularity of information (through multiple time periods). Additionally, the effect of adding equity in decision-making is explored.

### 2.4.1 Computational Efficiency

Prohibitive computational times can often be a barrier to model adoption in real life. In our case, the planning period is fairly large (a whole year), and therefore, no computational time limit was adopted for Gurobi. All the four models, for both service standards and given default values of parameters, converged in less than 2 hours for a range of $q$ values, indicating that the development of time-efficient heuristics was not required. The model coverage values with their computational times are provided in Table 2.1.

Table 2.1: Computational Efficiency

| Service Standard | $q$ | Model Coverage (%) | | | | Computational Time (sec) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MP-R | MP-D | SP-R | SP-D | MP-R | MP-D | SP-R | SP-D |
| SS1 | 3 | 12.57 | 15.76 | 12.9 | 15.76 | 32 | 6 | 2 | 1 |
| | 6 | 24.37 | 27.41 | 24.64 | 28.19 | 109 | 15 | 3 | 1 |
| | 9 | 33.52 | 37.93 | 34.62 | 38.8 | 416 | 17 | 3 | 1 |
| | 12 | 41.9 | 47.02 | 42.64 | 48.36 | 444 | 33 | 2 | 1 |
| | 15 | 49.28 | 53.9 | 49.37 | 55.6 | 44 | 23 | 2 | 1 |
| | 20 | 55.45 | 62.31 | 55.21 | 62.99 | 49 | 8 | 2 | 1 |
| | 25 | 59.92 | 67.19 | 59.09 | 67.4 | 44 | 4 | 2 | 1 |
| | 30 | 62.28 | 68.36 | 62.1 | 68.56 | 5 | 4 | 1 | 1 |
| | 35 | 62.28 | 68.5 | 62.28 | 68.71 | 2 | 1 | 1 | 1 |
| SS2 | 3 | 35.37 | 44.85 | 39.09 | 46.31 | 5736 | 152 | 38 | 2 |
| | 6 | 62.69 | 69.99 | 64.75 | 72.53 | 3937 | 152 | 64 | 1 |
| | 9 | 75.8 | 82.18 | 77.26 | 83.02 | 1342 | 26 | 44 | 1 |
| | 12 | 82.6 | 86.23 | 83.28 | 87.84 | 569 | 40 | 67 | 1 |
| | 15 | 87.01 | 89.6 | 87.19 | 90.91 | 694 | 44 | 32 | 1 |
| | 20 | 90.73 | 93.38 | 90.88 | 93.95 | 274 | 108 | 20 | 1 |
| | 25 | 92.58 | 94.73 | 92.76 | 95.68 | 45 | 48 | 9 | 1 |
| | 30 | 93.21 | 95.23 | 93.36 | 96.39 | 31 | 2 | 14 | 1 |
| | 35 | 93.38 | 95.29 | 93.53 | 96.63 | 2 | 1 | 1 | 1 |

Note:
SS1 is providing 90% coverage reliability in a response time of 4 minutes
SS2 is providing 95% coverage reliability in a response time of 10 minutes

The effect of adding additional time periods is found to be more profound than the effect of adding robustness to the formulation. On average, for SS1, adding robustness increases computational time by 5.2 times, whereas adding additional time-period increases computational times by 37.0 times. For SS2, these values are 24.5 times and 49.5 times, respectively. The primary reason behind this is the number of constraints added to the model. A multi-period formulation requires the facility transfer variables $z$ which adds $2 \cdot |J| \cdot |T \backslash \{1\}|$ facility matching equality constraints

along with a facility relocation budget constraint. Additionally, $|T| - 1$ simultaneous coverage reliability constraints are also added which further deteriorates computational performance. On the other hand, adding robustness adds more variables and constraints to the model, but the constraints are computationally simpler. The accessibility matrix $A$ is more sparse for the SS1 models than SS2 models, which leads to better computational performance.

The addition of multiple periods to the formulation decreases the model coverage by a little amount (0.8% on average). This is because the satisfaction of multiple coverage reliability constraints is required for demand point coverage. As expected, adding robustness decreases the model coverage by a significant amount (4.9% on average) as a consequence of accounting for worst-case scenarios.

### 2.4.2 Value of developing multi-period formulation and adding robustness

The value of using more information (through adding robustness and multiple time periods) in a model is evaluated in this section utilizing a Monte-Carlo simulation-based (MCS) framework. Generally, adding robustness to a formulation reduces the model coverage but should provide for better real-life performance, thereby reducing the gap between what is expected (model coverage) and what happens (simulated coverage). Similarly, having potentially different facility location layouts in different time periods should boost simulated coverage.

An MCS framework is proposed to quantify the value of using additional information. In an MCS scenario $s$, the time period is $t^s$, and $n = 1000$ values of wind directions and speeds are randomly generated. In our case, an MCS scenario $s$ can be thought of as a day of the year, and $n$ is the number of wind speed and direction observations made throughout the day. Therefore, for multi-period formulations, $t^s =$ 'Summer' with probability $\frac{183}{365}$ and $t^s =$ 'Winter' with probability $\frac{182}{365}$. For single-period formulations, $t^s =$ 'Whole Year' with probability 1. Depending on the value of $t^s$, the wind speeds are generated as in Algorithm 2.1 and wind directions are chosen as per the distributions of the time period. These angles and speeds are combined with the originally projected delivery angles and nominal drone delivery speed to find effective drone speed. The effective drone speeds are then utilized to determine the realizations of the probability of failure for the scenarios ($\tilde{p}^s$).

The solutions obtained from the robust and the deterministic formulations for the variable $y$ are denoted by $y^*$. The new values for the variable $x$ (denoted by $\tilde{x}$) and the actual coverage are calculated using $y^*$ and $\tilde{p}^s$. For multi-period formulation, the facility location layout is determined by the simulation time period $t^s$. A total of 100 MCS scenarios are evaluated and the algorithm for the described MCS is detailed in Algorithm 2.2.

**Algorithm 2.2** Monte Carlo simulation for evaluating coverage

**Input:** Input number of MCS scenarios ($MCS\_s$), number of wind speed and direction observations per scenario ($n$), probability distribution of time periods $t \in T$ ($\pi$), other model input parameters.

**Output:** $simulated\_coverage$

1: Solve the model and determine $y^*$, the optimum values of decision variable $y$
2: Determine $J_i$, the set of open and accessible facilities for each demand point $i \in I$
3: $s = 1$
4: $simulated\_coverage = zeros(MCS\_s)$ $s \leq MCS\_s$
5: Randomly select simulation time period $t^s$, such that $t^s = t$ with probability $\pi_t$
6: Generate $windspeeds[t^s]$, an array of $n$ elements, as in Algorithm 2.1
7: Generate $windangles[t^s]$, an array of $n$ elements, based on the probability distribution in time period $t^s$
8: Determine effective delivery angles and effective drone speeds using vector algebra.
9: For each $i, j$ combination, calculate $dist\_cov_{ij}^s$, an array of length $n$ describing distances covered by drones using effective delivery angles and effective drone speeds.
10: For each $i, j$ combination, calculate $\tilde{p}_{ij}^s = length(where(dist\_cov_{ij}^s < dist\_act[i,j]))/n$
11: $\tilde{w}^s = \log(\tilde{p}^s)$
12: $\tilde{x}^s = zeros(length(I))$ $i \in I$ $\sum_{j \in J_i} \tilde{w}_{ij}^s y_j^* \leq \beta$
13: $\tilde{x}_i^s = 1$
14: $simulated\_coverage[s] = \dfrac{\sum_{i \in I} c_i \tilde{x}_i^s}{\sum_{i \in I} c_i} \times 100$
15: $s \mathrel{+}= 1$

The simulated coverage values for all four models with default values are presented in Table 2.2. For SS1 (providing 90% coverage reliability in 4 minutes), extending to multi-period formulation improves average simulated coverage by 0.29 times on average for robust models, and by 0.41 times on average for deterministic models. Whereas for SS2 (providing 95% coverage reliability in 10 minutes), the improvements in average simulated coverage are by 0.02 times for robust models, and by 0.24 times for deterministic models. The improvements are higher when the response

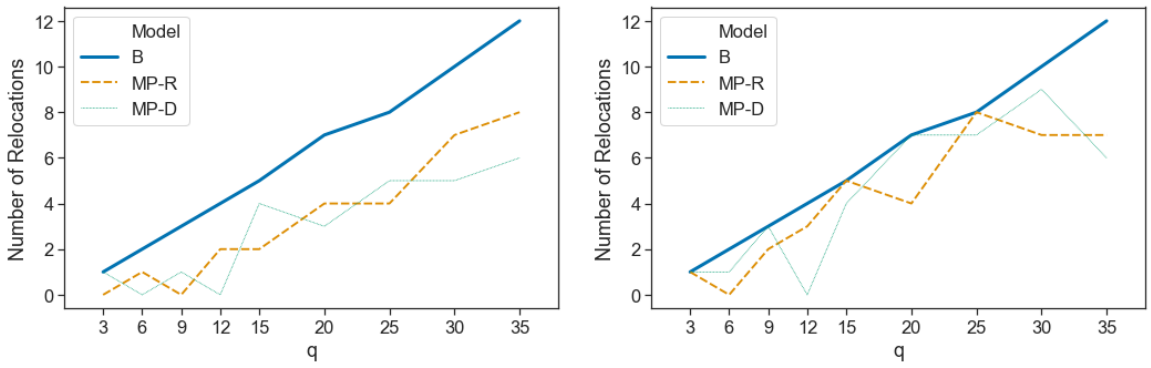
Table 2.2: Value of extending to multi-period formulation and adding robustness

| Model | $q$ | SS1 | | | | SS2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Model Cov. (%) | Simulated Cov. (%) | | | Model Cov. (%) | Simulated Cov. (%) | | |
| | | | Min | Ave | Max | | Min | Ave | Max |
| MP-R | 3 | 12.57 | 5.96 | 9.28 | 11.11 | 35.37 | 21.42 | 26.47 | 32.6 |
| | 6 | 24.37 | 13.89 | 16.59 | 19.55 | 62.69 | 36.59 | 46.01 | 56.38 |
| | 9 | 33.52 | 16.3 | 22.3 | 26.97 | 75.8 | 56.08 | 63.44 | 67.07 |
| | 12 | 41.9 | 24.52 | 29.96 | 35.67 | 82.6 | 66.18 | 72.25 | 79.38 |
| | 15 | 49.28 | 32.39 | 37.82 | 44.85 | 87.01 | 72.77 | 75.36 | 79.02 |
| | 20 | 55.45 | 35.58 | 45.26 | 53.67 | 90.73 | 72.62 | 78.47 | 85.97 |
| | 25 | 59.92 | 43 | 48.71 | 54.68 | 92.58 | 77.26 | 81.81 | 89.09 |
| | 30 | 62.28 | 48.99 | 52.58 | 56.05 | 93.21 | 77.98 | 83.21 | 88.86 |
| | 35 | 62.28 | 50.92 | 53.26 | 56.53 | 93.38 | 86.41 | 90.91 | 92.7 |
| MP-D | 3 | 15.76 | 5.81 | 8.56 | 10.25 | 44.85 | 23.27 | 24.75 | 28.07 |
| | 6 | 27.41 | 7.9 | 14.03 | 19.43 | 69.99 | 14.96 | 31.01 | 45.74 |
| | 9 | 37.93 | 14.66 | 21.79 | 28.22 | 82.18 | 45.05 | 51.95 | 60.25 |
| | 12 | 47.02 | 19.22 | 27.83 | 35.43 | 86.23 | 47.44 | 55.65 | 66.15 |
| | 15 | 53.9 | 24.76 | 32.04 | 39.81 | 89.6 | 41.87 | 51.32 | 64.99 |
| | 20 | 62.31 | 25.3 | 35.16 | 44.31 | 93.38 | 50.48 | 60.82 | 71.48 |
| | 25 | 67.19 | 34.45 | 40.98 | 47.5 | 94.73 | 68.33 | 74.49 | 78.55 |
| | 30 | 68.36 | 37.22 | 46.33 | 55.69 | 95.23 | 67.04 | 77.36 | 84.77 |
| | 35 | 68.5 | 40.88 | 48.01 | 58.19 | 95.29 | 77.83 | 83.27 | 90.67 |
| SP-R | 3 | 12.9 | 1.1 | 3.54 | 7.87 | 39.09 | 19.34 | 28.89 | 32.93 |
| | 6 | 24.64 | 10.37 | 11.96 | 18.06 | 64.75 | 36.92 | 42.25 | 52.18 |
| | 9 | 34.62 | 14.36 | 17.66 | 24.67 | 77.26 | 52.89 | 62.33 | 68.47 |
| | 12 | 42.64 | 23.81 | 28.73 | 34.92 | 83.28 | 58.88 | 65.28 | 74.91 |
| | 15 | 49.37 | 26.91 | 34.26 | 43 | 87.19 | 68.98 | 77.46 | 83.37 |
| | 20 | 55.21 | 29.05 | 40.63 | 50.63 | 90.88 | 70.98 | 78.71 | 84.86 |
| | 25 | 59.09 | 40.05 | 45.95 | 53.93 | 92.76 | 77.18 | 80.4 | 82.93 |
| | 30 | 62.1 | 42.01 | 50.33 | 57.33 | 93.36 | 76.28 | 81.65 | 89.06 |
| | 35 | 62.28 | 46.84 | 52.95 | 58.94 | 93.53 | 87.34 | 89.74 | 91.87 |
| SP-D | 3 | 15.76 | 3.93 | 4.53 | 6.23 | 46.31 | 18.98 | 22.57 | 26.82 |
| | 6 | 28.19 | 3.72 | 6.44 | 9.92 | 72.53 | 22.56 | 29.9 | 37.4 |
| | 9 | 38.8 | 7 | 11.68 | 16.78 | 83.02 | 32.06 | 41.54 | 47.94 |
| | 12 | 48.36 | 14.75 | 22.49 | 31.35 | 87.84 | 30.78 | 38.76 | 44.73 |
| | 15 | 55.6 | 19.19 | 25.42 | 32.75 | 90.91 | 29.14 | 38.72 | 46.28 |
| | 20 | 62.99 | 21.22 | 29.61 | 38.53 | 93.95 | 30.13 | 40.03 | 48.81 |
| | 25 | 67.4 | 32.03 | 40.82 | 47.94 | 95.68 | 54.02 | 57.29 | 63.29 |
| | 30 | 68.56 | 33.97 | 42.47 | 50.83 | 96.39 | 61.29 | 67.56 | 74.76 |
| | 35 | 68.71 | 39.6 | 48.4 | 54.08 | 96.63 | 72.82 | 77.38 | 80.39 |

Cov. = Coverage
SS1 is providing 90% coverage reliability in a response time of 4 minutes
SS2 is providing 95% coverage reliability in a response time of 10 minutes

time is short because the importance of choosing the right set of facility locations increases. An explanatory factor would be that multi-period formulation allows for more flexibility by allowing changing facility locations for different periods. The extent of facility relocation is depicted in Figure 2.3. The results reveal that at least 40% of the relocation budget is used when 15 or more facilities are opened. To further investigate the role of facility relocation, consider the visualization of facility location by season for MP-R SS2 model with $q = 15$, as an example, in Figure 2.4. Based on the wind patterns in Portland (see Figure 2.2), we expect the facilities in the summer season to provide better coverage reliability to demand locations in the west and/or north directions of them. As a result, the facility locations should be skewed a little bit towards the eastern and/or southern region of the operational area. Similarly, the locations in the winter season should be skewed a little bit towards the western and/or northern region. For our considered example, we indeed note that the centroid of facility locations opened in summer only is to the east of the centroid of facility locations opened in winter only, which is in agreement with our hypothesis.

For SS1, the improvements in average simulated coverage achieved by adding robustness to multi-period and single-period formulations are by 0.14 times and 0.28 times, respectively. For SS2, the improvements in average simulated coverage are by 0.23 times and 0.51 times for multi-period and single-period formulations, respectively. The improvement by adding robustness to a multi-period formulation is lower as more detailed information has been accounted which leads to lower variability in




(a) SS1 (b) SS2

Figure 2.3: Facility relocations in multi-period formulation ($B$ represents maximum allowable facility relocations)

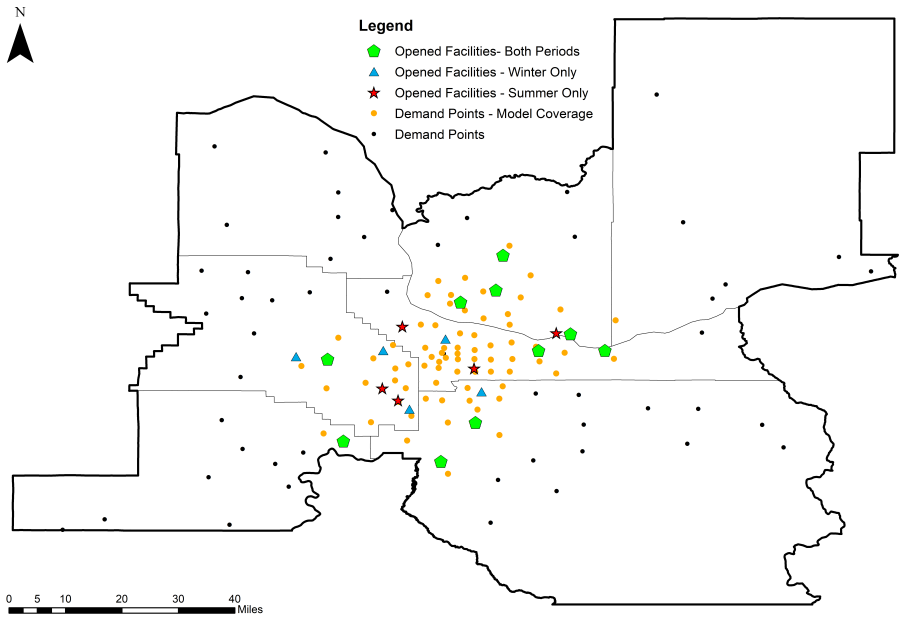



Figure 2.4: Facility relocation and model coverage for MP-R (SS2; $q = 15$)

data in each time period. Similarly, the variability in distance traveled by drone would increase with an increase in response time which leads to greater variability in failure probabilities. Therefore, the benefit obtained by adding robustness is greater when response times are longer. Overall, going from a single-period deterministic (SP-D) formulation to a multi-period robust (MP-R) formulation leads to an average simulated coverage improvement of 0.60 times and 0.54 times for SS1 and SS2, respectively. Figures 2.5 and 2.6 show model solution and an MCS simulation solution (having simulated coverage close to the average value) for SP-D and MP-R, respectively, for SS2 and $q = 15$. Accommodating uncertainty in decision-making leads to the consolidation of facilities towards the central core of the Portland Metro Area. Shorter travel distances lead to better coverage reliability in in the MP-R model.

Figure 2.7 shows the ratio of average simulated coverage to the model coverage (ASC-to-MC). The closer the values to 1 the better, as it indicates that the expected performance is close to real-life simulated scenarios. Accounting for robustness and/or extending to multi-period formulation leads to better outcomes on this metric. The ratio has a generally positive correlation with increasing values of $q$. This is expected, as with more opened facilities, the access to a demand point improves, and therefore, the coverage reliability also improves.

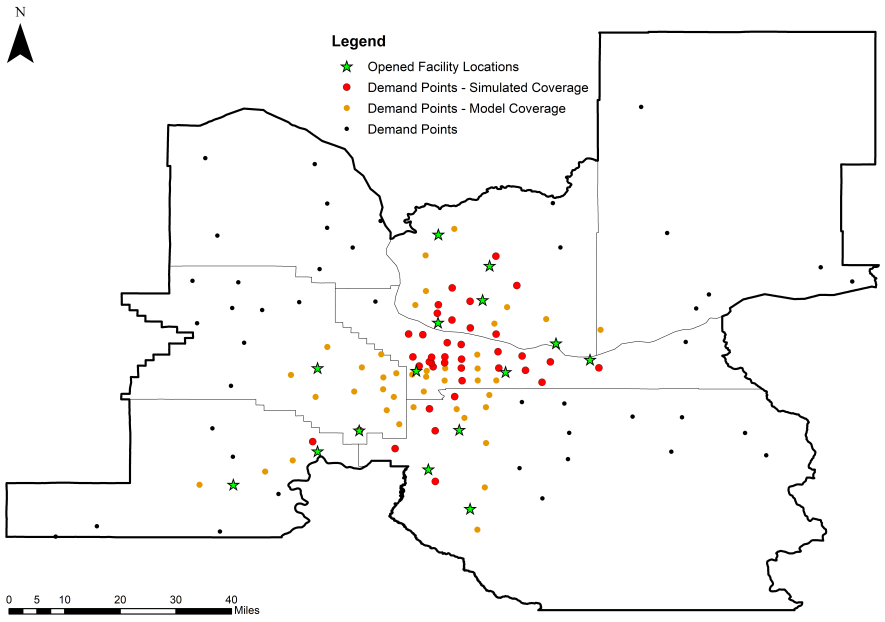



Figure 2.5: Opened Facility Locations and Demand Point Coverage for SP-D (SS2; $q = 15$)

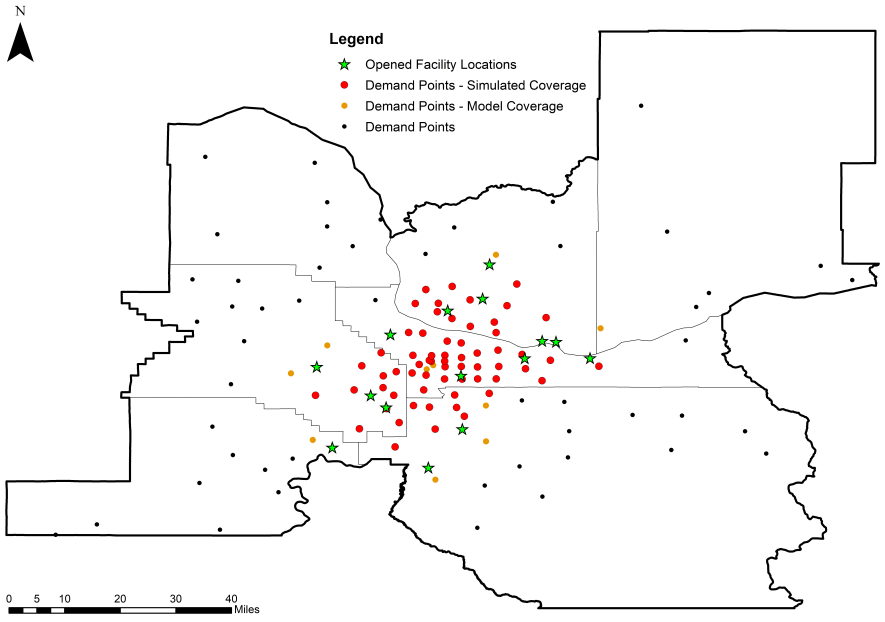
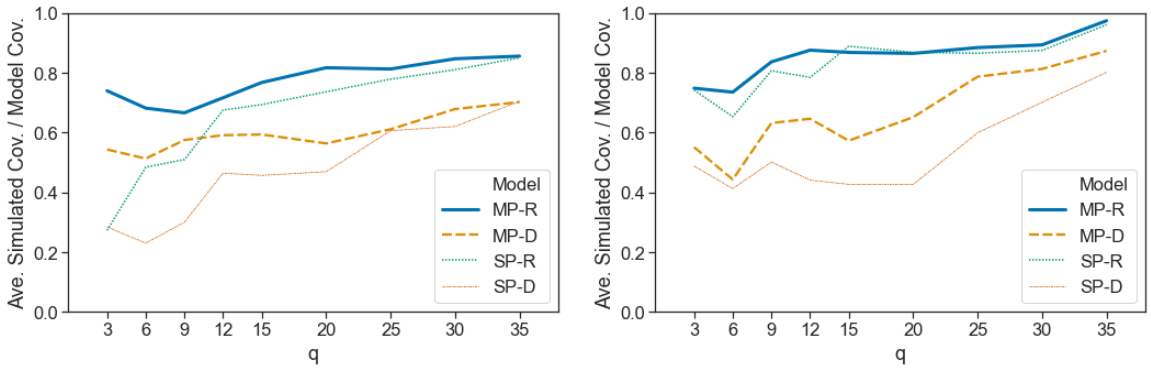



Figure 2.6: Opened Facility Locations and Demand Point Coverage for MP-R (SS2; $q = 15$; $\Gamma_i^t = 1$)




(a) SS1 (b) SS2

Figure 2.7: Ratio of average simulated coverage to model coverage

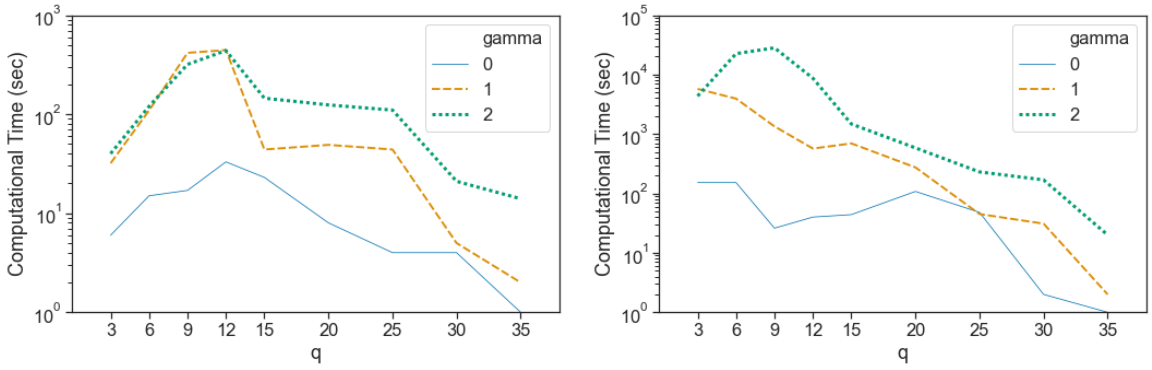
Table 2.3 shows the sensitivity of increasing conservatism on the coverage. For SS1, increasing robustness by increasing $\Gamma_i^t$ from 1 to 2 does not change the average simulated coverage much (-0.01 times). For SS2, this results in slightly better average simulated coverage (0.04 times). Computational times on the other hand typically increased with an increase in the budget of robustness. However, all models still converged in 8 hours, which is still not much considering a planning period of one year. Figure 2.8 shows the variation of computational times with the budget of robustness.

Table 2.3: Sensitivity to increasing conservatism in decision-making for multi-period formulation

| $\Gamma_i^t$ | $q$ | SS1 | | | | SS2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Model | Simulated Cov. (%) | | | Model | Simulated Cov. (%) | | |
| | | Cov. (%) | Min | Ave | Max | Cov. (%) | Min | Ave | Max |
| 0 | 3 | 15.76 | 5.81 | 8.56 | 10.25 | 44.85 | 23.27 | 24.75 | 28.07 |
| | 6 | 27.41 | 7.90 | 14.03 | 19.43 | 69.99 | 14.96 | 31.01 | 45.74 |
| | 9 | 37.93 | 14.66 | 21.79 | 28.22 | 82.18 | 45.05 | 51.95 | 60.25 |
| | 12 | 47.02 | 19.22 | 27.83 | 35.43 | 86.23 | 47.44 | 55.65 | 66.15 |
| | 15 | 53.90 | 24.76 | 32.04 | 39.81 | 89.60 | 41.87 | 51.32 | 64.99 |
| | 20 | 62.31 | 25.30 | 35.16 | 44.31 | 93.38 | 50.48 | 60.82 | 71.48 |
| | 25 | 67.19 | 34.45 | 40.98 | 47.50 | 94.73 | 68.33 | 74.49 | 78.55 |
| | 30 | 68.36 | 37.22 | 46.33 | 55.69 | 95.23 | 67.04 | 77.36 | 84.77 |
| | 35 | 68.50 | 40.88 | 48.01 | 58.19 | 95.29 | 77.83 | 83.27 | 90.67 |
| 1 | 3 | 12.57 | 5.96 | 9.28 | 11.11 | 35.37 | 21.42 | 26.47 | 32.60 |
| | 6 | 24.37 | 13.89 | 16.59 | 19.55 | 62.69 | 36.59 | 46.01 | 56.38 |
| | 9 | 33.52 | 16.30 | 22.30 | 26.97 | 75.80 | 56.08 | 63.44 | 67.07 |
| | 12 | 41.90 | 24.52 | 29.96 | 35.67 | 82.60 | 66.18 | 72.25 | 79.38 |
| | 15 | 49.28 | 32.39 | 37.82 | 44.85 | 87.01 | 72.77 | 75.36 | 79.02 |
| | 20 | 55.45 | 35.58 | 45.26 | 53.67 | 90.73 | 72.62 | 78.47 | 85.97 |
| | 25 | 59.92 | 43.00 | 48.71 | 54.68 | 92.58 | 77.26 | 81.81 | 89.09 |
| | 30 | 62.28 | 48.99 | 52.58 | 56.05 | 93.21 | 77.98 | 83.21 | 88.86 |
| | 35 | 62.28 | 50.92 | 53.26 | 56.53 | 93.38 | 86.41 | 90.91 | 92.70 |
| 2 | 3 | 12.57 | 5.96 | 9.28 | 11.11 | 33.85 | 23.96 | 25.98 | 31.62 |
| | 6 | 22.35 | 8.61 | 15.13 | 20.89 | 59.92 | 41.21 | 48.19 | 55.27 |
| | 9 | 30.99 | 16.12 | 21.36 | 25.74 | 73.60 | 57.54 | 63.83 | 69.49 |
| | 12 | 39.18 | 26.88 | 29.39 | 32.48 | 80.51 | 73.48 | 76.81 | 79.35 |
| | 15 | 45.65 | 32.99 | 38.29 | 45.11 | 85.40 | 75.27 | 79.28 | 82.39 |
| | 20 | 52.23 | 39.72 | 46.24 | 52.62 | 89.81 | 81.85 | 85.18 | 87.84 |
| | 25 | 56.59 | 43.12 | 49.16 | 53.87 | 92.01 | 82.54 | 86.01 | 86.86 |
| | 30 | 60.07 | 48.72 | 51.70 | 55.13 | 93.00 | 85.04 | 88.45 | 91.00 |
| | 35 | 61.92 | 51.37 | 54.68 | 58.46 | 93.38 | 87.19 | 89.53 | 90.20 |

Significant improvements in range of variation in simulated coverage as well as in the ratio ASC-to-MC were found, especially for larger values of $q$. These results are as expected as accounting for more amount of uncertainty should lead to reduced

model coverage (due to increased conservatism) and less variability in results (due to reduced probability of constraint violation). Therefore, finding a trade-off by changing the budget of robustness ($\Gamma_i^t$) can help improve the simulated coverage values, and reduce its gap from the model coverage. For example, Figure 2.9 shows variation in model and average simulated coverage with increasing value of $\Gamma_i^t$ for MP-R SS1 model with $q = 35$. It can be noticed that the gap between the model and average simulated coverage is the minimum when $\Gamma_i^t = 4$.




(a) SS1 (b) SS2

Figure 2.8: Computational times for varying values of $\Gamma_i^t$ in multi-period formulation

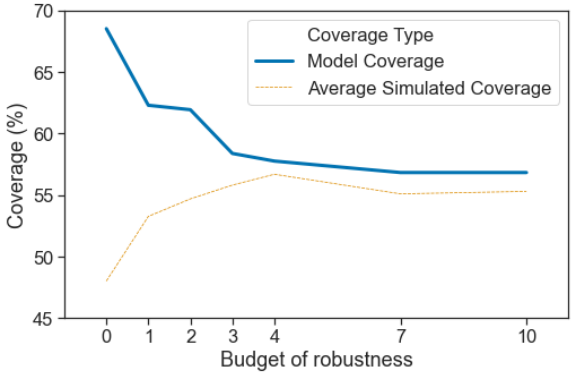



Figure 2.9: Model and Average Simulated Coverage with increasing values of budget of robustness $\Gamma_i^t$ (MP-R SS1 with $q = 35$)

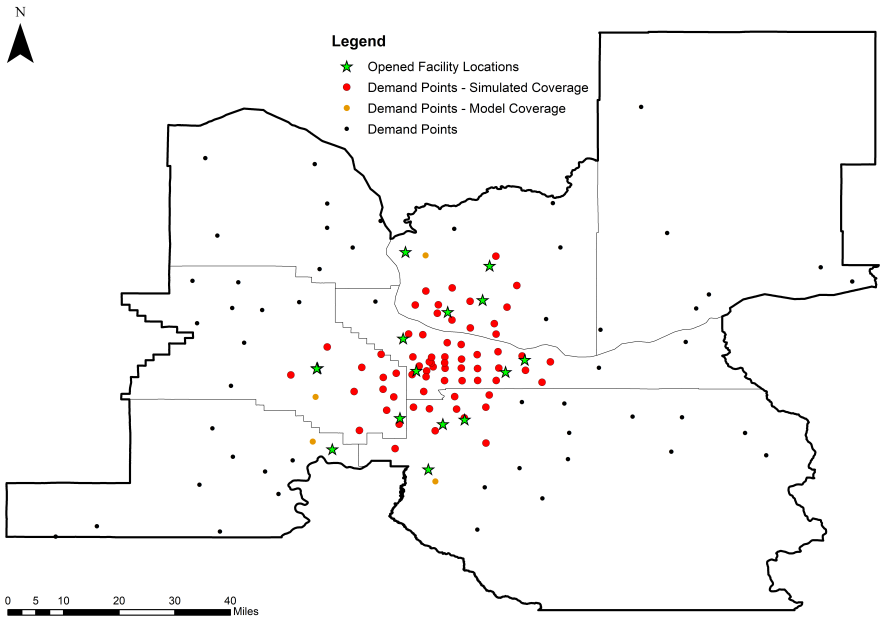



Figure 2.10: Opened Facility Locations and Demand Point Coverage for MP-R (SS2; $q = 15$; $\Gamma_i^t = 2$)

Figure 2.10 shows model solution and an MCS simulation solution (having simulated coverage close to the average value) for MP-R with $\Gamma_i^t = 2$ (SS2 and $q = 15$). Increasing conservatism further consolidates facilities around the central core compared to the case when $\Gamma_i^t = 1$ in MP-R, leading to better outcomes in terms of simulated coverage.

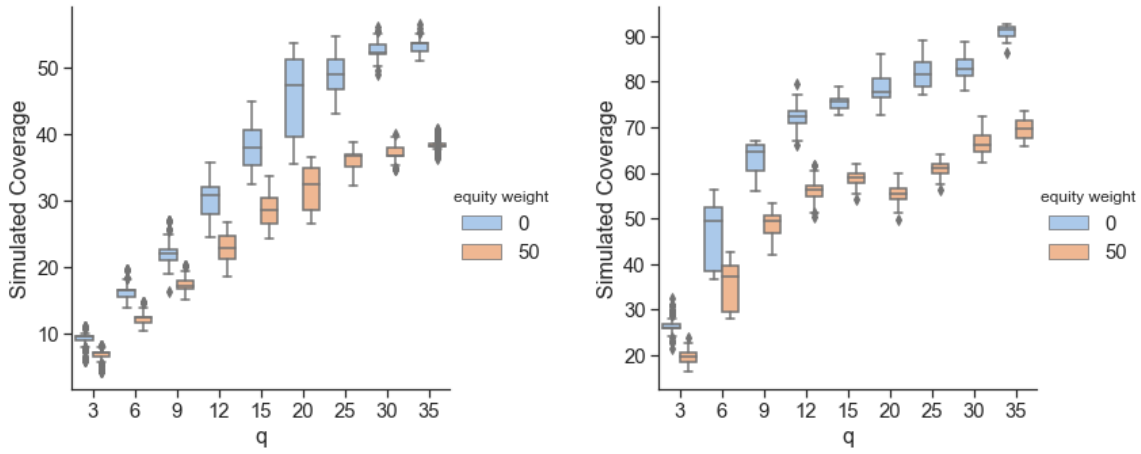
### 2.4.3 Incorporating equity in decision-making

For the previous sections, the coverage importance was just based on the normalized population of the demand points. However, it is possible to incorporate equity-related weights to determine coverage importance. For our case study, it can be considered that the facility locations that are not opened still have a defibrillator available onsite,

just that they can not be transported. Therefore, distance to the nearest potential facility location could be considered as a metric of equity, as in Mesa et al. (2003). The larger the minimum distance to a potential facility location from a demand point, the less equitable it is, and the more coverage importance it should get. The normalized inequity metric is calculated as:

- Distance to closest possible drone launch site from demand point $i$: $mindist(i)$
- Normalized inequity metric of demand point $i$
  $= \left\lceil \dfrac{100 \times mindist(i)}{\text{maximum value of } mindist(i)} \right\rceil$

For calculating the coverage importance metric, 50% weightage is assumed for both, the normalized population parameter, and the normalized inequity metric. Other parameters are set to their default values. The summary of results is shown in Figure 2.11. The simulated coverage values when equity is included are much lower than the values when equity is not included. When equity is included, the demand points far away from potential facility locations are given more importance, but, most of them can not even be accessed in the target response times (i.e. $|S_i|$ is a very small number). The spatial distribution of facilities and the demand coverage when equity is included is shown in figure 2.12 for the MP-R SS2 model with $q = 15$. It can be noticed that the figures 2.6 and 2.12 are very similar, covering almost the same demand points and most facilities opened at the same spot. A primary reason for this is the distribution of facility locations and demand points. The demand points




(a) SS1 (b) SS2

Figure 2.11: Effect of incorporating equity on simulated coverage

outside the more densely populated central core are located too far away from the potential facility locations. For the response times used in our case study, it does not make a practical difference if equity is included or not. However, for longer response times, equity inclusion could be beneficial (longer response times lead to larger values of $|S_i|$, which make it easier to meet service reliability target for all demand points).

(Aringhieri et al., 2017) state that equity is still one of the most challenging concerns for emergency medical services. More comprehensive methodologies that explicitly address equity concerns should be explored. Previous works in facility location have addressed equity by using metrics based on distance, exclusion, and conditional value-at-risk in model formulation Sirupa (2021).

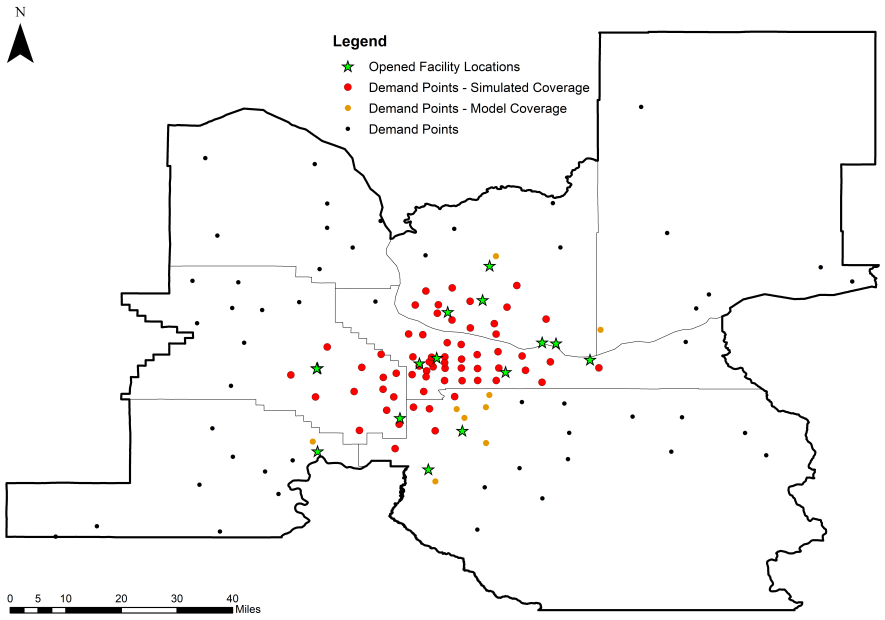



Figure 2.12: Opened Facility Locations and Demand Point Coverage for MP-R with equity inclusion (SS2; $q = 15$; $\Gamma_i^t = 1$)

## 2.5 Conclusion

This paper proposes a robust multi-period maximum covering facility location problem with coverage reliability (MP-R). MP-R is a generalized variant of the robust uncertain set covering the problem proposed by Lutter et al. (2017). The problem incorporates uncertainty in travel times via chance constraints and uses robust optimization using polyhedral uncertainty sets to tackle uncertainty. More conservative solutions can be obtained by increasing the value of parameter $\Gamma_i^t$.

A case study of the use of unmanned aerial vehicles (UAVs) or drones to deliver defibrillators in the Portland Metro Area is proposed. The uncertainty in drone travel times is a product of natural variability in wind speeds and directions. In

Portland, the wind characteristics (speed and direction) change drastically between the summer months (April to September) and winter months (October to March). Therefore, multiple periods are thought of as a discretization of recurring planning intervals (here, one whole year). We evaluate the effect of extending from a single period formulation (a whole year) to a multi-period formulation (two different time periods: summer and winter).

The value of adding robustness and extending to a multi-period formulation was evaluated utilizing a novel Monte-Carlo simulation scheme. The results highlighted that utilizing a multi-period formulation was particularly beneficial when response time thresholds were short or when uncertainty is not accounted for in the model. On the other hand, adding robustness to the deterministic models was more beneficial for single-period formulations or when response time thresholds were longer. Combining these different strengths led to an increase in average simulated coverage of MP-R by 57% compared to the deterministic single-period formulation (SP-D). Geographically, accounting for uncertainty (in MP-R) consolidates the facility locations towards the dense central core of the metro area compared to more spread out locations in SP-D. A more compact facility layout in MP-R improves the level of service in the central core of the metro area leading to superior simulated coverage.

For the MP-R model, a sensitivity analysis on the facility relocation cost budget showed very minor changes in model coverage as well as simulated coverage values. This implies that simply providing the model with more detailed information by dis-

cretizing over the planning period (even when facility relocation is not allowed) is helpful rather than providing the average information of the planning period. From our case study, when the response times are shorter, we recommend that an existing SP-D model should be extended to MP-R (i.e., incorporating uncertainty and discretizing to multiple periods). When the response times are longer, only incorporating uncertainty in the SP-D model is sufficient and multiple periods are not necessary.

The presented formulation can be used to analyze equity gaps and the need for additional resources. Analysis of distance-based equity inclusion in the objective yielded poorer coverage values. Equity inclusion increases the coverage importance of demand points further away from potential drone launch sites, but response times used in our study were too short for these points to be covered reliably. Geographically, equity inclusion did not affect the facility locations and demand point coverage significantly. However, for longer response times than used in this study, equity inclusion could be beneficial.

Even with the MP-R model providing the best performance, a significant gap exists between model coverage and the simulated coverage values. A major contributing factor is the assumption of independence among the failure probabilities. While some of the gap can be addressed by adjusting the budget of uncertainty and increasing the number of opened facilities, there is still a need to account for correlation in failure probabilities. Additionally, the study assumed that all the accessible open

facilities respond to the demand while not considering the possible unavailability of a drone at a located launch site. Future studies should also focus on including capacity considerations at located launch sites.

# 3 Maximum Profit Facility Location and Dynamic Resource Allocation for Instant Delivery Logistics

Cite this chapter as:

Darshan R. Chauhan, Avinash Unnikrishnan, Stephen D. Boyles, (2022), "*Maximum Profit Facility Location and Dynamic Resource Allocation for Instant Delivery Logistics*", Transportation Research Record: Journal of the Transportation Research Board, DOI: 10.1177/03611981221082574.

Author contributions:

The authors confirm contribution to the paper as follows: study conception and design: DRC, AU, SDB; mathematical formulation: DRC, AU; solution algorithms: DRC, AU; data collection and coding: DRC; analysis and interpretation of results: DRC, AU, SDB; draft manuscript preparation: DRC; finalizing manuscript and revisions: DRC, AU, SDB. All authors reviewed the results and approved the final version of the manuscript.

## 3.1 Introduction

E-commerce usage has become ever-ubiquitous now, especially due to social isolation requirements during the COVID-19 pandemic. A shift towards digital shopping has

resulted in double-digit retail e-commerce growth rates (32.1% growth from 2019 Q4 to 2020 Q4) compared to single-digit in total retail growth (6.9% growth from 2019 Q4 to 2020 Q4) (U.S. Census Bureau, 2021). In recent years, the delivery time thresholds for online purchases have also become intensive, with various e-commerce platforms providing same-day delivery options. Now, even options for 2-hour (Walmart, Amazon Prime Now, Walmart, Space NK) and 1-hour delivery (Instacart Express, Shipt, Alibaba Fresh Hema, buymie.eu) exist, with the industry gearing towards an instant (30-minute or better) delivery goal (Amazon Prime Air, Getir, Wolt). One of the viable alternatives for instant delivery right now is a UAV/drone. With the numerous large corporations including Amazon, Walmart, UPS, DHL, and Kroger heavily investing in drone technology, the growth in this sector has surpassed previous forecasts (Aouad, 2019). The higher operational speed and better cost-effectiveness of drones compared to traditional ground vehicles (Chiang et al., 2019) would be beneficial as extant logistics systems are stressed with increased demand from e-commerce growth.

This work delves into facility location and resource allocation for including instant delivery logistics into a company's operations. We assume that the instant deliveries (or, time-sensitive deliveries) are fulfilled through a battery-operated drone. While the non-time-sensitive orders can either be fulfilled by a drone from the located facilities or a truck from the central warehouse. The system consists of a set of facilities that can act as both "dark" stores (or micro fulfillment centers), and drone operations sites. During the planning stage, the facilities are located and resources

(product and battery capacity) are allocated such that it maximizes the total profit based on the deterministic information available. Once the facilities are set up, during the operational stage, the orders are received and real-time decisions are made regarding which facility and mode of transport would be used for fulfillment on an order-wise basis. Therefore, the goal in the operational stage is the adaptive learning of allocation of each order to maximize the cumulative profits while respecting resource capacity constraints.

Powell (2019) summarizes the work done by various communities on stochastic optimization. Focusing on online sequential decision-making communities, Markov decision processes, Q-learning, and approximate dynamic programming are formed based on state transition functions like Bellman's equation. These methods learn over time through multiple iterations over these states. In our problem, we model consumption of non-replenishable resources over time, and therefore, a state of the system does not recur. Additionally, the above methods use maximization of terminal reward as the objective function which is not the case for our problem. These nuances make the above approaches not suitable for our problem.

Multi-armed bandits is an online decision-making framework that maximizes the cumulative reward over the learning period (Powell, 2019). Multi-armed bandits can also be equipped with "context" that provide information available before making decisions, and "knapsacks" that can account for globalized resource consumption associated with the decisions (Slivkins, 2019). The above characteristics make multi-

armed bandits a suitable approach for our problem.

The key contributions of this study are: (i) formulating the instant delivery logistics problem as a two-stage problem – offline facility location with online resource allocation; (ii) while most of the previous logistics research has focused on time-aggregated dynamic resource allocation (Gu et al., 2020), we consider dynamic resource allocation at an order-level, which could be potentially beneficial for reducing delivery times (because of no lag in decision-making); and (iii) exploring a multi-armed bandit approach to effectively learn how to allocate orders to facilities in real-time and maximize the cumulative profits, and comparing it with other strategies.

The rest of the paper is organized as follows: the next section covers the relevant literature spanning the fields of facility location for stochastic demand, and dynamic resource allocation problems. Next, the problem description and formulation along with the dynamic resource allocation strategies are discussed. Later, computational experiments are conducted on test datasets. The final section concludes the work and provides avenues for further research.

## 3.2 Literature Review

In this section, we focus on primarily on facility location problems and dynamic resource allocation.

### 3.2.1 Facility Location Problems

Facility location has been one of the classical Operations Research problems and is one of the first prominent decisions that impact tactical and operational strategies for all organizations. An extensive number of books as well as review articles have been dedicated to facility location research: (Daskin, 2011; Farahani et al., 2010; ReVelle and Eiselt, 2005). Further, application-specific facility location reviews for humanitarian relief (Dönmez et al., 2021), healthcare and emergency location (Ahmadi-Javid et al., 2017), and urban-based applications (Farahani et al., 2019) are also available.

Mukundan and Daskin (1991) is one of the earliest works to explicitly consider maximizing profit (others implicitly considered maximizing profit as an alternative to minimizing cost). They consider joint location and sizing problems while considering cover-based constraints for facilities based on their size. Profit is defined as the difference between revenue and cost. We extend the problem considered by Mukundan and Daskin (1991) in two ways: firstly, by considering a continuum of facility size. This is achieved by converting costs in the objective functions to budget constraints. Therefore, the objective function in our work consists of only revenue-based terms. Secondly, we enforce capacity constraints derived from product and battery capacity allocations. Further, our problem does not consider coverage as a function of investment level as the facilities considered here are "dark-stores" which only cater to internet-based orders. The range of the facility is determined by explicitly modeling

energy consumption in battery-operated drones.

Ambulance location literature is rich in two-stage facility location models where ambulances are located offline and their allocations to demand points (and ambulance relocation) are made in real-time (Gendreau et al., 2001; Schmid, 2012). For the real-time allocation, either offline policies are used as in Gendreau et al. (2001), or adaptive online policies using methods like approximate dynamic programming are developed as in Schmid (2012). The above works model allocation decision for ambulances to a request and then, their relocation decisions. The unavailability of ambulances for a request due to being busy or excessive travel time is modeled, but the researchers do not consider the modeling of non-replenishable resource consumption (like cost budgets). In this work, facility location decisions are made offline and we develop an adaptive online policy using a multi-armed bandits approach for allocation of requests to appropriate fulfillment facility. Additionally, the unavailability of drones (due to range considerations) and trucks (for instant fulfillment requests) is modeled, and non-replenishable resource consumption related to routing costs and product capacity at individual facilities is considered.

### 3.2.2 Dynamic Resource Allocation

Resource allocation problems are widely observed: assigning a vehicle to a demand point in vehicle routing problems (VRP), a personnel to a job, etc. In the context of VRP, Bektaş et al. (2014) concisely defines dynamic problems as problems in which

information is revealed gradually over time, rather than all at once (static problems). Here, we use the same definition for resource allocation problems, and use the words "dynamic" and "online" interchangeably.

In operations research, dynamic resource allocation problems are generally tackled by formulating them as multi-stage stochastic programs (MSSP) (Guo et al., 2020). Several AI-based techniques have also been explored for online resource allocation problems which include Q-learning (Jiang et al., 2019; Yu et al., 2021), multi-armed bandits (Badanidiyuru et al., 2013; Villar et al., 2015), online algorithms (Devanur et al., 2019), and approximate dynamic programming (ADP) (Yu et al., 2019). Recently, (Powell, 2019) summarized the commonalities of various communities of stochastic and/or dynamic optimization and noted that adaptive learning algorithms based on dynamic programming (Q-learning, ADP, stochastic dual dynamic programming for MSSP) fall under the category of state-dependent problems with a terminal reward objective. Additionally, they observed that contextual bandits lend themselves well to several state-dependent problems with a cumulative reward objective. In this study, we explore a multi-armed bandit framework for dynamic resource allocation. Badanidiyuru et al. (2013) first proposed a multi-armed framework with universal budget constraints, which explicitly accommodates budget constraints in decision-making. Specifically, we use linear contextual bandits with knapsacks proposed by Agrawal and Devanur (2016) and extrapolate results to accommodate restricted "arm" availability arising from drone range constraints.

Focusing on delivery-related applications, Mallick et al. (2017) discuss recommendation systems for a carrier in truckload freight exchange marketplaces with constraints on arriving back at the origin within the planning period. Guo et al. (2020) studied drone-truck combined instant delivery logistics (90-minute grocery delivery) and proposed to solve it as a multi-stage stochastic program. Here, all the orders in a 30-minute interval were considered for allocation and routing. The truck and drone resources were available at each time interval for allocation (i.e., they are replenished). Similar assumptions have been made in humanitarian logistics applications (Yu et al., 2019, 2021), where a constant amount of resources to be distributed (food, medical kits) are made available every time period. However, in our study, a total product inventory and battery capacity are allocated to each open facility for the planning period and these are not replenished during the operational stage. Additionally, the above research addresses catering demand in each time interval. Our study differs by allocating demand at an order level, which would improve delivery time performance for instant orders, and is also adequate considering the limited payload capacity of drones (Moshref-Javadi and Winkenbach, 2021).

Availability-related constraints are adequately tackled in dynamic fleet management literature (Lin et al., 2018), where drivers available at any time period are varying. However in this study, we assume that a sufficient amount of drones are placed at facilities and trucks are central warehouses that vehicle-related availability constraints and congestion effects can be safely ignored, to emphasize more on

the allocation of facility and mode of delivery to a request. (Shavarani et al., 2019) studied a congested facility location problem for drone delivery with a case study implementation in San Francisco. The results show that for the system with 30 minute wait time, drone acquisition costs only accounted for about 12% of recharging infrastructure setup and operation costs. We consider that the drone acquisition costs are already considered in the drone operations facility opening and operation costs.

## 3.3 Problem Description

This section presents the maximum profit facility location problem for online demand satisfaction which is applicable for a logistics company employing UAVs/drones for last-mile delivery. We propose a two-stage framework wherein the planning stage, we locate facilities and allocate resources (product and battery capacity) to them with an a priori estimate of consumer demand. With the facilities located, and resources allocated, in the operational stage, the consumers place orders dynamically and we use multi-armed bandits for making real-time decisions to maximize the cumulative revenue while consuming resources. Note that the resources allocated at the end of the planning stage are not replenished in the operational stage. We consider two types of consumers: time-sensitive and regular. For time-sensitive customers, the demand must be met instantly, else, the demand is lost. For the regular customers, the demand must be fulfilled but instant deliveries are not mandatory.

### 3.3.1 Stage 1: Planning Stage

The online demand satisfaction problem is defined over a finite planning horizon. There are two modes of delivery available: drone delivery and traditional ground-based delivery. The drone-based deliveries are achieved by locating facilities that act as both products holding storerooms as well as drone launching stations. The ground-based deliveries are fulfilled using trucks located at a central warehouse. We

assume that there are enough drones/trucks at each facility that congestion effects can be neglected. The time-sensitive orders are fulfilled instantly (for example, 30 minutes or less), and the regular orders are fulfilled within a predetermined level of service (example, 1-day delivery or 2-day delivery). The time-sensitive orders are, therefore, assumed to be satisfied only using drones, while the regular orders can be delivered by either drones or trucks.

Let $G$ denote the set of demand points. During the planning stage, each demand point can be considered as a small geographical region (for example, ZIP code, Census tract, etc.). Let $n_g^S$ and $n_g^R$ denote the anticipated number of time-sensitive and regular orders from point $g \in G$, respectively. Let $F_g$ denote the set of all deliveries to point $g \in G$, i.e. $F_g = \{1, \ldots, n_g^S, n_g^S + 1, \ldots, n_g^S + n_g^R\}$. During the facility location (or, planning) stage, we do not take temporal aspects into consideration, and therefore, can consider both types of deliveries together.

Let $o_{gf}$ denote the order weight of $f^{th}$ $(\in F_g)$ order from demand point $g \in G$. We assume each order can weigh up to $o_{max}$. During the planning stage, $o_{gf}$ could take the form of using representative weights for different weight categories (therefore, obtaining the number of orders for different weight categories), and during the operational stage, the demand can assume the continuum of values up to $o_{max}$. The parameters $c_{gf}^D$ and $c_{gf}^T$ denote the estimated profit for satisfying the $f^{th}$ order of demand point $g \in G$ using drones and trucks, respectively.

The drone-based deliveries are carried out from potential facility locations spread

through the service area. The set of all potential facility locations is represented by $H$. A maximum of $p$ number of facilities can be opened. This parameter is prescribed considering leasing costs, recharging infrastructure setup costs, drone acquisition costs, and other related costs. A total of $\alpha$ amount of commodity and $\beta$ amount of battery capacity (for drones) are available for distribution among open facilities. If a facility is opened, then a minimum of $\alpha_{min}$ and $\beta_{min}$ amount of product and battery capacity must be allocated to it.

We assume that the drones only make one-to-one deliveries (from facility to demand point and back), as drones would be allocated at an order level during the operational stage. Let, $b_{ghf}$ be the battery consumption order to travel from a facility $h \in H$ to a point $g \in G$ and back while delivering the $f^{th}$ order, and $B_{drone}$ be the battery capacity of the drone. To be on a bit on the conservative side, we simplify $b_{ghf}$ to $b_{gh}$ (the battery consumption obtained using order weight of $o_{max}$). The set $G_h = \{g \in G : b_{gh} \leq B_{drone}\}$ describes the set of demand points that are accessible from a facility $h \in H$.

The truck-based deliveries are satisfied from a central warehouse. During the planning horizon, there are a maximum of $|M|$ truck trips available ($M$ denotes the set of truck trips) which are determined based on the predetermined level of service, for example, an average of 2 truck trips per day over the planning period of 300 days yields $|M| = 600$. The capacity of each truck (in the number of packages that can be delivered in each trip) is $B_{truck}$ which is determined based on actual truck capacity,

operator working hours constraints, etc. Each truck trip can be considered as a traveling salesman problem, and therefore, truck trip routing costs can be estimated using continuous approximation (Langevin et al., 1996). For a warehouse located '$d_0$' distance away from the center of the service area (of size $A$), the continuous approximation trip cost is given as:

$$m^{th} \text{ Truck Trip Cost} = k_1 n_m + k_2 d_0 \left(\mathbf{1}_{n_m>0}\right) + k_3 \sqrt{n_m A}$$

where, $n_m$ is the number of customers serviced on the $m^{th} (\in M)$ trip, $A$ is the size of service area, and $k_1, k_2, k_3$ are the proportionality constants. The term $(\mathbf{1}_{n_m>0})$ is the Heaviside step function with value 1 when $n_m > 0$, and 0 otherwise. The first term represents time costs (related to customer service times), and the second and third terms combined are distance related costs.

As the truck trip costs are concave, the best allocation is by consolidating as many assignments on a trip as possible. It is given that each truck can serve up to $B_{truck}$ customers per trip, and a total routing cost budget is $B_{routing}$. Then, the best allocation occurs when the first $m^*$ truck trips each serves $B_{truck}$ customers, and the remaining routing budget is utilized on the $(m^*+1)^{th}$ truck trip. The value of $m^*$ is given as:

$$m^* := \left\lfloor \frac{B_{routing}}{k_1 B_{truck} + k_2 d_0 + k_3 \sqrt{B_{truck} A}} \right\rfloor$$

Therefore, instead of incorporating a non-linear non-convex routing cost constraint, we can incorporate a simpler (and slightly conservative) linear constraint by limiting the total number of truck-based deliveries to $\omega = (\min\{|M|, m^*\} \cdot B_{truck})$.

Now, we discuss the decision variables for the planning stage optimization problem. The binary variable $y_h$ is 1 if a facility is opened at location $h \in H$. Let the variables $u_h$ and $z_h$ denote the product and battery capacity allocated to facility $h \in H$, respectively. The binary variable $x_{hgf}$ is 1 if $f^{th}$ $(\in F_g)$ order of point $g \in G$ is met by facility $h \in H$ using drone delivery, and the binary variable $w_{gf}$ be 1 if $f^{th}$ $(\in F_g)$ order of point $g \in G$ is met using truck delivery. Finally, the facility location problem can be described as:

$$\max_{u,w,x,y,z} \quad \sum_{h\in H}\sum_{g\in G}\sum_{f\in F_g} c^D_{gf} x_{hgf} + \sum_{g\in G}\sum_{f\in F_g} c^T_{gf} w_{gf} \tag{3.3.1}$$

$$\sum_{h\in H} y_h \leq p \tag{3.3.2}$$

$$\sum_{h\in H} u_h \leq \alpha \tag{3.3.3}$$

$$u_h \leq \alpha y_h \quad \forall\ h \in H \tag{3.3.4}$$

$$u_h \geq \alpha_{min} y_h \quad \forall\ h \in H \tag{3.3.5}$$

$$\sum_{g\in G_h}\sum_{f\in F_g} o_{gf} x_{hgf} \leq u_h \quad \forall\ h \in H \tag{3.3.6}$$

$$\sum_{h\in H} z_h \leq \beta \tag{3.3.7}$$

$$z_h \leq \beta y_h \quad \forall\ h \in H \tag{3.3.8}$$

$$z_h \geq \beta_{min} y_h \quad \forall \ h \in H \tag{3.3.9}$$

$$\sum_{g \in G_h} \sum_{f \in F_g} b_{gh} x_{hgf} \leq z_h \quad \forall \ h \in H \tag{3.3.10}$$

$$\sum_{h \in H} x_{hgf} + w_{gf} \leq 1 \quad \forall \ f \in F_g, g \in G \tag{3.3.11}$$

$$\sum_{g \in G} \sum_{f \in F_g} w_{gf} \leq \omega \tag{3.3.12}$$

$$x_{hgf} \in \{0, 1\} \quad \forall \ f \in F_g, g \in G, h \in H \tag{3.3.13}$$

$$w_{gf} \in \{0, 1\} \quad \forall \ f \in F_g, g \in G \tag{3.3.14}$$

$$y_h \in \{0, 1\} \quad \forall \ h \in H \tag{3.3.15}$$

$$u_h, z_h \geq 0 \quad \forall \ h \in H \tag{3.3.16}$$

The objective function (3.3.1) aims to maximize the cumulative profit achieved by the delivery system. Equation (3.3.2) constrains the number of open facilities to a maximum of $p$. Equations (3.3.3) and (3.3.7) ensure that a total amount of product and battery capacity allocated is less than $\alpha$ and $\beta$ (the overall budgets), respectively. Equations (3.3.4) and (3.3.5) ensure that product is only inventoried at open facilities, and that at least a minimum of $\alpha_{min}$ amount of product is inventoried when a facility is opened. Equations (3.3.8) and (3.3.9) enforce similar constraints to battery capacity, i.e., battery capacity can only be allocated to open facilities and at least a minimum of $\beta_{min}$ amount of battery capacity is allocated at an open facility. Equation (3.3.6) ensures that no more demand than the product inventory at a facility

is met. Similarly, equation (3.3.10) constrains the battery consumption to be less than the battery capacity of the facility. Note that equations (3.3.6) and (3.3.10) take drone delivery range into consideration. Equation (3.3.11) ensures that a particular order can be satisfied by at most one facility using a drone or by using truck-based delivery. Constraint (3.3.12) limits the total number of truck based deliveries to $\omega$. Equations (3.3.13)-(3.3.16) are variable definitions.

After finding an optimal solution, it is modified by allocating the remaining slack in product (equation (3.3.3)) and battery capacity (equation (3.3.7)). The slack is allocated such that it maximizes the minimum value of product and battery capacity at an open facility.

### 3.3.2 Stage 2: Operational Stage

At the end of stage 1, we have a solution for the planning stage problem, given by the tuple $(x^*, w^*, y^*, u^*, z^*)$. Let, the set of all opened facilities be represented by set $H' := \{h \in H : y_h^* = 1\}$. The product inventory and battery capacity located at an open facility $h \in H'$ are given by $u_h^*$ and $z_h^*$, respectively, which would be used for drone-based deliveries. The regular (i.e., non-time-sensitive) orders can also be fulfilled through the central warehouse using truck-based delivery (a maximum of $\omega$ number of regular order deliveries can be fulfilled). Order fulfillment leads to consumption of these resources (product, battery, truck) which are not replenished during the entirety of the operational stage. The goal during the operational stage

is to maximize cumulative profit by allocating orders arriving in an online manner to either one of the located facilities for drone delivery, or to the central warehouse for a truck delivery (only for regular orders), such that no more than available resources are utilized.

During the operational stage, uncertainties stem from various sources: the probability of an order being from demand location $g \in G$, the probability of an order being time-sensitive, the order weight distribution from demand location $g \in G$, the battery consumption distribution while using drones for delivering an order, and the profit distributions of time-sensitive and regular orders. However, in any market, these values cannot be known with absolute certainty and need to be learned by implementing field trials. We explore three strategies for this online operational stage problem: first, a multi-armed bandits-based approach; second, an allocation heuristic designed from the solution of the planning stage optimization problem; and third, two random choice heuristics based on the random choice among available options for an order. Based on the context of our problem, we use the words "profits" and "rewards" synonymously here.

Currently, there is no option of non-fulfillment of an order (possibly due to network congestion) and can be a part of future research. As a performance measurement here, we study the maximization of cumulative profit until the first resource constraint is violated considering the allocation of each order independent of its profit and resource consumption. Alternatively put, the episode ends at the first instance of a

facility (or central warehouse) running out of a resource. Instant delivery logistics would typically be adopted for a relatively small geographical area (like a metropolitan area). In such cases, orders received after the first instance of exhaustion of resources could lead to denial of service based on geography, which cannot be the case for practical applications. Also, this stopping criterion would indicate the need for resources to be replenished for uninterrupted service.

**Multi-armed bandits**

Multi-armed bandits are a reinforcement learning framework wherein the agent learns by exploring given set of options (a.k.a. "arms") such that the cumulative reward achieved is maximized. Here, specifically, we use linear contextual bandits with knapsacks (linCBwK), proposed by (Agrawal and Devanur, 2016). The linCBwK allows us to choose only from a subset of arms (as all open facilities are not available to each order placed) while accounting for constraints consisting of the product inventory and battery consumption budgets at each facility, and overall truck routing.

A linCBwK problem has five components to it. The first is the $K$ number of arms or actions. Here, these actions represent options for each order: drone delivery from a located facility, or truck delivery from the central warehouse. Therefore, we have $K = |H'| + 1$ arms, and $[K] := \{h \; \forall \; h \in H'\} \cup \{truck\}$ is the set of all arms.

The second is time horizon or total number of decision-making events $T$. Here, each time/event $t \in \{1, 2, \ldots, T\}$ represents an order placed in real-time by demand

point $g_t \in G$ (assumed i.i.d. to unknown distribution $\mathcal{D}^g$, abbreviated as $g_t \overset{iid}{\sim} \mathcal{D}^g$). Let, $\lambda_t$ (derived from $\lambda_t | g_t \overset{iid}{\sim} \mathcal{D}^{\lambda}_{g_t}$) be 1 if the order is time-sensitive and requires instant delivery, and 0, otherwise. Note that the demand points outside the drone-based coverage region can only place regular orders. Then, for each $h \in H'$, taking the current environmental factors into account, we observe the battery consumption, $b^t_{g_t h}$ ($b^t_{g_t h} | g_t \overset{iid}{\sim} \mathcal{D}^b_{g_t h}$) for all $h \in H'$. All of the above information is available before making the allocation decision for event $t$.

The third is the context. Context represents all information that we have prior to making a decision. At each event $t$, we observe an $m$-dimensional context vector for each arm $a \in [K]$, represented by $\mathbf{x}_t(a) \in [0,1]^m$. Let the context matrix $X_t := \{\mathbf{x}_t(a) \forall a \in [K]\} \in [0,1]^{m \times K}$. For our case, we observe a $m = K$ dimensional context for each arm. The $K \times K$ diagonal context matrix is constructed as:

$$X_t[h,h] := \begin{cases} 1 & \text{if } b^t_{g_t h} \leq B_{drone} \\ 0 & \text{otherwise} \end{cases}, \quad \forall \ h \in H'$$

$$X_t[truck, truck] := \begin{cases} 1 & \text{if } \lambda_t = 0 \\ 0 & \text{otherwise} \end{cases}$$

Agrawal and Devanur (2016) state that when the context matrix is a $K$-dimensional identity matrix, linCBwK emulates the bandits with knapsacks (BwK) problem (Badanidiyuru et al., 2013). We extrapolate this result to consider a BwK problem with restricted arm availability. The above definition of context only allows arms with context equal to 1 to be available for selection. While defining context, we ensure that the *truck* arm is only available for regular deliveries. Additionally, we consider that the demand points outside the drone-based coverage can only place regular orders. The above definition allows for the availability of at least one arm for selection by the algorithm.

The fourth component of linCBwK problem is reward. At time $t$, a scalar reward $r_t(a_t) \in [0, 1]$ is realized after playing action $a_t \in [K]$. Without loss of generality, we assume that the reward for fulfilling a time-sensitive delivery is $c^S \in (0, 1]$, and a regular delivery is $c^R \in [0, c^S)$ irrespective of which action is chosen. This assumption can be relaxed to model rewards that are a function of the ordering demand point $g_t$ and the action $a_t$ chosen.

The fifth component of linCBwK is knapsacks constraints, or globalized budget constraints. For our problem, there are $d = (2 \cdot |H'| + 1)$ universal knapsack constraints. The first $|H'|$ constraints represent the product consumption at facility $h \in H'$ with budgets $u_h^*$, the second $|H'|$ constraints represent the battery consumption at facility $h \in H'$ with budgets $z_h^*$, and the last knapsack constrains the total number of truck deliveries to a budget $\omega$ (as described in Stage 1). If at time $t$, the *truck* arm is

chosen, then, $\mathbf{0}$ amount of product ($U_h$), $\mathbf{0}$ amount of battery ($B_h$) resources, and 1 unit of truck delivery resources are used. If the truck-delivery arm is not chosen, then let $h_t$ denote the chosen facility for fulfillment of demand. When $h_t$ is chosen, $(o^t_{g_t} \cdot \mathbf{e}_{h_t})$ amount of product, $(b^t_{g_t h_t} \cdot \mathbf{e}_{h_t})$ amount of battery resource, and 0 unit of truck delivery resources are consumed. Here, $\mathbf{e}_{h_t}$ is a $|H'| \times 1$ matrix with value 1 for the row where $h = h_t$, and 0, otherwise. Let $I^{truck}_t$ be 1 if the truck-delivery arm is chosen at time $t$, and 0, otherwise.

The bandit optimization problem that we are tackling here is given as:

$$\max_{\mathbf{e}, I^{truck}} \quad \sum_{t=1}^{T} \Bigg[ \Big\{ c^S \lambda_t + c^R (1 - \lambda_t) \Big\} (1 - I^{truck}_t) + c^R (1 - \lambda_t) I^{truck}_t \Bigg] \tag{3.3.17}$$

$$\sum_{t=1}^{T} o^t_{g_t} \cdot \mathbf{e}_{h_t} \cdot (1 - I^{truck}_t) \leq u^*_h \quad \forall \ h \in H' \tag{3.3.18}$$

$$\sum_{t=1}^{T} b^t_{g_t h_t} \cdot \mathbf{e}_{h_t} \cdot (1 - I^{truck}_t) \leq z^*_h \quad \forall \ h \in H' \tag{3.3.19}$$

$$\sum_{t=1}^{T} I^{truck}_t \leq \omega \tag{3.3.20}$$

At time $t$, upon selection of arm $a_t$, let $\mathbf{v}_t(a_t)$ be the $d$-dimensional resource consumption vector. As a reminder, we allow only payloads up to $o_{max}$. Also note that for chosen arm, the value of $b^t_{g_t h_t}$ is always less than or equal to $B_{drone}$. Therefore, we can use the above values to normalize resource consumption vector $\mathbf{v}_t(a_t)$ in the range [0,1], a requirement to implement linCBwK. The first set of transformations

are given as:

Transformations I:

PC knapsacks: $o^t_{g_t} \leftarrow \frac{o^t_{g_t}}{o_{max}}$ $\quad u^*_h \leftarrow \frac{u^*_h}{o_{max}}$

BC knapsacks: $b^t_{g_t h_t} \leftarrow \frac{b^t_{g_t h_t}}{B_{drone}}$ $\quad z^*_h \leftarrow \frac{z^*_h}{B_{drone}}$

where, PC refers to product consumption, and BC refers to battery consumption. The other required transformation for linCBwK is a uniform value of budget for each knapsack constraint. Therefore, we scale each knapsack so that its budget to the lowest value after the above transformations. The new budget, $B$, is given as:

$$B = \min\Big\{ \min\{u^*_h : h \in H'\}, \min\{z^*_h : h \in H'\}, \omega\Big\}$$

The transformations to make the budget the same for all knapsack constraints are:

Transformations II:

PC knapsacks: $o^t_{g_t} \leftarrow \frac{B}{u^*_h} o^t_{g_t}$; $\quad u^*_h \leftarrow B$

BC knapsacks: $b^t_{g_t h_t} \leftarrow \frac{B}{z^*_h} b^t_{g_t h_t}$; $\quad z^*_h \leftarrow B$

Truck knapsack: $I^{truck}_t \leftarrow \frac{B}{\omega} I^{truck}_t$; $\quad \omega \leftarrow B$

We make the following two assumptions about context, rewards, and resource consumption vectors in linCBwK (Agrawal and Devanur, 2016):

- In every round $t$, the tuple $\{\mathbf{x}_t(a), r_t(a), \mathbf{v}_t(a)\}_{a=1}^{K}$ is generated from an unknown distribution $\mathcal{D}$, independent of everything in previous rounds. The procedure used for generating contexts, rewards, and resource consumption for our instant delivery logistics problem satisfies this assumption.

- There exists an unknown vector $\boldsymbol{\mu}_* \in [0,1]^{m\times 1}$ and a matrix $W_* \in [0,1]^{m\times d}$ such that for every arm $a$, given contexts $\mathbf{x}_t(a)$, and history $H_{t-1}$ before time $t$,

$$\begin{aligned} \mathbb{E}[r_t(a)|\, x_t(a), H_{t-1}] &= \boldsymbol{\mu}_*^{\mathsf{T}} x_t(a) \\ \mathbb{E}[\mathbf{v}_t(a)|\, x_t(a), H_{t-1}] &= W_*^{\mathsf{T}} x_t(a) \end{aligned} \tag{3.3.21}$$

The decision-making flow of the linCBwK algorithm consists of five major steps: observing the context matrix $X_t$, obtaining optimistic estimates of $\boldsymbol{\mu}_*$ and $W_*$ using the $l_2$-regularized norms of previously observed values of rewards and resource consumption, arm selection using an expected reward penalized with expected resource consumption, realizing the values of reward and resource consumption for the selected arm, and finally, updating the penalty weights using multiplicative weight update (MWU) algorithm (Agrawal and Devanur, 2016). The detailed algorithm is in Algorithm 3.1. Like all multi-armed bandit algorithms, the performance is measured by the complexity of the cumulative regret. For linCBwK, the regret is measured from the optimal static policy (Agrawal and Devanur, 2016), obtained from solving

a static stochastic optimization problem.

**Algorithm 3.1** Algorithm for linCBwK

**Input:** $B$, $T_0$, $T$, $(1-\delta)$ confidence level, MWU algorithm parameter $\epsilon$
**Output:** $a_t, r_t, \mathbf{v}_t$

1: Input parameters:
2: Compute $Z$ which satisfies assumptions presented in (Agrawal and Devanur, 2016)
3: Initialize $B' = B - T_0$, $T' = T - T_0$
4: Initialize $t = 1$, $\boldsymbol{\theta}_{1,j} = \frac{1}{1+d}$, $\forall\, j \in \{1, 2, \ldots, d\}$ $t \leq T'$
5: Observe context $X_t$
6: For every $a \in [K]$, compute $\tilde{\boldsymbol{\mu}}_t(a)$ and $\tilde{W}_t(a)$ (the optimistic estimates of $\boldsymbol{\mu}_*$ and $W_*$) as:

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_t(a) &\coloneqq \arg\max_{\boldsymbol{\mu} \in C_{t,0}} \mathbf{x}_t(a)^\mathsf{T} \boldsymbol{\mu}, \quad \text{where, } \hat{\boldsymbol{\mu}}_t \;:=\; M_t^{-1} \sum_{i=1}^{t-1} \mathbf{x}_i(a_i) r_i(a_i)^\mathsf{T} \\ \text{where, } C_{t,0} &\coloneqq \left\{ \boldsymbol{\mu} \in \mathbb{R}^{m \times 1} : \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t\|_{M_t} \leq radius_t \right\} \end{aligned} \tag{3.3.22}$$

$$\begin{aligned} \tilde{W}_t(a) &\coloneqq \arg\min_{W \in \mathcal{G}_t} \mathbf{x}_t(a)^\mathsf{T} W \boldsymbol{\theta}_t, \quad \text{where, } \hat{W}_t \;:=\; M_t^{-1} \sum_{i=1}^{t-1} \mathbf{x}_i(a_i) \mathbf{v}_i(a_i)^\mathsf{T} \\ \text{where, } \mathcal{G}_t &\coloneqq \left\{ \mathbb{R}^{m \times d} : \mathbf{w}_j \in C_{t,j} \right\} \\ \text{where, } C_{t,j} &\coloneqq \left\{ \mathbf{w} \in \mathbb{R}^{m \times 1} : \|\mathbf{w} - \hat{\mathbf{w}}_{tj}\|_{M_t} \leq radius_t \right\} \end{aligned} \tag{3.3.23}$$

7: Play arm $a_t \coloneqq \arg\max_{a \in [K]} \mathbf{x}_t(a)^\mathsf{T} \left( \tilde{\boldsymbol{\mu}}_t(a) - Z \tilde{W}_t(a) \boldsymbol{\theta}_t \right)$
8: Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$
9: If for some $j \in \{1, \ldots, d\}$, $\sum_{t' \leq t} \mathbf{v}_{t'}(a_{t'}) \cdot \mathbf{e}_j \geq B$, then EXIT. (Note: $\mathbf{e}_j$ is a $d \times 1$ matrix with value 1 for $j^{th}$ row, and 0, otherwise)
10: Update $\boldsymbol{\theta}_{t+1}$ using MWU algorithm, and with $g_t(\boldsymbol{\theta}_t) \coloneqq \boldsymbol{\theta}_t \cdot \left( \mathbf{v}_t(a_t) - \frac{B}{T} \mathbf{1} \right)$, as: for all $j \in \{1, 2, \ldots, d\}$

$$\begin{aligned} \boldsymbol{\theta}_{t+1,j} &= \frac{w_{t,j}}{1 + \sum_j w_{t,j}} \\ \text{where } w_{t,j} &= \begin{cases} w_{t-1,j}(1+\epsilon)^{g_{t,j}} & \text{if } g_{t,j} > 0, \\ w_{t-1,j}(1-\epsilon)^{-g_{t,j}} & \text{if } g_{t,j} \leq 0 \end{cases} \end{aligned} \tag{3.3.24}$$

11: $t \mathrel{+}= 1$

**Planning Stage Optimization Allocation (PSOA) Heuristic**

The planning stage optimization allocation (PSOA) heuristic uses the insights obtained from the solution of the optimization problem solved in Stage 1 to derive a policy. Particularly, $2|G|$ different policies are derived depending upon the demand point $g \in G$ placing an order, and the order being time-sensitive or not. The policies are given as:

For all $g \in G, h \in H$:

$$p_{g,\lambda}^{PSOA}(h) = \begin{cases} q_h^1(x^*) & \text{if, } \sum_{a \in H'} \sum_{f \in F} x_{agf}^* > 0,\ \lambda = 1 \\ q_h^2(x^*, w^*) & \text{if, } \sum_{a \in H'} \sum_{f \in F} x_{agf}^* > 0,\ \lambda = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.3.25}$$

$$q_h^1(x^*) = \frac{\sum_{f \in F} x_{hgf}^*}{\sum_{a \in H'} \sum_{f \in F} x_{agf}^*}$$

$$q_h^2(x^*, w^*) = \frac{\sum_{f \in F} \frac{n_g^R}{n_g^S + n_g^R} x_{hgf}^*}{\sum_{a \in H'} \sum_{f \in F} \frac{n_g^R}{n_g^S + n_g^R} x_{agf}^* + \sum_{f \in F} w_{gf}^*}$$

For all $g \in G$:

$$p_{g,\lambda}^{PSOA}(truck) = \begin{cases} 1 - \sum_{h \in H'} p_{g,\lambda}^{PSOA}(h) & \text{if, } \lambda = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.3.26}$$

where, $\lambda = 1$ for time-sensitive deliveries, and $\lambda = 0$ for regular deliveries. Note that as the type of orders (time-sensitive or regular) are not differentiated in the Stage 1 optimization problem, some demand points which can be covered through drone-based deliveries may only be served using the truck delivery option. As a result, these demand point could place a time-sensitive order and the PSOA heuristic would not know what to do. In such cases, the PSOA heuristic collects a reward of 0, and consumes **0** units of all the resources (product, battery, and truck-delivery). The algorithm for PSOA is presented in Algorithm 3.2.

**Algorithm 3.2** PSOA Heuristic

**Input:** $B$, $T_0$, $T$
**Output:** $a_t, r_t, \mathbf{v}_t$

Initialize parameters $B' = B - T_0$, and $T' = T - T_0$.
Initialize $t = 1$ $t \leq T'$
Observe ordering demand point $g_t$, time-sensitivity of order $\lambda_t$, and context $X_t$.
Select an arm $a_t \in [K]$ chosen randomly with probability of choosing $a_t$ is $p^{PSOA}_{g_t,\lambda_t}(a_t)$ $a_t \in H'$ and $X_t[a_t, a_t] = 1$
Play the selected arm $a_t$
Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$ $\lambda_t = 1$
Randomly select and play an arm $a_t$ such that $X_t[a_t, a_t] = 1$
Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$
Play the *truck* arm, i.e, $a_t = truck$
Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$
If for some $j \in \{1, \ldots, d\}$, $\sum_{t' \leq t} \mathbf{v}_{t'}(a_{t'}) \cdot \mathbf{e}_j \geq B$, then EXIT. (Note: $\mathbf{e}_j$ is a $d \times 1$ matrix with value 1 for $j^{th}$ row, and 0, otherwise)
$t += 1$

### Random Choice (RC) Heuristic

The random choice (RC) heuristic chooses one of the available options randomly in a weighted manner upon observing the context based on random choice among the

available alternative at each time $t$. The nominal probability of choosing the "*truck*" arm for a regular order is $p^{RC}_{truck} = \frac{\omega}{\sum_{g \in G} n^R_g}$. The algorithm is presented in Algorithm 3.3.

**Algorithm 3.3** RC Heuristic

**Input:** $B$, $T_0$, $T$, $(1-\delta)$ confidence level, MWU algorithm parameter $\epsilon$
**Output:** $a_t, r_t, \mathbf{v}_t$

1: Input parameters $B'$, and $T'$.
2: Initialize $t = 1$ $t \leq T'$
3: Observe ordering demand point $g_t$, time-sensitivity of order $\lambda_t$, and context $X_t$.
4: Calculate the set of facilities available for drone deliveries, i.e., $H_{avail} := \{h \in H' \,|\, X_t[h,h] = 1\}$ $|H_{avail}| > 0$ $\lambda_t = 1$
5: Play one of the available facility arms each with probability $\frac{1}{|H_{avail}|}$
6: Play the arm "*truck*" with probability $p^{RC}_{truck}$, and one of the available facility arms each with probability $\frac{1}{|H_{avail}|}(1 - p^{RC}_{truck})$
7: Play arm "*truck*"
8: Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$
9: If for some $j \in \{1, \ldots, d\}$, $\sum_{t' \leq t} \mathbf{v}_{t'}(a_{t'}) \cdot \mathbf{e}_j \geq B$, then EXIT. (Note: $\mathbf{e}_j$ is a $d \times 1$ matrix with value 1 for $j^{th}$ row, and 0, otherwise)
10: $t += 1$

### Blind Random Choice (BRC) Heuristic

The blind random choice (BRC) heuristic works like the RC heuristics, except that it does not have access to even the input parameters of the problem $(n^S, n^R, \omega)$. Therefore, at any time $t$, the BRC heuristic chooses one of the available arms randomly in a unweighted manner (i.e., each of the available arms has an equal probability of being selected).

## 3.4 Computational Experiments

The analysis is conducted on standard p-median test instances, adopted from Osman and Christofides (1994), each consisting of 50 locations that act both as demand points (represented by set $G$) and potential facility locations (represented by set $H$) on a randomly generated on a 100 $\times$ 100 grid (here, units are assumed to be kilometers). For the current planning period, the anticipated number of times-sensitive ($n_g^S$) and regular ($n_g^R$) deliveries are random integers in the interval [8,12] and [8,12], respectively. The estimated demand for each order is randomly selected from a discrete uniform distribution from 0.5 kg to 2.25 kg in the interval of 0.25 kg. Euclidean distances are used for distance computations. The battery consumption for a trip from facility $h \in H$ to demand point $g \in G$ and back is calculated as in Figliozzi (2017) assuming a payload of $o_{max}$ = 5 lbs (2.27 kg). The overall energy efficiency and the lift-to-drag ratio of the drone are 0.66 and 2.89, respectively. The battery capacity of the drone is 1410 Wh, and a maximum battery utilization factor of 0.8 is used. Thereby, the effective battery capacity of the drone ($B_{drone}$) is 1128 Wh. The values of parameters $\alpha$, $\beta$, $\alpha_{min}$, $\beta_{min}$, and $p$ are chosen to be 2500, $800 \cdot B_{drone}$, 800, $350 \cdot B_{drone}$, and 3, respectively. Considering the truck routing budget, the maximum number of orders that can be fulfilled by truck delivery (i.e., $\omega$) is determined to be 400. We do not consider congestion effects in the current study and assume that enough drones/trucks are available at each operational facility.

The solution of the planning stage problem determines the initial state of the operational stage problem (i.e. $K$, $d$, $m$, and $B$ can be calculated). For bandit learning using linCBwK, the $(1-\delta)$ confidence interval for estimating unknown parameters is taken to be 95%. The total number of orders $(T)$ is assumed to be 1000, and the initial learning iterations $(T_0)$ is assumed to be $m\sqrt{T}$ (rounded to the nearest integer). The above value of $T_0$ ensures that the linCBwK algorithm maintains the regret bound provided by Agrawal and Devanur (2016). The online learning parameter, $\epsilon$, is assumed to be $\sqrt{(d+1)/T}$, as proposed by Agrawal and Devanur (2016).

Unknown to the linCBwK algorithm, for the nominal case, we assume that the estimated values of time-sensitive and regular deliveries used in the planning stage are off by at most $\rho^S = 30\%$ and $\rho^R = 10\%$ compared to the actual simulation values observed during the operational stage. Therefore, for the simulations, the probability of a demand point ordering and the order being time-sensitive is determined by calculating simulation values of time-sensitive $(\tilde{n}_g^S)$ and regular $(\tilde{n}_g^R)$ deliveries from demand point $g \in G$ is set to:

For all $g \in G$:

$$\tilde{n}_g^S \in Uniform\left[\frac{1}{1+\rho^S} n_g^S, \frac{1}{1-\rho^S} n_g^S\right]$$
$$\tilde{n}_g^R \in Uniform\left[\frac{1}{1+\rho^R} n_g^R, \frac{1}{1-\rho^R} n_g^R\right]$$
$$P(g_t = g) = \frac{\tilde{n}_g^S + \tilde{n}_g^R}{\sum_{g \in G} \tilde{n}_g^S + \tilde{n}_g^R}$$

$$P(\lambda_{g_t} = 1 \,|\, g_t = g) = \begin{cases} \dfrac{\tilde{n}_g^S}{\tilde{n}_g^S + \tilde{n}_g^R} & ;\ g \text{ can be served using drones} \\ 0 & ;\ g \text{ cannot be served using drones} \end{cases}$$

At time $t \in \{1, 2, \ldots, T\}$, the ordering demand point $g_t = g$ with probability $P(g_t = g)$, and the order is time-sensitive with probability $P(\lambda_{g_t} = 1 \,|\, g_t = g)$. Unknown to the algorithms, we define parameters $\phi_g^1, \phi_g^2 \in Uniform(0.5, 5)\ \forall\ g \in G$. The demand $(o_{g_t})$ is randomly chosen in the interval $[0, o_{max}]$ from the beta distribution $o_{max} \cdot Beta(\phi_{g_t}^1, \phi_{g_t}^2)$, where, $o_{max}$ is the maximum weight of an order. The battery consumption at time $t$ (i.e., $b_{g_t h}^t$), between a demand point $g_t \in G$ and facility $h \in H'$ is assumed to vary in the interval $[b_{gh} - \hat{b}_{gh}, b_{gh} + \hat{b}_{gh}]$, where, $b_{gh}$ is the nominal battery consumption (used in the planning stage), and $\hat{b}_{gh}$ is the maximum variation is battery consumption. The value of $\hat{b}_{gh}$ is assumed to be an integer in the interval $[0.1b_{gh}, 0.3b_{gh}]$. Similar assumptions are made in Chauhan et al. (2021).

During the simulations, each instance is run 10 times to account for randomness in demand, time sensitivity, order weight, and battery consumption generation. Table 3.1 shows the cumulative reward achieved. All instances opened 2 facilities for drone delivery and had a truck-delivery option for regular orders. The linCBwK provides the best rewards, slightly over 7% additional rewards with respect to the trailing PSOA heuristic. This result is as expected as the linCBwK dynamically updates the expected rewards and resource consumption, and weighs them appropriately for decision-making. The BRC heuristic follows PSOA, and RC has the worst outcome

with respect to cumulative rewards. We hypothesized that the RC heuristic would perform better than BRC heuristic because of the weighted probability while choosing the delivery option. With a lower number of arms, for our computational experiments, the BRC heuristic uses the truck-based delivery option less intensively than RC which improves its performance. The same trend is observed for the successful number of allocations as seen in Figure 3.1.

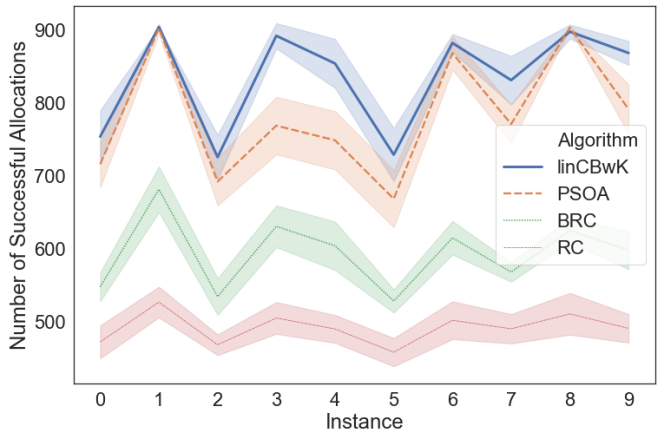



Figure 3.1: Number of successful allocations: average line with the standard deviation band ($T = 1000, T_0 = 95$)

Table 3.1: Cumulative reward obtained through successful allocations ($T = 1000, T_0 = 95$)

| Instance | linCBwK | | | PSOA | | | BRC | | | RC | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min | Ave | Max | Min | Ave | Max | Min | Ave | Max | Min | Ave | Max |
| 0 | 364.5 | 400.6 | 426.8 | 347.9 | 375.8 | 407.6 | 268.8 | 290 | 306.2 | 228 | 249.2 | 279.2 |
| 1 | 521.9 | 533.4 | 543.5 | 471.5 | 511.4 | 533.1 | 345.5 | 369.6 | 395.8 | 266.6 | 288.6 | 315.5 |
| 2 | 357.2 | 377.4 | 408.1 | 334.6 | 360 | 393.6 | 259.1 | 280.5 | 306.8 | 231.6 | 246.7 | 264.8 |
| 3 | 453.8 | 493.8 | 533.6 | 366.1 | 410.5 | 435 | 301.9 | 343.2 | 366 | 239.9 | 272.2 | 281.8 |
| 4 | 432.9 | 461.4 | 485.6 | 374.3 | 399.9 | 440.4 | 288.4 | 324.1 | 355.4 | 245.6 | 264.6 | 293.3 |
| 5 | 342.8 | 375.3 | 412.1 | 316.3 | 346.3 | 386 | 260.6 | 276.9 | 288.2 | 220.1 | 238.6 | 254 |
| 6 | 464.5 | 475.7 | 501 | 442.2 | 470.3 | 514.4 | 311.1 | 333.3 | 354.1 | 251.2 | 271.7 | 318.8 |
| 7 | 401.8 | 438.9 | 469.8 | 378.4 | 408.2 | 436.8 | 293.3 | 305.2 | 323 | 235.9 | 259.6 | 282.6 |
| 8 | 462.9 | 487.1 | 506.5 | 483.3 | 507.9 | 530 | 322.3 | 338.7 | 353.1 | 241.2 | 273.9 | 302.7 |
| 9 | 455.9 | 465.4 | 480 | 386 | 419.6 | 456.2 | 288.5 | 321.8 | 346.9 | 246.3 | 264.9 | 287.5 |
| Overall | 342.8 | 450.9 | 543.5 | 316.3 | 421 | 533.1 | 259.1 | 318.3 | 395.8 | 220.1 | 263 | 318.8 |

The values $\rho^S$ and $\rho^R$ show the maximum deviation of the estimate used in the Stage 1 optimization problem from the simulated values used in Stage 2. The pair $(\rho^S, \rho^R) = (0.0, 0.0)$ implies that there was no estimation error during the planning stage. Figure 3.2 shows the effect of uncertainty on the performance of the heuristics, based on all 10 instances. As the amount of observed uncertainty increases, all heuristics perform slightly better. A likely reason for this observation is that the higher diversity in the demand generation provides more opportunities to facilities that are used less often. This would delay the resource consumption violation at a more intensively used facility.

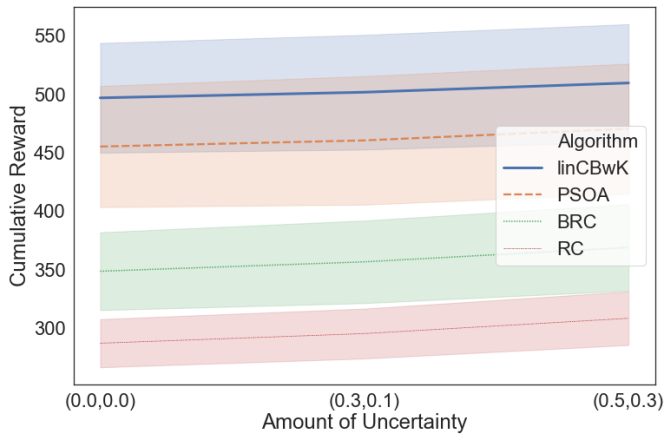
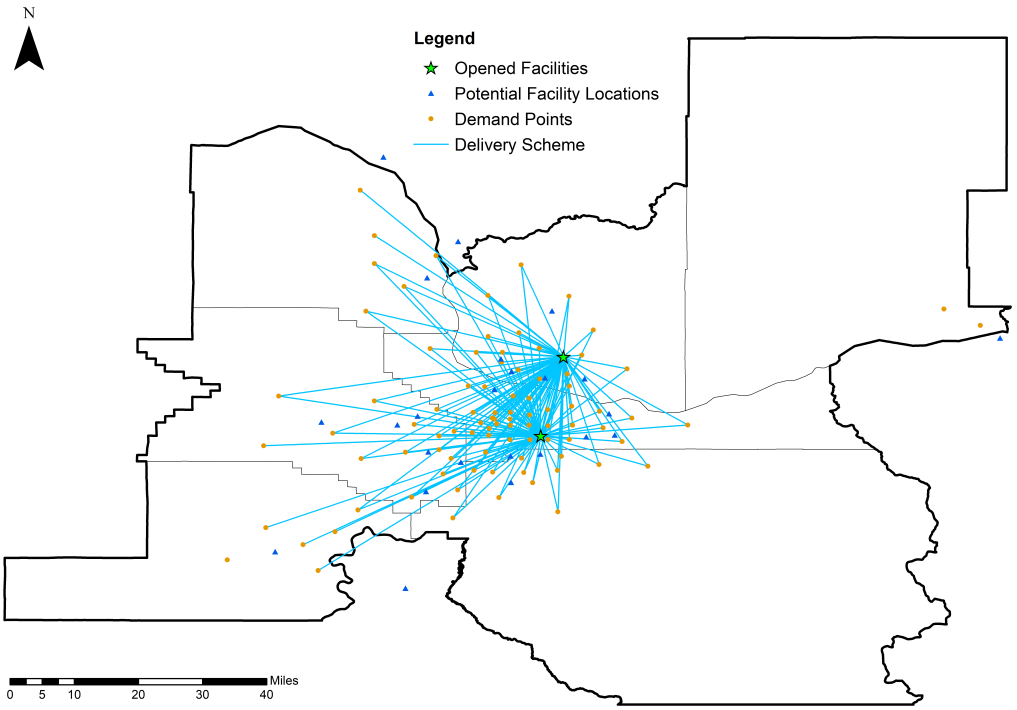



Figure 3.2: Cumulative rewards with varying the amount of uncertainty in estimating the number and type of deliveries ($\rho^S$,$\rho^R$): average line with the standard deviation band ($T = 1000, T_0 = 95$)

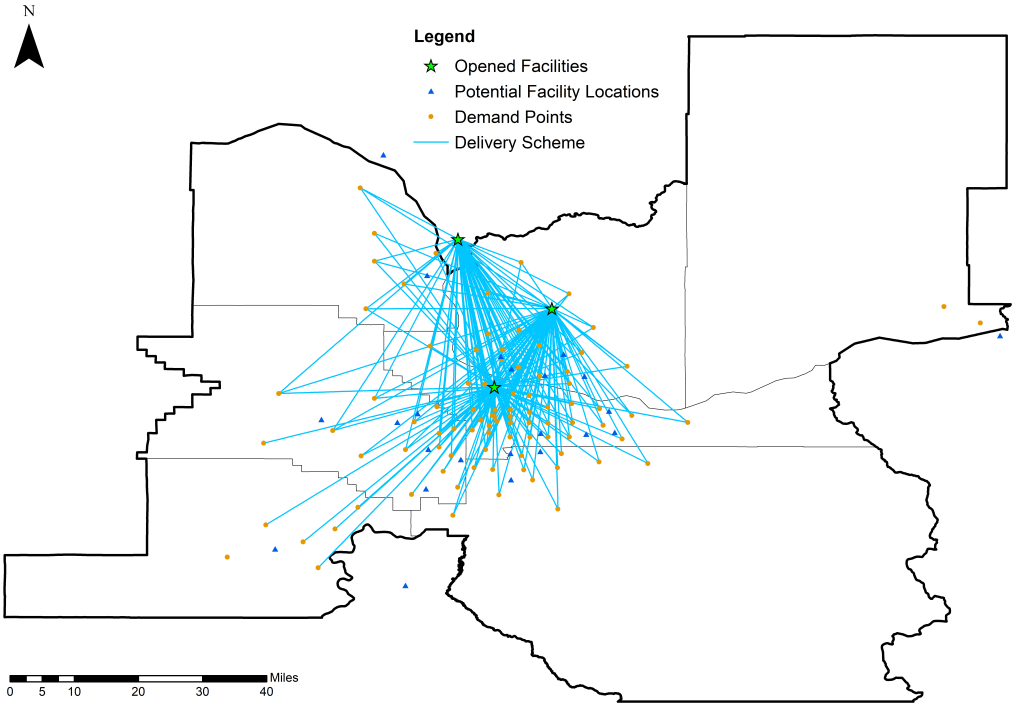
### 3.4.1 Portland Metro Area Case Study

For the Portland Metro Area case study, we consider Walmart expanding its ser-

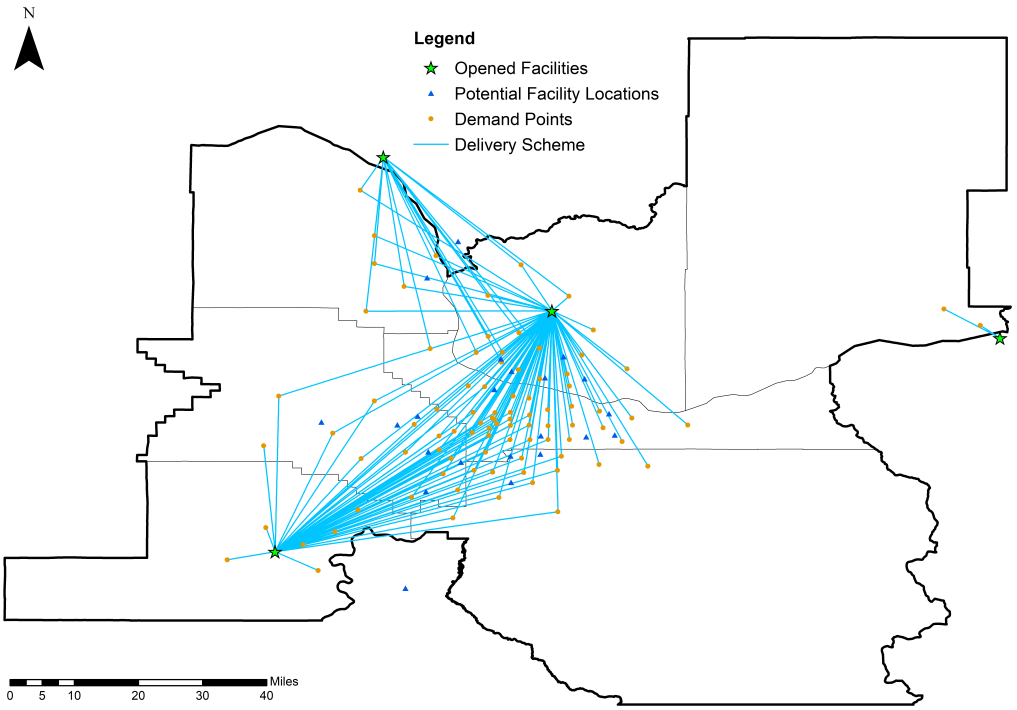
vice options to offer instant delivery to its Walmart+ service subscribers (similar to Amazon's Prime subscription). The 26 Walmart stores in and around the Portland Metro are considered the potential drone-based fulfillment center candidates. The 90 centroids of the ZIP Code Tabulation Areas (ZCTAs) in the Portland Metro Area that can be serviced by drones are considered as the demand locations. The locations of Walmart stores and ZCTAs used in the study are available at https://github.com/drc1807/MPFL_DRA. The latest estimate of Walmart+ subscriber base in the USA is 60.78 million (PYMTS, 2021). Assuming geographically uniform subscriber base in the US and monthly ordering behavior results in 13840 anticipated daily deliveries in Portland Metro Area. We consider a planning period of one day. The proportion of time-sensitive deliveries at each demand point is randomly distributed in the interval [0.4,0.7]. The values of parameters $\alpha$ and $\beta$ is set to 28000 and $13000 \cdot B_{drone}$, respectively. For the operational stage, the total number of orders ($T$) is set to 15000, and the amount of uncertainty is chosen as $\rho^S = 30\%$ and $\rho^R = 10\%$. We explore three cases by changing the number of opened facilities ($p$) from 2 to 4. The values of $\alpha_{min}/o_{max}$, $\beta_{min}/B_{drone}$, and $\omega$ are selected to be the smallest multiple of 50 greater than $(p+1)^{0.5}T^{0.75}$. These values ensure that the regret bound conditions for linCBwK mentioned by (Agrawal and Devanur, 2016) are met. Other parameters are the same as described for the p-median instances. The solutions of the planning stage optimization problem are shown in Figure 3.3.


N
Legend
Opened Facilities
Potential Facility Locations
Demand Points
Delivery Scheme
Miles
0 5 10 20 30 40

(a) $p = 2$


N
Legend
Opened Facilities
Potential Facility Locations
Demand Points
Delivery Scheme
Miles
0 5 10 20 30 40

(b) $p = 3$




(c) $p = 4$

Figure 3.3: Planning stage optimization problem solutions for Portland Metro Area

Figure 3.4 shows the variation in cumulative rewards achieved by various algorithms with the number of opened facilities. Here as well, the linCBwK algorithm performs best. A slight decrease in cumulative reward with increasing $p$ is expected as the effective number of deliveries used for all algorithms, $T'$ $(= T - T_0)$, decreases with increasing $p$. In this regard, PSOA gives a stable performance. However, a significant improvement in accumulated profits is experienced by linCBwK, BRC, and RC for $p = 3$. A primary reason for the improved performance is the availability of all facilities for almost every order (see Figure 3.3(b)) as well as proportionate distribution of product and battery resources among the facilities. As $p$ increases,

the minimum amount of resource allocation at facilities also increases. This means that the facilities available to a smaller proportion of demand points have more redundant capacity, and the more readily available facilities face scarcity of resources. This causes a drop in performance of linCBwK, BRC, and RC when $p = 4$ due to the outlying facility serving only two demand points. By numbers, linCBwK beats the second-best approach by 11.2%, 13.2%, and 25.1%, on average, as $p$ increases from 2 to 4, respectively.

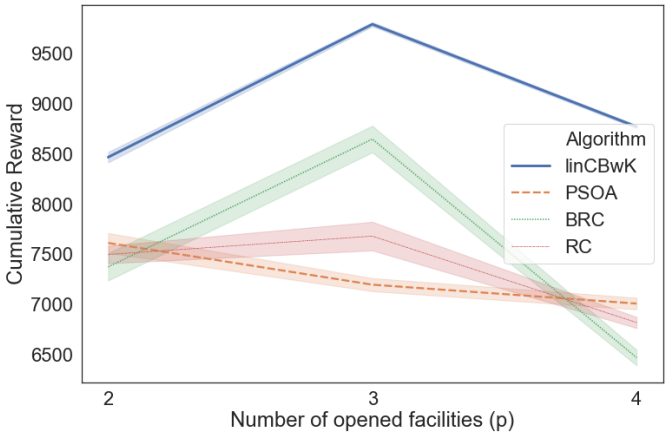



Figure 3.4: Cumulative rewards with varying the number of opened facilities ($p$): average line with the standard deviation band ($T = 15000,\ T_0 = (p+1)\sqrt{T}$)

## 3.5 Conclusions

This paper investigates a facility location and online demand allocation problem applicable to a logistics company expanding to instant delivery using UAV/drones. The problem consists of two stages: a planning stage, and an operational stage. During the planning stage, the company wishes to locate micro-fulfillment centers which

serve the dual purpose of product storage and drone operations. We present a profit-maximizing mixed-integer linear program that accounts for product capacity, battery capacity, and routing cost constraints. During the operational stage, the orders arrive in an online manner and real-time decisions are made for the satisfaction of demand with an objective of maximizing cumulative profits while respecting the resource budget constraints. To the best of the authors' knowledge, this work is the first application in logistics considering non-replenishable resource consumption constraints in real-time decision-making.

We explore a multi-armed bandit framework that explicitly accounts for global knapsack constraints. We extrapolate results from extant literature to account for restricted "arm" availability in the framework arising from drone range constraints. The multi-armed bandit framework is compared with a heuristic policy derived from the planning stage optimization solution (PSOA heuristic), and two heuristics based on random choice. The analysis on standard test instances shows that the multi-armed bandit framework beats the second-best PSOA heuristic by accumulating 7% more profits, on average. An application of this problem to Portland Metro Area with a larger time horizon yields similar results with the multi-armed bandit framework performing the best, beating the second-best approach by at least 11.2%.

The present work can be expanded in various aspects. Currently, the model does not accommodate the non-fulfillment of time-sensitive orders or provide an incentive to switch to regular orders. This may be an important feature due to possible net-

work congestion issues or drone availability issues. In this study, we assumed that enough drones are available at each facility, which may not be the case in many applications. Availability-related constraints are effectively tackled in the dynamic fleet management literature and can be a possible extension of this work.

# 4 Equitable Relief Prepositioning and Distribution in Post-Disaster Scenarios

Cite this chapter as:

Darshan R. Chauhan, Avinash Unnikrishnan, Stephen D. Boyles, (2022), "*Equitable Relief Prepositioning and Distribution in Post-Disaster Scenarios*", Working Paper.

Author contributions:

The authors confirm contribution to the paper as follows: study conception and design: DRC, AU, SDB; mathematical formulation: DRC, AU; solution algorithms: DRC; data collection and coding: DRC; analysis and interpretation of results: DRC; draft manuscript preparation: DRC; finalizing manuscript and revisions: DRC, AU, SDB. All authors reviewed the results and approved the final version of the manuscript.

## 4.1 Introduction

Natural disasters are outcomes of severe weather events, or natural hazard events (e.g., volcanic eruptions, earthquakes, etc.), that lead to significant effects on society; specifically, situations in which society is overwhelmed and the end results are considerable impacts on societal health and safety, the economy, and the environment itself (Banholzer et al., 2014; Prasad and Francescutti, 2017). In 2020, although the

number of fatalities was down compared to the previous year (2019), the number of natural disasters increased by about 14% with a total estimated cost in losses of approximately $268 billion (AON, 2021; Insurance Information Institute, 2021a). Of the top five events, in terms of losses, three occurred in the United States, with estimated losses of $30.8 billion (Insurance Information Institute, 2021b).

In the United States, specifically, there were approximately 250 deaths and $119 billion in economic losses due to natural disasters in 2020 (AON, 2021). These losses were a result of a variety of natural disasters, including hurricanes, wildfires, tornadoes, winter storms, severe convective storms, and flooding. With such losses, natural disaster preparedness has become a priority for the U.S. Department of Homeland Security (U.S. Department of Homeland Security, 2021). Of particular note is the dedication paid to responses in the event that a natural disaster occurs, where a primary focus is on meeting the needs of a community that is impacted by a natural disaster. This focus consists of emergency preparedness (e.g., first responders) to save lives, protect property, protect the environment itself, and provide other basic needs required by a community that is at the center of a natural disaster. To accommodate and provide these basic needs of communities, resilient infrastructure and the ability to continue to ship and deliver goods as needed are vital.

Although the type of natural disaster varies by geographical location, all communities in the United States can be impacted and experience substantial losses. It is not possible to prevent a natural disaster, but the ability to mitigate losses can

be achieved by being better prepared when a natural disaster occurs. The founding motivation for this study is the state of Oregon in the United States preparing for the long-overdue Cascadia Subduction Zone (CSZ) earthquake. The odds for a great earthquake (M8.7-M9.3) are 7-15% and a very large earthquake (M8.3-M8.6) are about 37% in the next fifty years (Oregon Seismic Safety Policy Advisory Commission, 2013). The impact of the earthquake is expected to be catastrophic with fatality estimates ranging from 1,250 to greater than 10,000 and direct and indirect economic losses of $30 billion. The humanitarian logistics research has claimed the first 72 hours after the disaster as the golden period for saving lives Sheu (2007, 2014); Yu et al. (2021). After the potential CSZ earthquake in Oregon, it is expected that even restoration to minimally operable (for medical and humanitarian aid only) backbone network connectivity would take 24-72 hours. Considering the widespread impact, scarcity of resources would also be experienced, which makes the equitable distribution of the limited available resources imperative.

In this study, we model a novel equitable resource distribution problem (ERDP) for post-disaster scenarios. As the network connectivity might be limited, we consider that a drone-only infrastructure is set up for providing medical and humanitarian aid. We propose a two-stage framework and explicitly consider that the resources are scarce compared to the demand. The first stage models the pre-disaster stage. It considers that the demand would be catered at relief shelters and the decisions regarding the location of distribution centers and prepositioning amounts must be made

with the objective of minimizing inequity. The occurrence of the disaster triggers the second stage of the model. In the second post-disaster stage, the relief shelters place requests for resources in an online manner. The decision-making agency needs to allocate the requests to distribution centers for satisfaction to minimize inequity while acknowledging the resource budget constraints at the distribution centers, and the transportation-related constraints. We apply this model to a case study based in Portland Urban Metro in Oregon in preparation for the anticipated CSZ earthquake.

The rest of the article is organized as follows: Section 4.2 reviews the relevant literature and describes the novelty of this research. Section 4.3 provides the formulation of the proposed two-stage model. Section 4.4 first describes the Portland case study and analyzes the computational performance of the proposed approach with two other heuristic methods. Finally, Section 4.5 concludes the research by summarizing the finding and providing directions for future research.

## 4.2 Literature Review

Our proposed Equitable Resource Distribution Problem (ERDP) consists of a two-stage formulation for determining the distribution center locations and prepositioning amounts in the pre-disaster phase, and equitable online request allocation in the post-disaster phase. The literature review consists of three parts. As our application falls under the umbrella of humanitarian logistics, the first part is devoted to that. The second part presents a review of online resource allocation applications in the field

of logistics. And finally, the third part reviews the equity considerations in facility location and resource allocation problems.

### 4.2.1 Humanitarian Logistics

Humanitarian logistics (HL) refers to efficient and cost-effective planning, inventory control, and distribution of goods or materials for the primary purpose of alleviating the suffering of vulnerable populace (Thomas and Kopczak, 2005). The past two decades have seen a surge in HL research and several review papers have synthesized research findings (Altay and Green III, 2006; Leiras et al., 2014; Minas et al., 2020). Additionally, several narrow-focus reviews on facility location (Boonmee et al., 2017; Dönmez et al., 2021), vehicle routing (Anuar et al., 2021), and network restoration and recovery operations (Çelik, 2016) decisions for HL are also available.

Holguín-Veras et al. (2012) further segregates the field of HL based on the event planning horizon. The first is humanitarian operations that focus on long-term disaster recovery and humanitarian assistance, where the primary goal is mostly operational efficiency. Whereas the other is humanitarian operations focusing on short-term disaster response which are characterized by a stark difference in operational environments: "... chaotic settings where urgent needs, life-or-death decisions and scarce resources are the norm." The second type of operation, described as post-disaster humanitarian logistics (PD-HL), forms the focus of the current study.

Two of the important criteria in PD-HL are capturing the dynamic (or, time-

dependent) nature of the problem, and capturing uncertainty. The dynamic nature is typically captured using multi-period formulations (Loree and Aros-Vera, 2018; Pérez-Rodríguez and Holguín-Veras, 2016; Rawls and Turnquist, 2012). The uncertainty is typically captured using two-stage stochastic programs (Aghajani and Torabi, 2019; Tofighi et al., 2016; Torabi et al., 2018), and recently two-stage robust programs (Zhu et al., 2022). However, research considering both of the criteria together is relatively scarce with (Rawls and Turnquist, 2012) modeling the multi-period stochastic model for HL operations. However, the problem is modeled as a single-level problem, and therefore, the characteristic of uncertainty realization (typically, found in two-stage formulations) is absent. In this study, we consider a two-stage formulation that models the facility location and inventory decisions in the pre-disaster first stage, and the resource allocation decisions are modeled in a dynamic manner in the post-disaster second stage.

Additionally, as the road infrastructure can be unavailable up to several days post-disaster, research efforts have been made to consider the possibility of drone deliveries for disaster management and humanitarian relief. The interested reader referred to Rejeb et al. (2021) and Daud et al. (2022) for an extensive review on this topic. For this study, we assume that the requests are fulfilled using a drone-only infrastructure. The effect of uncertainty in drone parameters is generally associated with accessibility loss. Various approaches have been proposed in the literature. Kim et al. (2019) model uncertain flight distance for drones and enforce a reliability

constraint on successful return, and the function is payload-independent. Chauhan et al. (2021) model uncertainty in battery consumption and initial battery availability with payload-dependent battery consumption function. Zhu et al. (2022) consider the effect of uncertain demand on drone battery consumption. Chauhan et al. (2022a) models drone range as a function of wind speed and direction distributions for a probabilistic representation of accessibility. In this study, we consider an approach similar to Chauhan et al. (2022a) wherein the accessibility between a distribution center and a relief shelter is probabilistically defined and is unknown to the decision-making agent initially.

### 4.2.2 Online Resource Allocation

Resource allocation problems are ubiquitous with applications in widespread fields like communication, cloud infrastructure, internet, energy markets, transportation, inventory management, logistics, and others. Here, we focus on problems where information is gradually revealed over time and are called online (or, dynamic) resource allocation problems.

Classical operations research-based methods model dynamic resource allocation problems as multi-stage stochastic problems Guo et al. (2020). Several AI-based methods have also been explored for dynamic resource allocation: Q-learning (Yu et al., 2019, 2021), multi-armed bandits (Badanidiyuru et al., 2013; Chauhan et al., 2022b; Villar et al., 2015), approximate dynamic programming (Yu et al., 2019), and

online algorithms (Devanur et al., 2011). Powell (2019) summarized the commonalities among various fields of stochastic optimization and argues that the principles of bandit problems should become a core dimension of mainstream stochastic optimization. In this study, we propose an equitable multi-armed bandit framework based on the linear contextual bandits with knapsacks problem proposed by Agrawal and Devanur (2016).

Focusing the work on Yu et al. (2021) propose an online resource allocation problem in the context of PD-HL in the first 72-hours of the disaster event. They consider a local relief shelter that receives a constant amount of resources in each time period (assumed to be one hour) which are distributed among the affected areas. They assume that the affected areas produce the same demand in each time period. They propose a Q-learning approach that learns through iterating over the entire time period multiple times. We also propose an online resource allocation problem for PD-HL in the first 72-hour period. However, we consider a prior planning stage for prepositioning available supplies at multiple distribution centers which must deliver supplies to the relief shelters. The demand from each relief shelter can be different and arrives in a dynamic manner. The primary difference is that learning and performance happen together, and can only be performed one time, as would be the case in actual disasters. We achieve these online learning and resource allocation decisions by adopting a multi-armed bandits approach.

Recently, Chauhan et al. (2022b) proposed a similar two-stage framework for

facility location and dynamic resource allocation, and used a multi-armed bandit approach. They considered the application premise of instant delivery logistics with the objective of maximizing profits. While our proposed framework is similar, we consider the application of resource distribution in post-disaster conditions with the equity-based objective. Several other nuanced differences are discussed later in Section 4.3.

### 4.2.3 Equity

Equity among the demand nodes is always a crucial aspect of planning in HL applications. Dönmez et al. (2021) observed that equity has been primarily considered in the HL literature using two main inequity metrics: demand satisfaction (or, service level) and accessibility. One of the ways to incorporate demand-based equity is by either ensuring complete demand satisfaction (Aslan and Çelik, 2019; Kim et al., 2019; Yahyaei and Bozorgi-Amiri, 2019) or considering deprivation costs (Cotes and Cantillo, 2019; Loree and Aros-Vera, 2018; Paul and Wang, 2019; Rivera-Royero et al., 2020). Holguín-Veras et al. (2013) and Shao et al. (2020) provide an extensive literature on modeling of deprivation costs for PD-HL.

Another common methodology is a Rawlsian approach (Rawls, 2020) in which the worst-off element is controlled. For example, in the case of demand equity, minimum demand satisfaction levels can be set, or in the case of accessibility, a maximum travel distance threshold can be adopted. Most commonly the accessibility-based constraints

are modeled in a Rawlsian manner to account for network discontinuity and loss of infrastructure and are commonly applied in coverage-based models. Alternatively, the Rawlsian approach can also be applied to objective functions, for example, minimizing the maximum travel distance as in the case of modeling emergency medical services.

Several research efforts in humanitarian relief operations have explored social welfare utility functions as an indicator of fairness and equity (Balcik et al., 2014; Campbell et al., 2008; Huang et al., 2012; Lien et al., 2014). All of these works consider that an agency has a social welfare function available and how that choice of objective function has a significant impact on resource allocation decisions. In contrast, Rey et al. (2018) do not make any assumptions for a social welfare function and instead propose envy-freeness as a fairness criterion for humanitarian relief operations based on the idea of envy (Foley, 1966; Varian, 1974). Envy-freeness models fair allocation of resources considering all people as equal and needing equal access. In our study, we adopt envy-based criteria for fairness and aim to minimize the total deviation from fair allocation in a stochastic system.

## 4.3 Problem Description and Formulation

### 4.3.1 Stage 1: Facility Location and Relief Supply Prepositioning in Pre-Disaster Phase

Consider an agency planning to provide medical aid in post-disaster scenarios in a service area. The planning horizon of the operations is assumed to be the first 72-hours after the disaster event. The agency has determined the locations of the relief shelters (RSs) from where the relief operations would be carried out. Let $\mathcal{G}$ (indexed using $g$) represent the set of all RS locations. The planning agency believes that there will be a $D_g$ number of medical aid packages of uniform weight and size (henceforth, referred to as "product") will be required at RS $g \in \mathcal{G}$. It is assumed that an adequate number of medical personnel are readily available at (or, can be transported to) each RS location.

The agency has access to a total of $\alpha$ amount of product, and budgets $P^F$ and $P^D$ for setting up the supplies at distribution centers (DCs) and their transportation post-disaster, respectively. It is known that $c_h^F$ and $c_{gh}^D$ are the cost for setting up a DC $h \in \mathcal{H}$ and the unit-cost for transporting product from DC $h \in \mathcal{H}$ to RS $g \in \mathcal{G}$, respectively. Without loss of generality, we assume that $c_h^F, c_{gh}^D \in [0,1] \; \forall \; g \in \mathcal{G}, h \in \mathcal{H}$. In the effort for disaster preparedness, the agency wishes to determine which distribution centers to set up and how much relief supplies to preposition at each distribution center.

A major consideration for the agency is the equitable or fair distribution of products post-disaster, primarily due to the scarcity of resources in comparison to the total demand. We choose the envy-freeness principle to guide the resource allocation decisions since it is a widely used resource allocation scheme for the public sector and social welfare applications (Foley, 1966; Rey et al., 2018).

Depending on the disaster considered, ground transportation may not be available for use in response to the disaster. In this study, we consider that only the aerial transportation option is available for use. Specifically, for this study, we employ battery-operated UAVs/drones. Additionally, for the sake of simplicity in model formulation, we assume that the amount of drone battery capacity available is much greater than the total amount of product available. Alternatively put, we consider that the product is the strictly limiting resource, and not drone battery capacity. Therefore, we do not consider constraints referring to the allocation of drone battery capacity at located facilities. However, depending on the application, they can be easily integrated into the decision-making process (see, Chauhan et al. (2022b) as an example).

**Envy-Free Fair Distribution**

We now formally define envy as used in this study. We know that $D_g$ is the demand of RS $g \in \mathcal{G}$. We say that an RS $g \in \mathcal{G}$ envies another RS $g' \in \mathcal{G}$ if envy $\mathcal{E}_{gg'} > 0$. We

define envy $\mathcal{E}_{gg'}$ of node $g$ to $g'$ as:

$$\mathcal{E}_{gg'} \coloneqq \min\{D_g, R_{g'}\} - R_g \qquad \forall\ g, g' \in \mathcal{G} \tag{4.3.1}$$

where, $R_g$ defines the amount of product received at RS $g \in \mathcal{G}$. Hence, the RS $g \in \mathcal{G}$ envies RS $g' \in \mathcal{G}$ if and only if it experiences unsatisfied demand (i.e., $D_g > R_g$) and it is receives less product than RS $g' \in \mathcal{G}$ (i.e., $R_{g'} > R_g$) (Foley, 1966; Varian, 1974).

For our problem, we are given a total of $\alpha$ amount of product that needs to be distributed among all RSs. We say that each RS $g \in \mathcal{G}$ should receive $F_g$ amount of product for the system to be envy-free (or, equitable). Algorithm 4.1 describes the process to obtain the envy-free fair distribution $\mathbf{F} \coloneqq \{F_g\ \forall\ g \in \mathcal{G}\}$, and is similar to the approach presented in Rey et al. (2018).

**Algorithm 4.1** Envy-Free Fair Distribution

**Input:** Set $\mathcal{G}$, $D_g\ \forall\ g \in \mathcal{G}$, $\alpha$
**Output:** Set $\mathbf{F}$

$\mathcal{G}' \coloneqq$ set $\mathcal{G}$ sorted in non-decreasing order of $D_g$
$F_g \leftarrow 0 \quad \forall\ g \in \mathcal{G}\ |\mathcal{G}'| > 0$
$g \leftarrow \mathcal{G}'[0]\ \alpha/|\mathcal{G}'| \leq D_g$
$F_g \leftarrow \alpha/|\mathcal{G}'|$
$F_g \leftarrow D_g$
$\alpha \leftarrow \alpha - F_g$
$\mathcal{G}' \leftarrow \mathcal{G}' \backslash \{g\}$
$\mathbf{F} \coloneqq \{F_g\ \forall\ g \in \mathcal{G}\}$
**return F**

## Summary of Nomenclature

*Sets and Indices*

| | |
|---|---|
| $\mathcal{G}$ | Set of all relief shelter locations (indexed as $g \in \mathcal{G}$) |
| $\mathcal{H}$ | Set of all potential distribution center locations (indexed as $h \in \mathcal{H}$) |

*Parameters*

| | |
|---|---|
| $\alpha$ | total amount of product (units) available for prepositioning ($\alpha \in \mathbb{Z}^+ \cup \{0\}$) |
| $\alpha_{min}$ | minimum amount of product (units) prepositioning required at an opened distribution center ($\alpha_{min} \in \mathbb{Z}^+ \cup \{0\}$) |
| $P^F$ | budget available for setting up product supply at distribution centers ($P^F \geq 0$) |
| $P^D$ | budget available for the transportation of product from distribution centers to relief shelters ($P^D \geq 0$) |
| $c_h^F$ | cost of setting up distribution center $h \in \mathcal{H}$ ($c_h^F \in [0,1]$) |
| $c_{gh}^D$ | unit cost of transporting product to relief shelter $g \in G$ from distribution center $h \in \mathcal{H}$ ($c_{gh}^D \in [0,1]$) |
| $D_g$ | estimated demand (units) at relief shelter $g \in \mathcal{G}$ ($D_g \in \mathbb{Z}^+ \cup \{0\}$) |
| $F_g$ | envy-free fair allocation determined for relief shelter $g \in \mathcal{G}$ using Algorithm 4.1 ($F_g \geq 0$) |
| $q_{gh}$ | one-way travel distance to RS $g \in \mathcal{G}$ from DC $h \in \mathcal{H}$ ($q_{gh} \geq 0$) |
| $Q$ | one-way delivery range of the vehicle ($Q \geq 0$) |
| $A_{gh}$ | 1, if RS $g \in \mathcal{G}$ can be accessed by DC $h \in \mathcal{H}$ (i.e., $q_{gh} \leq Q$), and $\mathsf{M}$ (a very large number), otherwise. |

*Decision Variables*

$u_{gh}$ product received at relief shelter $g \in \mathcal{G}$ from distribution center $h \in \mathcal{H}$

$y_h$ 1, if distribution center $h \in \mathcal{H}$, and 0, otherwise

$z_h$ amount of product prepositioned at distribution center $h \in \mathcal{H}$

$E_g$ deviation from fair allocation at relief shelter $g \in \mathcal{G}$

**Formulation**

We consider an objective of minimizing the total deviation of the resource allocation plan from the envy-free fair allocation, as shown in Equation 4.3.2.

$$\min_{x,y,z,E} \quad \sum_{g \in \mathcal{G}} E_g \tag{4.3.2}$$

Equations (4.3.3) and (4.3.4) represent the cost budget constraints for setting up DCs and transportation of product from DCs to RSs, respectively.

$$\sum_{h \in \mathcal{H}} c_h^F y_h \leq P^F \tag{4.3.3}$$

$$\sum_{g \in \mathcal{G}} \sum_{h \in \mathcal{H}} c_{gh}^D u_{gh} \leq P^D \tag{4.3.4}$$

Equation (4.3.5) completely prepositions the product at facilities. Equations (4.3.6) and (4.3.7) ensure that the product is allocated at only opened DCs and that

the minimum prepositioning requirements are met.

$$\sum_{h \in \mathcal{H}} z_h = \alpha \tag{4.3.5}$$

$$z_h \leq \alpha y_h \qquad \forall\ h \in \mathcal{H} \tag{4.3.6}$$

$$z_h \geq \alpha_{min} y_h \qquad \forall\ h \in \mathcal{H} \tag{4.3.7}$$

Equation (4.3.8) constrains the amount of product distributed from a DC to be less than the amount of product prepositioned there. Similarly, the equation (4.3.9) ensures that a RS does not receive more product than its demand.

$$\sum_{g \in \mathcal{G}} A_{gh} u_{gh} \leq z_h \qquad \forall\ h \in \mathcal{H} \tag{4.3.8}$$

$$\sum_{h \in \mathcal{H}} A_{gh} u_{gh} \leq D_g \qquad \forall\ g \in \mathcal{G} \tag{4.3.9}$$

Equations (4.3.10) and (4.3.11) measure deviation of product received at a RS from its fair allocation.

$$E_g \geq F_g - \sum_{h \in \mathcal{H}} u_{gh} \qquad \forall\ g \in \mathcal{G} \tag{4.3.10}$$

$$E_g \geq -F_g + \sum_{h \in \mathcal{H}} u_{gh} \qquad \forall\ g \in \mathcal{G} \tag{4.3.11}$$

Finally, Equations (4.3.12)-(4.3.15) represent variable definitional constraints.

$$y_h \in \{0,1\} \qquad \forall\ h \in \mathcal{H} \tag{4.3.12}$$

$$z_h \geq 0 \qquad \forall\ h \in \mathcal{H} \tag{4.3.13}$$

$$u_{gh} \geq 0 \qquad \forall\ g \in \mathcal{G}, h \in \mathcal{H} \tag{4.3.14}$$

$$E_g \geq 0 \qquad \forall\ g \in \mathcal{G} \tag{4.3.15}$$

Based on the structure of the problem, choosing cost-related parameters intelligently can allow the modeling of specific situations. Consider the following cases:

(i) Choosing $c_h^F$ as 1. This changes the interpretation of Equation 4.3.3 to ensuring that a maximum of $P^F$ facilities can be opened.

(ii) Choosing $c_{gh}^D$ as 1 and $P^D > \alpha$. This assumption essentially ensures that transportation cost constraint (Equation 4.3.4) is never a binding constraint at optimality (and, therefore, not a limiting resource). When $P^D = \alpha$, transportation cost will be a limiting constraint along with product prepositioning constraint (Equation 4.3.5). This situation also provides an alternative interpretation as follows: during the distribution of product from DC to RS, one complete battery pack is utilized by the drone (serving in a radius $Q$ from an opened DC). This is a conservative assumption as multiple drone trips might be possible. Then, $P^D$ refers to the total number of drone battery packs available. Additionally when $P^D = \alpha$, we assume that $z_h$ number of drone battery packs are located at

each facility.

Note that we model the planning stage problem using a deterministic formulation. However, the approach can be generalized to consider parameter uncertainty (demand, costs, accessibility) by using stochastic programming or robust optimization methodologies.

### 4.3.2 Stage 2: Stochastic Distribution of Relief Supplies in Post-Disaster Phase

The Stage 1 problem deals with the planning for a natural disaster. Therefore, before the disaster event is observed, we know the location-allocation plan for each distribution center (DC), distribution details, and the lowest deviation from fair allocation possible, i.e., the solution tuple $\{u^*, y^*, z^*, E^*\}$ is known. Let, $\mathcal{H}' \coloneqq \{h \in \mathcal{H} : y_h^* = 1\}$ be the set of all opened DCs. The medical care package inventory at DC $h \in \mathcal{H}'$ is given by $z_h^*$. As budgeted in the planning stage, the transportation cost budget for package distribution is given by $P^D$.

The planning horizon for the operational stage (Stage 2) problem is 72-hours. This planning period is alternatively represented by the total number of requests received during the event horizon and is represented using $T$ (set $\mathcal{T} \coloneqq \{1, 2, \ldots, T\}$. We abuse the notation a little bit here and use $t$ to also represent the time when the $t^{th}$ request ($\in \mathcal{T}$) arrives. During the operational stage, a major source of uncertainty is about the demand generation at relief shelters (RSs). We assume the probability that

RS $g \in \mathcal{G}$ places a request at time $t \in \mathcal{T}$ is drawn from an unknown static distribution $\mathcal{P}$. Another source of uncertainty is the delivery range $Q$ of the delivery vehicle. For ground-based vehicles, this would be affected by the post-disaster event loss of infrastructure in the road transportation network. For unmanned aerial vehicles (UAVs) (or, drones), the delivery range is affected by weather conditions, mostly by wind speed and directions (Glick et al., 2022). Additionally, there might be uncertainties related to transportation costs for package distribution. As the requests for medical care packages arrive in an online manner, the uncertainty is realized stochastically. For the online operational stage problem, we explore three methodologies: first, a multi-armed bandits-based approach; second, a random-choice heuristics; and third, a deterministic allocation policy.

The objective of the operational stage problem is the minimization of the sum of deviations from envy-free fair distribution while respecting the package consumption constraints at each opened DC. Later, we show that the operational stage objective is equivalent to the maximization of the cumulative number of successful allocations of incoming requests. This is especially an attractive property as almost all multi-armed bandit frameworks are designed for maximizing cumulative rewards.

The organization of this section is as follows: first, we discuss the methodology for incorporating and evaluating fairness, and then, the methodologies for online resource allocation are discussed.

**Incorporating Equity**

In this study, we incorporate equity in decision-making by adjusting the non-skipping probability of a request placed by RS $g \in \mathcal{G}$. Let $g_t \in \mathcal{G}$ denote the RS placing a request at time $t \in \mathcal{T}$, $\hat{D}_g^t$ denote the number of requests received from RS $g \in \mathcal{G}$ up to time and including $(t-1) \in \mathcal{T}$, and demand generation probability distribution $\hat{\mathcal{P}}^t$ be the realization of the unknown static probability distribution $\mathcal{P}$ at time $t \in \mathcal{T}$. We calculate $\hat{\mathcal{P}}^t$ as:

$$\hat{\mathcal{P}}^t(g_t = g) \coloneqq \frac{\hat{D}_g^t}{\sum_{g \in \mathcal{G}} \hat{D}_g^t} \equiv \frac{\hat{D}_g^t}{t} \qquad \forall\ g \in \mathcal{G} \tag{4.3.16}$$

Let $\tilde{D}_g^{T-t}$ denote the expected number of requests that would be placed by the RS $g \in \mathcal{G}$ in the remaining planning horizon $T-t$. We estimate $\tilde{D}_g^{T-t}$ as:

$$\tilde{D}_g^{T-t} \coloneqq \max\left\{ \left\lceil \frac{T-t}{t} \hat{D}_g^t \right\rceil, 1 \right\} \equiv \max\left\{ \left\lceil (T-t) \cdot \hat{\mathcal{P}}^t(g_t = g) \right\rceil, 1 \right\} \qquad \forall\ g \in \mathcal{G} \tag{4.3.17}$$

Note that $\tilde{D}_g^{T-t}$ is a function of the realized demand generation probability distribution $\hat{\mathcal{P}}^t$. The above definition allows $\tilde{D}_g^{T-t}$ to have a minimum value of 1 for stability. Let $\hat{z}_h^t$ denote the number of remaining medical care packages at DC $h \in \mathcal{H}'$ at the start of time $t \in \mathcal{T}$, and $\tilde{F}_g^t$ denote the envy-free fair allocation of remaining packages at $t \in \mathcal{T}$ for the remainder of the planning horizon. Let $\lambda_g^t$ denote the non-

skipping probability of a request placed by RS $g \in \mathcal{G}$ at time $t \in \mathcal{T}$. The non-skipping probabilities at time $t \in \mathcal{T}$ are calculated as:

$$\lambda_g^t = \frac{\tilde{F}_g^t}{\tilde{D}_g^{T-t}} \qquad \forall\ g \in \mathcal{G} \tag{4.3.18}$$

Intuitively, this means that by skipping a $(1 - \lambda_g^t)$ proportion of demand from the RS $g \in \mathcal{G}$ (i.e., $\tilde{D}_g^{T-t}$), we would remain with an expected $\tilde{F}_g^t$ requests, which if successfully allocated would lead to an envy-free allocation. The expression can be easily modified to include considerations like the probability of failure to reach RS $g \in \mathcal{G}$ (multiply by the reciprocal of the estimated failure probability value).

We assume that that $\lambda_g^t$ are updated after every $T_{FU}$ requests, and we call this event as fairness update. Fairness update is presented in Algorithm 4.2. We initialize the non-skipping probabilities as: $\lambda_g^1 = F_g/D_g\ \forall\ g \in \mathcal{G}$ using the planning stage (Stage 1) problem parameters.

**Algorithm 4.2** Fairness Update

**Input:** Set $\mathcal{G}$, time $t \in \mathcal{T}\backslash\{1\}$, fairness update frequency $T_{FU}$, observed demands $\hat{D}_g^t\ \forall g \in \mathcal{G}$, total remaining resources $\sum_{h \in \mathcal{H}'} \hat{z}_h^t$

**Output:** Non-skipping probabilities $\lambda_g^t\ \forall\ g \in \mathcal{G}$

1: **if** $t \% T_{FU} = 0$ **then**
2: Calculate estimated remaining demands $\tilde{D}_g^{T-t}\ \forall g \in \mathcal{G}$ using Equation (4.3.17).
3: Calculate envy-free fair allocations $\tilde{F}_g^t\ \forall g \in \mathcal{G}$ using Algorithm 4.1 (inputs $\mathcal{G}$, $\tilde{D}_g^t\ \forall g \in \mathcal{G}$, $\sum_{h \in \mathcal{H}'} \hat{z}_h^t$)
4: Update non-skipping probabilities $\lambda_g^t \coloneqq \tilde{F}_g^t / \tilde{D}_g^{T-t} \quad \forall\ g \in \mathcal{G}$ (Equation (4.3.18))
5: **else**
6: Do not update non-skipping probabilities, i.e., $\lambda_g^t = \lambda_g^{t-1} \quad \forall\ g \in \mathcal{G}$
7: **end if**

**Proposition 1** *When $t$ is large enough, the non-skipping probabilities $\lambda_g^t \ \forall g \in \mathcal{G}$ converge to their (unknown) expected values.*

**Proof.** As demands are randomly sampled from the unknown static distribution $\mathcal{P}$, we know that the expected value of $\hat{D}_g^t$ is $t \cdot \mathcal{P}(g_t = g)$. Using this we estimate the realized demand generation probability distribution $\hat{\mathcal{P}}^t$ using Equation (4.3.16). As the demands are randomly sampled from $\mathcal{P}$, it can be shown that $\hat{\mathcal{P}}^t$ converges to $\mathcal{P}$ for a large enough $t$. Note, as the sample represents a random sample of a multinomial distribution, we obtain the same estimate for $\hat{P}^t$ using maximum likelihood theory.

Extending the observations, the expected value of remaining demand $\tilde{D}_g^{T-t}$ is given as $(T-t) \cdot \mathcal{P}(g_t = g)$. As the variable $\tilde{D}_g^{T-t}$ can only assume integral values, we estimate it using Equation (4.3.17). Additionally, a minimum value of 1 is adopted for numerical stability (as division by 0 is not defined). Therefore, for a large enough $t$, the estimated $\tilde{D}_g^{T-t}$ also converges to its expected value. As $\tilde{D}_g^{T-t}$ converges, the estimation of fair distribution of remaining resources at time $t$, $\tilde{F}_g^t$, would also converge to its true expected value.

The terminal expected values of demands at location $g \in \mathcal{G}$ is given as $\bar{D}_g^T = T \cdot \mathcal{P}(g_t = g)$. Therefore, the expected value of terminal envy-free fair distribution (denoted using $\bar{F}_g^T$) can be calculated using expected demands $T \cdot \mathcal{P}(g_t = g) \ \forall g \in \mathcal{G}$ and total available packages $\alpha$. Therefore, by satisfying an expected $\bar{F}_g^T / \bar{D}_g^T$ proportion of requests placed by location $g \in \mathcal{G}$ we can achieve fair distribution. We denote this expected non-skipping probability value for each demand location $g \in \mathcal{G}$

as $\bar{\lambda}_g$. We extend the observation by assuming that the total available packages are used uniformly across the planning horizon. Therefore, at time $t \in \mathcal{T}$, the remaining expected demand at location $g \in \mathcal{G}$ would be $\bar{D}_g^{T-t} = (T-t) \cdot \mathcal{P}(g_t = g)$, and the expected value of total packages available would be $(T-t) \cdot \alpha$. Let, $\bar{F}_g^t$ denote the envy-free fair distribution of remaining resources at time $t$. Let $\lambda_g^t$ be the estimate of $\bar{\lambda}_g$ at time $t$ (calculated using (4.3.18)). The following relationship holds:

$$\lambda_g^t := \frac{\tilde{F}_g^t}{\tilde{D}_g^{T-t}} \approx \frac{\bar{F}_g^t}{\bar{D}_g^{T-t}} \equiv \frac{\bar{F}_g^T}{\bar{D}_g^T} = \bar{\lambda}_g \qquad \forall\ g \in \mathcal{G}, t \text{ is large enough}$$

□

**Corollary 1** *The objective of maximizing the cumulative successful allocation in the operational stage is equivalent to the proposed objective of minimizing the sum of deviation from fair allocation.*

**Proof.** From Proposition 1, we know that the non-skipping probabilities converge to their expected value which results in the minimization of the sum of deviations from fair distribution. Therefore, the successful allocation of the $\lambda_g^t$ proportion of demand generated by location $g \in \mathcal{G}$ that is not skipped would also lead to the minimization of cumulative deviation from fair distribution. Now, consider the demand requests that are skipped to be an unsuccessful allocation. Therefore, more generally, the maximization of cumulative successful allocation is equivalent to minimizing the sum of deviations from fair allocation.

□

**Equitable Linear Contextual Bandits with Knapsacks**

Multi-armed bandits are a reinforcement-learning framework that maximizes the cumulative rewards at the end of the time horizon by exploring the "arms" (i.e., a set of options). The problems, therefore, inherently exhibit an explore-exploit tradeoff. Here, specifically, we use a modified version of linear contextual bandits with knapsacks (linCBwK) proposed by Agrawal and Devanur (2016), which we call equitable linCBwK (E-linCBwK). Like linCBwK, E-linCBwK allows to explicitly account for budget constraints (which represent utilization of medical care packages at each DC, and the overall transportation cost constraint) while allowing to maximize rewards by choosing a subset of options. Recently, Chauhan et al. (2022b) used the linCBwK framework to model dynamic resource allocation of demands to facilities for drone-based instant delivery logistics operations with the objective of maximizing profits. A major difference is that E-linCBwK models the minimization of total deviation from fair distribution instead of maximizing profits. We also show that the E-linCBwK maintains the same regret bound as linCBwK. Typically, the linCBwK ends at the first instance of a budget constraint violation (i.e., the first instance when either the medical care packages exhaust at a DC, or the transportation cost constraint is violated). However, deliveries are still possible until the medical care packages are available at a DC and the transportation cost constraint is not violated and should be

considered given the critical nature of our problem (resource allocation immediately after a disaster event). We consider this extension for E-linCBwK but do not provide a new regret bound for the extension.

An E-linCBwK problem consists of five components, the same as a linCBwK problem. The first is the $K$ number of arms or actions. Here, these actions represent options for each request, i.e., drone delivery from an opened DC. Therefore, we have $K = |\mathcal{H}'|$ arms and $[K] := \{h \ \forall \ h \in \mathcal{H}'\}$ is the set of all arms. Additionally, the algorithm consists of an implicitly modeled "no-op" arm which represents that no action must be taken (or equivalently, the request should be skipped).

The second is the planning horizon of the problem or total number of decision-making events represented using $T$. Here, each time/event $t \in \mathcal{T} \coloneqq \{1, 2, \ldots, T\}$ represents an order placed in real-time by demand point $g_t \in G$ (which is assumed i.i.d. from unknown distribution $\mathcal{P}$, abbreviated as $g_t \overset{iid}{\sim} \mathcal{P}$). Due to stochasticity in weather conditions, let $A^t_{g_th}$ ($A^t_{g_th}|g_t, h \overset{iid}{\sim} \mathcal{D}^A_{g_th}$) be 1 if the request placed by RS $g_t$ can fulfilled by drone delivery from DC $h \in \mathcal{H}'$, and 0, otherwise. Therefore, the set of DCs that can access the RS $g_t$ at time $t$ is given as $H'^t_{g_t} \coloneqq \{h \in \mathcal{H}' : A^t_{g_th} = 1\}$. Let $c^t_{g_th}$ ($c^t_{g_th}|g_t, h \overset{iid}{\sim} \mathcal{D}^c_{g_th}$) be the transportation cost for the satisfying the request placed by RS $g_t$ at time $t$ from DC $h \in \mathcal{H}'$ (w.l.o.g., we assume that $c^t_{g_th} \in [0, 1]$). Additionally, at time $t$, let $\hat{z}^t_h$ represent the remaining medical care packages at DC $h \in \mathcal{H}'$. All of the above information is available before making the allocation decision for event $t$.

The third is the context. Context is a representation of the information that is available prior to making a decision. We observe an $m$-dimensional context vector for each arm $a \in [K]$ for each event $t$, represented by $\mathbf{x}_t(a) \in [0,1]^m$. Let the context matrix $X_t := \{\mathbf{x}_t(a)\ \forall a \in [K]\} \in [0,1]^{m \times K}$. For our application, we observe an $m = K$ dimensional context for each arm. The $K \times K$ diagonal context matrix is constructed as:

$$X_t[h,h] := \begin{cases} 1 & \text{if } \hat{z}_h^t > 0 \text{ and } h \in \mathcal{H}'^t_{g_t} \\ 0 & \text{otherwise} \end{cases}, \quad \forall\ h \in \mathcal{H}' \tag{4.3.19}$$

Agrawal and Devanur (2016) state that when the context matrix is a $K$-dimensional identity matrix, linCBwK emulates the bandits with knapsacks (BwK) problem (Badanidiyuru et al., 2013). Similar to Chauhan et al. (2022b), we extrapolate this result to consider a BwK problem with restricted arm availability for E-linCBwK. The above definition of context only allows arms with context equal to 1 to be available for selection. Alternatively, choosing an arm with context equal to 0 is equivalent to choosing the "no-op" arm. We consider that an arm $a \in [K]$ is available if it has packages available and can successfully deliver a package to RS $g_t$.

The reward is the fourth component of an E-linCBwK problem. At time $t$, the agent receives a scalar reward $r_t(a_t) \in [0,1]$ after playing action $a_t \in [K]$. The objective of the bandit problem is to maximize cumulative rewards. Per Corollary 1,

we can represent our considered objective of minimizing the sum of deviation from fair allocation to maximizing the cumulative successful allocations. We define a successful allocation as choosing an arm $a_t$ such that $X_t[a_t, a_t] = 1$. Without loss of generality, we assume that the reward for a successful allocation is 1.

The fifth component of E-linCBwK is globalized budget constraints (or, knapsack constraints). For our application, there are $d = (|\mathcal{H}'| + 1)$ universal knapsack constraints. The first $|\mathcal{H}'|$ constraints represent the medical care package consumption at facility $h \in \mathcal{H}'$ with budgets $z_h^*$. The last knapsack represents the transportation cost constraint and has a budget value of $P^D$. At time $t$, let DC $h_t$ represent the arm chosen by the algorithm. If $X_t[h_t, h_t] = 1$, $\mathbf{e}_{h_t}$ amount of medical care packages are consumed, and the cost for transportation is $c^t_{g_t h_t}$. Here, $\mathbf{e}_{h_t}$ is a $|H'| \times 1$ vector with value 1 where $h = h_t$, and 0, otherwise. If $X_t[h_t, h_t] = 0$, the action is equivalent to using the "no-op" arm and a total of $\mathbf{0}$ units of medical care packages are consumed at DCs, and no transportation cost is incurred.

The bandit optimization problem that we are tackling here is given as:

$$\max_{I} \quad \sum_{t \in \mathcal{T}} 1 \cdot X_t[h, h] \cdot I_t^h \tag{4.3.20}$$

$$\text{s.to.} \quad \sum_{t \in \mathcal{T}} 1 \cdot X_t[h, h] \cdot I_t^h \leq z_h^* \qquad \forall\ h \in \mathcal{H}' \tag{4.3.21}$$

$$\sum_{t \in \mathcal{T}} \sum_{h \in \mathcal{H}'} c^t_{g_t h} \cdot X_t[h, h] \cdot I_t^h \leq P^D \tag{4.3.22}$$

$$\sum_{h \in \mathcal{H}'} I_t^h = 1 \qquad \forall\ t \in \mathcal{T} \tag{4.3.23}$$

$$I_t^h \in \{0,1\} \qquad \forall\ h \in \mathcal{H}', t \in \mathcal{T} \tag{4.3.24}$$

The objective (Equation (4.3.20)) is to maximize the cumulative successful allocation. Equation (4.3.21) represents medical care package consumption at opened DCs, and Equation (4.3.22) represents the transportation cost budget constraint. Equations (4.3.21) and (4.3.22) together represent the knapsack constraints for our E-linCBwK problem. Equations (4.3.23) and (4.3.24) represent the basic decision space for choosing the fulfillment option.

At time $t$, upon selection of arm $a_t$, let $\mathbf{v}_t(a_t)$ be the $d$-dimensional resource consumption vector. One of the requirements for implementing linCBwK is that the vector $\mathbf{v}_t(a_t)$ must be in the range [0,1]. This condition is satisfied as each request consumes up to 1 unit of the medical care package, and in the planning stage, w.l.o.g, we assumed that the transportation costs are in the range [0,1]. Another requirement for linCBwK is a uniform budget value for all the knapsack constraints. Therefore, we scale each knapsack so that its budget is equal to the lowest value of all the budgets. The new budget, $B$, is given as:

$$B = \min\left\{ \min\{z_h^* : h \in H'\}, P^D \right\} \tag{4.3.25}$$

The transformations to make the budget the same for all knapsack constraints are:

Package consumption knapsacks: $$1 \cdot X_t[h,h] \leftarrow \frac{B}{z_h^*} \cdot 1 \cdot X_t[h,h]; \qquad z_h^* \leftarrow B \tag{4.3.26}$$

Transportation cost knapsack: $$c_{g_t h}^t \cdot X_t[h,h] \leftarrow \frac{B}{\omega} \cdot c_{g_t h}^t \cdot X_t[h,h]; \qquad P^D \leftarrow B \tag{4.3.27}$$

There are three major changes in E-linCBwK compared to linCBwK:

- Foremost is the objective in consideration. The E-linCBwK considers minimization of the sum of deviation from the fair distribution. Using Corollary 1, we show that the equivalent objective is maximizing the cumulative successful allocations which can be modeled using the existing linCBwK. Critical to achieving the equity-based objective is the modeling of non-skipping probabilities which requires certain requests to be skipped. While the "no-op" arm is generally available in the linCBwK framework, mandatory skipping is not considered.

- The second is the termination criteria. The linCBwK ends at the first instance of a knapsack constraint being violated. This stems from an inherent assumption that choosing an arm leads to resource consumption for all the knapsack constraints, i.e., "globalized" knapsacks. However, we deal with the case of "stratified" knapsacks instead of "globalized" knapsacks, meaning that the resource consumption vector is usually very sparse, and the indices of zero

resource consumption are known for all arms. Specifically, consider a case when we know that transportation cost would be the last constraint to be violated (i.e, $P^D \geq \alpha$). In such a case, the resource allocation can go on until medical care packages are exhausted at all DCs instead of the first instance of medical care package exhaustion at any opened DC. This is incorporated using a time-varying context distribution, as detailed in Equation (4.3.19). Note that the context distribution is static until the first package consumption knapsack is violated.

- The third is tackling budget loss while estimating the penalty parameter $Z$. In linCBwK, the first $T_0$ requests are dedicated to the exploration of the solution space and determining the penalty parameter $Z$. The Lagrangian-like parameter $Z$ is used to calculate the reward estimates penalized with estimated resource consumption for the later $T' = T - T_0$ requests. A budget of $B_0 = T_0$ is dedicated during the initial $T_0$ so as to model a constraint-free bandit problem. Later, only a budget of $B' = B - B_0$ is available for the remaining $T'$ requests. Given the critical nature of our application, we consider that all the remaining resources after the first $T_0$ requests are available for the later $T'$ requests so that no resources are wasted.

The E-linCBwK decision-making procedure has six major steps. The first is observing the context matrix $X_t$, and the second is obtaining optimistic estimates of $\boldsymbol{\mu}_*$ and $W_*$ using the $l_2$-regularized norms of previously observed values of rewards

and resource consumption. The third is arm selection using an expected reward penalized with expected resource consumption and the fourth is realizing the values of reward and resource consumption for the selected arm. The fifth step is updating the penalty weights using the multiplicative weight update (MWU) algorithm (Agrawal and Devanur, 2016). Until the fifth step, the procedure is very similar to linCBwK. And finally, the sixth step is applying the fairness update (Algorithm 4.2). The detailed algorithm is in Algorithm 4.3. Like all multi-armed bandit algorithms, the performance is measured by the complexity of the cumulative regret. For linCBwK, the regret is measured from the optimal static policy (Agrawal and Devanur, 2016), obtained from solving a static stochastic optimization problem. We show that the regret bound E-linCBwK until the violation of the first knapsack constraint, we show that the problem can be modeled using a linCBwK framework.

**Proposition 2** *Until the first instance of a knapsack constraint violation, E-linCBwK can be modeled as the linCBwK problem, originally proposed by Agrawal and Devanur (2016).*

**Proof.** The following two assumptions about context, rewards, and resource consumption vectors are made for linCBwK (Agrawal and Devanur, 2016):

- In every round $t$, until the first instance of a knapsack constraint violation, the tuple $\{\mathbf{x}_t(a), r_t(a), \mathbf{v}_t(a)\}_{a=1}^{K}$ is generated from an unknown distribution $\mathcal{D}$, independent of everything in previous rounds. The procedure used for gener-

ating contexts, rewards, and resource consumption for our equitable resource distribution problem for post-disaster scenarios satisfies this assumption.

- There exists an unknown vector $\boldsymbol{\mu}_* \in [0,1]^{m\times 1}$ and a matrix $W_* \in [0,1]^{m\times d}$ such that for every arm $a$, given contexts $\mathbf{x}_t(a)$, and history $H_{t-1}$ before time $t$,

$$\mathbb{E}[r_t(a)|\, x_t(a), H_{t-1}] = \boldsymbol{\mu}_*^{\mathsf{T}} x_t(a), \qquad \mathbb{E}[\mathbf{v}_t(a)|\, x_t(a), H_{t-1}] = W_*^{\mathsf{T}} x_t(a) \quad (4.3.28)$$

To prove the proposition, we first show that the uncertain parameters in the E-linCBwK also satisfy the above assumptions. Secondly, we show that the budget loss upgrade would only improve the regret bound. Finally, we show a way to incorporate the E-linCBwK into the linCBwK framework.

We already described that for any request $t \in \mathcal{T}$, the realization of demand point ($g_t$), accessibility ($A^t_{gh}$), and transportation costs ($c^t_{gh}$) are drawn from an unknown independent distribution. The only remaining parameters are the non-skipping probabilities, $\lambda^t_g$. Using Proposition 1, we can show that $\lambda^t_g$ adequately estimate the expected non-skipping probabilities, $\bar{\lambda}_g$, which are unknown constants. Therefore, the parameters $\lambda^t_g$ are also drawn a static unknown distribution.

In linCBwK, during the first $T_0$ requests which are used to estimate the penalty parameter $Z$, a budget of $B_0 = T_0$ is allocated. For the remaining $T' = T - T_0$ requests, only a budget of $B' = B - B_0$ is available. For E-linCBwK, we transfer the remaining budgets after the $T_0$ requests to the budget for the next $T'$ requests.

We again rescale the budgets to a uniform value of $B''$ after the first $T_0$ requests. Therefore, the following relationship is true, $B'' \geq B'$, and the regret bound for the new problem would only improve.

For E-linCBwK, we assumed that the demand points are generated from an unknown distribution $\mathcal{P}$. Now, consider the following reformulation consisting of $|\mathcal{G}|+1$ demand points, of the fist $|\mathcal{G}|$ are the same as the original problem and the last demand point is always inaccessible. We denote the new set of demand points using $\mathcal{G}'$. During the operational stage, consider that the demand points $g' \in \mathcal{G}'$ are sampled from an unknown static distribution $\mathcal{P}'$ which follows the following relationship with the original problem.

$$\mathcal{P}'(g_t = g) = \lambda_g^t \cdot \mathcal{P}(g_t = g) \qquad \forall\ g \in \mathcal{G}$$

$$\mathcal{P}'(g_t = |\mathcal{G}'|) = 1 - \sum_{g \in \mathcal{G}} \mathcal{P}'(g_t = g)$$

Now, for the new problem, as the last point is always inaccessible, the context matrix is always a zero matrix, and therefore, would always result in the selection of the "no-op" arm which is mathematically equivalent to skipping a request.

□

**Corollary 2** *Until the first instance of a knapsack constraint violation, E-linCBwK has at least the same regret bound as the linCBwK problem.*

**Proof.** A direct consequence of Proposition 2.

□

**Algorithm 4.3** E-linCBwK Algorithm

**Input:** $K$, $m$, $d$, $B$, $T$, $(1-\delta)$ confidence level, MWU algorithm parameter $\epsilon$
**Output:** Decisions $I$

1: Obtain decisions $I$ for time $t = \{1, 2, \ldots, T_0\}$, parameter $Z$, and non-skipping probabilities $\lambda_g^{T_0}$ $g \in \mathcal{G}$, remaining packages at DCs $\hat{z}_h^{T_0}$ $h \in \mathcal{H}'$, and remaining transportation cost budget $\hat{P}_{T_0}^D$ using Algorithm 4.4.
2: Initialize $t = 1$, $\boldsymbol{\theta}_{1,j} = \frac{1}{1+d}$, $\forall\, j \in \{1, 2, \ldots, d\}$, $rad_t = \sqrt{m \log\left(\frac{d+tmd}{\delta}\right)} + \sqrt{m}$
3: Initialize $B' = B - T_0$, $T' = T - T_0$ $t \leq T'$
4: Observe the RS placing the request, $g_{T_0+t}$. Update $\hat{D}^{T_0+t} = \hat{D}^{T_0+t-1} + \mathbf{e}_{g_{T_0+t}}$
5: **if** $rand[0,1] \geq \lambda_{g_t}^t$ **then**
6: The context $X_t$ is a $K \times K$ matrix of zeros
7: **else**
8: Observe context $X_t$
9: **end if**
10: For every $a \in [K]$, compute $\tilde{\boldsymbol{\mu}}_t(a)$ and $\tilde{W}_t(a)$ (the optimistic estimates of $\boldsymbol{\mu}_*$ and $W_*$) as:

$$\begin{aligned}
&\tilde{\boldsymbol{\mu}}_t(a) \coloneqq \arg\max_{\boldsymbol{\mu} \in C_{t,0}} \mathbf{x}_t(a)^\intercal \boldsymbol{\mu}, \quad \text{where, } \hat{\boldsymbol{\mu}}_t \;:=\; M_t^{-1} \sum_{i=1}^{t-1} \mathbf{x}_i(a_i) r_i(a_i)^\intercal \\
&\text{where, } C_{t,0} \coloneqq \left\{\boldsymbol{\mu} \in \mathbb{R}^{m\times 1} : \|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}_t\|_{M_t} \leq rad_t\right\} \\
&\text{where, } M_t = identity(K) + \sum_{i=1}^{t-1} \mathbf{x}_i(a_i)\mathbf{x}_i(a_i)^\intercal
\end{aligned} \tag{4.3.29}$$

$$\begin{aligned}
&\tilde{W}_t(a) \coloneqq \arg\min_{W \in \mathcal{G}_t} \mathbf{x}_t(a)^\intercal W \boldsymbol{\theta}_t, \quad \text{where, } \hat{W}_t \;:=\; M_t^{-1} \sum_{i=1}^{t-1} \mathbf{x}_i(a_i)\mathbf{v}_i(a_i)^\intercal \\
&\text{where, } \mathcal{G}_t \coloneqq \left\{\mathbb{R}^{m\times d} : \mathbf{w}_j \in C_{t,j}\right\} \\
&\text{where, } C_{t,j} \coloneqq \left\{\mathbf{w} \in \mathbb{R}^{m\times 1} : \|\mathbf{w} - \hat{\mathbf{w}}_{tj}\|_{M_t} \leq rad_t\right\}
\end{aligned} \tag{4.3.30}$$

Play the arm $a_t \coloneqq \arg\max_{a\in[K]} \mathbf{x}_t(a)^\mathsf{T}\left(\tilde{\boldsymbol{\mu}}_t(a) - Z\tilde{W}_t(a)\boldsymbol{\theta}_t\right)$, Update $I^t = \mathbf{e}_{a_t}$
Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$
Update $\hat{z}^{T_0+t} = \hat{z}^{T_0+t-1} - \mathbf{e}_{a_t} \cdot X_t[a_t, a_t]$ and $\hat{P}^D_{T_0+t} -= c^{T_0+t}_{g_{T_0+t}a_t} \cdot X_t[a_t, a_t]$
Update $\boldsymbol{\theta}_{t+1}$ using MWU algorithm, and with $g_t(\boldsymbol{\theta}_t) \coloneqq \boldsymbol{\theta}_t \cdot \left(\mathbf{v}_t(a_t) - \frac{B}{T}\mathbf{1}\right)$, as:

$$\begin{aligned} \boldsymbol{\theta}_{t+1,j} &= \frac{w_{t,j}}{1+\sum_j w_{t,j}}, \\ \text{where, } w_{t,j} &= \begin{cases} w_{t-1,j}(1+\epsilon)^{g_{t,j}} & \text{if } g_{t,j} > 0, \\ w_{t-1,j}(1-\epsilon)^{-g_{t,j}} & \text{if } g_{t,j} \le 0. \end{cases}, \quad \forall\, j \in \{1,2,\ldots,d\} \end{aligned} \tag{4.3.31}$$

Apply fairness update using Algorithm 4.2 with inputs $\mathcal{G}$, $t$, $T_{FU}$, $\hat{D}^t_{g_t} \forall g \in \mathcal{G}$, $\sum_{h\in H'} \hat{z}^t_h$
$t += 1$

**Algorithm 4.4** E-linCBwK Algorithm $-$ $Z$ Computation

**Input:** $K$, $m$, $d$, $B$, $T$, $T_{FU}$, $(1-\delta)$ confidence level

**Output:** Decisions $I$ for times $t \in \{1,2,\ldots,T_0\}$, parameter $Z$, non-skipping probabilities $\lambda^{T_0}_g\ \forall g \in \mathcal{G}$, remaining packages at DCs $\hat{z}^{T_0}_h\ h \in \mathcal{H}'$, and remaining transportation cost budget $\hat{P}^D_{T_0}$

Definition: $M$-norm of a vector $\boldsymbol{\mu}$ is $\|\boldsymbol{\mu}\|_M \coloneqq \sqrt{\boldsymbol{\mu}^\mathsf{T} M \boldsymbol{\mu}}$, where $M$ is PSD matrix.
Initialize $t = 1$, $T_0 = \lceil m\sqrt{T} \rceil$, $\gamma = \frac{2mT}{T_0}\sqrt{T_0 \log(T_0) \log\left(\frac{T_0 d}{\delta}\right)}$,
Initialize $\hat{D}^0_g = 0\ \forall g \in \mathcal{G}$, $\lambda^1_g = F_g/D_g\ \forall g \in \mathcal{G}$, $\hat{z}^0_h = z^*_h\ \forall h \in \mathcal{H}'$, $\hat{P}^D_0 = P^D\ t \le T_0$
Observe the RS placing the request, $g_t$. Update $\hat{D}^t = \hat{D}^{t-1} + \mathbf{e}_{g_t}\ rand[0,1] \ge \lambda^t_{g_t}$
The context $X_t$ is a $K \times K$ matrix of zeros
Observe context $X_t$
Calculate best arm play probability distribution $p_t$ as:

$$p_t \coloneqq \arg\max_{p\in\Delta^{[K]}} \|X_t p\|_{M_t^{-1}}, \quad \text{where, } M_t = identity(K) + \sum_{i=1}^{t-1} X_i p_i \tag{4.3.32}$$

Play arm $a_t = \arg\max_{a\in[K]} \|x_t(a)\|_{M_t^{-1}}$, Update $I^t = \mathbf{e}_{a_t}$
Observe reward $r_t(a_t)$ and resource consumption $\mathbf{v}_t(a_t)$
Update $\hat{z}^t = \hat{z}^{t-1} - \mathbf{e}_{a_t} \cdot X_t[a_t, a_t]$ and $\hat{P}^D_t -= c^t_{g_t a_t} \cdot X_t[a_t, a_t]$
Construct estimate $\hat{\boldsymbol{\mu}}_t, \hat{W}_t$ of $\boldsymbol{\mu}_*, W_*$ as:

$$\hat{\boldsymbol{\mu}}_t \coloneqq M_t^{-1} \sum_{i=1}^{t-1} (X_i p_i) r_i(a_i), \qquad \hat{W}_t \coloneqq M_t^{-1} \sum_{i=1}^{t-1} (X_i p_i) \mathbf{v}_i(a_i)^\mathsf{T} \tag{4.3.33}$$

12: Apply fairness update using Algorithm 4.2 with inputs $\mathcal{G}$, $t$, $T_{FU}$, $\hat{D}^t_{g_t} \forall g \in \mathcal{G}$, $\sum_{h \in H'} \hat{z}^t_h$

13: $t+ = 1$

14: Calculate $\hat{OPT}^\gamma$ as:

$$\hat{OPT}^\gamma := \begin{aligned} \max_{\pi} \quad & \frac{T}{T_0} \sum_{i=1}^{T_0} \hat{\boldsymbol{\mu}}_i^\mathsf{T} X_i \pi(X_i) \\ \text{s.to.} \quad & \frac{T}{T_0} \sum_{i=1}^{T_0} \hat{W}_i^\mathsf{T} X_i \pi(X_i) \leq B + \gamma \end{aligned} \tag{4.3.34}$$

15: Set $Z = 2 \left[ \frac{\hat{OPT}^{2\gamma} + 2\gamma}{B} \right]$

## 4.4 Computational Experiments

Natural disasters are a part of any region. In the state of Oregon in the United States, a major threat is the expected Cascadia Subduction Zone earthquake with odds of 7-15% for a great earthquake (M8.7-M9.3) and about 37% for a very large earthquake (M8.3-M8.6) in the next fifty years (Oregon Seismic Safety Policy Advisory Commission, 2013). For our analysis, we consider that the decision-making agency is based in the Portland Urban Metro Area as determined by the region's Urban Growth Boundary (abbreviated as Portland UGB). As per the 2020 decennial census, more than 1.5 million people reside in the Portland UGB.

The State of Oregon has set up Seismic Rehabilitation Grant Program to support seismic rehabilitation of critical buildings, particularly school districts and emergency facilities. Since its inauguration in 2019, 110 grants (cumulative value of over $220 million) have been awarded. In addition to the seismic resiliency of structures, the

availability of large open spaces would be important for setting up relief shelters (RSs) for carrying out medical relief operations post-disaster. For this study, we consider that RSs are set up at high schools and TriMet transit centers (local transportation authority) which apart from the above factors, are also strategically located and locally recognizable. During the planning stage (Stage 1), we assume that medical care is provided to people from the nearest RS, or alternatively, people access the nearest RS post-disaster and that the entire population is reachable. We consider hospitals in the region as potential distribution center sites for prepositioning medical care packages. These are locations are mapped in Figure 4.1, and detailed information is available openly at `https://github.com/drc1807/ERD_PDS`.

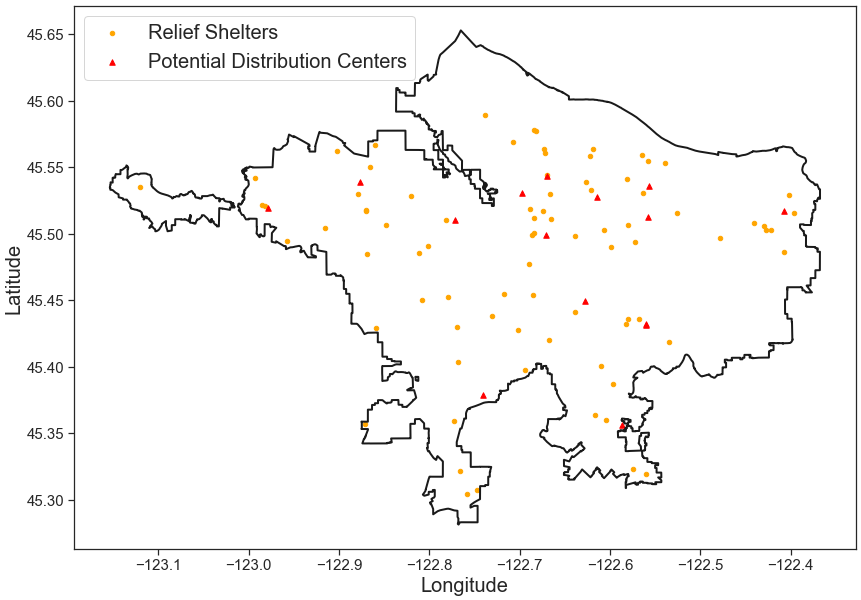



Figure 4.1: Portland Metro Area Urban Growth Boundary

According to the Oregon Seismic Resilience Plan 2013 (Oregon Seismic Safety Policy Advisory Commission, 2013), for the Portland UGB region, Tier 1 seismic

lifeline routes[1] would become minimally operational[2] in 24-72 hours of the seismic event. In the disaster response literature, 72 hours is also considered critical time saving most lives (Sheu, 2007, 2014; Yu et al., 2021). Therefore, we consider a 72-hour time horizon for the study. As ground transportation infrastructure would likely be unavailable during this period, we consider that unmanned aerial vehicles (UAVs) or drones are employed to deliver medical care packages during this time horizon. We assume that all medical care packages to be transported from DCs to RSs have the same weight of 5 kg each. We consider that the deliveries are carried out using an electric battery-operated drone, as studied in Figliozzi (2017) and Chauhan et al. (2021), with a battery capacity of 777 Wh and a range of 36 km (22.37 miles) at a 5 kg payload. Assuming a maximum battery utilization factor of 80%, we get the effective one-way range of the drone to be 14.4 km (8.9 miles). During the first 72-hour period, we assume that 150,000 medical care packages would be required (approximately 10% of the population). However, based on funding limitations, medical personnel availability, and other factors, we consider that only 120,000 medical care packages are available for distribution to RSs in the time period and that a minimum of 20,000 medical care packages should be prepositioned at each opened facility.

For the sake of simplicity, we make the following two generalizable assumptions:

(i) Firstly, we consider that the cost of setting up a distribution center ($c_h^F$) is

[1]Tier 1 is a small backbone system that allows access to all vulnerable regions, major population centers, and areas considered vital for rescue and recovery operations.

[2]A minimum level of service is restored, primarily for the use of emergency responders, repair crews, and vehicles transporting food and other critical supplies

1, and therefore, the budget available for setting up distribution centers ($P^F$) corresponds to the maximum number of facilities that can be opened.

(ii) Secondly, we assume the unit cost of transportation from any DC to RS which corresponds to the utilization of one complete drone battery pack on each trip. This is a conservative assumption as multiple drone trips might be possible while utilizing a single battery pack (refer to Chauhan et al. (2021)). Further, we consider that the total budget available for transportation ($P^D$) (which is equivalent to the total number of drone battery packs available) is equal to the number of medical care packages available for distribution. It is implied that the number of drone battery packs prepositioned at a facility is equal to the number of medical care packages prepositioned there.

The solution of the planning stage problem determines the initial state of the operational stage (Stage 2) problem (i.e., the parameters $K$, $m$, and $B$ can be determined). The generalizable assumption (ii) allows us to ignore the overall transportation cost constraint in the online allocation problem. Therefore, $d = K$ instead of $(K + 1)$. For the operational stage, we consider that a total of $T = 150,000$ requests arrive stochastically. At the time $t \in \{1, 2, 3, \ldots, T\}$, the RS placing the request is determined from a probability distribution $\mathcal{P}$, which is unknown to the algorithms. For the case study, we assume that the unknown distribution $\mathcal{P}$ is indicative of the intensity of the natural disaster, and therefore, affects the radius around RS in which medical care is provided. In general, the greater the intensity of the disaster, the more

people are injured and the infrastructure affected, and therefore, the lower the radius from where medical care can be accessed at an RS. For the planning stage, we assume that the people access the nearest RS or that the people are accessed from the nearest RS (we call this "closest") (covers the entire population: 1.56 million). For the operational stage, we assume the intensity of the natural disaster is such that people within 2 miles can access an RS or can be accessed from RS (we call this "radius 2.0") (covers 1.46 million people). These distributions are shown graphically in Figure 4.2. We consider that fairness updates are applied after every 2,000 requests (i.e., about every hour; the average number of requests obtained every hour is 2083.3).

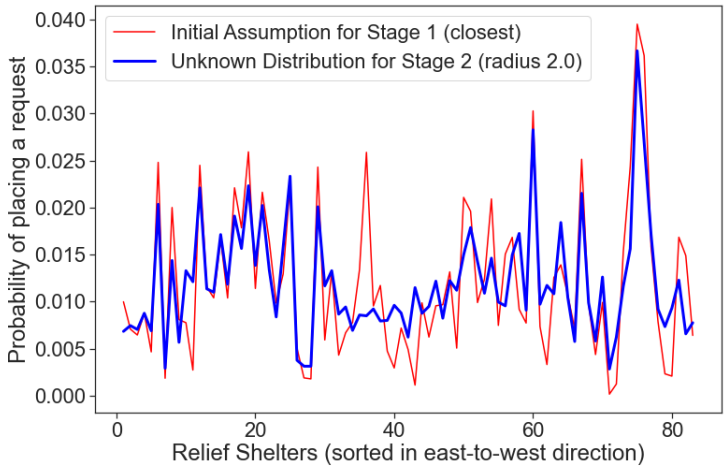



Figure 4.2: Demand Distributions

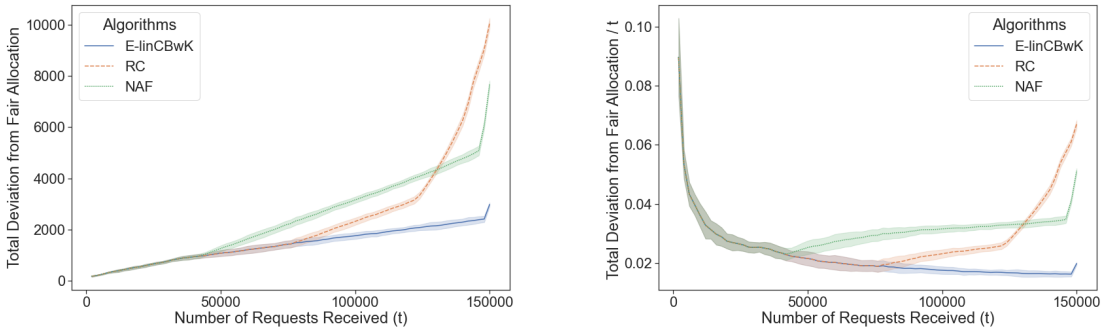
For bandit learning using E-linCBwK, in accordance to Agrawal and Devanur (2016), the $(1-\delta)$ confidence interval for estimating the reward and resource consumption is taken to be 95%, the online learning parameter $(\epsilon)$ is assumed to be 0.1, and the initial exploration iterations $(T_0)$, which are used to estimate the penalizing

parameter $Z$, are assumed to be $\lceil m\sqrt{T} \rceil$.

E-linCBwK is compared to two other heuristic methods. The first is a random choice (RC) policy which chooses one of the accessible DCs randomly if the request is not skipped. The second is a nearest available facility (NAF) allocation policy which allocates the demand to the nearest accessible DC if the request is not skipped. If there are no accessible DCs for a request, then the request is skipped for both RC and NAF policies. During the simulations, the instance is run 10 times to account for inherent randomness in the demand generation. We define regret ($R(t)$) as total deviation from the envy-free fair distribution that can be achieved up to request number $t \in \{1, 2, \ldots, T\}$, and therefore, is calculated whenever the fairness update is applied. Additionally, we define the regret-to-request ratio as the ratio of regret $R(t)$ to the number of requests received $t$. Figure 4.3 shows the evolution of regret of the three algorithms during the problem horizon. The results show that the E-linCBwK algorithm is able to achieve a strongly sub-linear regret to almost the end of the planning horizon.

Next, we investigate the value of introducing non-skipping probabilities by comparing E-linCBwK with a scenario that does not allow for request skipping, i.e., all requests are available for allocation. Figure 4.4 shows the regret evolution in both the cases. It can be noted that not allowing for skipping increases the deviation from fair allocation dramatically as requests are increased. However, the E-linCBwK No Skip policy observes an inflection point when the request $t = \alpha = 120,000$ requests

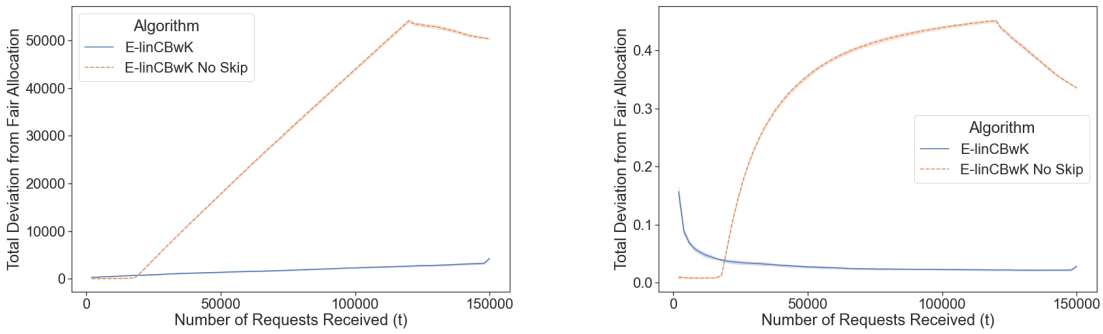
are received. Even though all requests are open to being allocated, the demands are under-satisfied by the E-linCBwK No Skip algorithm because of the usage of the "no-op" arm, whereas, the fair allocation would require complete demand satisfaction up to this point (as all requests can be potentially allocated). After the $t = \alpha$ requests mark, the fair allocation requirement proportions would drop for most demand points, and therefore, close the gap leading to improvement in the objective. The massive gap between the E-linCBwK and the E-linCBwK No Skip policies further highlights the importance of our contribution to introducing non-skipping probabilities.




(a) Regret

(b) Regret-to-request Ratio

Figure 4.3: Empirical Regret of the Algorithms (average line with standard deviation band)




(a) Regret (b) Regret-to-request Ratio

Figure 4.4: Empirical Regret of the Algorithms (average line with standard deviation band)

### 4.4.1 Sensitivity to Post-Disaster Relief Shelter Demand Distributions

Post-disaster relief shelter demand distribution is an indicator of the intensity of the natural event. Greater intensity of the natural disaster would lead to increased injury rates as well as the increased loss of infrastructure, and therefore, greater would be the impairment of accessibility. For the greatest intensity, we consider that the accessibility during the planning horizon (72-hour period after the event) would be limited to a 0.5-mile radius around each relief shelter (we call this "radius 0.5"). Similarly, other scenarios are radius 1.0 and radius 2.0. As a benchmark, we also consider the case that all people access the closest relief shelter post-disaster (we call this "closest"), which is considered as the demand distribution for the planning stage (stage 1) problem for all the cases. The demand distributions are shown in Figure 4.5. All the other parameters are the same as used in the nominal scenario.

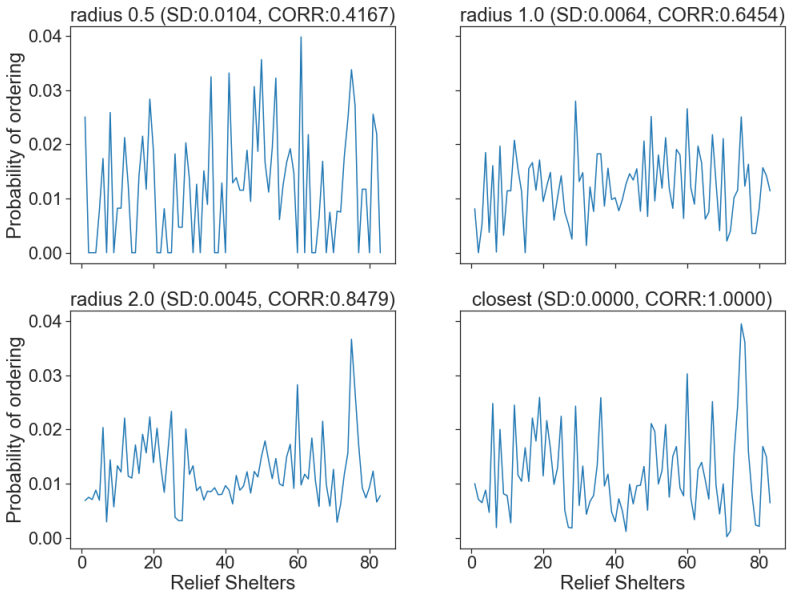
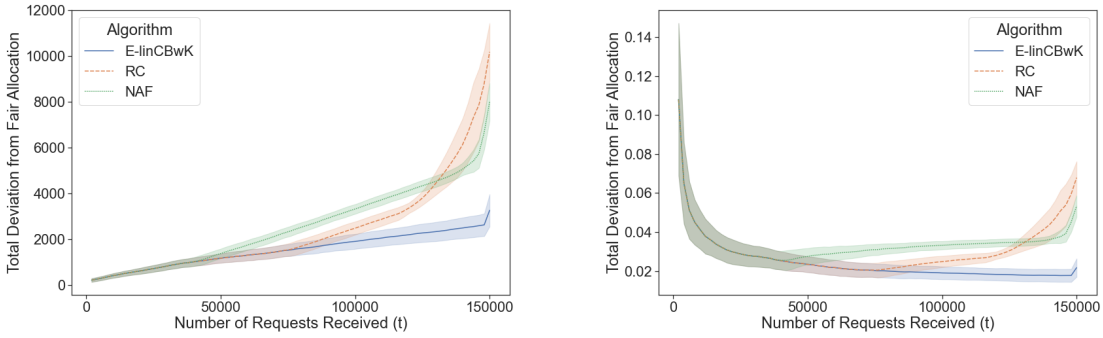



Figure 4.5: Demand Distributions for Sensitivity Analysis. The standard deviation (SD) is measured for the difference of the distribution from "closest" demand distribution. The Pearson correlation coefficient (CORR) shows correlation between the distribution and "closest" demand distribution (average line with standard deviation band).

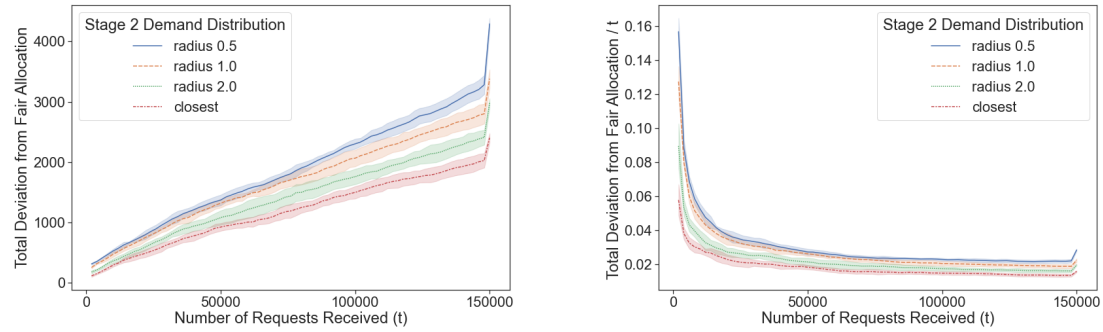
Figure 4.6 shows the evolution of the regret for the algorithms for all the four post-disaster demand distribution scenarios. The trends remain similar to the nominal scenario, with E-linCBwK performing the best, followed by NAF and then, RC. Figure 4.7 shows the regret evolution for E-linCBwK for the four different demand distribution scenarios. As the standard deviation of the operational stage (stage 2) demand distribution increases (and correlation reduces) from the planning stage (stage 1) demand distribution, the accumulated regret increases, which is as expected. While the algorithm maintains sublinear regret, the change of unknown stage 2 dis-

tribution from "closest" (as planned in stage 1) to "radius 0.5" (maximum difference from stage 1 demand distribution assumption) results in a regret increase of about 78%.




(a) Regret (b) Regret-to-request Ratio

Figure 4.6: Empirical regret of all the algorithm for all post-disaster demand distribution scenarios combined (average line with standard deviation band)




(a) Regret (b) Regret-to-request Ratio

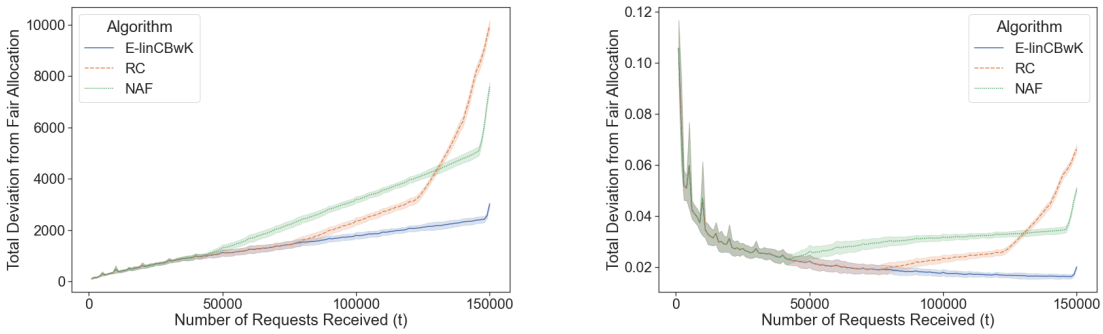
Figure 4.7: Empirical regret of E-linCBwK algorithm for all four post-disaster demand distribution scenarios (average line with standard deviation band)

### 4.4.2 Sensitivity to Fairness Update Frequency

This subsection evaluates the effect of changes in the frequency of fairness update iterations on the performance of the algorithms. The fairness update is an integral part of the algorithms as observes the ordering frequency of the relief shelters, and accordingly updates the beliefs about the remaining demand at each relief shelter, and how can the remaining resources be equitably distributed. The higher the fairness update frequency, the larger would be the difference between the sampled distribution from the actual stage 2 distribution. However, a greater number of opportunities are available to calibrate the remaining demand at relief shelters and update their non-skipping probabilities. For the nominal scenario, it was assumed that the fairness update is applied every 2,000 requests. In this section, we additionally evaluate the effect on regret when the fairness update is applied every 1,000 requests, 5,000 requests, and 10,000 requests. The rest of the parameters are the same as chosen for the nominal scenario.

Figure 4.8 shows the regret for the three algorithms for the four different frequencies of fairness update combined. The overall trend of regret evolution remains the same as the nominal scenario. Figure 4.9 shows the regret for the E-linCBwK algorithm for each fairness update frequency separately. The regret evolution for the fairness update frequencies up to every 5,000 requests is the same for the most part of the planning horizon. Even though the regret when the fairness update is applied

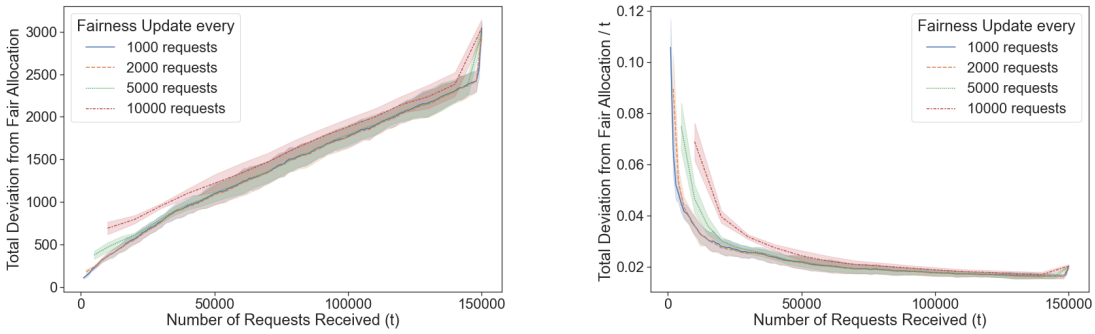
every 10,000 requests is a bit high during most of the planning horizon, the terminal regret achieved using the different fairness update frequencies is similar (see Figure 4.10). On average, the fairness update frequency of every 2,000 gives the least terminal regret, and the frequency of every 1,000 requests gives the most terminal regret (difference of 2%). One of the reasons that the results here show a minimal sensitivity to the fairness update frequency is the assumption of a static demand distribution for the operational stage problem. While this can be justified when time horizons are shorter, we believe this feature would play greater importance when the planning horizons are longer leading to time-dependent demand distributions. The study of time-dependent demand distributions is left as a future research endeavor.




(a) Regret

(b) Regret-to-request Ratio

Figure 4.8: Empirical regret of all the algorithm for all fairness update frequencies combined (average line with standard deviation band)




(a) Regret (b) Regret-to-request Ratio

Figure 4.9: Empirical regret of E-linCBwK algorithm for all four fairness update frequencies (average line with standard deviation band)

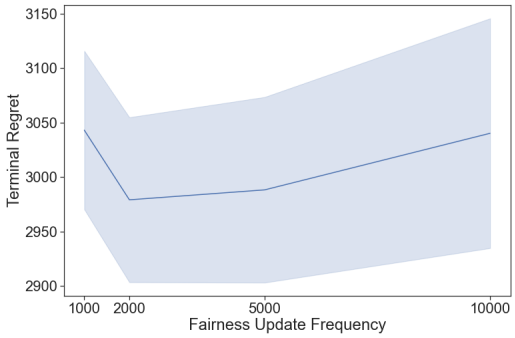


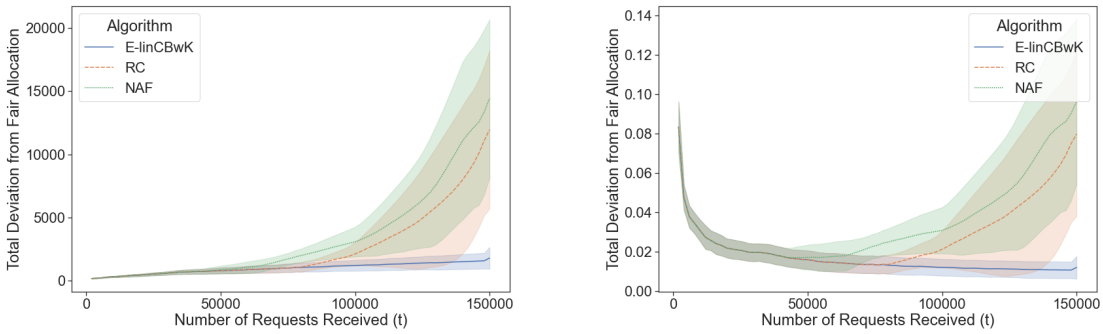

Figure 4.10: Terminal regret achieved using various fairness update frequencies (average line with standard deviation band)

### 4.4.3 Sensitivity to the Number of Opened Distribution Centers

This subsection investigates the effect of the number of opened facilities on the regret evolution of the algorithms. the key trade-off here is that increasing the number of opened facilities increases the initial accessibility to relief shelters by potentially providing them to be served with a couple of different distribution centers. While opening more facilities also reduces the number of packages prepositioned at a distri-

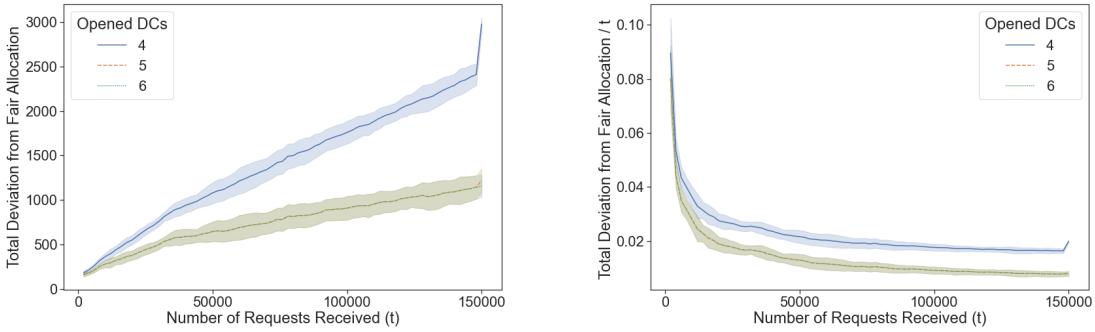
bution center, potentially leading to early exhaustion and hampering the accessibility in the later stages.

Figure 4.11 shows the regret evolution of algorithms for a different number of opened facilities combined. A difference here is that the RC algorithm provides better regret than NAF when a higher number of facilities are opened. However, their terminal regret is much greater than the terminal regret achieved by E-linCBwK. Figure 4.12 shows the regret evolution for E-linCBwK algorithm for each different number of opened distribution centers. On average, the terminal regret for E-linCBwK decreases by 59% when $P^F$ increases from 4 to 5, and further by 5% as $P^F$ increases from 5 to 6. The E-linCBwK algorithm maintains strongly sublinear regret for the entire time horizon when the number of opened facilities is 6, with an average terminal regret of 1159. Alternatively, the results mean that the E-linCBwK algorithm is able to achieve a total deviation from a fair allocation which is less than 1% of the total number of packages (120,000) distributed.




(a) Regret (b) Regret-to-request Ratio

Figure 4.11: Empirical regret of all the algorithm for different number of opened DCs combined (average line with standard deviation band)




(a) Regret

(b) Regret-to-request Ratio

Figure 4.12: Empirical regret of E-linCBwK algorithm for increasing number of opened DCs (average line with standard deviation band)

## 4.5 Conclusions

This research presents a novel two-stage facility location and dynamic resource allocation problem for relief prepositioning and their equitable distribution in post-disaster conditions. Without loss of generalization, we only consider scenarios where the total amount of resources available for distribution is less than the total demand for resources. We assume that complete fairness or equity is achieved by obtaining an envy-free allocation among the demands at relief shelter (RSs) locations. During the first stage, the planning stage, the distribution centers (DCs) are located and the relief supplies are prepositioned among them considering facility opening cost, transportation cost, and allocation constraints. We consider the objective of minimizing the total deviation from envy-free allocation under the assumption of demand distribution. During the second stage, the operational stage, the requests for resources arrive in an online manner. The centralized command makes an allocation decision

regarding which DC should be utilized or if the request should be rejected, with the objective of minimizing total deviation from envy-free fair allocation in an online manner.

The planning stage problem, representing the pre-disaster phase, is formulated as a mixed-integer linear program. The solution to the planning stage problem is used to initialize the operational stage problem, which onsets just after the disaster. To solve the dynamic resource allocation problem in the second stage, we propose a new multi-armed bandit framework, namely the equitable linear contextual bandits with knapsacks (E-linCBwK). An integral part of the E-linCBwK methodology is the non-skipping probability ($\lambda$) for requests arriving from each RS. We find the maximum likelihood estimates of the non-skipping probabilities. The estimates are then used to show that the E-linCBwK problem, which has an equity-based objective, can be modeled as an existing multi-armed bandit problem proposed by Agrawal and Devanur (2016) with an objective of maximizing the successful allocations. Consequently, we also obtain a sub-linear regret bound for the E-linCBwK algorithm.

The computational analysis is conducted on a case study based in Portland Urban Metro Area. A major threat to the region is the expected Cascadia Subduction Zone earthquake with odds of about 44-52% for an earthquake greater than magnitude 8.3 in the next fifty years. The analysis shows the E-linCBwK performs much better than the two other simpler heuristics based on random choice and nearest available facility assignment, respectively. The importance of introducing non-skipping probabilities

($\lambda$) is highlighted by comparing E-linCBwK with its No Skip version. The results show that the terminal total deviation from fair allocation is higher in the No Skip version by an order of magnitude. Sensitivity analyses on the variation in post-disaster demand distributions, fairness update frequencies, and the number of opened DCs show that E-linCBwK consistently achieves terminal total deviation from the fair allocation which is less than 3.5% of the total number of packages distributed (which is 120,000) for all cases.

This research could be expanded in various research directions in the future. The use of stratified knapsack constraints is proposed in this research (in contrast to the globalized knapsacks in the extant literature) which to allow the utilization of resources until all the facilities run out of them. However, in the current study, the algorithm regret bounds are proven until the violation of the first knapsack constraint. Future methodological research can dedicate efforts to effectively incorporating stratified knapsacks in the multi-armed bandit literature. Additionally, we consider do not consider the effect of network recovery in the 72-hour period. For Portland, the estimated recovery for a backbone network is less than 72 hours, and therefore, some network recovery can be expected. As road infrastructure becomes available new avenues for effective resource distribution both drones and trucks should also be investigated.

# 5 Linear Contextual Blocking Bandits with Context-dependent Delays: An Application to Real-time Electric Truckload Pick-up and Delivery

Cite this chapter as:

Darshan R. Chauhan, Avinash Unnikrishnan, Stephen D. Boyles, (2022), "*Linear Contextual Blocking Bandits with Context-dependent Delays: An Application to Real-time Electric Truckload Pick-up and Delivery*", Working Paper.

Author contributions: The authors confirm contribution to the paper as follows: study conception and design: DRC, AU, SDB; mathematical formulation: DRC, AU; solution algorithms: DRC; data collection and coding: DRC; analysis and interpretation of results: DRC; draft manuscript preparation: DRC; finalizing manuscript and revisions: DRC, AU, SDB. All authors reviewed the results and approved the final version of the manuscript.

## 5.1 Introduction

Trucking is the most dominant mode of freight transportation in the United States: collecting \$791.7 billion in gross revenue (80.1% of the nation's freight bill), and accounting for 11.84 billion tonnes of freight (72.5% of total freight tonnage) [American Trucking Association (2019)]. However, this dominance is not reflected in its use of

data analytics to improve the efficiency and performance of its business operations [Wang et al. (2018)]. Improvements in information and communication technologies provide opportunities for using real-time information for improving the efficiency and optimizing the performance of freight systems. For maximizing profits, a carrier needs to increase revenues and cut down costs. Most revenue maximization studies focus on optimizing load pricing and procurement in freight marketplaces while the optimization of load allocation and routing forms a major portion of research focusing on cost minimization. In this study, we focus only on the load allocation and routing aspect for a carrier. For further details on load pricing and procurement in freight marketplaces, the reader is referred to Figliozzi et al. (2003), Figliozzi et al. (2005), Huang and Xu (2013), Kuyzu et al. (2015), and Wang et al. (2018).

Unsurprisingly, the transportation sector is a major contributor to the anthropogenic U.S. greenhouse gas (GHG) emissions accounting for 29% [US EPA (2021)]. Emissions from light-duty, medium-duty, and heavy-duty trucks comprise 82% of all transportation sector emissions. To aid the emissions cause, some researchers have focused on minimizing emissions from conventional freight vehicles while routing [Figliozzi (2010), Bektaş and Laporte (2011), Jabali et al. (2012)], whereas others have focused on cleaner alternatives [Pelletier et al. (2016), Pelletier et al. (2019)]. With lower GHG emissions and noise, electric freight vehicles (EFVs) provide a more sustainable alternative for freight delivery [Pelletier et al. (2019)]. Despite their environmental benefits, EFVs face significant issues in integration in the goods delivery

schemes due to high initial investment cost, limited range and payload, and long recharging times. This study proposes the usage of EFVs for freight logistics considering their range and recharging time limitations.

In this study, we focus on a dynamic electric truckload pickup and delivery problem (E-TPDP) wherein each request arrives in a stochastic manner and has to be transported directly from its pickup to the delivery location. The loads are transported using electric freight vehicles (EFVs), and, only one request can be served at a time by a vehicle typically due to vehicle capacity constraints. Previous efforts on the truckload pickup and delivery problem (TPDP) employing conventional vehicles have tacked the problem using rolling horizon framework [Yang et al. (1999),Yang et al. (2004)], and approximate dynamic programming [Godfrey and Powell (2002), Topaloglu and Powell (2006), Simao et al. (2009), Wang et al. (2018)]. For the case of conventional vehicle TPDP, the route of the truck for each request is always known: the starting point of the truck to the pick-up location to the delivery location. As the range of the conventional truck is large, fuel stations are widely available, and refueling operations are quick, the time spent rerouting and refueling is considered to be negligible. But, in the case of EFVs, due to limited range, long recharging times, and relatively scarce charging infrastructure, recharging decisions affect route construction. In this work, we propose a novel multi-armed bandit (MAB) methodology to tackle this problem. The MAB formulation can be applied to TPDP directly, and to E-TPDP with an additional preprocessing step for route construction.

The literature on ride-hailing systems is also close to the TPDP. The interested reader is referred to Agatz et al. (2012) and Mourad et al. (2019) for a review of various ride-hailing/ride-sharing services and solution methods available. However, there is limited literature considering fleet electrification (example, Chen et al. (2016); Iacobucci et al. (2019); Kang et al. (2017)). Three of the closest works to our case here are Shi et al. (2019), Al-Kanj et al. (2020), and Kullman et al. (2021). All three works consider the operation of an electric ride-hailing fleet under a centralized command and consider decisions relating to the assignment of requests to a vehicle, recharging, and repositioning. Al-Kanj et al. (2020) and Kullman et al. (2021) additionally consider the option of rejecting a request, unlike Shi et al. (2019). In this study, we consider the operation of the fleet under a centralized command and consider the assignment and recharging-related decisions. We consider that the requests received in the E-TPDP are a result of successful procurement in spot truckload markets, and as a result, requests cannot be considered for rejection, and delay-based costs are incurred. The commonly cited technical challenges associated with fleet electrification are long recharging times and limited charging infrastructure (Kullman et al., 2021). These problems are exacerbated when considering the larger size and special equipment needs of EFVs. Additionally, the larger service areas (and therefore also, longer trip distances) make it imperative to consider recharging decisions in route construction and assignment. Generally, most approaches in ride-hailing consider aggregation of requests in a time interval and then decisions are made for all these

requests together (example, Shi et al. (2019) consider 6-min intervals and Al-Kanj et al. (2020) consider 15-min intervals). In this study, like Kullman et al. (2021), we consider decision-making at a request level to avoid any unnecessary delays.

Basu et al. (2019) proposed a new multi-armed bandit setting for resource allocation wherein an arm observes a known "blocked" time when it would be unavailable for selection, every time it is selected. This problem is named the blocking bandits problem. This is relevant to transportation resource utilization-related problems wherein the resource is unavailable when it is being used. Example applications include bike-sharing, ride-hailing, and truckload operations. Recently, many extensions of blocking bandits setting are proposed considering combinatorial action play (Atsidakou et al., 2021; Simchi-Levi et al., 2021), adversarial rewards Bishop et al. (2020), decaying rewards Simchi-Levi et al. (2021), stochastic delays Atsidakou et al. (2021); Simchi-Levi et al. (2021), and matroid constraints (Papadigenopoulos and Caramanis, 2021). Of particular interest is the extension considered by Basu et al. (2021). They propose contextual blocking bandits, where the player observes a context before the arm selection stage. The context represents the information available apriori to decision-making and affects the reward obtained. The delays are not dependent on the context. Their proposed methodology is tractable only when the set of possible contexts is small enough, and therefore, becomes prohibitive when the context has a high dimension, or when it captures continuous characteristics like time (which require fine discretization for practical purposes). Here, we propose a linear contex-

tual blocking bandits (linCBB) problem, where we consider that both the reward and "blocked" times are linearly dependent on the context (the linear relationship is not known). The linearity assumption helps tackle high-dimensional contexts with continuous components. We apply this methodology for load allocation in E-TPDP. The linCBB methodology can potentially also be extended to incorporate congestion in the instant drone delivery logistics, and the operation of ride-hailing services and instant food delivery platforms.

The rest of the article is organized as follows: Section 5.2 describes the problem and provides an online formulation of the problem. The remainder of the section formulates a load allocation policy using the proposed linear contextual blocking bandits with context-dependent delays framework and describes the route construction preprocessing step required when using electric freight vehicles. Section 5.3 describes the setting for the computational analysis of the proposed approach. Finally, section 5.4 concludes the article and provides directions for future work.

## 5.2 Problem Description and Formulation

A truckload carrier company owns and operates a fleet of $K$ electric freight vehicles (EFVs) (represented by the set $\mathcal{K}$) in a region. The carrier receives requests for truckload moves (hereafter, jobs) in real-time within the prespecified region of operation, and wishes to maximize its cumulative rewards over a time horizon of $T$ requests (represented by the set $\mathcal{T} := \{1, 2, \ldots, T\}$). The EFVs can serve only one load at

a time. Due to the limited range of EFVs, we also consider recharging decisions for EFVs. This problem is named the Real-Time Electric Truckload Pick-up and Delivery Problem (E-TPDP). We assume that the requests are a result of successful bids by the carrier company in spot markets. Spot market contracts are short-term contracts for serving unfilled or urgent shipments and form a significant portion of all truckload contracts (Wang et al., 2018). Load procurement by the carrier in the spot markets is outside the scope of the current study. Additionally, we also assume that we know the locations of charging infrastructure in the region and that there will not be any wait time at the charging stations. Optimal location and incorporation of congestion at charging stations can be a part of future studies. It is also assumed that we can compute battery consumption along a route based on the job characteristics. We first discuss the load allocation strategy formulated using linear contextual blocking bandits. This can be used for both E-TPDP and TPDP. Then, we discuss the recharging decision and route construction strategy required specifically for E-TPDP.

### 5.2.1 Load allocation using linear contextual blocking bandits

A job $t \in \mathcal{T}$ comes associated with a pick-up location, a delivery location, and load weight. We abuse the notation a little bit here, and also refer to the time that the job arrives as $t$. At time $t$, we know the earliest time when and the location where an EFV $a \in \mathcal{K}$ would become available, along with its expected state of charge or battery level. For now, we assume that recharging decisions and potential routes for EFVs

are known. Therefore, we consider that the job $t \in \mathcal{T}$ comes associated with a context $\mathbf{x}_t(a)$ ($\mathbf{x}_t(a) \in \mathbb{R}^m$, $\|\mathbf{x}_t(a)\|_2 \leq 1$) for each EFV $a \in \mathcal{K}$. The context includes information relating to recharging decision, route characteristics, load characteristics, traffic conditions, expected battery consumption, etc. We define $X_t := \{\mathbf{x}_t(a) \; \forall \; a \in \mathcal{K}\}$. We assume that the context $X_t$ observed in round $t$ is sampled from an unknown distribution, independent from previous rounds. Observing this context $X_t$, the carrier instructs EFV $a_t$ to serve the job which results in a reward of $\mu_{a_t,t}(\mathbf{x}_t(a_t)) \in [0,1]$ and a blocked time of $d_{a_t,t}(\mathbf{x}_t(a_t)) \in \mathbb{N}$. The blocked time represents the time EFV would be unavailable due to fulfillment of job request. Note that we consider both the reward and the blocked time to be context-dependent. So, given a context $X$, the optimal policy $\pi^*(X)$ for the described contextual blocking bandit problem (abbreviated as CBB; first proposed by Basu et al. (2021)) is obtained by solving:

$$\begin{aligned} \max_{\pi} \quad & \sum_{a \in \mathcal{K}} \mu_{a,t}(\mathbf{x}(a)) \pi_a(\mathbf{x}(a)) \\ & d_{a,t}(\mathbf{x}(a)) \pi_a(\mathbf{x}(a)) \leq 1 \quad \forall \; a \in \mathcal{K} \\ & \sum_{a \in \mathcal{K}} \pi_a(\mathbf{x}(a)) \leq 1 \\ & \pi_a(\mathbf{x}(a)) \geq 0 \quad \forall \; a \in \mathcal{K} \end{aligned} \tag{5.2.1}$$

The goal of the problem is to find a policy $\pi(X)$ that maximizes rewards under context $X$. The blocking time equation ensures that the (fluidized) average playing rate of arm $a$ under context $X$ is less than $1/d_{a,t}(\mathbf{x}(a))$.

As the possible number of contexts can be very large, we operate under a linear realizability assumption for rewards and blocking times. There exist unknown weight vectors $\theta_\mu^*, \theta_d^* \in \mathbb{R}^m$ ($\|\theta_\mu^*\|_2 \leq 1$ and $\|\theta_d^*\|_2 \leq 1$) such that:

$$\begin{aligned} \mathbb{E}(\mu_{a,t}|\mathbf{x}_t(a)) &= \mathbf{x}_t(a)^T\theta_\mu^* \quad \forall\ a \in \mathcal{K}, t \in \mathcal{T} \\ \mathbb{E}(d_{a,t}|\mathbf{x}_t(a)) &= c\mathbf{x}_t(a)^T\theta_d^* \quad \forall\ a \in \mathcal{K}, t \in \mathcal{T} \end{aligned} \tag{5.2.2}$$

where, $c$ is the scaling factor for blocking times. Substituting the linear realizability assumptions (5.2.2) into the formulation (5.2.1) results in:

$$\begin{aligned} \max \quad & \sum_{a\in\mathcal{K}} \mathbf{x}_t(a)^T\theta_\mu \pi_a(\mathbf{x}(a)) \\ & c\mathbf{x}_t(a)^T\theta_d \pi_a(\mathbf{x}(a)) \leq 1 \quad \forall\ a \in \mathcal{K} \\ & \sum_{a\in\mathcal{K}} \pi_a(\mathbf{x}(a)) \leq 1 \\ & \pi_a(\mathbf{x}(a)) \geq 0 \quad \forall\ a \in \mathcal{K} \end{aligned} \tag{5.2.3}$$

We refer to the above problem (5.2.3) as linear contextual blocking bandit problem with context-dependent delays, abbreviated as linCBB. We now penalize the

blocking time constraint which results in:

$$\begin{aligned} \max \quad & \sum_{a \in \mathcal{K}} (\mathbf{x}_t(a)^T \theta_\mu - c\lambda_a \mathbf{x}_t(a)^T \theta_d) \pi_a(\mathbf{x}(a)) + \sum_{a \in \mathcal{K}} \lambda_a \\ & \sum_{a \in \mathcal{K}} \pi_a(\mathbf{x}(a)) \leq 1 \\ & \pi_a(\mathbf{x}(a)) \geq 0 \quad \forall \ a \in \mathcal{K} \end{aligned} \tag{5.2.4}$$

where $\lambda_a$ are the penalty parameters. The above problem (5.2.4) can be seen as a variant of the linear contextual bandit problem (abbreviated as linCB), first described by Chu et al. (2011). In addition to estimating the unknown weight vectors ($\theta$) and the arm-play policy ($\pi$), as required for linCB, our problem linCBB also needs to learn the optimal values of the penalty parameters ($\lambda$). A similar application of penalty parameters is also adopted by Agrawal and Devanur (2016) for solving the linear contextual bandit with knapsacks problem, where the knapsack constraints were penalized. An algorithm based on upper confidence bounds to solve linCBB for a planning horizon of $T$ requests is described in Algorithm 5.1. The penalty parameters are learned through the multiplicative weight update, a fast and efficient online learning algorithm proposed by Arora et al. (2012). Note that Algorithm 5.1 can be applied directly for the TPDP, and can be applied to E-TPDP once the recharging decisions are made and the potential route for each EFV is constructed.

### 5.2.2 Recharging decisions and route construction for E-TPDP

For the case of conventional vehicle TPDP, the route of the truck for each request is always known: the starting point of the truck to the pick-up location to the delivery location. As the range of the conventional truck is large, fuel stations are widely available, and refueling operations are quick, the time spent on refueling is considered to be negligible. But, in the case of EFVs, due to limited range, long recharging times, and relatively scarce charging infrastructure, recharging decisions affect route construction. Therefore, route construction considering recharging is essential for E-TPDP.

A job received at time $t$ comes associated with a pick-up location $l_1^t$, drop-off location $l_2^t$, and load weight $w^t$. At time $t$, we know that the EFV $a \in \mathcal{K}$ would be available earliest at time $\tau^{a,t} \geq t$ and location $l_0^{at}$ with state of charge/ battery level $q_0^{at}$ (kWh). Each EFV $a \in \mathcal{K}$ is assumed to have a maximum battery capacity of $Q$ (kWh).

**Algorithm 5.1** linCBB-UCB

**Input:** $\alpha_\mu, \alpha_d \in \mathbb{R}^+$, $K \in \mathbb{N}$ (set represented by $\mathcal{K}$), $m \in \mathbb{N}$, $c_T \in \mathbb{R}^+$ (time-based cost)

**Output:** $\{a_t, \mu_{a_t,t}, d_{a_t,t}\}\ \forall\ t = 1, 2, \ldots, T$

1: $A \leftarrow I_m$ {the $m \times m$ identity matrix}
2: $b_\mu, b_d \leftarrow \mathbf{0}_m$ {$m$-dimensional vector of zeros}
3: Initialize $\lambda_a = \frac{Kc_T}{(K+1)}\ \forall\ a \in \mathcal{K}$, and $w_{a,0} = 1\ \forall\ a \in \mathcal{K}\ t = 1, 2, \ldots, T$
4: Update the set of available arms $\mathcal{K}_t$ $\mathcal{K}_t$ is not empty
5: $\theta_\mu^t \leftarrow A^{-1}b_\mu$, $\theta_d^t \leftarrow A^{-1}b_d$
6: Observe context $X_t \in \mathbb{R}^{m \times K}$ $a \in \mathcal{K}_t$
7: Computes the upper confidence bounds $p_{t,a}$ as:

$$p_{t,a} \leftarrow \left(\mathbf{x}_t(a)^T\theta_\mu^t + \alpha_\mu\sqrt{\mathbf{x}_t(a)^T A^{-1}\mathbf{x}_t(a)}\right) - c\lambda_a\left(\mathbf{x}_t(a)^T\theta_d^t - \alpha_d\sqrt{\mathbf{x}_t(a)^T A^{-1}\mathbf{x}_t(a)}\right)$$

8: Choose action $a_t = \arg\max_a p_{t,a}$, ties broken arbitrarily
9: Observe reward $\mu_{a_t,t}$ and the corresponding blocking time $d_{a_t,t}$
10: Update the online learning weights $w_{a,t}$ as:

$$w_{a,t} \leftarrow \begin{cases} w_{a,t-1}(1+\varepsilon)^{\frac{d_a}{c} - \mathbf{x}_t(a)^{\mathrm{T}}\theta_d^t} & \text{if } \dfrac{d_a}{c} - \mathbf{x}_t(a)^{\mathrm{T}}\theta_d^t \geq 0 \\ w_{a,t-1}(1-\varepsilon)^{\frac{d_a}{c} - \mathbf{x}_t(a)^{\mathrm{T}}\theta_d^t} & \text{if } \dfrac{d_a}{c} - \mathbf{x}_t(a)^{\mathrm{T}}\theta_d^t < 0 \end{cases} \quad \forall\ a \in \mathcal{K}$$

11: $\lambda_a \leftarrow \frac{Kc_T w_{a,t}}{1+\sum_{a\in\mathcal{K}} w_{a,t}}\quad \forall\ a \in \mathcal{K}$
12: $A \leftarrow A + \mathbf{x}_t(a_t)\mathbf{x}_t(a_t)^T$
13: $b_\mu \leftarrow b_\mu + \mathbf{x}_t(a_t)\mu_{a_t,t}$, $b_d \leftarrow b_d + \mathbf{x}_t(a_t)(d_{a_t,t}/c)$

There are $N$ charging stations (represented by set $[N]$) available to the carrier in its region of operation. We assume that each EFV has two charging opportunities while potentially serving the job request: at most once between the EFV's earliest available location $l_0^{at}$ and pick-up location $l_1^t$, and at most once between pick-up location $l_1^t$ and delivery location $l_2^t$. When an EFV decides to visit a charging station, we assume that it departs only after being fully charged at a constant rate of $\eta_Q$

(hr/kWh). We assume that the total recharging cost consists of a base fee of $c_{BQ}$ ($) and a charging fee of $c_{CQ}$ ($/kWh). Therefore, if an EFV arrives at a charging station with a battery level of $q$ (kWh) at time $\tau$, we assume that it departs at a battery level of $Q$ at time $\tau + (Q - q)\eta_Q$ and pays $[c_{BQ} + c_{CQ}(Q - q)]$ as the total recharging cost.

The recharging decision and route construction problem for an EFV $a \in \mathcal{K}$ is assumed to occur on a directed sub-graph $G_{a,t}$ of the operational region which consists of the following nodes: $l_0^{at}$, $l_1^t$, $l_2^t$, a layer of $N$ charging station nodes between $l_0^{at}$ and $l_1^t$, and a layer of $N$ charging station nodes between $l_1^t$ and $l_2^t$. Directed paths exist between $l_0^{at}$ and $l_1^t$, one directly (representing no recharging) and other $N$ through the charging stations. Similarly, directed paths exist between $l_1^t$ and $l_2^t$, one directly (representing no recharging) and other $N$ through the charging stations. Additionally, recharging operations are charging stations nodes can be representing by splitting the nodes and representing recharging operations over an arc. Note that an EFV travels with a zero payload between nodes $l_0^{at}$ and $l_1^t$, and with a payload $w^t$ between nodes $l_1^t$ and $l_2^t$.

Let, $c_D$ ($/mile) and $c_T$ ($/hr) represent the unit distance- and time-based cost, respective. We assume that the directed arcs in the graph $G_{a,t}$ are pre-computed shortest paths considering both distance- and time-based costs. It is assumed the we can accurately calculate the payload-dependent battery consumption on an arc of graph $G_{a,t}$. As distance and travel time are known for all arcs of $G_{a,t}$, we know

that the cost incurred on non-recharging arcs is the sum of distance- and time-based costs and the cost incurred on recharging arcs is the sum of total recharging cost and the time-based cost. It is an operational requirement that an EFV arrives at the delivery location with a minimum battery level of $Q_{thr}$. The graph $G_{a,t}$ is pictorially represented in Figure 5.1.

Finally, the recharging decision and route construction problem for an EFV can be modeled as a shortest path problem with battery threshold constraint. The nomenclature is summarized below:

*Sets and Indices*

| | |
|---|---|
| $V$ | Set of all nodes $(u, v \in V)$ |
| $V_Q$ | Set of nodes representing end of recharging operations $(V_Q \subset V)$ |
| $A'$ | Set of all arcs $((u, v) \in A')$ |
| $A_Q$ | Set of arcs representing recharging operations $(A_Q \subset A')$ |
| $FS_u$ | Set of nodes in the set $V$ which are in the forward star of node $u \in V$, i.e., set of all nodes such that there exists an arc $(u, v) \in A'$ |
| $RS_u$ | Set of nodes in the set $V$ which are in the reverse star of node $u \in V$, i.e., set of all nodes such that there exists an arc $(v, u) \in A'$ |

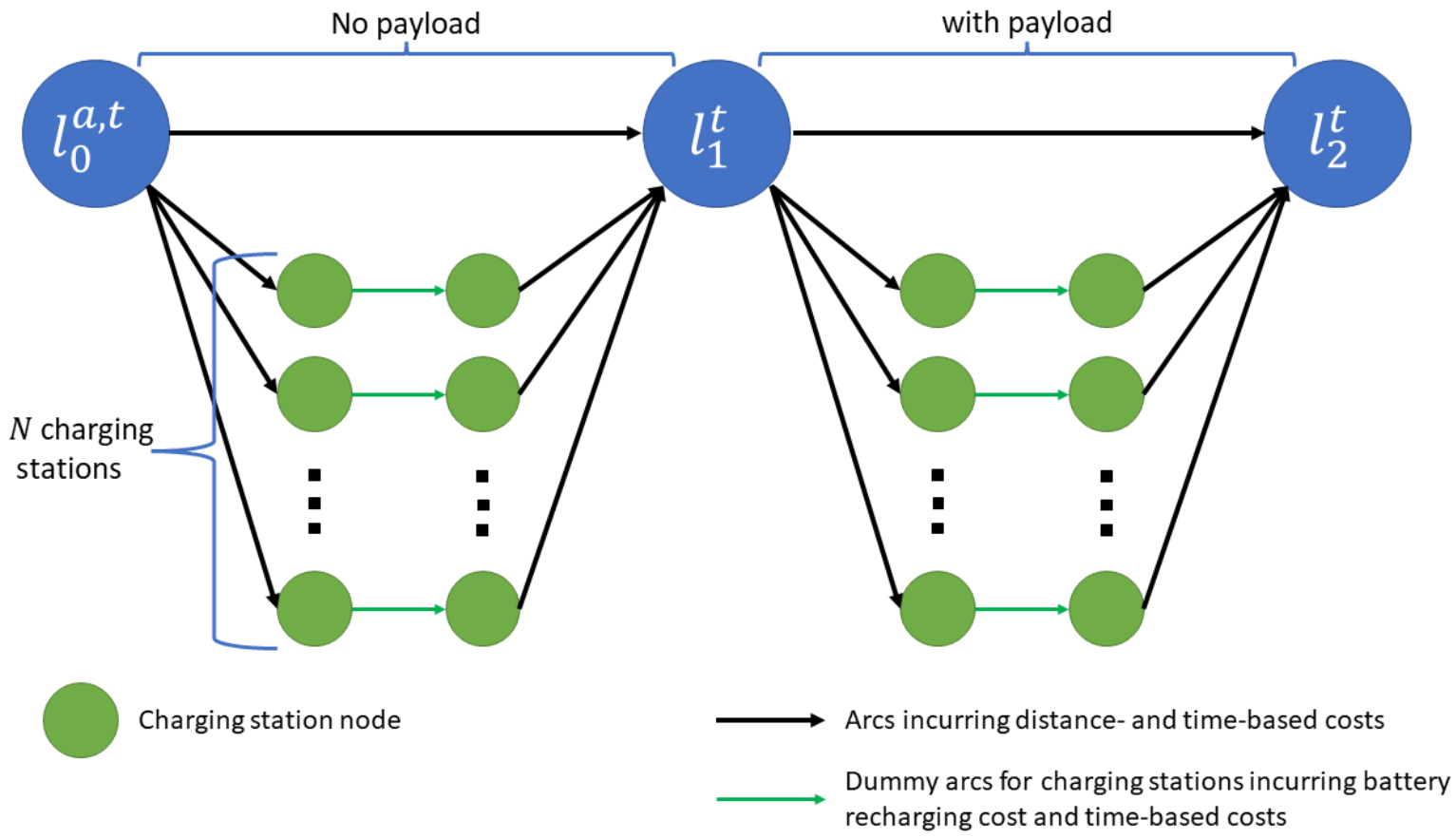



Figure 5.1: Graph $G_{a,t}$

*Parameters*

| | |
|---|---|
| $l_0^{at}$ | Node representing starting point of EFV |
| $l_1^t$ | Node representing pick-up location of the job request |
| $l_2^t$ | Node representing delivery location of the job request |
| $c_{uv}^{at,F}$ | Sum of distance- and time-based costs of traversing on arc $(u,v) \in A' \backslash A_Q$ ($) |
| $q_{uv}^{at,F}$ | Battery consumption on non-recharging arc $(u,v) \in A' \backslash A_Q$ (kWh) |
| $c_T$ | Unit time-based cost ($/hour) |
| $c_{BQ}$ | Base fee for recharging ($) |
| $c_{CQ}$ | Unit recharging cost at charging station ($/kWh) |
| $\eta_Q$ | Constant recharging rate at charging station (hour/kWh) |
| $q_0^{at}$ | Battery level of EFV at node $l_0^{at}$ (kWh) |
| $Q$ | Maximum battery capacity of the EFV (kWh) |
| $Q_{thr}$ | Minimum battery level required when reaching the delivery location (node $l_2^t$) (kWh) |

*Decision Variables*

| | |
|---|---|
| $y_{uv}^{at}$ | 1, if the arc $(u,v) \in A'$ is chosen to be traversed, and 0, otherwise |
| $z_u^{at}$ | Battery level of the EFV at node $u \in V$ (kWh) |

The formulation for the shortest path problem with battery threshold constraint is now presented.

$$CSP^{a,t}: \min_{y,z} \sum_{(u,v)\in A'\setminus A_Q} c^F_{uv} y^{at}_{uv} + \sum_{(u,v)\in A_Q} \left[c_T \eta_Q (z^{at}_v - z^{at}_u) + c_{BQ} + c_{CQ}(z^{at}_v - z^{at}_u)\right] y^{at}_{uv} \tag{5.2.5}$$

$$\text{s.to.} \quad \sum_{v\in FS_u} y^{at}_{uv} - \sum_{v\in RS_u} y^{at}_{vu} = \begin{cases} 1 & \text{if } u = l^{at}_0 \\ -1 & \text{if } u = l^t_2 \\ 0 & \text{if } u \in V\setminus\{l^{a,t}_0, l^t_2\} \end{cases} \tag{5.2.6}$$

$$z^{at}_{l^{at}_0} = q^{at}_0 \tag{5.2.7}$$

$$z^{at}_v = Q \quad \forall\ v \in V_Q \tag{5.2.8}$$

$$z_{l^t_2} \geq Q_{thr} \tag{5.2.9}$$

$$\sum_{u\in RS_v} (z_u - q^F_{uv}) \cdot y^{at}_{uv} = z^{at}_v \quad \forall\ v \in V\setminus\left\{V_Q \cup \{l^{a,t}_0\}\right\} \tag{5.2.10}$$

$$y^{at}_{uv} \in \{0,1\} \quad \forall\ (u,v) \in A' \tag{5.2.11}$$

$$z^{at}_u \geq 0 \quad \forall\ u \in V \tag{5.2.12}$$

Equation 5.2.5 represents the objective of minimizing cost. The first term consists of costs on non-recharging arcs, and the second term represents the recharging-related costs. Equation 5.2.6 represent the shortest path selection constraints. Equation 5.2.7 ensures that battery level at the starting node is $q^{at}_0$, and equation 5.2.8 ensures that the EFV departs with a fully charged battery after recharging. Equation 5.2.9 en-

sures that the battery level at the delivery location is greater than the minimum required battery level $Q_{thr}$. Equation 5.2.10 represents the battery level conservation constraints. It ensures that the EFV's battery levels are correctly computed at each node over the selected route. Equations 5.2.11 and 5.2.12 are variable definitions. Equation 5.2.12 ensures that an EFV would never run out of battery on the selected route. The non-linearity in the objective function and Equation 5.2.10 can be linearized by introducing new variables to yield a mixed-integer linear program (MILP), and then, can be solved using a MILP solver. When the problem is solved, we know the potential route for an EFV for the demand request. This information is utilized for generating context for the linCBB framework proposed in the previous section.

### 5.2.3 A note on battery consumption and recharging in EFVs

A plethora of research on battery consumption and recharging in electric vehicles is available. The energy consumption in EFVs is dependent on the power required to accelerate and overcome rolling resistance, aerodynamic drag, and grade resistance (Davis and Figliozzi, 2013). This power is dependent on the curb weight and the payload of the EFV. Concerning the payload, Pelletier et al. (2019) shows that the battery consumption in EFV is a linear function of the payload. While Al-Kanj et al. (2020) assumes that the battery consumption can be computed based on trip characteristics, and assumes no explicit relationship with payload. In this study, we follow the battery consumption relationship proposed by Pelletier et al. (2019),

however, any function for calculation of battery consumption can be adopted for modeling.

In the context of electric vehicle routing problem (E-VRP), Pelletier et al. (2019) assume that charging is not allowed on the vehicle tour, and that the vehicles are charged once and fully before they leave the depot. E-VRPs considering recharging can be divided on basis two recharging policies: full-recharging and partial recharging. The full-recharge policy literature consists of research that assumes constant recharging time [Conrad and Figliozzi (2011), Adler and Mirchandani (2014), Hof et al. (2017)] which may indicate battery swapping operations, and recharging at constant rate [Schneider et al. (2014), Goeke and Schneider (2015), Desaulniers et al. (2016), Hiermann et al. (2016)] where the charging time is dependent on initial battery level and the recharging rate. The partial recharging policy literature considers capability for partial recharging of electric vehicles and have adopted linear charging function [Felipe et al. (2014), Keskin and Çatay (2016), Desaulniers et al. (2016), Schiffer and Walther (2017), Al-Kanj et al. (2020)] and non-linear charging functions with partial charging policies [Montoya et al. (2017), Froger et al. (2019)]. This study assumes a full-recharging policy with linear recharging function. Incorporating partial recharging policy can be a part of future research.

In addition, we assume that the freight carrier has access to charging infrastructure in the region and that the locations of charging stations are known. The optimal location of charging stations in a region is out of the scope of the current study. We

also assume that an unlimited number of EFVs can recharge at a charging station. Similar assumptions are made in Montoya et al. (2017) and Al-Kanj et al. (2020).

## 5.3 Computational Experiments

For the computational experiments, we assume that a trucking carrier operates a fleet of $K$ electric freight vehicles (EFV) (represented by set $\mathcal{K}$) in a service area. The trucking company caters to requests for truckload pick-up and delivery, as procured from spot markets, which must be instantaneously serviced. The planning horizon of the problem is $T$ requests, and we assume that a request/job arrives every hour (i.e., constant inter-arrival time). We abuse the notation a little bit and denote both the request number and the time it arrives by $t \in \{1, 2, \ldots, T\}$.

The service area is assumed to be flat and rectangular with dimensions of 100 miles $\times$ 100 miles. The EFVs travel over Manhattan distances over the grid at a constant speed of 50 miles per hour. The origin ($l_1^t$) and destination ($l_2^t$) points for job $t$ are uniformly generated integer values over the grid, such that the Manhattan distance between them is at least $req\_dist_{min}$. We assume that $req\_dist_{min}$ is 50 miles. Initially, all EFVs are located at a central depot located at [50,50] and are fully charged. We consider that charging infrastructure is available at the central depot as well as at four additional locations: [20,20], [20,80], [80,20], and [80,80].

The fleet of EFVs is considered to be Freightliner eM2 electric box trucks [Freightliner (2021)] with a GVWR of up to 33,000 lb. As data about curb weight of eM2

truck is unavailable, we assume its value to be 16,500 lbs (about 3000 lbs higher than similar non-electric Freightliner truck). For any request, the payload weight ($w^t$) is assumed to be a uniformly distributed integer between 2000 lb and a maximum payload ($w^{max}$) of 16,500 lb.

The Freightliner eM2 trucks have a usable battery capacity of 315 kWh with a fully-loaded range of 230 miles. Additionally, with Freightliner's proprietary charging infrastructure, 80% of battery capacity can be charged in 60 minutes. Based on the data, for our experiments, we assume that the battery capacity of each truck ($Q$) is 250 kWh (≈80% of 315 kWh), and assume that it can be fully charged in 1 hour (i.e. charging rate $\eta_Q = 1/250$ hr/kWh). Based on the assumed value of $Q$, we get a fully-loaded range ($range_{full}$) of 180 miles (≈80% of 230 miles), and assume that the range at zero-payload ($range_{zero}$) is 250 miles. As in Pelletier et al. (2019), we assume that the battery consumption is linearly dependent on the payload. Specifically, we utilize the following relationship for calculating the payload-dependent battery consumption rate ($\phi(w^t)$):

$$\phi(w^t) = \frac{Q}{range_{zero}} + \frac{w^t}{w^{max}} \cdot \left( \frac{Q}{range_{full}} - \frac{Q}{range_{zero}} \right) \tag{5.3.1}$$

The payload dependent battery consumption ($q(w^t)$) over a travel distance of $dist$ (miles) can then be calculated as $\phi(w^t) \cdot dist$ (kWh).

The distance-based unit cost ($c_D$) and time-based unit cost ($c_T$) are derived from

the ATRI 2020 marginal cost estimates [Leslie and Murray (2021)]. The parameter $c_D$ is assumed to be \$0.60/mile and is obtained by discounting driver-based costs (\$0.74/mile) and fuel costs (\$0.31/mile) from the 2020 marginal cost estimate of \$1.65/mile. Similarly, the parameter $c_T$ is assumed to be \$30/hr and only considers the driver-related costs from the ATRI 2020 marginal cost estimate. The unit cost for the delay in arrival at the request origin location ($l_1^t$) from the moment the request is received is also assumed to be $c_T$. The base fee for recharging ($c_{BQ}$) at the charging station is considered to be \$2.50, and the charging fee ($c_{CQ}$) is considered to be \$0.24/kWh (twice the on-peak rate of \$0.12/kWh in Portland, OR, to consider additional charging infrastructure-related costs).

For catering a request/job $t$, there are phases: the request arrival phase, the allocation decision phase, and the post-decision phase.

A job received at time $t$ comes associated with a pick-up location $l_1^t$, drop-off location $l_2^t$, and load weight $w^t$. During the request arrival phase, we consider each truck $a \in \mathcal{K}$ for servicing the job $t$ from its earliest available location $l_0^{at}$ having remaining battery capacity of $q_0^{at}$. Battery-constrained shortest paths are constructed for each truck $a \in \mathcal{K}$ by solving the $CSP^{a,t}$ formulation (equations 5.2.5-5.2.12). Based on the solution, we extract the following information for each truck $a \in \mathcal{K}$ at time $t$: the total travel distance, denoted by $EFV\_dist^{at}$, and the time to reach the origin point of the request ($l_1^t$), denoted by $EFV\_time\_req\_origin^{at}$. We define the

context $\mathbf{x}_t(a)$ with $m = 2$ for arm $a \in \mathcal{K}$ as:

$$\mathbf{x}_t(a) := \frac{1}{\sqrt{2}} \begin{bmatrix} \dfrac{req\_dist_{min}}{EFV\_dist^{at}} \\ \\ \dfrac{time\_req\_origin_{max} - EFV\_time\_req\_origin^{at}}{time\_req\_origin_{max}} \end{bmatrix}, \quad \forall\ a \in K \quad (5.3.2)$$

where $time\_req\_origin_{max}$ represents the theoretical maximum time for an EFV to reach the origin of the request. For our case, that would correspond to the time required to travel the maximum possible distance in the service area along with the time required to fully charge the EFV once (as we assume that an EFV can only charge once before reaching the request origin node). The maximum shortest path distance considering a detour to a recharging station to reach the origin node from any location in the service area is 200 miles (EFV location and request origin location are on diagonally opposite ends, and as we consider travel over Manhattan distances, additional detour distance is zero). As we assume a constant travel speed of 50 miles per hour, the time spent on traveling is 4 hours. The maximum possible time for recharging is $eta_Q \cdot Q$, which for our case is 1 hour. Therefore, $time\_req\_origin_{max}$ is 5 hours.

The next is the allocation decision phase. The solution algorithms observe the context before making a decision. The linCBB-UCB algorithm chooses to play an available (i.e., unblocked arm) with the highest value of UCB index, as shown in Algorithm 5.1. The linCBB-UCB algorithm is compared against a simple greedy

rule that selects an available arm with the least value of arrival delay. This rule is named "greedy-AD". Alternatively put, greedy-AD chooses an available arm with the highest value of the second context argument.

Upon making the allocation decision, we arrive at the post-decision phase. During the post-decision phase, the reward and the delay for the chosen arm are realized. Unknown to the algorithms, the total costs for each arm $a \in \mathcal{K}$ for job $t$, denoted as $total\_cost^{at}$, are calculated from their shortest path solutions as:

$$total\_cost^{at} = \left( c^D + c^{charge} + \frac{c^T}{speed} \right) \cdot EFV\_dist^{at} + \rho \cdot EFV\_time\_req\_origin^{at} \tag{5.3.3}$$

where, *speed* is the constant traveling speed, here 50 miles per hour, $c^{charge}$ is the average cost of recharging the battery per unit distance, and $\rho$ denotes the penalty for delay in arrival at the pickup location of the request. For the nominal scenario, we consider $\rho = c_T$. The range of the EFV at average payload is 206 miles, and the cost of a full recharge is \$62.5 ($c^{BQ} + c^{CQ}Q = \$62.5$). For our case, the value of $c^{charge}$ comes out to be \$0.30/mile. Let, $total\_cost^t_{min}$ be the minimum total cost for serving job $t$, and $a_t$ be the arm chosen by an algorithm. Then, the realized reward $r_t$ and delay $d_t$ for that algorithm are:

$$r_t := \begin{cases} 1 & \text{if } \ total\_cost^{a_t t} = total\_cost^t_{min} \\ 0 & \text{otherwise} \end{cases} \tag{5.3.4}$$

$$d_t := \left\lceil \frac{EFV\_dist^{a_t t}}{speed} \right\rceil + random\{1,2\} \tag{5.3.5}$$

where, $random\{1,2\}$ takes the value 1 or 2 with equal probability. Additionally, we update the earliest available time ($EAT^{a,t}$), earliest available location ($l_0^{a,t}$), and remaining battery capacity ($q_0^{a,t}$) for each EFV $a \in \mathcal{K}$ before going forward to the next request $t+1$ as:

$$EAT^{a,t+1} = \begin{cases} EAT^{a,t} + \min\{1, d_t\} & \text{if } a = a_t \\ EAT^{a,t} + 1 & \text{otherwise} \end{cases}, \quad \forall\ a \in \mathcal{K} \tag{5.3.6}$$

$$l_0^{a,t+1} = \begin{cases} l_2^t & \text{if } a = a_t \\ l_0^{a,t} & \text{otherwise} \end{cases}, \quad \forall\ a \in \mathcal{K} \tag{5.3.7}$$

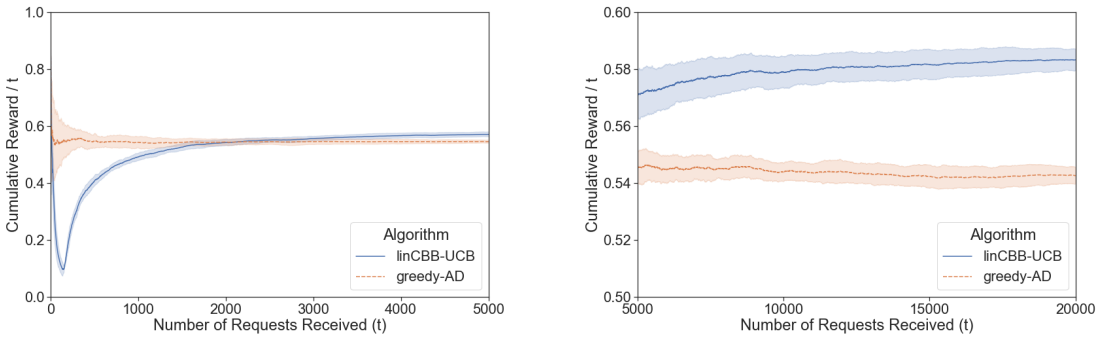
$$q_0^{a,t+1} = \begin{cases} z_{l_2^t}^{at*} & \text{if } a = a_t \\ q_0^{a,t} & \text{otherwise} \end{cases}, \quad \forall\ a \in \mathcal{K} \tag{5.3.8}$$

where, $z_{l_2^t}^{at*}$ is the battery level at the request destination location $l_2^t$ as obtained from the optimal solution of $CSP^{a,t}$ (equations 5.2.5-5.2.12).

It is to be noted that the request characteristics (pick-up location, delivery location, and payload weight), reward function, and delay function are considered to be the same for both the solution algorithms (linCBB-UCB and greedy-AD). However, due to different allocation decisions, both algorithms have different states, and therefore, the context, as well as the total costs for both the algorithms, would be

different.

The computational analysis is conducted considering a planning horizon of $T = 20,000$, and a fleet of $K = 10$ EFVs. Results are reported for ten instance runs. We assume the value of the online learning parameter $\varepsilon$ (for linCBB-UCB, see Algorithm 5.1) as 0.1. The growth of cumulative reward normalized by the number of requests received is shown in Figure 5.2. It can be noted that the linCBB-UCB reward growth as a function of time is better than linear growth, suggesting a slightly sub-linear regret of the linCBB-UCB algorithm. The greedy-AD reward growth on the other end is slightly sub-linear. This leads to a stronger performance of linCBB-UCB over greedy-AD with increasing time.

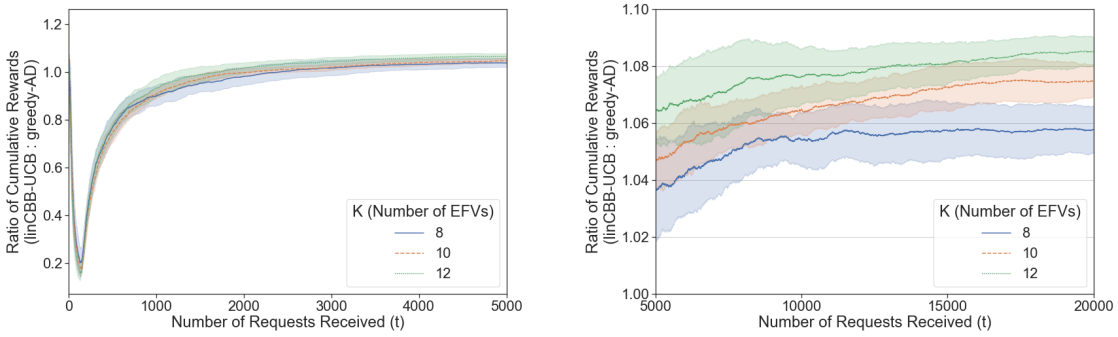



(a) $t \leq 5000$ (b) $t \geq 5000$

Figure 5.2: Ratio of cumulative reward normalized by the number of requests received ($t$) (average line with standard deviation band)

Next, a sensitivity analysis of the size of the EFV fleet is conducted. The results are summarized in Table 5.1 at various time horizons and fleet sizes. To reiterate, the reward is set to 1 if the vehicle that would observe the least cost is selected

(considering that all the past requests were served), and 0, otherwise. Therefore, there are two opposing factors to reward accrual. As the fleet size increases, trivially, the identification of the best possible EFV arm becomes difficult. On the other hand, the fraction of time that the best possible EFV is busy reduces thereby aiding the growth of reward. The results show that the latter is a more prominent contributor to reward growth. Figure 5.3 shows the growth of the ratio of cumulative rewards achieved by both algorithms for varying fleet sizes. It can be noted that for the varying fleet sizes of 8, 10, and 12, the linCBB-UCB algorithm gathers 5.7%, 7.5%, and 8.5% more rewards, respectively, compared to the greedy-AD algorithm.




(a) $t \leq 5000$ (b) $t \geq 5000$

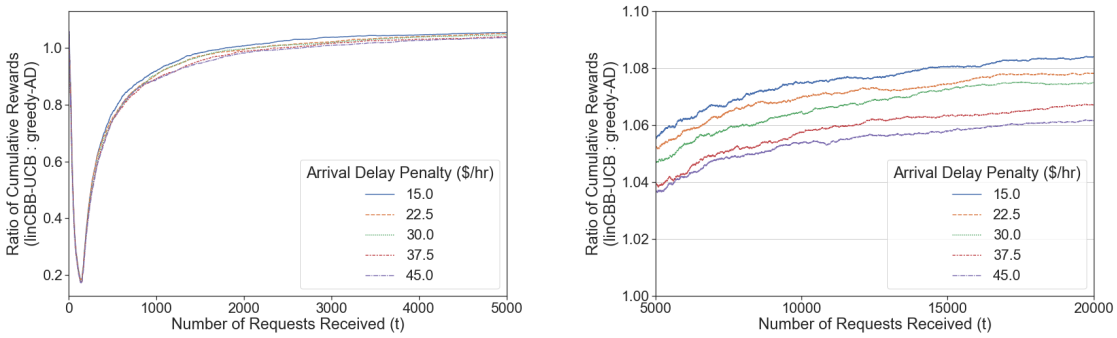
Figure 5.3: Growth of ratio of cumulative reward obtained by linCBB-UCB to greedy-AD with increasing number of requests (average line with standard deviation band)

Table 5.1: Sensitivity of Cumulated Rewards to EFV Fleet Size

| $K$ | | linCBB-UCB | | | | greedy-AD | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $T$ =5000 | $T$ =10000 | $T$ =15000 | $T$ =20000 | $T$ =5000 | $T$ =10000 | $T$ =15000 | $T$ =20000 |
| 8 | Min | 2419 | 4947 | 7384 | 9901 | 2309 | 4704 | 7043 | 9361 |
| | Ave | 2465 | 5001 | 7506 | 10030 | 2377 | 4744 | 7098 | 9483 |
| | Max | 2514 | 5076 | 7612 | 10164 | 2413 | 4800 | 7149 | 9575 |
| 10 | Min | 2781 | 5697 | 8615 | 11557 | 2672 | 5373 | 8027 | 10734 |
| | Ave | 2855 | 5790 | 8725 | 11666 | 2728 | 5437 | 8135 | 10852 |
| | Max | 2953 | 5855 | 8870 | 11831 | 2773 | 5478 | 8207 | 10927 |
| 12 | Min | 3074 | 6235 | 9410 | 12631 | 2844 | 5760 | 8708 | 11653 |
| | Ave | 3117 | 6303 | 9503 | 12724 | 2929 | 5857 | 8793 | 11727 |
| | Max | 3174 | 6379 | 9571 | 12819 | 2988 | 5936 | 8884 | 11806 |

Note: $K$ denotes the EFV fleet size

A sensitivity analysis on the penalty cost for the delay in pickup ($\rho$) for a request is presented in Table 5.2 for various time horizons and penalty values. The analysis considers variation in the range of 50%-150% of the nominal value (the nominal value of $\rho$ is the time-based cost $c_T$ = \$30/hr) and considers the fleet size ($K$) of 10. It is expected that an increase in the penalty would reduce the gap between the two algorithms, as greedy-AD explicitly prioritizes myopic arrival delay. Figure 5.4 shows the growth of the ratio of cumulative rewards achieved by both algorithms for varying arrival delay penalty values. The increase in the final reward obtained by the linCBB-UCB algorithm in comparison to the greedy-AD algorithm reduces from 8.4% to 6.2% as the penalty values increased from \$15/hr (50% of nominal value) to \$45/hr (150% of nominal value). However, linCBB-UCB still improves the reward collection ratio with increasing numbers of requests received.




(a) $t \leq 5000$ (b) $t \geq 5000$

Figure 5.4: Growth of ratio of cumulative reward obtained by linCBB-UCB to greedy-AD with increasing number of requests (line represents average of ten instance runs)

Table 5.2: Sensitivity of Cumulated Rewards to EFV Arrival Delay Penalty

| $\rho$ | | linCBB-UCB | | | | greedy-AD | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $T$ =5000 | $T$ =10000 | $T$ =15000 | $T$ =20000 | $T$ =5000 | $T$ =10000 | $T$ =15000 | $T$ =20000 |
| 15 | Min | 2779 | 5617 | 8541 | 11489 | 2620 | 5253 | 7854 | 10541 |
| | Ave | 2810 | 5700 | 8606 | 11538 | 2664 | 5303 | 7964 | 10645 |
| | Max | 2860 | 5780 | 8678 | 11625 | 2704 | 5364 | 8128 | 10811 |
| 22.5 | Min | 2807 | 5696 | 8559 | 11534 | 2651 | 5307 | 7979 | 10670 |
| | Ave | 2843 | 5759 | 8672 | 11602 | 2703 | 5384 | 8071 | 10762 |
| | Max | 2876 | 5837 | 8743 | 11660 | 2754 | 5512 | 8199 | 10905 |
| 30 | Min | 2781 | 5697 | 8615 | 11557 | 2672 | 5373 | 8027 | 10734 |
| | Ave | 2855 | 5790 | 8725 | 11666 | 2728 | 5437 | 8135 | 10852 |
| | Max | 2953 | 5855 | 8870 | 11831 | 2773 | 5478 | 8207 | 10927 |
| 37.5 | Min | 2777 | 5703 | 8612 | 11562 | 2711 | 5411 | 8064 | 10783 |
| | Ave | 2850 | 5773 | 8726 | 11678 | 2742 | 5459 | 8206 | 10944 |
| | Max | 2920 | 5842 | 8840 | 11824 | 2775 | 5542 | 8292 | 11047 |
| 45 | Min | 2838 | 5734 | 8676 | 11620 | 2707 | 5420 | 8133 | 10867 |
| | Ave | 2855 | 5791 | 8739 | 11705 | 2754 | 5495 | 8262 | 11026 |
| | Max | 2872 | 5841 | 8792 | 11789 | 2788 | 5564 | 8346 | 11098 |

Note: $\rho$ denotes the penalty for delay in arrival at the request pickup location ($/hr)

## 5.4 Conclusions

Recently, Basu et al. (2019) proposed a blocking bandits problem wherein the arm becomes "blocked" upon play and remains unavailable for the subsequent rounds defined by the delay. As one of the extensions of the framework, Basu et al. (2021) proposed contextual blocking bandits problem wherein the reward obtained is context-dependent and the delays are context-independent. In this study, we extend these frameworks by considering that both rewards and delays are linearly dependent on the context. The proposed extension allows the consideration of a very large number of contexts in a scalable manner. We also propose an upper confidence bound policy (called linCBB-UCB) to solve the linear contextual blocking bandits with context-dependent delays problem.

We apply this framework to the dynamic truckload pickup and delivery problem while considering fleet electrification. We consider that the requests are a result of successful procurement in spot truckload markets, and only request allocations need to be made via a central command. Due to the larger size of electric freight vehicles and longer trip distances, the fleet electrification challenges of long recharging times and limited compatible charging infrastructure are more profound in our case. As a result, the recharging decisions must be considered in route construction as well as request allocation.

The computational experiments are considered on a synthetic service area of

200×200 miles which contains five uncapacitated charging stations. In the nominal scenario, we consider that the fleet consists of 10 electric freight vehicles (EFVs), and the request allocation decisions are made by a central command while incurring distance-based, time-based, recharging, and delay costs. The cumulative reward represents the number of requests for which the least cost EFV is chosen (after the costs are realized), irrespective of their availability status. The performance of the linCBB-UCB algorithm is compared against a greedy allocation policy (called greedy-AD) that myopically selects the best available EFV which minimizes arrival delay at the pickup location. The computational experiments elicit that cumulative rewards increase faster than linearly suggesting a sub-linear regret. Whereas for greedy-AD policy, cumulative rewards in a sub-linear manner, indicating poorer performance. For $T = 20,000$ total requests and a fleet size of 10, linCBB-UCB obtains 7.5% more rewards than greedy-AD policy, on average. Sensitivity analyses on fleet size and arrival delay penalty are also conducted.

There are several ways in which the current research can be expanded. The current study does not incorporate the relocation feature when an EFV is not serving a request. The impact of the feature could be significant especially as more service request history is accumulated. A minor extension could be incorporating multiple charges between two key locations which become essential for larger service areas. An extension to a combinatorial bandit setting could allow the selection of multiple arms at a single instance which could result in improved fleet management.

# 6 Summary and Directions for Future Research

The transportation industry is on the cusp of a revolution. Within the past decade, the industry saw the rise of electric vehicles, connected vehicles and infrastructure, bike share systems and shared micro-mobility, ride-hailing/ride-sharing platforms, and crowdsourced delivery. And now, we head to a future where automated vehicles, autonomous delivery robots, and drone-based deliveries will be a reality. In addition to these technologies being potentially transformative for human life, they also hold an opportunity to aid efforts in realizing national priorities like reducing greenhouse gas emissions and congestion. The use of transportation network modeling and decision analytics can help us realize the most benefit from the interactions of these technological innovations with the transportation networks. Uncertainty is intrinsic to such systems due to natural stochasticity or insufficient information. Incorporating and tackling uncertainty in network modeling formed the core of this dissertation. More narrowly, this dissertation discussed decision-making in freight logistics systems utilizing emerging transportation technologies like electric-unmanned aerial vehicles and electric trucks. More specifically, through four application scenarios, the planning and real-time resource allocation in such systems are studied while considering different sources of uncertainty.

Chapter 2 discussed a network planning problem utilizing unmanned aerial ve-

hicles (UAVs) (or drones) equipped with automatic external defibrillators (AEDs) to combat out-of-hospital cardiac arrests. The chapter presented a robust multi-period facility location model for locating drones to maximize service coverage. The coverage of a demand location was defined using a coverage reliability constraint modeled using a chance constraint. The primary source of coverage uncertainty for drones was due to environmental factors, namely wind speed and direction. The uncertainty was captured using two methodological frameworks. Primarily, the uncertainty was captured using the robust optimization (RO) framework utilizing polyhedral uncertainty sets. In addition to their superior computational tractability, polyhedral uncertainty sets were chosen also because controlling the conservatism (using a budget of uncertainty) in decision-making is very intuitive. Secondly, the planning period is split into multiple smaller periods to disaggregate some uncertainty that arises from seasonality to aid RO in tackling them.

The computational analysis was conducted for the Portland Metro area. The results showed that extending to a multi-period formulation, rather than using average information in a single period, was particularly beneficial for shorter response times or when uncertainty in coverage probabilities was not considered. Accounting for uncertainty in decision-making improved coverage significantly while also reducing variability in simulated coverage, especially when response times are longer. Going from a single-period deterministic formulation to a multi-period robust formulation boosts the simulated coverage values by 57%, on average.

The study made several assumptions, like independence in coverage probabilities and that all facilities in the designated radius respond to an event. For future research, correlation in coverage probabilities should be considered to improve model realism and performance. Additionally, queue-based models could be explored for considering the dynamic nature of cardiac events. From a computational standpoint, new algorithms should be explored to improve tractability especially when the number of periods is larger.

Chapters 3 and 4 discussed network planning problems along with their real-time operational component. Both chapters utilized drones as a means to deliver package items to the customers in two different contexts.

Chapter 3 presented an application consisting of a logistics company expanding its delivery options to include instant delivery (for example, within 30 minutes). It was considered that requests for instant delivery could only be satisfied using drones, while regular requests (for example, same-day or next-day) could be satisfied using either drones or conventional delivery trucks. During the planning stage, the company needs to decide which facilities to upgrade to accommodate new infrastructure and inventory due to the new fulfillment option. Once the company sets up the infrastructure, the operational stage begins wherein the customer starts placing orders in an online manner. The central command dynamically made decisions regarding if and from where the request needs to be fulfilled to maximize the cumulative profits while considering resource budget constraints at facilities. The planning stage prob-

lem was formulated as a mixed-integer linear program. A multi-armed bandits-based (MAB) methodology was adopted for the operational stage wherein the uncertainties arising from stochasticity in demand, request type, and drone battery consumption were considered.

The computational analysis was carried out first on the standard p-median instances. The proposed MAB methodology was compared to three other heuristics and resulted in 7% additional profit from the second-best approach, on average. A sensitivity analysis elicited that all four methodologies were able to effectively tackle increasing variation in demand. Then, an analysis was carried out for the Portland Metro area for a larger operational period. The MAB methodology again provided the best performance, beating the second-best approach by as much as 25%, on average.

Chapter 4 presented an application consisting of an agency planning for short-term disaster response by prepositioning relief packages pre-disaster and carrying out their distribution post-disaster. This problem was formulated as a two-stage problem, similar to Chapter 3, by considering the planning stage to be the pre-disaster phase and the operational stage to be the post-disaster phase. A prime difference here is that instead of maximizing profits, the agency here aims for the equitable distribution of scarce resources. As the primary motive of post-disaster humanitarian logistics is saving human lives, only horizontal equity was considered here. The fairness (or, equity) in distribution was modeled using the concept of envy. During the planning stage, distributions centers are set up and resources allocated to them for post-disaster

distribution. The incidence of the disaster triggers the operational stage, where the agency strived to achieve an envy-free resource distribution in the service area in a dynamic setting.

Similar to Chapter 3, the planning stage was modeled as a mixed-integer linear program to achieve fair distribution in a static sense. A new MAB methodology (called, equitable linear contextual bandits with knapsacks, or E-linCBwK) with an equity-based objective for the operational stage wherein the uncertainties arising from stochasticity in demand were considered. A critical component of the E-linCBwK was deciding which requests to deny to ensure equity while distributing scarce resources. This was captured using the non-skipping probabilities for requests arriving from each relief shelter, which were updated dynamically as more information was available. Additionally, the performance regret bound for the E-linCBwK is also derived.

The computational analysis was conducted on a case study based in the Portland Urban Metro area, motivated by the efforts undertaken by the state of Oregon to improve seismic resilience against the expected magnitude 9.0 earthquake off the Pacific coast. The analysis showed that E-linCBwK performs much better than the two other heuristics. The importance of introducing non-skipping probabilities was highlighted by comparing E-linCBwK with its version that did not allow request skipping. The results show that the terminal total deviation from fair allocation is higher in the "No Skip" version by an order of magnitude. Sensitivity analyses on the variation in post-disaster demand distributions, fairness update frequencies, and

the number of opened DCs showed that E-linCBwK consistently achieved terminal total deviation from the fair allocation which was less than 3.5% of the total number of packages distributed for all cases.

Chapters 3 and 4 proposed a novel two-stage methodology and a new multi-armed bandit problem. While some of the limitations in Chapter 3 are tackled in Chapter 4, there are still avenues for improvement. A primary assumption in both the studies is that enough drones are available such that drone availability is not an issue. However, for practical applications, it is important to consider such capacity constraints. Chapter 4 explicitly considered stratification of resource constraints by each facility in the operational stage, however, the regret bound is only derived until the first instance of constraint violation. Future research in multi-armed bandits should dedicate efforts to effectively incorporating stratified resource budget constraints. Specifically for Chapter 4, a drone-only infrastructure is considered. However, as time passes, road infrastructure is recovered and therefore, could provide new avenues for resource distribution.

Chapter 5 discussed truckload allocation for the dynamic pickup and delivery problem. We considered that a carrier operated a fleet of electric trucks in a region and the central command was tasked with allocating the truckload move requests, successfully procured from spot markets, in the region. A new MAB framework (called, linear contextual blocking bandits, or linCBB) was proposed for the truckload allocation problem that can handle the large context space in a scalable manner

while explicitly considering unavailability constraints. The challenges related to fleet electrification made it imperative to consider recharging decisions while routing and load allocation. The routing affected the context for linCBB, which in turn affected the rewards and the unavailable times. An upper confidence bound (UCB) algorithm was proposed to solve the linCBB problem.

The computational analysis was conducted on a synthetic service area of size $200 \times 200$ miles. The request allocation decisions are made by a central command while incurring distance-based, time-based, recharging, and delay costs. The performance of the linCBB-UCB algorithm was compared against a greedy allocation policy (called greedy-AD) that myopically selects the best available electric truck which minimizes arrival delay at the pickup location. The computational analysis elicited that linCBB-UCB achieved a sub-linear regret and obtained 7.5% more rewards than the greedy-AD policy, on average. Sensitivity analyses on fleet size and arrival delay penalty were also conducted. There are several opportunities to expand this research. Firstly, the relocation feature is not modeled when the electric trucks in not actively serving a request. This could have important implications especially as more service request history is accumulated. Further, a combinatorial bandit setting could be explored as it allows the selection of multiple arms at a single instance which could result in improved fleet management.

## Bibliography


J. D. Adler and P. B. Mirchandani. Online routing and battery reservations for electric vehicles with swappable batteries. *Transportation Research Part B: Methodological*, 70:285–302, 2014.

N. Agatz, A. Erera, M. Savelsbergh, and X. Wang. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295–303, 2012.

M. Aghajani and S. A. Torabi. A mixed procurement model for humanitarian relief chains. *Journal of Humanitarian Logistics and Supply Chain Management*, 2019.

S. Agrawal and N. Devanur. Linear contextual bandits with knapsacks. In *Advances in Neural Information Processing Systems*, pages 3450–3458, 2016.

A. Ahmadi-Javid, P. Seyedi, and S. S. Syam. A survey of healthcare facility location. *Computers & Operations Research*, 79:223–263, 2017.

L. Al-Kanj, J. Nascimento, and W. B. Powell. Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles. *European Journal of Operational Research*, 284(3):1088–1106, 2020.

N. Altay and W. G. Green III. Or/ms research in disaster operations management. *European journal of operational research*, 175(1):475–493, 2006.

American Trucking Association. Economics and industry data, 2019. https://www.trucking.org/economics-and-industry-data, last accessed on 09/07/2021.

W. K. Anuar, L. S. Lee, S. Pickl, and H.-V. Seow. Vehicle routing optimisation in humanitarian operations: A survey on modelling and optimisation approaches. *Applied Sciences*, 11(2):667, 2021.

AON. Weather, Climate & Catastrophe Insight. Technical report, AON, Chicago, IL, 2021. URL `http://catastropheinsight.aon.com`.

A. Aouad. Walmart has filed for 97 drone patents – nearly double the amount of Amazon's, 2019. https://www.businessinsider.com/walmart-files-double-amazons-drone-patents-2019-6, last accessed on 1/29/2021.

R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. T. van Essen. Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368, 2017. doi: 10.1016/j.cor.2016.09.016.

S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

E. Aslan and M. Çelik. Pre-positioning of relief items under road/facility vulnerability with concurrent restoration and relief transportation. *IISE Transactions*, 51(8): 847–868, 2019.

A. Atsidakou, O. Papadigenopoulos, S. Basu, C. Caramanis, and S. Shakkottai. Combinatorial blocking bandits with stochastic delays. In *International Conference on Machine Learning*, pages 404–413. PMLR, 2021.

M. Ayamga, S. Akaba, and A. A. Nyaaba. Multifaceted applicability of drones: A review. *Technological Forecasting and Social Change*, 167:120677, 2021. doi: 10.1016/j.techfore.2021.120677.

M. H. Azizan, C. S. Lim, W. L. W. Hatta, and L. C. Gan. Application of openstreetmap data in ambulance location problem. In *2012 Fourth International Conference on Computational Intelligence, Communication Systems and Networks*, pages 321–325. IEEE, 2012. doi: 10.1109/CICSyN.2012.66.

A. Badanidiyuru, R. Kleinberg, and A. Slivkins. Bandits with knapsacks. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 207–216. IEEE, 2013.

B. Balcik, S. Iravani, and K. Smilowitz. Multi-vehicle sequential resource allocation for a nonprofit distribution system. *IIE Transactions*, 46(12):1279–1297, 2014.

R. H. Ballou. Dynamic warehouse location analysis. *Journal of Marketing Research*, 5(3):271–276, 1968. doi: 10.1177/002224376800500304.

S. Banholzer, J. Kossin, and S. Donner. The Impact of Climate Change on Natural Disasters. In A. Singh and Z. Zommers, editors, *Reducing Disaster: Early Warning Systems for Climate Change*, chapter 2, pages 21–49. Springer, Dordrecht, Netherlands, 2014. ISBN 9789401785983. doi: 10.1007/978-94-017-8598-3.

A. Başar, B. Çatay, and T. Ünlüyurt. A taxonomy for emergency service station location problem. *Optimization letters*, 6(6):1147–1160, 2012. doi: 10.1007/s11590-011-0376-1.

S. Basu, R. Sen, S. Sanghavi, and S. Shakkottai. Blocking bandits. *arXiv preprint arXiv:1907.11975*, 2019.

S. Basu, O. Papadigenopoulos, C. Caramanis, and S. Shakkottai. Contextual blocking bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 271–279. PMLR, 2021.

T. Bektaş and G. Laporte. The pollution-routing problem. *Transportation Research Part B: Methodological*, 45(8):1232–1250, 2011.

T. Bektaş, P. P. Repoussis, and C. D. Tarantilis. Chapter 11: Dynamic vehicle routing problems. In P. Toth and D. Vigo, editors, *Vehicle Routing: Problems, Methods, and Applications, Second Edition*, pages 299–347. SIAM, 2014.

A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations research letters*, 25(1):1–13, 1999. doi: 10.1016/S0167-6377(99)00016-4.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton university press, 2009. doi: 10.1515/9781400831050.

D. Bertsimas and M. Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004. doi: 10.1287/opre.1030.0065.

D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011. doi: 10.1137/080734510.

N. Bishop, H. Chan, D. Mandal, and L. Tran-Thanh. Adversarial blocking bandits. *Advances in Neural Information Processing Systems*, 33:8139–8149, 2020.

C. Boonmee, M. Arimura, and T. Asada. Facility location optimization model for emergency humanitarian logistics. *International Journal of Disaster Risk Reduction*, 24:485–498, 2017.

S. Budge, A. Ingolfsson, and D. Zerom. Empirical analysis of ambulance travel times: the case of calgary emergency medical services. *Management Science*, 56(4):716–723, 2010. doi: 10.1287/mnsc.1090.1142.

A. M. Campbell, D. Vandenbussche, and W. Hermann. Routing for relief efforts. *Transportation science*, 42(2):127–145, 2008.

M. Çelik. Network restoration and recovery in humanitarian operations: Framework, literature review, and research directions. *Surveys in Operations Research and Management Science*, 21(2):47–61, 2016.

U. S. Census-Bureau and American-FactFinder. 2017 American Community Survey 5-Year Estimates. `https://factfinder.census.gov/`, 2018. Accessed: 3/8/2020.

D. Chauhan, A. Unnikrishnan, and M. Figliozzi. Maximum coverage capacitated facility location problem with range constrained drones. *Transportation Research Part C: Emerging Technologies*, 99:1–18, 2019. doi: 10.1016/j.trc.2018.12.001.

D. R. Chauhan, A. Unnikrishnan, M. Figliozzi, and S. D. Boyles. Robust maximum coverage facility location problem with drones considering uncertainties in battery availability and consumption. *Transportation Research Record*, 2675(2):25–39, 2021.

D. R. Chauhan, A. Unnikrishnan, M. A. Figliozzi, and S. D. Boyles. Robust multi-period maximum coverage drone facility location problem considering coverage reliability. *Transportation Research Record*, 2022a. Forthcoming.

D. R. Chauhan, A. Unnikrishnan, and S. D. Boyles. Maximum profit facility location and dynamic resource allocation for instant delivery logistics. *Transportation Research Record*, 2022b. doi: 10.1177/03611981221082574.

T. D. Chen, K. M. Kockelman, and J. P. Hanna. Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transportation Research Part A: Policy and Practice*, 94:243–254, 2016.

W.-C. Chiang, Y. Li, J. Shang, and T. L. Urban. Impact of drone delivery on sustainability and cost: Realizing the uav potential through vehicle routing optimization. *Applied energy*, 242:1164–1175, 2019.

W. Chu, L. Li, L. Reyzin, and R. Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artifi-*

*cial Intelligence and Statistics*, pages 208–214. JMLR Workshop and Conference Proceedings, 2011.

R. G. Conrad and M. A. Figliozzi. The recharging vehicle routing problem. In *Proceedings of the 2011 industrial engineering research conference*, page 8. IISE Norcross, GA, 2011.

N. Cotes and V. Cantillo. Including deprivation costs in facility location models for humanitarian relief logistics. *Socio-Economic Planning Sciences*, 65:89–100, 2019.

M. S. Daskin. *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons, 2011.

S. M. S. M. Daud, M. Y. P. M. Yusof, C. C. Heo, L. S. Khoo, M. K. C. Singh, M. S. Mahmood, and H. Nawawi. Applications of drone in disaster management: A scoping review. *Science & Justice*, 62(1):30–42, 2022.

B. A. Davis and M. A. Figliozzi. A methodology to evaluate the competitiveness of electric delivery trucks. *Transportation Research Part E: Logistics and Transportation Review*, 49(1):8–23, 2013.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010. doi: 10.1287/opre.1090.0741.

G. Desaulniers, F. Errico, S. Irnich, and M. Schneider. Exact algorithms for electric vehicle-routing problems with time windows. *Operations Research*, 64(6):1388–1405, 2016.

N. R. Devanur, K. Jain, B. Sivan, and C. A. Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 29–38, 2011.

N. R. Devanur, K. Jain, B. Sivan, and C. A. Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)*, 66(1):1–41, 2019.

Z. Dönmez, B. Y. Kara, Ö. Karsu, and F. Saldanha-da Gama. Humanitarian facility location under uncertainty: Critical review and future prospects. *Omega*, page 102393, 2021.

S. Enayati, M. E. Mayorga, H. K. Rajagopalan, and C. Saydam. Real-time ambulance redeployment approach to improve service coverage with fair and restricted workload for ems providers. *Omega*, 79:67–80, 2018. doi: 10.1016/j.omega.2017.08.001.

G. Erdoğan, E. Erkut, A. Ingolfsson, and G. Laporte. Scheduling ambulance crews for maximum coverage. *Journal of the Operational Research Society*, 61(4):543–550, 2010. doi: 10.1057/jors.2008.163.

E. Erkut, A. Ingolfsson, and G. Erdoğan. Ambulance location for maximum survival. *Naval Research Logistics (NRL)*, 55(1):42–58, 2008. doi: 10.1002/nav.20267.

FAA. Unmanned Aircraft Systems (UAS) BEYOND, 2021. URL `https://www.faa.gov/uas/programs_partnerships/beyond/`. last accessed on 2/14/2022.

R. Z. Farahani, M. SteadieSeifi, and N. Asgari. Multiple criteria facility location problems: A survey. *Applied mathematical modelling*, 34(7):1689–1709, 2010.

R. Z. Farahani, S. Fallah, R. Ruiz, S. Hosseini, and N. Asgari. Or models in urban service facility location: A critical review of applications and future developments. *European journal of operational research*, 276(1):1–27, 2019.

Á. Felipe, M. T. Ortuño, G. Righini, and G. Tirado. A heuristic approach for the green vehicle routing problem with multiple technologies and partial recharges. *Transportation Research Part E: Logistics and Transportation Review*, 71:111–128, 2014.

M. Figliozzi. Vehicle routing problem for emissions minimization. *Transportation Research Record*, 2197(1):1–7, 2010.

M. A. Figliozzi. Lifecycle modeling and assessment of unmanned aerial vehicles (drones) co2e emissions. *Transportation Research Part D: Transport and Environment*, 57:251–261, 2017. doi: 10.1016/j.trd.2017.09.011.

M. A. Figliozzi, H. S. Mahmassani, and P. Jaillet. Framework for study of carrier strategies in auction-based transportation marketplace. *Transportation Research Record*, 1854(1):162–170, 2003.

M. A. Figliozzi, H. S. Mahmassani, and P. Jaillet. Impacts of auction settings on the performance of truckload transportation marketplaces. *Transportation research record*, 1906(1):89–96, 2005.

D. K. Foley. *Resource allocation and the public sector*. PhD thesis, Yale University, 1966.

Freightliner. Freightliner eM2, 2021. https://freightliner.com/trucks/em2/, last accessed on 12/28/2021.

A. Froger, J. E. Mendoza, O. Jabali, and G. Laporte. Improved formulations and algorithmic components for the electric vehicle routing problem with nonlinear charging functions. *Computers & Operations Research*, 104:256–294, 2019.

V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3):471–483, 2014. doi: 10.1016/j.ejor.2013.09.036.

A. A. Garner and P. L. van den Berg. Locating helicopter emergency medical service bases to optimise population coverage versus average response time. *BMC emergency medicine*, 17(1):31, 2017. doi: 10.1186/s12873-017-0142-5.

M. Gendreau, G. Laporte, and F. Semet. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel computing*, 27(12):1641–1653, 2001.

Z. Ghelichi, M. Gentili, and P. B. Mirchandani. Logistics for a fleet of drones for medical item delivery: A case study for louisville, ky. *Computers & Operations Research*, 135:105443, 2021. doi: 10.1016/j.cor.2021.105443.

T. B. Glick, M. A. Figliozzi, and A. Unnikrishnan. Case study of drone delivery reliability for time-sensitive medical supplies with stochastic demand and meteorological conditions. *Transportation Research Record*, 2676(1):242–255, 2022. doi: 10.1177/03611981211036685.

G. A. Godfrey and W. B. Powell. An adaptive dynamic programming algorithm for dynamic fleet management, i: Single period travel times. *Transportation Science*, 36(1):21–39, 2002.

D. Goeke and M. Schneider. Routing a mixed fleet of electric and conventional vehicles. *European Journal of Operational Research*, 245(1):81–99, 2015.

J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010. doi: 10.1287/opre.1090.0795.

B. L. Gorissen, İ. Yanıkoğlu, and D. den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015. doi: 10.1016/j.omega.2014.12.006.

Q. Gu, T. Fan, F. Pan, and C. Zhang. A vehicle-uav operation scheme for instant delivery. *Computers & Industrial Engineering*, 149:106809, 2020.

W. Guo, B. Atasoy, W. Beelaerts van Blokland, and R. R. Negenborn. Dynamic and stochastic shipment matching problem in multimodal transportation. *Transportation Research Record*, 2674(2):262–273, 2020.

L. Gurobi Optimization. Gurobi optimizer reference manual, 2020. URL `http://www.gurobi.com`.

G. Hiermann, J. Puchinger, S. Ropke, and R. F. Hartl. The electric fleet size and mix vehicle routing problem with time windows and recharging stations. *European Journal of Operational Research*, 252(3):995–1018, 2016.

J. Hof, M. Schneider, and D. Goeke. Solving the battery swap station location-routing problem with capacitated electric vehicles using an avns algorithm for vehicle-routing problems with intermediate stops. *Transportation research part B: methodological*, 97:102–112, 2017.

J. Holguín-Veras, M. Jaller, L. N. Van Wassenhove, N. Pérez, and T. Wachtendorf. On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, 30(7-8):494–506, 2012.

J. Holguín-Veras, N. Pérez, M. Jaller, L. N. Van Wassenhove, and F. Aros-Vera. On the appropriate objective function for post-disaster humanitarian logistics models. *Journal of Operations Management*, 31(5):262–280, 2013.

G. Q. Huang and S. X. Xu. Truthful multi-unit transportation procurement auctions for logistics e-marketplaces. *Transportation Research Part B: Methodological*, 47:127–148, 2013.

M. Huang, K. Smilowitz, and B. Balcik. Models for relief routing: Equity, efficiency and efficacy. *Transportation research part E: logistics and transportation review*, 48 (1):2–18, 2012.

R. Iacobucci, B. McLellan, and T. Tezuka. Optimization of shared autonomous electric vehicles operations with charge scheduling and vehicle-to-grid. *Transportation Research Part C: Emerging Technologies*, 100:34–52, 2019.

Insurance Information Institute. Facts + Statistics: Global Catastrophes, 2021a. https://www.iii.org/fact-statistic/facts-statistics-global-catastrophes#Top Five World Natural Catastrophes By Fatalities, 2020, last accessed on March 7, 2021.

Insurance Information Institute. Facts + Statistics: U.S. Catastrophes, 2021b. URL `https://www.iii.org/fact-statistic/facts-statistics-us-catastrophes`.

O. Jabali, T. Van Woensel, and A. De Kok. Analysis of travel times and co2 emissions in time-dependent vehicle routing. *Production and Operations Management*, 21(6): 1060–1074, 2012.

Z. Jiang, J. Gu, W. Fan, W. Liu, and B. Zhu. Q-learning approach to coordinated optimization of passenger inflow control with train skip-stopping on a urban rail transit line. *Computers & Industrial Engineering*, 127:1131–1142, 2019.

N. Kang, F. M. Feinberg, and P. Y. Papalambros. Autonomous electric vehicle sharing system design. *Journal of Mechanical Design*, 139(1):011402, 2017.

M. Keskin and B. Çatay. Partial recharge strategies for the electric vehicle routing problem with time windows. *Transportation research part C: emerging technologies*, 65:111–127, 2016.

D. Kim, K. Lee, and I. Moon. Stochastic facility location model for drones considering uncertain flight distance. *Annals of Operations Research*, 283(1):1283–1302, 2019. doi: 10.1007/s10479-018-3114-6.

N. D. Kullman, M. Cousineau, J. C. Goodson, and J. E. Mendoza. Dynamic ride-hailing with electric vehicles. *Transportation Science*, 2021.

G. Kuyzu, Ç. G. Akyol, Ö. Ergun, and M. Savelsbergh. Bid price optimization for truckload carriers in simultaneous transportation procurement auctions. *Transportation Research Part B: Methodological*, 73:34–58, 2015.

A. Langevin, P. Mbaraga, and J. F. Campbell. Continuous approximation models in freight distribution: An overview. *Transportation Research Part B: Methodological*, 30(3):163–188, 1996.

A. Leiras, I. de Brito Jr, E. Q. Peres, T. R. Bertazzo, and H. T. Y. Yoshizaki. Literature review of humanitarian logistics research: trends and challenges. *Journal of Humanitarian Logistics and Supply Chain Management*, 2014.

M. Leonard. Amazon Prime Air gets FAA clearance for drone delivery on 'highly rural' test range, 2020. https://www.supplychaindive.com/news/amazon-prime-air-faa-clearance-drone-delivery-rural-test-range/584436/, last accessed on 2/14/2022.

A. Leslie and D. Murray. An Analysis of the Operational Costs of Trucking: 2021 Update, 2021. American Transportation Research Institute.

X. Li, Z. Zhao, X. Zhu, and T. Wyatt. Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research*, 74(3):281–310, 2011. doi: 10.1007/s00186-011-0363-4.

R. W. Lien, S. M. Iravani, and K. R. Smilowitz. Sequential resource allocation for nonprofit operations. *Operations Research*, 62(2):301–317, 2014.

K. Lin, R. Zhao, Z. Xu, and J. Zhou. Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1774–1783, 2018.

N. Loree and F. Aros-Vera. Points of distribution location and inventory management model for post-disaster humanitarian logistics. *Transportation Research Part E: Logistics and Transportation Review*, 116:1–24, 2018.

P. Lutter, D. Degel, C. Büsing, A. M. Koster, and B. Werners. Improved handling of uncertainty and robustness in set covering problems. *European Journal of Operational Research*, 263(1):35–49, 2017. doi: 10.1016/j.ejor.2017.04.044.

P. Mallick, S. Sarkar, and P. Mitra. Decision recommendation system for transporters in an online freight exchange platform. In *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, pages 448–453. IEEE, 2017.

J. A. Mesa, J. Puerto, and A. Tamir. Improved algorithms for several network location problems with equality measures. *Discrete applied mathematics*, 130(3):437–448, 2003. doi: 10.1016/S0166-218X(02)00599-1.

J. P. Minas, N. Simpson, and Z. Tacheva. Modeling emergency response operations: a theory building survey. *Computers & Operations Research*, 119:104921, 2020.

A. Montoya, C. Guéret, J. E. Mendoza, and J. G. Villegas. The electric vehicle routing problem with nonlinear charging function. *Transportation Research Part B: Methodological*, 103:87–110, 2017.

M. Moshref-Javadi and M. Winkenbach. Applications and research avenues for drone-based models in logistics: A classification and review. *Expert Systems with Applications*, page 114854, 2021.

A. Mourad, J. Puchinger, and C. Chu. A survey of models and algorithms for optimizing shared mobility. *Transportation Research Part B: Methodological*, 123:323–346, 2019.

S. Mukundan and M. S. Daskin. Joint location/sizing maximum profit covering models. *INFOR: Information Systems and Operational Research*, 29(2):139–152, 1991.

J. Naoum-Sawaya and S. Elhedhli. A stochastic optimization model for real-time ambulance redeployment. *Computers & Operations Research*, 40(8):1972–1978, 2013. doi: 10.1016/j.cor.2013.02.006.

NFPA. *NFPA 1710: Standard for the Organization and Deployment of Fire Suppression Operations, Emergency Medical Operations, and Special Operations to the Public by Career Fire Departments*. National Fire Protection Agency, 2020.

S. Nickel and F. Saldanha-da Gama. Multi-period facility location. In *Location science*, pages 303–326. Springer, 2019. doi: 10.1007/978-3-030-32177-2_11.

A. A. Nyaaba and M. Ayamga. Intricacies of medical drones in healthcare delivery: Implications for africa. *Technology in Society*, 66:101624, 2021. doi: 10.1016/j.techsoc.2021.101624.

Oregon Seismic Safety Policy Advisory Commission. The Oregon Resilience Plan - Reducing Risk and Improving Recovery for the Next Cascadia Earthquake and Tsunami. Technical report, 2013. URL https://www.oregon.gov/oem/Documents/Oregon_Resilience_Plan_Final.pdf.

I. H. Osman and N. Christofides. Capacitated clustering problems by hybrid simulated annealing and tabu search. *International Transactions in Operational Research*, 1 (3):317–336, 1994.

O. Papadigenopoulos and C. Caramanis. Recurrent submodular welfare and matroid blocking semi-bandits. *Advances in Neural Information Processing Systems*, 34, 2021.

J. A. Paul and X. J. Wang. Robust location-allocation network design for earthquake preparedness. *Transportation research part B: methodological*, 119:139–155, 2019.

S. Pelletier, O. Jabali, and G. Laporte. 50th anniversary invited article – goods distribution with electric vehicles: review and research perspectives. *Transportation science*, 50(1):3–22, 2016.

S. Pelletier, O. Jabali, and G. Laporte. The electric vehicle routing problem with energy consumption uncertainty. *Transportation Research Part B: Methodological*, 126:225–255, 2019.

N. Pérez-Rodríguez and J. Holguín-Veras. Inventory-allocation distribution models for postdisaster humanitarian logistics with explicit consideration of deprivation costs. *Transportation Science*, 50(4):1261–1285, 2016.

Philips. Side by side. step by step. philips heartstart onsite aed, 2022. URL https://www.documents.philips.com/assets/20170523/5777e26b79a1438b8138a77c014ef7d9.pdf. last accessed on 2/14/2022.

W. B. Powell. A unified framework for stochastic optimization. *European Journal of Operational Research*, 275(3):795–821, 2019.

A. S. Prasad and L. H. Francescutti. Natural Disasters. In S. R. Quah and W. Cockerham, editors, *The International Encyclopedia of Public Health*, pages 215–222. Elsevier, Waltham, MA, 2nd edition, 2017. ISBN 0791456218. doi: 10.5860/choice.46.07.1247.

A. Pulver and R. Wei. Optimizing the spatial location of medical drones. *Applied geography*, 90:9–16, 2018. doi: 10.1016/j.apgeog.2017.11.009.

A. Pulver, R. Wei, and C. Mann. Locating aed enabled medical drones to enhance cardiac arrest response times. *Prehospital Emergency Care*, 20(3):378–389, 2016. doi: 10.3109/10903127.2015.1115932.

PYMTS. Walmart+ Leverages Grocery Delivery; Gains 8 Million New Paid Subscribers, 2021. https://www.pymnts.com/news/retail/2021/walmart-leverages-grocery-delivery-gains-8-million-new-paid-subscribers/, last accessed on 5/28/2021.

C. G. Rawls and M. A. Turnquist. Pre-positioning and dynamic delivery planning for short-term response following a natural disaster. *Socio-Economic Planning Sciences*, 46(1):46–54, 2012.

J. Rawls. *A theory of justice*. Harvard university press, 2020.

A. Rejeb, K. Rejeb, S. Simske, and H. Treiblmaier. Humanitarian drones: A review and research agenda. *Internet of Things*, 16:100434, 2021.

C. S. ReVelle and H. A. Eiselt. Location analysis: A synthesis and survey. *European journal of operational research*, 165(1):1–19, 2005.

D. Rey, K. Almi'ani, and D. J. Nair. Exact and heuristic algorithms for finding envy-free allocations in food rescue pickup and delivery logistics. *Transportation Research Part E: Logistics and Transportation Review*, 112:19–46, 2018.

D. Rivera-Royero, G. Galindo, and R. Yie-Pinedo. Planning the delivery of relief supplies upon the occurrence of a natural disaster while considering the assembly process of the relief kits. *Socio-Economic Planning Sciences*, 69:100682, 2020.

J. Røislien, P. L. van den Berg, T. Lindner, E. Zakariassen, K. Aardal, and J. T. van Essen. Exploring optimal air ambulance base locations in norway using advanced mathematical modelling. *Injury prevention*, 23(1):10–15, 2017. doi: 10.1136/injuryprev-2016-041973.

M. Schiffer and G. Walther. The electric location routing problem with time windows and partial recharging. *European journal of operational research*, 260(3):995–1013, 2017.

V. Schmid. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European journal of operational research*, 219 (3):611–621, 2012.

V. Schmid and K. F. Doerner. Ambulance location and relocation problems with time-dependent travel times. *European journal of operational research*, 207(3): 1293–1303, 2010. doi: 10.1016/j.ejor.2010.06.033.

M. Schneider, A. Stenger, and D. Goeke. The electric vehicle-routing problem with time windows and recharging stations. *Transportation science*, 48(4):500–520, 2014.

J. Shao, X. Wang, C. Liang, and J. Holguín-Veras. Research progress on deprivation costs in humanitarian logistics. *International Journal of Disaster Risk Reduction*, 42:101343, 2020.

S. M. Shavarani, S. Mosallaeipour, M. Golabi, and G. İzbirak. A congested capacitated multi-level fuzzy facility location problem: An efficient drone delivery system. *Computers & Operations Research*, 108:57–68, 2019.

J.-B. Sheu. An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transportation Research Part E: Logistics and Transportation Review*, 43(6):687–709, 2007.

J.-B. Sheu. Post-disaster relief–service centralized logistics distribution with survivor resilience maximization. *Transportation research part B: methodological*, 68:288–314, 2014.

J. Shi, Y. Gao, W. Wang, N. Yu, and P. A. Ioannou. Operating electric vehicle fleet for ride-hailing services with reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4822–4834, 2019.

H. P. Simao, J. Day, A. P. George, T. Gifford, J. Nienow, and W. B. Powell. An approximate dynamic programming algorithm for large-scale fleet management: A case application. *Transportation Science*, 43(2):178–197, 2009.

D. Simchi-Levi, Z. Zheng, and F. Zhu. Online matching with reusable network resources and decaying rewards: New framework and analysis. *Available at SSRN 3981123*, 2021.

R. Sirupa. Incorporating conditional $\beta$-mean based equity metric in coverage based facility location problems. Master's thesis, Portland State University, 2021.

A. Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.

A. S. Thomas and L. R. Kopczak. From logistics to supply chain management: the path forward in the humanitarian sector. *Fritz Institute*, 15(1):1–15, 2005.

S. Tofighi, S. A. Torabi, and S. A. Mansouri. Humanitarian logistics network design under mixed uncertainty. *European Journal of Operational Research*, 250(1):239–250, 2016.

H. Topaloglu and W. B. Powell. Dynamic-programming approximations for stochastic time-staged integer multicommodity-flow problems. *INFORMS Journal on Computing*, 18(1):31–42, 2006.

S. A. Torabi, I. Shokr, S. Tofighi, and J. Heydari. Integrated relief pre-positioning and procurement planning in humanitarian supply chains. *Transportation Research Part E: Logistics and Transportation Review*, 113:123–146, 2018.

U.S. Census Bureau. Quaterly Retail E-commerce sales – 4th Quarter 2020, 2021. https://www2.census.gov/retail/releases/historical/ecomm/20q4.pdf, last accessed on 4/8/2021.

U.S. Department of Homeland Security. Natural Disasters, 2021. URL `https://www.dhs.gov/natural-disasters`.

U.S. Department of Transportation. National Strategic Freight Plan, 2020. https://www.transportation.gov/sites/dot.gov/files/2020-09/NFSP_fullplan_508_0.pdf; last accessed: 07/09/2021.

U.S. Department of Transportation, Bureau of Transportation Statistics. Transportation economic trends, 2021. https://www.bts.gov/product/transportation-economic-trends; last accessed: 07/09/2021.

U.S. Department of Transportation, Federal Highway Administration. Integrating Demand Management into the Transportation Planning Process: A Desk Reference,

2012. https://ops.fhwa.dot.gov/publications/fhwahop12035/fhwahop12035.pdf; last accessed: 07/09/2021.

US EPA. Fast facts on transportation greenhouse gas emissions, 2021. https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions; last accessed: 10/3/2021.

H. R. Varian. Equity, envy, and efficiency. *Journal of Economic Theory*, 9 (1):63–91, 1974. ISSN 0022-0531. doi: https://doi.org/10.1016/0022-0531(74)90075-1. URL https://www.sciencedirect.com/science/article/pii/0022053174900751.

A. K. Vatsa and S. Jayaswal. Capacitated multi-period maximal covering location problem with server uncertainty. *European Journal of Operational Research*, 289 (3):1107–1126, 2021. doi: 10.1016/j.ejor.2020.07.061.

S. S. Villar, J. Bowden, and J. Wason. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 30(2):199, 2015.

Y. Wang, J. M. D. Nascimento, and W. Powell. Reinforcement learning for dynamic bidding in truckload markets: an application to large-scale fleet management with advance commitments. *arXiv preprint arXiv:1802.08976*, 2018.

M. Yahyaei and A. Bozorgi-Amiri. Robust reliable humanitarian relief network design: an integration of shelter and supply facility location. *Annals of Operations Research*, 283(1):897–916, 2019.

J. Yang, P. Jaillet, and H. S. Mahmassani. On-line algorithms for truck fleet assignment and scheduling under real-time information. *Transportation Research Record*, 1667(1):107–113, 1999.

J. Yang, P. Jaillet, and H. Mahmassani. Real-time multivehicle truckload pickup and delivery problems. *Transportation Science*, 38(2):135–148, 2004.

L. Yu, H. Yang, L. Miao, and C. Zhang. Rollout algorithms for resource allocation in humanitarian logistics. *IISE Transactions*, 51(8):887–909, 2019.

L. Yu, C. Zhang, J. Jiang, H. Yang, and H. Shang. Reinforcement learning approach for resource allocation in humanitarian logistics. *Expert Systems with Applications*, 173:114663, 2021.

M. A. Zaffar, H. K. Rajagopalan, C. Saydam, M. Mayorga, and E. Sharer. Coverage, survivability or response time: A comparative study of performance statistics used in ambulance location models via simulation–optimization. *Operations research for health care*, 11:1–12, 2016. doi: 10.1016/j.orhc.2016.08.001.

T. Zhu, S. D. Boyles, and A. Unnikrishnan. Two-stage robust facility location problem with drones. *Transportation Research Part C: Emerging Technologies*, 137:103563, 2022.