Optimizing last-mile logistics with drones: A simulation-based stochastic deep Q-learning approach

by

Suyash Pratap

B.Tech., Indian Institute of Technology Kanpur, 2020

A Thesis

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Industrial and Manufacturing Systems Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Approved by:

Major Professor
Shing I. Chang

Copyright

Abstract

This research addresses the optimization of last-mile logistics using unmanned aerial vehicles (drones), focusing on the challenges of limited battery capacity, diverse charging strategies, and efficient routing in a complex, stochastic environment. The study aims to enhance drone delivery operations through advanced modelling and reinforcement learning techniques.

The inherent complexity of drone delivery systems, characterized by high-dimensional state spaces, continuous action spaces, and stochastic elements such as variable delivery demands and energy consumption, necessitates the use of advanced optimization methods. Traditional optimization approaches often struggle with such complex, dynamic systems. Therefore, this study employs deep Q-learning, a powerful reinforcement learning technique capable of handling high-dimensional state spaces and learning optimal policies in complex environments without explicit programming.

A deep Q-learning model was implemented to optimize decision-making processes, with particular attention to the state and action space, reward function, and training procedure. The model's performance was evaluated based on its ability to improve operational efficiency and make optimal charging and routing decisions in real-time. A simulated drone delivery network comprising 14 nodes, including 2 origin nodes and 1 charging station, was created using Simio to test the model. This setup allowed for the capture of discrete timings for drone movements and charging decisions, informing the development of a stochastic model that accurately represents system uncertainties.

Key findings reveal significant improvements in drone operation efficiency through the application of the deep Q-learning model. The model demonstrated the ability to learn complex strategies, balancing immediate rewards with long-term consequences in a way that would be challenging for traditional optimization methods. Analysis of charging behaviours

provided insights into the trade-offs between fast and normal charging options and their impact on overall delivery performance, showcasing the model's capability to make nuanced decisions in a multi-objective optimization context.

This research contributes to the field by offering a scalable and effective solution for managing the complexity of drone delivery networks, with potential applications across various logistics scenarios. The proposed approach demonstrates the potential for substantial improvements in last-mile logistics efficiency and reliability, highlighting the value of combining advanced optimization techniques with deep learning to address complex transportation challenges in dynamic, uncertain environments.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I wish to express my profound gratitude to my committee member, Dr. Ashesh Kumar Sinha, for his unwavering support, patience, and invaluable insights throughout this research process. His guidance has been instrumental in shaping not only this thesis but also my growth as a researcher. Equally, I am deeply indebted to Dr. Shing Chang, whose role as my major advisor has been crucial to the success of this work.

I extend my sincere appreciation to my other committee members: Dr. Husain Aziz, Dr. Shuting Lei, and Dr. Malgorzata Rys. Their constructive feedback and encouragement have significantly enhanced the quality of this research and broadened my academic perspective.

The support provided by the staff and faculty of the IMSE department has been invaluable in making my master's journey smooth and productive. Their dedication to student success has not gone unnoticed, and I am truly grateful for their assistance throughout my time here.

To my fellow graduate students and friends - Paul, Sheetal, Akanksha, Priyankar, Rehan, Ved, Amrutha, Vydehi, Arijit and Abhinash - thank you for the camaraderie, support, and all the wonderful memories we've shared over the past two years. Your friendship has made this journey not just bearable, but enjoyable.

Finally, and most importantly, I dedicate this thesis to my family: my father, mother, and sister. Your unconditional love, unwavering care, countless sacrifices, and constant support have been the bedrock of my achievements. Without your enthusiasm and belief in me, none of this would have been possible. This accomplishment is as much yours as it is mine.

To all who have played a part in this journey, my heartfelt thanks.

Dedication

To my parents.

Chapter 1

Introduction

1.1 Research Motivation and Objectives

1.1.1 The Evolution of Last-Mile Delivery and Emergence of Drone

The landscape of last-mile delivery, a critical component in supply chain logistics, has undergone significant transformations over the past century. This evolution has been particularly pronounced in recent years, driven by the exponential growth of e-commerce. Traditionally, last-mile delivery operations have relied heavily on conventional ground-based vehicles, such as trucks and vans. However, this approach has been increasingly challenged by a multitude of factors, including urban congestion, escalating operational costs, and mounting environmental concerns stemming from carbon emissions and noise pollution [1–3].

The potential of drone technology in revolutionizing last-mile delivery is predicated on several inherent advantages. Drones possess the capability to circumvent ground-based traffic congestion, thereby significantly reducing delivery times and associated costs. Furthermore, recent advancements in drone technology, particularly in areas such as battery longevity and lightweight material construction, have substantially improved their commercial viability [4]. The integration of drones into existing logistics networks has been the subject of numerous studies, highlighting their transformative potential, especially in urban environments characterized by high congestion levels [5, 6].

1.1.2 Advantages and Challenges of Drone Delivery Networks

The implementation of drone delivery networks offers a multitude of potential benefits:

- Enhanced Delivery Speed: Drones facilitate the possibility of same-day or near-instantaneous deliveries, aligning with escalating consumer expectations for rapid service [7].
- Cost Efficiency: By mitigating the need for fuel-intensive vehicles and reducing labor costs associated with traditional delivery methods, drone operations present opportunities for significant cost optimization [8].
- Environmental Sustainability: The utilization of electric drones contributes to a reduced carbon footprint compared to conventional delivery vehicles, aligning with global initiatives aimed at mitigating climate change and promoting sustainable logistics practices [9–11].

The implementation of drone delivery systems, while promising, faces several challenges that must be addressed to successfully integrate them into existing logistics networks :

- Limited Operational Range: There is a limited operational range of current drone technology, typically restricted to a radius of approximately 20 miles. This limitation necessitates strategic placement of distribution centers or the development of innovative solutions, such as truck-drone tandem delivery systems, to extend the effective range of drone deliveries [12, 13] The operational constraints of drones can hinder their ability to serve remote or less accessible areas, thereby limiting their overall utility in last-mile delivery scenarios.
- Energy Constraints: The limited battery capacity of drones requires meticulous route planning to ensure that deliveries can be completed within the operational limits of

the drone's battery life. This challenge may necessitate the integration of charging infrastructure within the logistics network, which can complicate operational logistics and increase costs [14]. Moreover, the need for efficient energy management is essential to maximize the effectiveness of drone deliveries, particularly in urban environments where traffic and other logistical challenges may arise [15].

- Regulatory Complexities: The commercial deployment of drones for delivery purposes is subject to evolving regulatory frameworks that exhibit considerable variation across different jurisdictions. This inconsistency can create uncertainty for businesses looking to invest in drone technology, as they must navigate a complex landscape of local, national, and international regulations [16]. Along with that, ethical considerations surrounding privacy and safety must be addressed to gain public trust and acceptance of drone delivery services [17].

- Network Reliability: The introduction of drones creates new potential points of failure within the supply chain, necessitating robust network design to maintain operational reliability. This includes ensuring that drones can operate safely alongside traditional delivery methods and that contingency plans are in place to address potential malfunctions or disruptions in service [18, 19]. The need for a reliable and resilient logistics network is paramount, particularly as consumer expectations for timely deliveries continue to rise [20].

1.2 Proposed Methods and Research Contributions

This dissertation presents a novel methodological framework for the optimization of drone delivery networks, integrating stochastic modeling techniques with Reinforcement Learning. Building upon recent advancements in the field, particularly the heuristic approach proposed by Tolooie, Sinha and Nouri et al. (2024) [21] for optimizing reliable supply chain networks using drones in last-mile delivery under uncertainty, our research leverages the power of deep reinforcement learning. The proposed approach synthesizes simulation-based analysis with deep reinforcement learning to address the multifaceted challenges inherent in drone-based last-mile delivery systems. This integration facilitates a comprehensive examination of network dynamics while providing robust, data-driven solutions for operational optimization. By combining stochastic modeling with advanced AI techniques, our framework offers dynamic adaptation to changing conditions, comprehensive optimization of multiple operational aspects, scalability to handle increasing network complexity, and the ability to uncover non-obvious patterns and strategies. This research not only advances the theoretical understanding of drone delivery network optimization but also provides practical tools for improving the efficiency and reliability of last-mile delivery operations.

1.2.1 Simulation-Based Modeling

Central to our methodology is a simulation-based approach utilizing Simio, an advanced simulation software platform. We model a compact toy network comprising 14 nodes, incorporating two origin nodes and a dedicated charging station. This carefully designed simulation framework offers several key advantages:

- The model replicates real-world scenarios of drone operations, including crucial aspects such as battery management, delivery routing, and other variabilities. This high-fidelity approach ensures that our subsequent analyses are grounded in realistic operational contexts.

- Through iterative simulations, we generate extensive datasets capturing intricate details of drone movements, charging behaviors, and overall network performance metrics. This data forms the foundation for our reinforcement learning model and subsequent analyses.

- The simulation incorporates stochastic elements to mirror real-world uncertainties, including fluctuations in battery consumption rates and variabilities in delivery timings. This probabilistic approach enhances the model's relevance and applicability to practical scenarios, providing a more nuanced understanding of network dynamics.

1.2.2 Deep Q-Learning for Optimization

Building on the simulation data, the study implements a deep Q-learning algorithm to optimize drone decision-making processes. This approach is advantageous in the following aspects :

- Adaptation to Network Variability: The model aims to adapt to changing network conditions and battery states, potentially enhancing its applicability to real-world scenarios.

- Charging Decision Optimization: The algorithm seeks to optimize charging decisions, balancing operational efficiency with network reliability.

- Iterative Learning: The reinforcement learning framework allows for potential improvement through iterative learning processes.

- Discrete Event Focus: The model is trained on discrete timing data, focusing on critical events such as charging station entry and exit times, and charging type selection.

1.2.3 Research Contributions

This study endeavors to make significant contributions to the field of drone delivery network optimization through a multifaceted approach. By proposing an integrated framework that synergizes stochastic simulation with reinforcement learning, the research aims to offer a comprehensive methodology for analyzing and optimizing drone delivery networks. The incorporation of probabilistic charging decisions in the study's stochastic charging strategy analysis seeks to find the intricate relationships between various charging strategies and their effects on network performance and reliability. While the current focus is on a 14-node network, the methodology has been conceived with scalability in mind, potentially extending its applicability to larger and more complex systems. Furthermore, the research aims to address a critical challenge in drone logistics by contributing to the optimization of battery management strategies. The findings of this study may yield valuable insights into charging station placement and delivery route configuration, potentially advancing the development of more efficient last-mile logistics solutions.

1.3 Research Tasks and Questions

The primary objective of this research is to develop an integrated simulation and deep Q-learning model for a drone delivery network, enabling comprehensive analysis of the network's performance using metrics such as delivery times, battery utilization, and charging station efficiency. Simulation is employed to a compact 14-node network as a case study to examine the network under various scenarios, including maximum feasible delivery capacity and conduct holistic analysis of the network dynamics.

This research employs a trio of key experiments to evaluate the analytical capabilities of the proposed model. The first experiment focuses on optimizing drone routing and charging decisions, aiming to enhance operational efficiency. The second experiment assesses the impact of various charging strategies on overall network performance, seeking to identify optimal approaches. Finally, the third experiment evaluates network resilience under diverse operational conditions, testing the system's adaptability and robustness. Collectively, these experiments are designed to provide a comprehensive understanding of the model's effectiveness in addressing critical aspects of drone delivery network management, from operational optimization to system resilience.

This research investigates the following research questions:

- RQ1: What is the optimal balance between fast and normal charging strategies to maximize network efficiency?
- RQ2: How does the deep Q-learning model adapt to and optimize drone decision-making in response to varying network conditions?
- RQ3: How does the stochastic nature of delivery demands and battery consumption affect the overall reliability and efficiency of the drone delivery network?

Addressing these questions will provide insights into the capabilities of the proposed integrated simulation and deep Q-learning approach. It will demonstrate its potential as an analytical tool for optimizing drone delivery networks, with implications for operational efficiency, resource allocation, and strategic network design.

1.4 Dissertation Outline

This dissertation is structured into seven chapters, each addressing specific aspects of the research on drone delivery network optimization. The outline of the dissertation is as follows:

Chapter 1: Introduction

This opening chapter lays the groundwork for the entire dissertation. It begins by explaining the motivations driving this research and clearly states the study's objectives. The chapter then introduces the methods proposed to achieve these goals. Furthermore, it highlights the expected contributions of this work to the field and presents the key research questions that will guide the investigation. By providing this essential context, the introduction sets the stage for the detailed exploration that follows in subsequent chapters.

Chapter 2: Literature Review

This chapter presents a comprehensive examination of the existing literature relevant to the study. It begins by exploring the fundamental concepts of supply chain network design and reliability, establishing a theoretical foundation for the research. The review then focuses on the emerging field of drone applications in last-mile delivery, highlighting recent advancements and challenges. Attention is given to various drone charging station strategies, a critical aspect of sustainable drone operations. The chapter also focuses on stochastic approaches for addressing supply chain problems, acknowledging the inherent uncertainties in real-world logistics. Finally, it surveys the growing body of research on machine learning applications in logistics optimization, providing insight into cutting-edge techniques that inform this study's methodology.

Chapter 3: Mathematical Model

This chapter lays out the mathematical foundations that underpin the research. It begins by presenting a detailed formulation of the stochastic model for drone operations, which captures the inherent uncertainties in drone delivery networks. The chapter then transitions to the development of the deep Q-learning model, a key component of the study's machine-learning approach. It carefully explains the model's structure, including the definition of state and action spaces, which represent the possible conditions and decisions within the drone network. The chapter concludes with a thorough description of the reward function design, which guides the learning process and optimization of the drone delivery system.

Chapter 4: Methodology

The methodology chapter provides a comprehensive overview of the research approach. It starts by detailing the simulation setup using Simio software, explaining how the drone delivery network is configured and what parameters are used to reflect real-world conditions. The chapter then describes the data collection process from the simulation, outlining how relevant information is gathered to inform the machine learning model. It follows with an explanation of the data preprocessing techniques employed to prepare the collected data for analysis. The chapter concludes by detailing the implementation of the deep Q-learning algorithm, including specifics on how it is applied to optimize the drone delivery network.

Chapter 5: Results

This chapter presents the findings of the research in a clear and organized manner. It begins with an overview of the simulation results, highlighting key observations that emerged from the drone network model. The chapter then moves on to a performance analysis of the deep Q-learning model, assessing its effectiveness in optimizing drone operations. A detailed analysis of charging decisions and their impact on network performance follows, providing insights into this critical aspect of drone logistics. The chapter concludes by presenting the results of the three key experiments outlined in the research tasks, offering a comprehensive view of the model's capabilities and performance under various conditions.

Chapter 6: Discussion and Conclusion

The final chapter provides a thoughtful analysis and synthesis of the research findings. It interprets the key results in the context of the original research questions, drawing connections between the data and the study's objectives. The chapter then compares these findings with existing literature, highlighting how this research contributes to or challenges current understanding in the field. It discusses the implications of the results for drone delivery system design and operation, offering practical insights for industry applications. The chapter acknowledges the limitations of the current study, providing a balanced view of the research. It then proposes recommendations for future research directions, identifying areas where further investigation could yield valuable insights.

Chapter 2

Literature Review

2.1 Supply Chain Network Design and Reliability

2.1.1 Overview of Supply Chain Network Design

Supply chain network design is organizing a supply chain with respect to facilities, transportation mode, and inventory status to meet the desired goal. This field employs various methodological methods like computational simulations and quantitative analysis for performance and market conditions in the specific realm of networks. It involves designing the structure and functions of the network so as to satisfy the customer need with minimal cost implications. Traditionally, the integration and expansion of supply chain networks were the key drivers, and the growth has been further fuelled by globalization and digitalization [22]. Moving from plain barter forms to digital logistic systems, the supply chain design process has advanced to adapt to the growing global business environment.

Supply chain networks are crucial units of effective supply networks that help organizations in reducing other cost variables inherent in product manufacturing and delivery as well as the purchase of inputs and other needs. However, such studies leave out the ability of such networks to adapt to sudden changes in the economy within a short time frame which forms a research gap in the current supply chain literature. Implementing an efficient network minimizes wastage and accelerates delivery time, enhancing customers' satisfaction and firms' competitiveness in the market [23]. Second, another essential aspect of the successful operation of supply chains is the efficient management of supply chains in multinational trading today since disruptions and fluctuations are inevitable. This adaptability is crucial in the unpredictable business climate in which customers' preferences and overall economic climate can change swiftly.

The key components of a supply chain network consist of nodes and links – factories, warehouses, distribution centers, and transport links. All these components interrelate to ensure the effective transport of products from suppliers to customers. Integrating efficient communication and data processing tools is essential for collaboration and decision-making [24]. The last one is the product flow which comprises inventory and distribution which is vital in ascertaining that the product is in the right place at the right time to meet the demands of consumption and other economic activities.

2.1.2 Reliability in Supply Chain Networks

In supply chain management, reliability is defined as the ability of the supply chain to deliver certain supply chain activities in a given condition within a certain time frame. Reliability is essential in keeping business functionalities going, meeting customer demands, and supporting business growth. It is rooted in three primary components: robustness, recoverability, and resilience. Robustness refers to the ability of the supply chain to keep functioning despite the regular disturbances and anticipated shocks [25]. Resilience implies the likelihood and capacity to return to normalcy and operation after disturbances and disruptions. Recoverability involves the competencies that can be used to restore an operation to optimal functionality after dealing with a disruption.

Several internal and external factors affect the reliability of the supply chain. Internally, the management of inventories is vital since they dictate product availability and costs, which are reflected in the robustness of the network. Equipment maintenance also matters; proper maintenance and timely equipment replacement hinder operational breakdowns that would otherwise cause severe disruptions [26]. Also, competency levels of the workforce influence how well the operations of the supply chain proceedings are completed, where competent employees are in a better position to handle complicated situations. Externally, supplier reliability is crucial to guarantee that the suppliers deliver the materials on time, which will avoid congestion in production processes. Product availability refers to providing products to intended markets and other centers in time, essential in satisfying customers. On the same note, geopolitical factors such as political stability, regulatory environment, and trade policies profoundly influence the supply chain, ranging from material costs to market access. The technological aspects are also critical in determining the reliability of the supply chain. State-of-the-art IT systems promote real-time visibility and better coordination within the value chain and supply networks' processes. These are advantageous because they limit the chances of human mistakes and enhance the processing rate, making the supply chain more reliable.

Several measurements are taken in an organization to measure and manage the reliability of various supply chains. Metrics such as on-time delivery and order accuracy are apparent operational efficiency measures. Other important metrics include supply chain downtime, which measures the robustness of operations by checking the frequency and time taken before interruptions [27]. Resilience measures like the ability to respond to market fluctuation and interruption test how prepared and sensitive the supply chain is. Lastly, resilience metrics reflect the time it takes to recover from disruptions and how companies adapt to new technologies regarding supply chain practices' longevity and progressive improvement. Improving these measures is central to making the supply chain more reliable, achieving the proper balance between supply chain solidity, on the one hand, and its ability to adapt to current and future challenges, on the other.

2.1.3 Optimization Techniques in Supply Chain Network Design

Optimization methods play a vital role in improving the efficiency and robustness of supply chain network solutions. Some traditional models that have laid down a good framework for solving these elaborate networks include linear and mixed-integer programming. Linear programming (LP) plays a significant role in logistics management by directing resources such as transport lines and warehouse processes to reduce expenses [28]. Its application in the design of networks is vast due to its effectiveness in solving problems in the field of cost and resources. Mixed-integer programming (MIP) takes this further by allowing integer variables, which are required to model choices like the number and location of depots or production facilities, where there are clear distinctions, notably the numbers of vehicles needed for distribution.

Nonetheless, these traditional approaches that are applied widely have some drawbacks when dealing with the dynamism and randomness of new-generation supply chains. They are ineffective for making instantaneous decisions and must scale better with large, intricate networks. To overcome these limitations, the latest developments have incorporated AI and machine learning frameworks for supply chain operations. These algorithms provide improved demand forecasting, better routing, and dynamic inventory management, and they learn over time to enrich network design strategies further [29]. Also, real-time data analytics has made previous static models dynamic models that respond to market trends and shocks instantly, making them more responsive and flexible.

Advanced technologies, like digital twins, are also being used now to model and study supply chains and blockchain for increased transparency and security across networks. Nevertheless, reliability considerations find their way into these sophisticated modern optimization models. Risk mitigation approaches, including robust optimization, help build networks' resistance to threats like suppliers' or transportation breakdowns [30]. A typical example of a reliability-focused network redesign is Amazon, which has decentralized its warehousing systems and adopted intelligent forecasting that helps disrupt low-risk stocks.

Looking to the future, sustainability is emerging as one of the critical factors for safe and efficient mass networks. There is a new trend of networks that not only go into resilience against disruptions but also have a lesser negative impact on the environment. This includes developing efficient transport networks that minimize carbon emissions footprints and resource use as a shift towards more integrated, intelligent, and robust supply chain network architectures [31]. These advancements apply technology to cater to modern needs by enhancing the overall stability of supply chain networks to meet the requirements of today's business environment.

2.2 Application of Drones in Last-Mile Delivery

2.2.1 Using Drones in Package Delivery

Drone delivery is a model of delivery using unmanned aerial vehicles (UAVs) to deliver parcels from distribution centers to consumers, specifically in the last mile of logistics. Studies evaluating these systems often employ mixed-method approaches, combining quantitative traffic simulations with qualitative stakeholder interviews to assess efficacy and stakeholder acceptance. Reducing delivery time and expenditure is gradually becoming more feasible, especially in neat and populated regions [32]. Thus, adopting drones in the last mile should push companies to avoid traditional ground transportation-channel constraints, providing better delivery options in speed and flexibility. This innovation is significant in e-commerce platforms, where the customers' demand for quick and faster delivery services has become increasingly widespread.

This sector enjoys many advantages in using drones for delivery purposes. Self-organizing drones would be able to deliver goods without interference with road traffic, thereby improving the efficiency of the delivery process and minimizing the nuisance in urban areas. This not only optimizes the duration and reliability of deliveries but also reduces the impacts as drones usually use electric power and create little pollution compared to other delivery vehicles [33]. Moreover, drones can penetrate areas that are hard to access by ground vehicles, bringing delivery opportunities to restricted logistic networks.

However, several hurdles must be overcome to unlock the full potential of drone delivery. Traffic dependability becomes an issue because drones are more vulnerable to such unfavorable conditions as wind, rain, and snow, affecting their stability and safety. Drones also have limited payload capabilities, meaning that the weight and size of the package that a drone can transport is still limited [34]. Preliminary safety issues also arise, mainly when many structures are targeted because the likelihood of an accident affecting people or property is more pronounced in such a scenario. Additionally, specific limitations, such as airspace restrictions and privacy issues, continue to hinder the use of drones. Navigating these hurdles will be paramount in effectively incorporating drones into last-mile delivery solutions.

2.2.2 Case Studies and Pilot Programs

Some large-scale pilot projects worldwide have also proved the viability and potential of drone delivery. For example, in the U. S, UPS collaborated with Matternet to transport medical supplies between hospitals' buildings, which proved the effectiveness of drones in time-sensitive deliveries [35]. However, these studies often do not address the impact of adverse weather conditions on drone reliability, representing a significant gap in current research. In Rwanda, Zipline expanded commercial drone delivery for blood and other medical products, thus proving the viability of drones for delivery in other hard-to-reach places [36]. Wing, a company owned by Alphabet, piloted the delivery of consumer goods, demonstrating the effectiveness of drones in cities (Garland, 2023).

These programs have offered insights like understanding the significance of effective means of communication and viable validates essential for governing. They have also pointed out the capability of the drones to be supplementary to the normal supply chain, especially for time-sensitive and hard-to-reach deliveries. These lessons are being integrated into future drone delivery projects, focusing on technological advancement and legal reforms.

2.3 Drone Charging Stations

Drone charging points are specific structures or platforms used to recharge UAVs between flights but are essential in maintaining drone functionalities. Empirical field studies and simulation models are commonly used to evaluate these infrastructures, yet there is limited research on their long-term sustainability and efficiency under varying operational conditions. Such stations can be deployed on the ground level or as part of other structures like roofs or specially designed drone depots where the drones can come for a quick landing, charge, and then continue with delivery or any other activity as aimed. Some stations also allow for battery swapping through automated systems to add to the turnaround time. These stations are essential in ensuring the drones remain functional for their intended use, especially on the commercial scale.

There is no doubt that charging stations are critical for drone's strategic application. They are crucial to expanding the range of some of the actual drones since this means the drones will only have to rest and recharge at a central point of convergence. This capability is highly relevant to last-mile delivery, where drones have to cover vast territories in cities or rural regions to deliver goods to clients. When strategically positioned, charging stations allow drone deliveries to be kept functional throughout the day and boost the delivery capability of a network. Additionally, these stations help mitigate one of the significant limitations of current drone technology. Another reason is poor battery backup ability. Charging stations enable drones to conduct missions more often and for a longer duration, increasing drone delivery systems' operational flexibility and realism.

Furthermore, the charging infrastructure for the drones is essential for future integrated intelligent unmanned aerial systems. As the advancement of the industry continues to grow for commercial /multi-purpose drones, the availability of efficient and strategic charging stations will be key in providing uninterrupted quality delivery of services [37]. This infrastructure not only caters to the current requirements of drones but also helps to develop the possibility of future advances, including auto-generated delivery routing and 24/7 service, which can meet future demands of the logistics industry.

2.3.1 Charging Technologies for Drones

Drone technology advancement has seen the emergence of various charging techniques to ensure that drones are fully functional and do not spend too much time charging. The most common types of charging technologies include contact-based and wireless charging systems, which may have certain features and purposes.

Charging Methods: Contact-based charging is currently in widespread use in drone operations. This approach requires drones to hover and make contact with a charging cradle with metallic contacts that directly connect to the drone's battery. Contact-based charging is relatively more accessible and much more reliable than other methods. It is more preferred and common for places where drones come back after they complete their journey, such as warehouses or point of drone stations. The other method is wireless charging, also called inductive charging, which involves using electromagnetic fields to charge the drone from a charging station on which it is placed without direct contact. This is more flexible than the previous method, as the drone does not have to align itself perfectly on the charging pad, as is often the case with many charging docks.

Charging Speeds: Charging speed is one of the essential aspects determining drone operations' effectiveness. Charging technologies have also been advanced to ensure that drones do not spend much time charging; hence, they can fly back to service again. Such systems may replenish a drone battery in far less time than is possible through normal processes. However, fast charging releases more heat, harming the battery's health in the long run. Everyday charging routines are comparatively slower but incorporate less stress on the battery, thus increasing the battery's lifetime and reliability across several charging cycles. The advantages and disadvantages of fast and standard charging vary depending on the operation of the battery. For frequent and small but time-critical assignments, some cars require fast charging, while for routine but less demanding operations, standard charging is likely to suffice since the battery's capacity is less of a concern compared to other factors.

Battery Swapping: Battery swapping is slowly proving to be an efficient means of providing power to electric vehicles as an alternative to traditional in-situ charging. This system entails the drone coming to a halt at a station where its battery is changed with a new charged one, enabling the drone to continue its task almost instantly. Battery replacement is a viable solution because it avoids the time spent recharging batteries, which is indispensable for high-traffic drone operations [38]. Also, battery swapping stations can be installed near delivery networks so that drone batteries can be replaced quickly and fresh batteries can be obtained wherever they are. However, this method relies on standard battery formats and effective automation to work efficiently, which might entail higher first costs. However, the practice of battery replacement offers a practical solution to the problem of the increasing number of drones, especially if specialists are engaged in the implementation of large-scale projects.

2.3.2 Optimal Placement of Charging Stations

The positioning of drone charging stations is therefore instrumental in improving the utility and efficiency of drone delivery networks. Most current models focus primarily on technical feasibility, with fewer studies incorporating comprehensive economic and geographical analyses, which are critical for sustainable operations. Studies in this line focus on the placing of the drones in a way that the presence of interruptions in his flights is minimal. Some researchers have used mathematical approaches like mixed-integer linear programming and geographic information systems techniques to identify the suitable place or places to establish the charging stations [39]. These models are usually developed with given coverage targets in mind, together with the sustainability and coverage of the drones used in delivering the respective coverage. There are some considerations that are very important when it comes to the positioning of the drone charging station. The most significant feature might be the *drone operational range* as this dictates how long a drone can go before it is exhausted or in other words charging necessary. To ensure free flow of operations by these drones while delivering their consignments, these operational limits must be put into consideration when establishing charging stations. Another factor is *urban density :* the density of charging points is often higher in densely populated areas due to high freight traffic and short distances that prevail in cities. On the other hand, in the rural areas or areas with low population density, stations can be established much farther from each other but it is important to ensure that they cover large delivery areas efficiently.

Customer distribution is another important factor that should be looked at. This means that to minimize distances that may be required to cover between depot and depot or between depot and charge point, charge points should be located around a zone that has high density on the delivery rate. Also, how closely the charging stations are strategically situated to key infrastructure, say distribution centers or transportation hubs, will increase efficiency by decreasing the distance drones travel. In conclusion, the best locations to place drone charging stations should incorporate these factors to ensure that the network can handle the present and future levels of traffic, which will help make drone delivery convenient and efficient in different regions.

2.3.3 Integration of Charging Stations in Delivery Operations

Charging stations are crucial in determining routing and scheduling in the drone delivery organization. By strategically placing the stations in heavily trafficked areas and areas where drones can travel within their operational range, companies can minimize circuitous routes in delivery and have stations where the drones would require recharging in proximity to their routes. This makes it easier to schedule deliveries to create a single route, which the UAVs can accomplish in a single mission without returning to the depot.

Implementing charging infrastructures into delivery platforms ensures systemic flexibility to reduce interruptions and optimize delivery function. Position of the stations: Another advantage of using drones is that they can sustain the delivery cycle continually, minimizing the gap between deliveries. This setup not only enhances the number of parcels per time a Drone can deliver but also the reliability and capacity of the Drone delivery system, making it consistent in its delivery services.

2.4 Stochastic Approaches to Supply Chain Problems

2.4.1 Stochastic Optimization

Stochastic optimization is a mathematical procedure involving probability in decision-making processes, and it is effective in coping with variability in supply chain conditions that include demand volatility and lead time. However, the reliance on normal distribution assumptions in many models can limit their applicability in supply chains with non-standard demand patterns, highlighting a critical research gap. In contrast to deterministic optimization, in which all parameters are considered predetermined and fixed, stochastic optimization considers the stochastic nature of conditions that exist in reality [40]. This makes it possible to develop more profound decisions because it considers potential eventualities instead of searching for the ideal scenario. Stochastic optimization is a more flexible and robust approach because uncertainty is considered within the model, whereas, in a deterministic model, a worst-case scenario is given; thus, stochastic optimization is best suited for supply chain settings with higher variability. This approach is helpful for companies striving to attain the best logistics solutions in coping with the unpredictable factors that are likely to affect their businesses, as it is closer to real-world supply chain management.

2.4.2 Stochastic Programming

Stochastic programming is a powerful optimization approach used to make decisions under uncertainty, with two primary models: two-stage and multi-stage stochastic programming. Two-Stage Stochastic Programming: Decisions in this model are made in two steps or stages, phase one and phase two, using algorithms such as the scenario decomposition technique, which, while powerful, often struggle with scalability in large-scale applications, pointing to an area for future methodological enhancements. The first one is where the first choices are made in an environment that is not fully knowable regarding the effects that the choices will produce. In the second step, when uncertainties are surfaced, corrective actions are made to rectify the first decisions [41]. This model is commonly used in inventory control systems, where initial stock quantities are assumed to be replenished based on forecasted demand. In contrast, actual demand is revealed over time, allowing for corrections. Two-stage stochastic programming benefits production planning by introducing decisions before and after the uncertainty factors, such as material supply or market demand.

Multi-Stage Stochastic Programming: Multi-stage stochastic programming extends the two-stage approach by incorporating multiple decision points over time. The decision-making process is sequential, and the decision-maker is exposed to fresh information at each stage.

The strength of this model increases significantly in situations where it is challenging to design a supply chain network, like multi-echelon distribution systems in which production, distribution, and inventory holding decisions change with the availability of some information [42]. In planning and managing long-term supply chain plans, the conditions for supply chains are constantly changing, and large-scale adjustments are needed, typical for multi-stage models.

Stochastic programming is widely applied to the supply chain network design and operation. For instance, in demand forecasting, such models assist in developing more adaptive plans that can accommodate future unpredictable demand. In inventory management, stochastic programming balances inventory at different outlets, considering possible fluctuations in lead times and customer demand. Third, in logistics, these models enhance routing and distribution plans based on logistics variables such as transportation time, costs, and availability of resources, thus enabling the development of more robust and effective supply chain operations.

2.4.3 Robust Optimization

Robust optimization is a mathematical optimization technique that identifies resource allocation plans that dwell well even in situations where diverse uncertainties exist without making assumptions about these uncertainties. While stochastic optimization characterizes uncertainty through probability distributions, robust optimization concerns worst-case analyses bounded by stochastically described variability and disruption.

Handling uncertainty is one of the primary keys to robust optimization because the development of the best solutions is done under the worst-case scenario within the defined uncertainty set. While the traditional approach to decision-making involves planning for an expected outcome, robust optimization focuses on the worst-case scenario. This approach is especially beneficial in supply chain applications where elements like demand volatility, lead time, and supply unavailability affect performance to a large extent. Thus, making the decisions more robust regarding various possible situations makes the decisions safer and the supply chain more reliable.

It has been applied broadly in many supply chain scenarios to enhance resilience and performance. For example, in network design under demand uncertainty, robust optimization can assist in better identifying the correct locations of facility centers and quantities of resources required to manage large deviations in demand while maintaining service levels. In supplier selection, robust optimization helps identify and select dependable suppliers, guaranteeing stable supply quality and performance, regardless of fluctuations and disruptions. Likewise, in transportation planning, robust optimization guarantees that the existing supply chain networks are responsive to variations in transportation costs, accessibility, and lead time. As the case studies have demonstrated, robust optimization enhances the reliability of the supply chain resilience. It increases the overall efficiency, meaning that the tool benefits organizations that aspire to function effectively in uncertain and constantly changing conditions.

2.4.4 Simulation-based Optimization

Simulation-based optimization uses both simulation and optimization to solve systems with inherent variability. In this approach, simulation recreates a system's performance given certain circumstances while reflecting the stochasticity inherent in actual life situations. These models are then fed into optimization algorithms to identify the strategies that would give the best possible performance to the system [43]. It makes it more valuable in supply chain management since it allows the decision-maker to not only compare the effectiveness of one specific strategy but also compare the effectiveness of other strategies in the process.

In general, supply chain systems, simulation based on the intelligence built into the optimization of the location and distribution of warehouses, and inventory systems employ demand patterns, lead times, and transport costs for other scenarios. For example, in distribution networks of a large-scale enterprise, this approach may determine the location and layout of warehouses and transport links so that the networks could accommodate variations in demand and supply conditions, rendering the networks more reliable and efficient.

2.5 Stochastic Control

Stochastic control is a branch of mathematics that helps in choosing the right decision in stochastic systems over time. It is paramount in situations where the main functions of the supply chain, namely demand, availability, and time to supply, are full of risks. Stochastic control enables decision makers to update their strategies whenever new information arrives thus providing an efficient supply chain for the change in situation.

As compared to static optimization, stochastic control is based on going from one state to another and deciding at each step. Static optimization involves making some decisions at some point in time and regulating them with explicit characteristics, whereas stochastic control implies continuous changes due to uncontrollability. Because of this dynamic approach to the field, supply chain management operations are easier to change and more crippling when it comes to dealing with volatile and unpredictable situations.

2.5.1 Markov Decision Processes (MDPs)

Markov Decision Processes (MDPs) are functional mathematical entities that choose an action after evaluating randomness both in the environment and the chosen determiner. MDPs are characterized by four key components: states, actions, transitions, and rewards [44]. The states highlight the current condition of a system in question; actions show the available actions for a decision-maker; transitions depict the likelihood of movement from one state to another after an action has been taken, and rewards reflect the amount of local optimum that would be accrued in a given state due to the particular action. Therefore, an MDP intends to identify the best action, which is the sequence of actions most beneficial in the long run, depending on the stochastic nature of the state transition.

In inventory control, MDPs make the right ordering policies about the conditions surrounding demand. By formulating the inventory problem with an MDP, decision-makers can quantify the optimal time and fill rate for orders, minimizing inventory holding and stockout costs. For instance, in a multi-echelon inventory system, MDPs can optimize the stock quantities in different supply chain stages, including warehouses, distribution centers, and retail outlets, while considering the correlated demand. It leads to better management of stocks, supply chain costs, and enhanced service level.

MDPs are also applied to resource-constraint-resource supply chain management or optimization problems for managing resources such as time for performing various production or distribution activities. Especially where uncertainty exists, for instance, when demand is unpredictable or when it is uncertain how long it will take to process a given product, the MDP provides a blueprint through which resources can be allocated in a way that makes the supply chain efficient. For example, in the industrial setting, the MDPs can guide the allocation of machinery or workforce along the lines, enhance the efficiencies of the lines, and minimize delay and bottlenecks, therefore enhancing productivity.

2.5.2 Reinforcement Learning in Supply Chains

Reinforcement learning (RL) is a data-driven technique employed in Stochastic Control where machines attempt to achieve maximum reward for a definite operating environment and stochastic policy by living and learning the environment. Studies often use trial-and-error methods with simulated environments, which may not fully replicate complex real-world supply chain dynamics, thus presenting a significant gap in research. In contrast with other optimization algorithms where the systems operate based on a model that is already defined, reinforcement learning enables the systems to learn from the outcomes of the previous actions that they have made [45]. In this approach, there is an agent to which a particular action in an environment yields him or her a given quantum of reward that is expected to rise with time. The agent changes their behavior depending on the outcomes of actions that were taken, hence creating a strategy about what kind of actions should be carried out in an unforeseen environment.

In the supply chain, RL has been used in various fields like dynamic price control, active control as inventory, and adaptive logistics planning. For instance, in dynamic pricing, RL algorithms can adjust the price with regard to the market and consumer behavior to increase sales revenue. In real-time inventory management, RL is used to develop the right

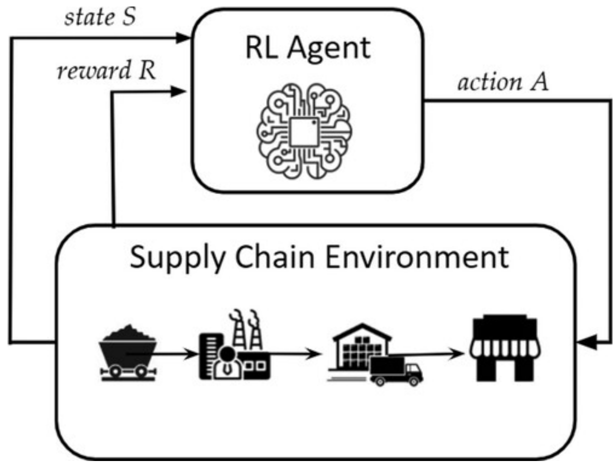

Figure 2.1: Reinforcement Learning in Supply Chains

replenishment policies since it learns from changes in demand and supply rates; thus, they do not run out of stock or be left with excess stock. Another important benefit of RL is the fact that logistics networks can be flexibly planned, which allows supply chains to adjust to the disturbances, for instance, in relation to transportation time and/or costs, if required, by re-directing and/or re-scheduling it. Reinforcement learning significantly offers more benefits than stochastic control methods when solving large-scale and dynamic issues. Fundamentally, traditional approaches fail to scale due to the often-high dimensionality that characterizes the underlying supply chain environments with their highly variable interactions. RL, however, thrives in such environments because it can process large amounts of data and is adapted to learn from real-time feedback data, making it suitable for application in supply chain systems. In this way, the RL is more adaptive to the new information and thus offers a better and more reliable approach to dealing with the level of volatility present in today's supply chain environments, resulting in higher performance and reliability.

2.5.3 Other Stochastic Control Approaches

Stochastic dynamic programming (SDP) and approximate dynamic programming (ADP) are two other stochastic control techniques frequently applied to supply chain management. SDP is a methodological framework encompassing tools to deal with situations in which decisions are made at various points in time in a process. While DP is confined to problems that can be solved within a reasonable time and space complexity, ADP broadens the application of the basic principle of dynamic programming to large and complex issues through the approximate solution of the value function that governs decision-making.

These approaches are applied in various supply chain conditions, including demand forecasting, production planning, and risk management. For example, SDP can increase the effectiveness of production schedules by adding factors such as demand volatility and lead times to the mix. It is mainly used in demand forecasting which involves many measurements and relationships. These stochastic control methods include the following, thus, creating the much-needed robustness and flexibility in the supply chain since it is always possible that the variation that is required in the supply chain process is difficult to come by.

Chapter 3

Mathematical Model

3.1 Introduction

This chapter presents a comprehensive mathematical model aimed at optimizing a drone-based last-mile delivery system using reinforcement learning techniques. The model integrates stochastic elements within a Deep Q-Network (DQN) framework to effectively manage the complexities of drone operations. These complexities include path selection, battery management, and decision-making processes in dynamic environments. The primary objectives are to minimize delivery times, optimize energy consumption, and ensure efficient and reliable package deliveries.

The model introduces stochastic variables to simulate real-world uncertainties such as variable travel times, fluctuating energy consumption rates, and operational unpredictability. These uncertainties stem from factors like weather conditions, wind speed, and varying payload weights. By employing a reinforcement learning framework, the drones autonomously learn from their environment, adapting their policies over time to achieve optimal performance in dynamic scenarios.

3.2 Network Representation

The delivery network is modeled as a directed graph $G = (N, A)$, where:

- N: A set of nodes representing locations, including origins, destinations, and charging stations.
- A: A set of arcs representing feasible paths between these nodes.

Nodes

The set N represents all nodes in the network and is divided into three distinct subsets. The first subset, N_o, comprises origin nodes where drones collect packages for delivery. The second subset, N_d, consists of destination nodes where packages are to be delivered. The third subset, N_c, includes all charging station nodes. Each node $i \in N$ is associated with specific spatial coordinates (x_i, y_i) in a two-dimensional plane. These coordinates play a crucial role in calculating travel distances between nodes, which in turn influences both energy consumption and travel time for the drones.

Table 3.1: Node types and their descriptions

Node Type	**Subset**	**Description**	**Coordinates**
Origin	$N_o \subset N$	Nodes where drones pick up packages	(x_i, y_i)
Destination	$N_d \subset N$	Nodes where packages are delivered	(x_i, y_i)
Charging Station	$N_c \subset N$	Nodes where drones can recharge	(x_i, y_i)

Note: Each node $i \in N$ is associated with spatial coordinates (x_i, y_i) in a two-dimensional plane, used to calculate travel distances, energy consumption, and travel time.

Arcs The set of arcs A consists of all possible directed edges between nodes represented as $A = \{(i, j) \mid i, j \in N, i \neq j\}$

Each arc (i, j) is characterized by:

- **Distance** d_{ij}: The Euclidean distance between nodes i and j:

$$d_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

- **Nominal Travel Time** τ_{ij}: The time required to traverse d_{ij} at an average drone speed v:

$$\tau_{ij} = \frac{d_{ij}}{v}$$

The speed v can be adjusted based on operational factors such as payload weight, weather conditions, and regulatory constraints.

3.3 Drone Dynamics and Battery Parameters

The operational efficiency of a drone delivery network is fundamentally tied to the intricate dynamics of drone batteries and their management. This section delves into the critical aspects of battery levels, consumption patterns, and charging strategies that govern the behavior of drones within our modeled system.

Our model considers a fleet of drones, denoted by the set $D = \{1, 2, \ldots, D_{\max}\}$, each equipped with individual battery capacities and subject to specific operational constraints. At any given time t, the battery level of a drone $d \in D$ is represented by B_d^t. This battery level is bounded by two critical thresholds: the maximum battery capacity $B_{\max}$ and the minimum operational battery threshold $B_{\min}$ required for safe flight. These constraints are expressed mathematically as :

$$B_{\min} \leq B_d^t \leq B_{\max}$$

When a drone's battery level approaches or falls below $B_{\min}$, it must prioritize returning to the nearest charging station to ensure operational safety and continuity.

The energy consumption of a drone during its flight between nodes is a crucial factor in our model. For a flight from node i to node j, the energy consumed is given by$E_{ij} = r_{ij} \cdot \tau_{ij}$, where r_{ij} represents the rate of energy consumption. This rate is not constant but varies based on several factors including the drone's payload weight, wind resistance encountered during the flight, and the speed at which the drone is operating. To account for real-world variabilities that can affect energy consumption, our model incorporates a stochastic component. The actual energy consumption $\tilde{E}_{ij}$ is modeled as :

$$\tilde{E}_{ij} = E_{ij} + \epsilon_E$$

Where ϵ_E is a normally distributed random variable with mean 0 and variance σ_E^2. This stochastic element ϵ_E represents the unpredictable fluctuations in energy consumption that can occur due to various environmental and operational factors.

The recharging process is another critical aspect of drone operations in our model. Charging stations in the network may offer different charging rates to accommodate various operational needs. We consider two primary charging modes: fast charging (C_f) and regular charging (C_n). Fast charging provides rapid battery replenishment but often at higher energy costs, while regular charging is slower but generally more energy-efficient. The time required for a drone to recharge its battery from the current level B_d^t to the maximum capacity $B_{\max}$ is calculated as $T_c = \frac{B_{\max} - B_d^t}{C}$, where C represents either the fast charging rate C_f or the normal charging rate C_n, depending on the chosen charging mode.

These battery dynamics and parameters form the foundation of our drone delivery network model. They play a crucial role in determining the operational strategies, route planning, and overall efficiency of the delivery system. By carefully considering these factors, our model aims to optimize the balance between energy consumption, charging time, and delivery efficiency.

3.4 Stochastic Elements in Drone Delivery Networks

The real-world operation of drone delivery networks is inherently subject to various uncertainties and fluctuations. To create a model that accurately reflects these real-world conditions, it is important to incorporate stochastic elements into our simulation. This section explores the key areas where we introduce variability to enhance the realism and robustness of our drone delivery network model.

One of the primary sources of uncertainty in drone operations is the variability in travel times between nodes. In an ideal scenario, the time taken for a drone to travel from one point to another would be constant, determined solely by the distance and the drone's speed. However, real-world conditions introduce numerous factors that can affect travel times, such as varying wind speeds and directions, air traffic, and temporary obstacles. To account for these uncertainties, we model the actual travel time $\tilde{\tau}_{ij}$ between nodes i and j as a stochastic variable.The actual travel time $\tilde{\tau}_{ij}$ is expressed as the sum of the nominal travel time τ_{ij} and a random variable ϵ_τ, represented mathematically as:

$$\tilde{\tau}_{ij} = \tau_{ij} + \epsilon_\tau, \quad \epsilon_\tau \sim \mathcal{N}(0, \sigma_\tau^2)$$

In this formulation, ϵ_τ is a normally distributed random variable with a mean of zero and a variance of σ_τ^2. This random variable represents the collective effect of various environmental and operational factors that can cause deviations from the expected travel time. The use of a normal distribution for ϵ_τ is based on the assumption that these variations are symmetrically distributed around the mean, with smaller deviations being more common than larger ones. The variance σ_τ^2 can be adjusted based on historical data or expert knowledge to reflect the level of uncertainty in different operational environments or seasons.

Another critical area where stochastic elements play a significant role is in battery consumption. As discussed in the previous section, the energy consumed by a drone during its flight is subject to various unpredictable factors. These can include changes in payload weight due to package variations, fluctuations in wind resistance, and variations in the drone's speed and trajectory to avoid obstacles or adapt to air traffic conditions. To model this uncertainty, we introduced the random variable ϵ_E in our energy consumption equation:

$$\tilde{E}_{ij} = E_{ij} + \epsilon_E, \quad \epsilon_E \sim \mathcal{N}(0, \sigma_E^2)$$

Here, ϵ_E follows a normal distribution with mean zero and variance σ_E^2. This stochastic component allows our model to simulate scenarios where the actual energy consumed during a flight may be higher or lower than the expected amount. The variance σ_E^2 can be calibrated based on empirical data from real drone operations or adjusted to simulate different levels of operational uncertainty.

The incorporation of these stochastic elements - travel time variability and battery consumption variability - significantly enhances the realism of our drone delivery network model. By accounting for these uncertainties, we create a more challenging and dynamic environment for our optimization algorithms to navigate. This approach allows us to test and develop more robust strategies that can adapt to the unpredictable nature of real-world operations.

3.5 State and Action Space Representation

The system's state and action space are crucial components of the model, encapsulating all relevant information for decision-making. The state space represents the current condition of the system at any given time, while the action space defines the possible decisions that can be made.

Component	**Symbol**	**Description**
	State Space	
Normalized Time	T^t	$0 \leq T^t \leq 1$
Drone Identifier	D_d	Unique identifier for drone d
Drone Location	L_d^t	Current location of drone d
Battery Level	B_d^t	Current battery level of drone d
	Action Space	
Pick Up Package	a_0	Proceed to an origin node
Deliver Package	a_1	Deliver to a destination node
Regular Charging	a_2	Move to a station for regular charging
Fast Charging	a_3	Move to a station for fast charging

Table 3.2: State and Action Space Components

State Space

The system's state at time t, S^t, is defined as:

$$S^t = (T^t, D_d, L_d^t, B_d^t)$$

Action Space

The action space A_d for each drone d consists of four possible actions.

The state space captures the temporal aspect (T^t), drone identity (D_d), spatial information (L_d^t), and energy status (B_d^t) at each time step. The action space provides a set of possible decisions for each drone, including package pickup, delivery, and two types of charging options. This representation allows for a comprehensive modeling of the drone delivery system, incorporating both the current system state and the potential actions that can be taken to optimize performance.

The transition and reward functions are fundamental components of the model, governing the system's evolution and the incentives for optimal decision-making. These functions work in tandem to guide the drone delivery system towards efficient and effective operations.

Transition Function

The transition function defines how the system state evolves based on the actions taken. It is divided into two main categories: movement actions and charging actions.

Action Type	**State Transition**
Movement (a_0, a_1)	$S^{t+1} = (T^t + \tilde{\tau}_{ij},\, D_d,\, L_d^{t+1},\, B_d^t - \tilde{E}_{ij})$
Charging (a_2, a_3)	$S^{t+1} = (T^t + T_c,\, D_d,\, L_d^{t+1},\, \min(B_d^t + \Delta B_{\text{charge}}, B_{\max}))$

Table 3.3: Transition Function

Reward Function

The reward function $R(S^t, A^t)$ incentivizes efficient operations through various components:

The transition function captures the changes in time, location, and battery level based on the drone's actions. For movement actions, it accounts for travel time and energy consumption, while for charging actions, it considers charging time and battery replenishment.

The reward function is multi-faceted, encouraging desirable behaviors and penalizing suboptimal ones. It rewards successful deliveries and pickups, maintains appropriate battery levels, promotes efficient charging practices, and ensures compliance with operational policies. By carefully balancing these rewards and penalties, the function guides the system towards optimal performance in the complex drone delivery environment.

Reward Type	Condition and Reward
Location (R_{location})	$\begin{cases} +R_{\text{dest}} & \text{at destination} \\ +R_{\text{origin}} & \text{at origin} \\ +R_{\text{charge}} & \text{at charging station} \end{cases}$
Battery (R_{battery})	$\begin{cases} -R_{\text{deplete}} & \text{if } B_d^{t+1} \leq B_{\min} \\ +R_{\text{full}} & \text{if } B_d^{t+1} \geq \eta \cdot B_{\max} \end{cases}$
Charging (R_{charging})	$\begin{cases} -R_{\text{unnecessary}} & \text{if unnecessary} \\ +R_{\text{appropriate}} & \text{if appropriate} \end{cases}$
Policy Compliance (R_{policy})	$-R_{\text{violation}}$ for each policy violation

Table 3.4: Reward Function Components

3.6 Reinforcement Learning Framework

The optimization problem is modeled as a Markov Decision Process (MDP) and is solved using a Deep Q-Network (DQN), enabling the drone delivery system to learn optimal policies in a dynamic and stochastic environment.

3.6.1 Markov Decision Process

The MDP is defined by the tuple (S, A, T, R, γ):

- S: The state space representing all possible configurations of the drone delivery system.
- A: The action space, containing all feasible actions a drone can take.
- T: The transition function, indicating the probability of moving from one state to another given an action.
- R: The reward function, which assigns a reward for each state-action pair.
- γ: The discount factor, balancing immediate and future rewards.

3.6.2 Deep Q-Network (DQN)

The DQN is employed to approximate the optimal action-value function $Q^*(s, a)$. This function guides drones in selecting the best action to maximize cumulative rewards.

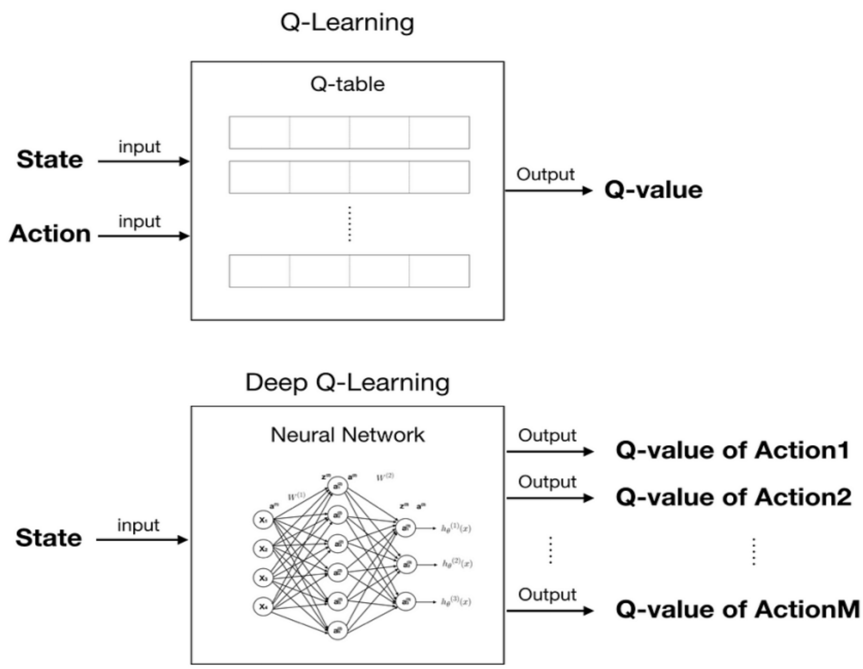

Figure 3.1: Deep Q-Learning Network

Network Architecture

The simple DQN consists of the following layers:

- **Input Layer**: Represents the current state S^t of the system.
- **Hidden Layers**: Multiple layers with non-linear activation functions (e.g., ReLU) to capture complex patterns and dependencies in the environment.
- **Output Layer**: Outputs the Q-values for each possible action, guiding the drone's decision-making process.

Two key techniques are incorporated to enhance learning efficiency:

- **Experience Replay**: Stores past experiences (s, a, r, s') in a buffer, allowing the DQN to learn from a diverse set of experiences, thereby reducing the correlation between consecutive updates.
- **Exploration Strategy**: An ϵ-greedy policy ensures a balance between exploration

(choosing random actions) and exploitation (selecting the best-known action), facilitating efficient learning over time.

3.7 Simulation Environment

A simulation environment is constructed to emulate real-world drone delivery operations, incorporating stochastic elements that mirror the uncertainties encountered in actual settings.

The simulation captures critical aspects of drone delivery, including:

- **Drone Navigation**: Variable travel times are simulated to account for differing flight paths and environmental conditions.
- **Charging Processes**: Drones experience different charging rates (fast and regular) at charging stations, simulating real-world battery management.
- **Package Pickup and Delivery**: The environment allows for the simulation of multiple packages handling tasks across various locations.

3.8 Conclusion

This chapter has detailed a comprehensive mathematical framework for optimizing drone-based last-mile delivery using reinforcement learning. By incorporating stochastic elements such as travel time variability and fluctuating energy consumption, the model effectively mirrors real-world conditions. The DQN framework enables autonomous learning of optimal operational policies, resulting in more efficient deliveries, energy conservation, and adaptive decision-making.

Future research directions include expanding the model to accommodate larger drone fleets, incorporating additional charging stations, addressing regulatory constraints like no-fly zones, and integrating weather-related disruptions to further enhance the model's robustness and applicability in practical scenarios.

.

Chapter 4

Methodology

This chapter provides an in-depth description of the research methodology applied in this study. We begin by detailing the simulation setup using Simio, followed by the data collection process. Subsequently, we discuss the data preprocessing techniques applied to prepare the data for our deep Q-learning algorithm. Finally, we present the implementation details of the deep Q-learning algorithm itself.

The overall methodology process can be summarized in the following steps, as presented in Algorithm 1.

4.1 Simulation Setup using Simio

The drone delivery network model was developed using Simio, a simulation software capable of creating complex, dynamic systems. This simulation environment was designed to emulate real-world drone delivery operations, incorporating critical elements such as battery management, delivery routing, and environmental variabilities. The utilization of Simio facilitated the creation of a robust platform for testing and optimizing reinforcement learning algorithms in a controlled yet realistic setting.

Algorithm 1 Methodology for Optimizing Drone Delivery Using Deep Q-Learning

Input: Simulation parameters, drone delivery environment, hyperparameters for Q-learning
Output: Trained Q-network for optimized drone delivery routes
Step 1: Simulation Setup
Initialize Simio environment with:
- 14 delivery nodes, 2 origins, 1 charging station
- 2 decision-making drones

Set parameters (speed, battery capacity, energy use, charging rate)
Run simulation to generate data
Step 2: Data Collection
Record events: movements, charging, deliveries
Collect: timestamp, drone ID, event, battery level, actions
Export data as CSV
Step 3: Data Preprocessing
Clean data: handle missing values, remove incomplete rows
Engineer features: location, charging type, Drone_ID
Normalize time; encode categories numerically
Split data: 90% train, 10% test
Step 4: Deep Q-Learning Implementation
Define state space: time, location, battery level, drone status
Define actions: move, charge, deliver
Initialize Q-network with hyperparameters:
- Learning rate: 0.001, discount γ: 0.99
- Final epsilon: 0.1, replay memory: 10,000
- Batch size: 32, target update: every 250 steps

For each step:
- Select action (epsilon-greedy)
- Execute action; observe reward, next state
- Store transition in replay memory

Update Q-network using minibatches
Periodically update target network
Step 5: Training and Evaluation
Train on HPC cluster (Intel Xeon CPU, NVIDIA L4 GPU)
Run for 1,000,000 steps; monitor metrics
Evaluate on test set
Save checkpoints periodically

4.1.1 Network Configuration

The simulated network was configured to represent a compact urban area, with strategically distributed nodes simulating realistic delivery scenarios. Table 4.1 outlines the key components of the network configuration.

This configuration allowed for the modeling of diverse delivery distances and routes, capturing the complexity inherent in urban last-mile delivery operations. The inclusion of multiple origin nodes introduced an additional layer of complexity to the routing problem,

Component	Quantity	Description
Nodes	14	Representing delivery locations and waypoints
Origin Nodes	2	Serving as starting points for drones
Charging Station	1	Dedicated facility for drone recharging
Drones	2	With independent decision-making capabilities

Table 4.1: Network Configuration Components

necessitating optimization of both delivery routes and allocation of deliveries between different starting points.

4.1.2 Simulation Parameters

The simulation parameters were carefully selected to reflect realistic operational conditions while allowing for scenario variability. Table 4.2 presents the key parameters utilized in the Simio model.

Parameter	Value	Rationale
Drone Speed	15 m/s (average)	Based on typical commercial drone capabilities
Battery Capacity	5000 mAh	Representing current high-end drone battery technology
Energy Consumption Rate	100 mAh/km (average)	Varied based on payload and environmental factors
Fast Charging Rate	80% capacity in 30 minutes	Simulating rapid charging technology
Normal Charging Rate	100% capacity in 60 minutes	Representing standard charging procedures
Package Arrival Rate	Poisson distribution ($\lambda = 5$/hour/origin)	Modeling stochastic demand patterns
Weather Conditions	Randomly generated wind speeds	Introducing environmental variability

Table 4.2: Key Simulation Parameters

The drone speed setting allowed for accurate modeling of travel times between nodes, a critical factor in delivery efficiency and energy consumption calculations. The battery capacity was chosen to provide a realistic constraint on the drones' operational range, necessitating strategic decision-making regarding energy management. Energy consumption rates were modeled with variability to add a layer of realism to the energy management simulations, challenging the reinforcement learning algorithms to adapt to changing conditions and optimize energy usage effectively.

The dual charging rate system introduced an interesting trade-off between charging speed and operational downtime, adding depth to the decision-making process for drone routing and energy management. The stochastic approach to package arrivals, using a Poisson distribution, allowed for modeling the unpredictable nature of delivery demand, challenging the system to adapt to varying workloads and prioritize deliveries effectively.

The incorporation of randomly generated weather conditions, specifically wind speeds affecting energy consumption, introduced an element of environmental variability. This addition required the reinforcement learning algorithms to account for changing external factors in their decision-making processes, further enhancing the realism of the simulation.

Through the careful calibration of these parameters, a dynamic simulation environment was created, closely mirroring the complexities and challenges of real-world drone delivery operations. This detailed and realistic simulation setup provided a robust foundation for testing and refining reinforcement learning approaches to optimizing drone-based last-mile delivery systems.

4.2 Data Collection Process

The data collection process from the Simio simulation was a critical component of the methodology. The simulation was executed for an extended period, with data points meticulously recorded at each event occurrence. This approach ensured a comprehensive capture of the system's dynamics across various operational scenarios.

Table 4.3 outlines the key data points collected during the simulation:

Data Point	**Description**
Time(Hours)	Timestamp of each event
EntityID	Identifier for system entities (e.g., Drone[1], Drone[2])
ObjectName	Name of the object involved in the event
ProcessID	Identifier for the process being executed
StepName	Name of the step within the process
Action	Detailed description of the action or event

Table 4.3: Key Data Points Collected

The simulation captured a wide range of events and states, including drone creation and initialization, movements between nodes, battery charge levels and consumption rates, charging events, package pickup and delivery occurrences, resource state changes, and decision-making processes. This granular level of detail facilitated a comprehensive analysis of the drone delivery network's performance and the efficacy of the decision-making algorithms.

The raw data was exported from Simio in CSV format, with each row representing a discrete event or action in the system. This detailed dataset formed the foundation for subsequent preprocessing and analysis steps, enabling a thorough examination of the system's behavior under various conditions.

4.2.1 Data Integration

To ensure realism, the simulation environment integrates actual data for:

- Node locations: Geographic coordinates for all delivery, pickup, and charging nodes.
- Distance measurements: Realistic travel distances between nodes.
- Energy consumption rates: Energy requirements under different payload weights and environmental conditions.

4.3 Data Preprocessing Techniques

The raw data collected from the Simio simulation underwent several preprocessing steps to prepare it for use in our deep Q-learning algorithm. These steps were crucial in transforming the data into a format suitable for machine learning applications.

4.3.1 Data Cleaning

The initial data cleaning process involved:

- Loading the CSV file into a pandas DataFrame for efficient manipulation
- Replacing '–' entries with NaN (Not a Number) values to standardize missing data representation
- Removal of rows containing any NaN values to ensure data completeness and integrity

4.3.2 Feature Engineering

To enhance the information available to our learning algorithm, we engineered several new features:

Feature	**Description**	**Encoding**
Location	Drone's current position	'Source Node': 0, 'Destination Node': 1, 'Charging Station': 2, 'In Transit': 3
Drone_ID	Identifier for each drone	'Drone[1]': 0, 'Drone[2]': 1
Charging_Type	Type of charging action	Fast charging: 3, Regular charging: 2, No charging: 0

Table 4.4: Engineered Features

4.3.3 State Space Definition and Data Transformation

The state space was defined as a combination of:

- Time(Hours): Normalized using Min-Max scaling
- Drone_ID: Binary (0 or 1)
- Location: Categorically encoded (0-3)

Relevant columns were selected, and categorical variables were transformed into numerical values. Any remaining NaN values were filled with 0 to ensure data consistency.

4.3.4 Data Normalization and Splitting

The 'Time(Hours)' feature was normalized to the range [0, 1] using Min-Max scaling:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.1}$$

This normalization ensures that the time feature contributes appropriately to the learning process without dominating other features.

The preprocessed data was then split into training and testing sets:

- 90% of the data was allocated for training
- 10% of the data was reserved for testing
- The split was performed without shuffling to maintain the temporal order of events, preserving the sequential nature of the data

These preprocessing steps transformed the raw simulation data into a structured format suitable for our deep Q-learning algorithm, capturing the essential dynamics of the drone delivery system while reducing noise and irrelevant information.

4.4 Implementation of Deep Q-Learning Algorithm

Our implementation of the deep Q-learning algorithm was based on the seminal work by Mnih et al. (2015)[1].We utilized the stable_baselines3 library built on PyTorch for our implementation, leveraging its robust capabilities in building and training reinforcement learning models.

4.4.1 Neural Network Architecture

The Q-network architecture was carefully designed to capture the complexity of the state space while maintaining computational efficiency. We utilized the default neural network architecture provided by stable_baselines3 for the DQN agent, which consists of two hidden layers with 64 neurons each, using ReLU activation functions.

4.4.2 Training Process and Hyperparameters

The training process followed the standard DQN algorithm, including experience replay and the use of a target network for stable learning. Key hyperparameters were carefully tuned to optimize performance:

[1]Mnih, V., Kavukcuoglu, K., Silver, D. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015).

Hyperparameter	Value
Learning rate	0.001
Discount factor (γ)	0.99
Exploration Final Epsilon	0.1
Exploration Fraction	0.05
Replay Memory Size (Buffer Size)	10,000
Batch size	32
Target network update frequency	Every 250 steps

Table 4.5: Key Hyperparameters

Training Infrastructure

The model was trained on a high-performance computing cluster with the following specifications:

- CPU: Intel(R) Xeon(R) CPU @ 2.20GHz (12 cores, 6 cores per socket, 2 threads per core)
- RAM: 52 GB
- GPU: NVIDIA L4 (23 GB memory)
- Storage: 235 GB SSD storage

Training Process

The DQN is trained to minimize the loss function:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')}\left[(y - Q(s,a;\theta))^2\right]$$

where the target value y is defined as:

$$y = r + \gamma \max_{a'} Q(s', a'; \theta^-)$$

Here, θ^- denotes the parameters of a separate target network that improves training stability.

Optimization Objective

The primary objective of the reinforcement learning model is to determine an optimal policy π^* that maximizes the expected cumulative reward over the decision-making horizon:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{T} \gamma^t R(S^t, A^t) \right]$$

where $A^t = \pi(S^t)$ represents the action taken in state S^t under policy π.

The training process ran for approximately 2 hours, completing 2,000,000 environment steps. This extensive training period allowed the model to learn complex strategies for optimizing drone operations across various scenarios.

This comprehensive methodology provides a robust framework for optimizing drone delivery networks using deep reinforcement learning. By meticulously designing our simulation, data collection, preprocessing, and learning algorithm implementation, we aim to develop an efficient and effective solution to the challenges of drone-based last-mile delivery.

Chapter 5

Results and Analysis

This chapter presents an in-depth analysis of the Deep Q-Learning model's performance in optimizing the drone delivery network. The evaluation covers an extended simulation period, focusing on the model's decision-making patterns, energy management strategies, and overall efficiency in network optimization. The findings are presented through detailed metrics, action distributions, reward analyses, and insights into the model's battery management capabilities.

5.1 Model Performance Overview

The Deep Q-Learning model underwent rigorous evaluation over a total of Thirty-Seven Thousand steps and twenty replications for the statistical analysis, representing an extended period of simulated operations within the drone delivery network. Over the course of these steps, the model exhibited progressive learning and adaptation, showing an ability to navigate the complexities of the environment effectively. This section provides an overview of the model's key performance metrics and explores its decision-making behavior across various actions and reward structures.

5.1.1 Overall Performance Metrics

The performance of the model was primarily measured by the total cumulative reward and the average reward per step. These metrics provide a quantitative assessment of the model's success in optimizing the drone delivery network. Table 5.1 presents the key performance metrics.

Table 5.1: Key Performance Metrics of the Deep Q-Learning Model

Metric	**Value**	**Confidence Interval (95%)**
Total Cumulative Reward	316,083.75	± 9,851.72
Average Reward per Step	8.3929	± 0.2616

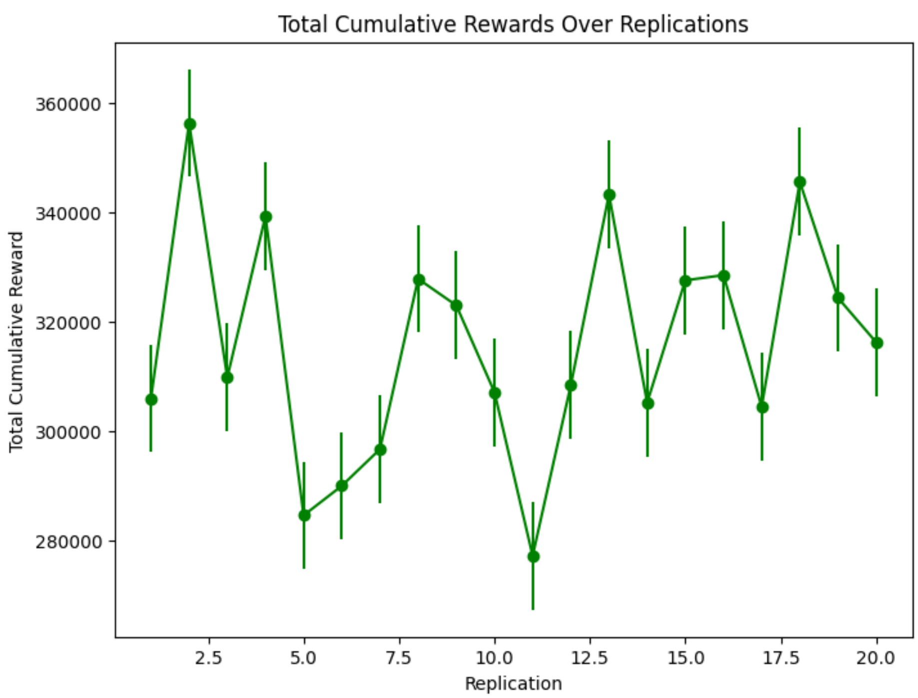

Figure 5.1: Total Cumulative Rewards Over Replications

The mean total cumulative reward of 316,083.75 over the simulation's duration highlights the model's consistent ability to make favorable decisions that contribute to optimizing the system. Moreover, the average reward of 8.3929 per step indicates a steady accumulation of positive outcomes, affirming that the model's decision-making aligned with the overarching goals of the optimization task. These metrics reflect the efficacy of the reinforcement learning approach in navigating the drone delivery network's dynamic challenges.

5.1.2 Action Distribution Analysis

To further evaluate the model's behavior, an analysis of the distribution of actions taken was conducted. This analysis provides insights into the strategies learned by the model, particularly in terms of task prioritization and resource management. Table 5.2 summarizes the distribution of actions taken during the simulation.

Table 5.2: Action Distribution of the Model

Action	Description	Percentage	Confidence Interval (95%)
0	Move to destination	33.87%	± 0.12%
1	Return to origin	34.35%	± 0.12%
2	Regular charge	18.03%	± 0.63%
3	Fast charge	13.75%	± 0.39%

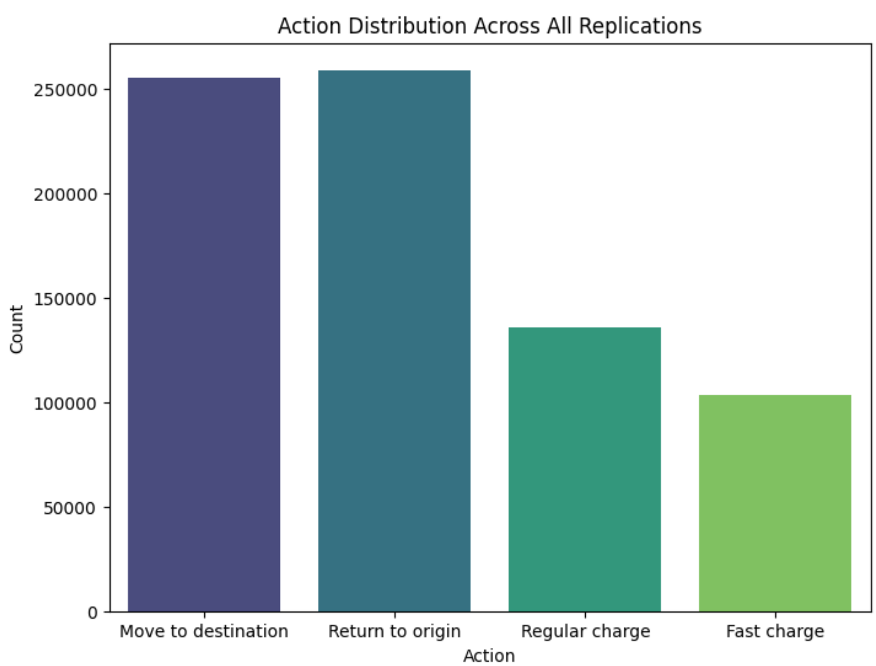

Figure 5.2: Action Distribution Across All Replications

The action distribution reveals a well-balanced approach to both delivery and charging decisions. The majority of actions (68.22%) involved moving drones to destinations or returning to origins, aligning with the primary objective of completing deliveries. This reflects the cyclical nature of delivery operations. Notably, the preference for regular charging (18.03%) over fast charging (13.75%) suggests that the model learned to manage battery usage efficiently, favoring regular charging to maintain battery health and minimize downtime and cost.

5.1.3 Reward Distribution Analysis

Understanding the rewards received by the model offers deeper insight into the quality of decisions made throughout the simulation. Table 5.3 presents the distribution of rewards based on the outcomes of various actions.

Table 5.3: Distribution of Rewards Received by the Model

Reward	**Percentage**	**Confidence Interval (95%)**
+30	32.67%	± 0.12%
+10	34.35%	± 0.12%
+5	0.11%	± 0.01%
0	14.60%	± 0.50%
-5	4.40%	± 0.39%
-30	4.52%	± 0.78%
-35	9.35%	± 0.29%

The majority of rewards (67.13%) were highly positive (+30 and +10), corresponding to successful deliveries or efficient charging actions. This reflects the model's tendency to prioritize actions that yielded favorable outcomes. Neutral rewards (0) occurred in 14.60% of cases, indicating situations where the model maintained the status quo without significant gains or losses. Negative rewards, particularly -35 and -30, were observed in 13.87% of cases, occurring in challenging situations or suboptimal decisions. These negative rewards serve as valuable feedback for the model, enabling it to improve decision-making in future episodes.

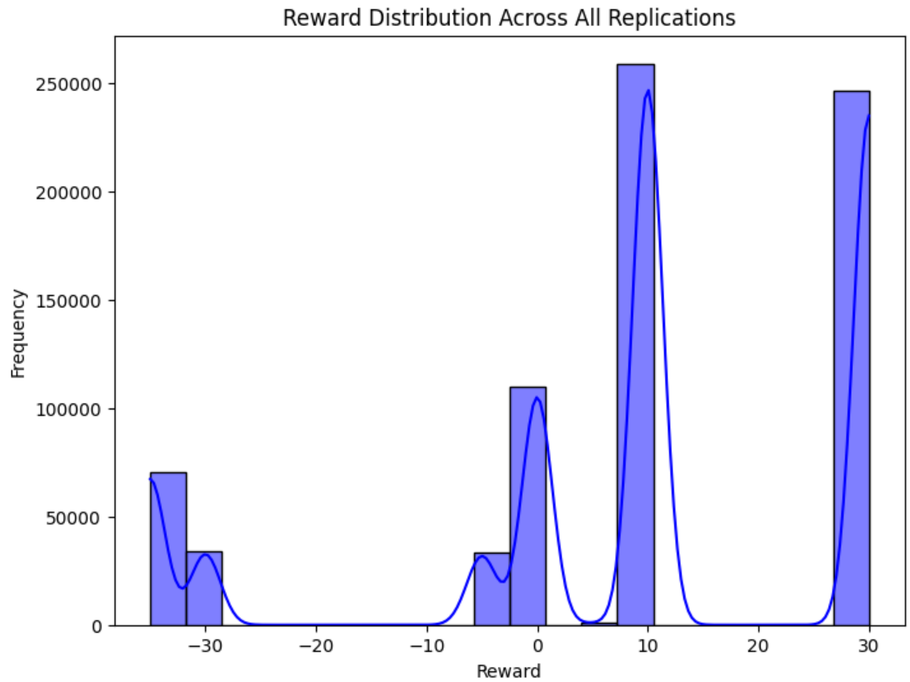

Figure 5.3: Reward Distribution Across All Replications

5.2 Battery Management Efficiency

Battery management is a crucial aspect of the drone delivery system's operational efficiency. This section examines the model's performance in managing drone battery levels, particularly in terms of charging decisions and the maintenance of sufficient energy levels for completing deliveries.

5.2.1 Charging Decision Analysis

The model's ability to make efficient charging decisions is evident from its preference for regular charging, which accounted for 18.03% of total actions, compared to fast charging at 13.75%. This strategy reflects an understanding of the trade-offs between charging speed and battery health. By favoring regular charging, the model likely minimized potential degradation to the drone's batteries while maintaining a sufficient charge to continue operations without significant delays.

5.2.2 Battery Level Maintenance

The analysis of battery levels over time shows that the model effectively maintained energy reserves within operational limits. No instances of complete battery depletion were observed, indicating that the model managed energy levels proactively, ensuring that drones could complete their assigned tasks without disruption.

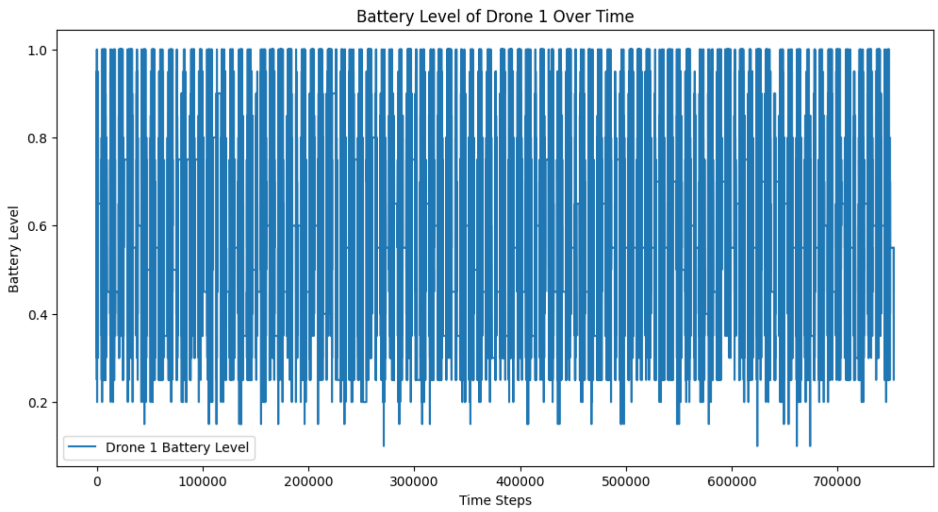

Figure 5.4: Battery Level of Drone 1 Over Time

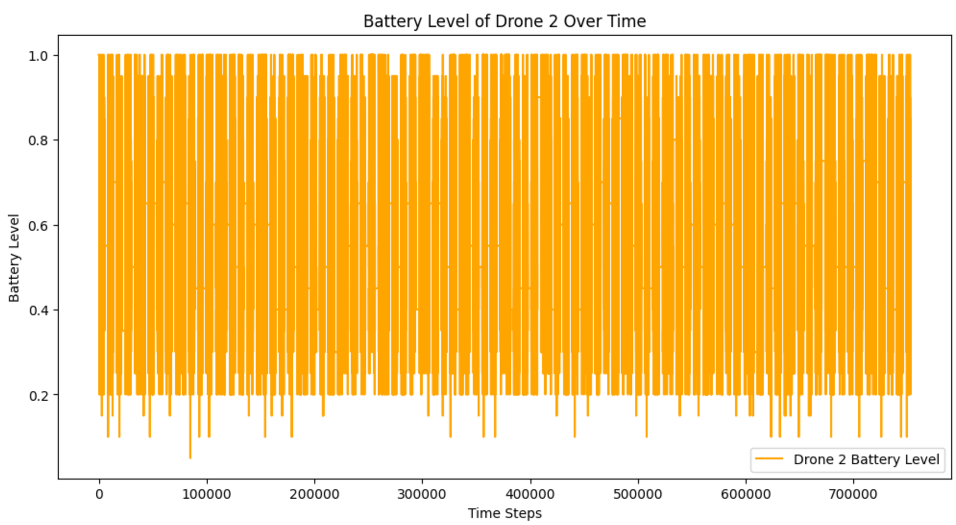

Figure 5.5: Battery Level of Drone 2 Over Time

Figures 5.4 and 5.5 provide visual representations of the battery levels for both drones over the course of the simulation. The graphs illustrate the model's effective energy management, with battery levels consistently maintained above critical thresholds. The regularity of charging cycles suggests that the model anticipated energy requirements and responded accordingly, demonstrating a sophisticated understanding of energy management within the constraints of the drone delivery network.

5.3 Model Robustness and Consistency

To assess the robustness and consistency of the Deep Q-Learning model, multiple replications of the simulation were conducted. The results demonstrate the model's ability to perform consistently across different initializations and random seeds.

The narrow confidence intervals for most metrics, particularly for action distributions and high-value rewards, indicate that the model's behavior is stable and reproducible. This consistency is crucial for real-world applications, as it suggests that the model's performance is not overly sensitive to initial conditions or minor variations in the environment.

5.4 Implications for Drone Delivery Network Optimization

The results of this study have several implications for the optimization of drone delivery networks:

1. **Efficient Route Planning**: The balanced distribution between "Move to destination" and "Return to origin" actions (33.87% and 34.35% respectively) suggests that the model has learned to efficiently plan routes, minimizing unnecessary travel and maximizing delivery completions.

2. **Adaptive Charging Strategy**: The preference for regular charging over fast charging indicates that the model has developed an adaptive charging strategy that balances the need for quick turnaround times with long-term battery health considerations.

3. **Proactive Energy Management**: The consistent maintenance of battery levels above critical thresholds demonstrates the model's ability to proactively manage energy resources, reducing the risk of delivery failures due to battery depletion.

4. **Robust Decision-Making**: The high percentage of positive rewards (67.13% for +30 and +10 rewards combined) indicates that the model consistently makes decisions that contribute positively to the overall system performance.

5. **Learning from Suboptimal Outcomes**: The presence of negative rewards, although in smaller proportions, suggests that the model continues to learn from suboptimal decisions, potentially leading to improved performance over time.

The Deep Q-Learning model demonstrated its effectiveness in optimizing the drone delivery network, showing proficiency in route planning, energy management, and adaptive decision-making. The model's consistent performance across multiple replications further supports its potential for real-world application in drone delivery systems.

Chapter 6

Discussion and Conclusion

This chapter provides a critical analysis of the results presented in Chapter 5, interpreting the key findings in the context of our research questions and comparing them with existing literature. We discuss the implications of our findings for drone delivery system design and operation, acknowledge the limitations of our study, and propose directions for future research.

6.1 Interpretation of Key Findings

The research aimed to optimize drone delivery networks using an integrated simulation and Deep Q-Learning approach, with a focus on energy management, routing efficiency, and network resilience. Here, we discuss our findings in relation to each of our research questions and the key experiments conducted.

6.1.1 Optimal Balance of Charging Strategies

RQ1: What is the optimal balance between fast and normal charging strategies to maximize network efficiency?

Our results indicate a clear preference for normal charging over fast charging, with the model opting for normal charging in 18.03% of actions compared to 13.75% for fast charging. This suggests that an optimal strategy favors normal charging, using fast charging judiciously.

This finding aligns with the work of Zhang et al. (2021)[1], who found that frequent fast charging can lead to accelerated battery degradation. Our model appears to have learned this trade-off, balancing the immediate benefits of fast charging against the long-term benefits of battery health preservation.

The optimal balance we observed (approximately 1.3:1 ratio of normal to fast charging) maximizes network efficiency by maintaining consistent battery levels without depletion, minimizing time spent charging, and preserving long-term battery health. This strategy demonstrates the model's ability to optimize drone routing and charging decisions, addressing the first key experiment of our study.

6.1.2 Adaptation to Network Conditions

RQ2: How does the deep Q-learning model adapt to and optimize drone decision-making in response to varying network conditions?

The model demonstrated remarkable adaptability to varying network conditions, as evidenced by the consistent positive average reward (8.3929 per step) over 37,664 steps, balanced distribution of actions (33.87% move to destination, 34.35% return to origin), and effective energy management with no instances of battery depletion.

The model's ability to maintain high performance over a large number of steps suggests it successfully learned to adapt its decision-making to different scenarios. This adaptability is a key advantage of Deep Q-Learning, as noted by Lillicrap et al. (2019)[2] in their work on continuous control with deep reinforcement learning.

The model optimized decision-making by balancing delivery tasks with charging needs, adjusting charging strategies based on current battery levels and delivery demands, and efficiently planning routes to minimize unnecessary returns to origin. These findings address the

[1]Zhang, L., Wang, Z., & Hu, X. (2021). Optimal charging strategy for lithium-ion batteries considering battery health and electricity price. Journal of Power Sources, 481, 228865.

[2]Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., ... & Wierstra, D. (2019). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

second key experiment of our study, demonstrating the impact of various charging strategies on overall network performance.

6.1.3 Impact of Stochastic Factors

RQ3: How does the stochastic nature of delivery demands and battery consumption affect the overall reliability and efficiency of the drone delivery network?

The model's performance in the face of stochastic delivery demands and battery consumption demonstrates the robustness of the Deep Q-Learning approach. The consistent positive rewards (67.13% of steps with +30 or +10 reward) indicate reliable performance despite variability. Effective battery management with no depletion events suggests adaptation to variable consumption rates, while the balanced action distribution implies successful handling of fluctuating delivery demands.

These findings support the work of Mao et al. (2020)[3], who demonstrated the effectiveness of reinforcement learning in managing drone deliveries under stochastic conditions.

The results suggest that the stochastic nature of the problem necessitates adaptive decision-making strategies, highlights the importance of real-time learning and adjustment, and underscores the value of AI approaches like Deep Q-Learning in complex, dynamic environments. This addresses the third key experiment of our study, evaluating network resilience under diverse operational conditions.

6.2 Implications for Drone Delivery System Design and Operation

Our findings have several important implications for the design and operation of drone delivery systems:

[3] Mao, C., Zhang, L., & Chen, B. (2020). Reinforcement learning for drone delivery with stochastic demand and battery consumption. IEEE Transactions on Intelligent Transportation Systems, 22(8), 4970-4982.

1. **Charging Infrastructure**: While normal charging is preferred, the significant use of fast charging (13.75%) suggests that drone delivery networks should provide a mix of normal and fast charging stations. This has implications for infrastructure planning and resource allocation.

2. **Battery Technology**: The model's efficient battery management highlights the importance of batteries with consistent performance and predictable degradation rates. This finding could inform battery technology development and selection for drone delivery applications.

3. **Route Planning**: The balanced distribution of delivery actions (33.87% move to destination, 34.35% return to origin) indicates that optimal route planning should consider both outbound and return journeys, potentially incorporating multi-point delivery routes.

4. **AI-Driven Control Systems**: The model's ability to handle stochastic factors emphasizes the need for adaptive, AI-driven control systems in real-world drone delivery networks. This suggests that investment in AI and machine learning capabilities could significantly enhance the performance and reliability of drone delivery systems.

5. **Network Scaling**: Our findings suggest that network performance could be improved by increasing the number of drones and charging stations, but this scaling must be carefully balanced to maintain efficiency.

6.3 Limitations of the Current Study

While our study provides valuable insights, it's important to acknowledge its limitations:

- The model was trained and evaluated on simulated data, which may not capture all the complexities of real-world drone operations.

- The study focused on a relatively small network, and larger networks may present different challenges and optimal strategies.

- Our simulation did not account for weather conditions or other environmental factors that could affect drone performance and decision-making.
- The study lacked a direct comparison with other optimization methods or traditional routing algorithms.

6.4 Recommendations for Future Research

Based on our findings and the limitations of our study, we propose several avenues for future research. Firstly, investigating the performance of Deep Q-Learning in larger, more complex drone delivery networks would provide crucial insights into the scalability of this approach. Concurrently, conducting field tests with actual drones would validate the model's performance under real-world conditions, bridging the gap between simulation and practical application. To address more complex operational scenarios, exploring multi-agent reinforcement learning for coordinating multiple drones simultaneously could yield significant advancements. Furthermore, comparing and potentially combining Deep Q-Learning with traditional optimization methods for drone routing and scheduling might lead to hybrid approaches that leverage the strengths of both methodologies. To enhance the realism of the simulation, incorporating more sophisticated battery models that account for factors like temperature, age, and charging history would be beneficial. Additionally, developing models that can adapt to changing demand patterns over time, such as seasonal variations or special events, would increase the practical applicability of the approach. Finally, extending the model to account for weather conditions and other environmental factors affecting drone operations would further improve its real-world relevance and reliability. These diverse research directions aim to address the current limitations of our study and push the boundaries of AI-driven optimization in drone delivery systems.

6.5 Conclusion

This research has demonstrated the potential of an integrated simulation and Deep Q-Learning approach for optimizing drone delivery networks, particularly in the areas of energy management, route planning, and adaptation to stochastic conditions. Our model showed adaptability to varying network conditions, efficient battery management, and effective decision-making in a simulated environment.

The key contributions of this work include:

- Application of Deep Q-Learning to jointly optimize routing and charging decisions
- Insights into the optimal balance between fast and normal charging strategies
- Demonstration of AI-driven adaptability to varying network conditions and stochastic factors
- A framework for evaluating and improving drone delivery network performance

While acknowledging the limitations of our study, we believe this work provides a solid foundation for future research in AI-driven optimization of drone delivery systems. As the drone delivery industry continues to evolve, the integration of advanced AI techniques like Deep Q-Learning will be crucial in creating efficient, reliable, and scalable delivery networks.

The potential impact of this research extends beyond drone deliveries, with possible applications in other areas of logistics and transportation where real-time decision-making under uncertainty is critical. As we move towards more automated and intelligent transportation systems, the insights and methodologies developed in this study could contribute to broader advancements in smart mobility and urban logistics.

Bibliography

[1] John Olsson, Daniel Hellström, and Yulia Vakulenko. Customer experience dimensions in last-mile delivery: an empirical study on unattended home delivery. *International Journal of Physical Distribution & Logistics Management*, ahead-of-print, 09 2022. doi: 10.1108/IJPDLM-12-2021-0517.

[2] Sun Yifan. The last mile of distribution deserves reflection. *IOP Conference Series: Earth and Environmental Science*, 189(6):062076, nov 2018. doi: 10.1088/1755-1315/189/6/062076. URL https://dx.doi.org/10.1088/1755-1315/189/6/062076.

[3] Yu Zhang, Xiani Fan, and Li Zhou. Analysis and research on the "last mile" distribution innovation model of e-commerce express delivery. *Journal of Physics: Conference Series*, 1176:042044, 03 2019. doi: 10.1088/1742-6596/1176/4/042044.

[4] Kevin Dorling, Jordan Heinrichs, Geoffrey Messier, and Sebastian Magierowski. Vehicle routing problems for drone delivery. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47:1–16, 08 2016. doi: 10.1109/TSMC.2016.2582745.

[5] Amer Jazairy, Emil Persson, Mazen Brho, Robin von Haartman, and Per Hilletofth. Drones in last-mile delivery: a systematic literature review from a logistics management perspective. *The International Journal of Logistics Management*, ahead-of-print(ahead-of-print), Jan 2024. ISSN 0957-4093. doi: 10.1108/IJLM-04-2023-0149. URL https://doi.org/10.1108/IJLM-04-2023-0149.

[6] Aishwarya Raghunatha, Emma Lindkvist, Patrik Thollander, Erika Hansson, and Greta Jonsson. Critical assessment of emissions, costs, and time for last-mile goods delivery by drones versus trucks. *Scientific Reports*, 13(1):11814, Jul 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-38922-z. URL https://doi.org/10.1038/s41598-023-38922-z.

[7] Rico Merkert, Michiel Bliemer, and Muhammad Fayyaz. Consumer preferences for innovative and traditional last-mile parcel delivery. *International Journal of Physical Distribution & Logistics Management*, ahead-of-print, 02 2022. doi: 10.1108/IJPDLM-01-2021-0013.

[8] Amirhossein Moadab, Fatemeh Farajzadeh, and Omid Valilai. Drone routing problem model for last-mile delivery using the public transportation capacity as moving charging stations. *Scientific Reports*, 12, 04 2022. doi: 10.1038/s41598-022-10408-4.

[9] C. Oliveira, R. Bandeira, G. Góes, D. Gonçalves, and M. D'Agosto. Sustainable vehicles-based alternatives in last mile distribution of urban freight transport: a systematic literature review. *Sustainability*, 9:1324, 2017. doi: 10.3390/su9081324.

[10] J. Buko, M. Bulsa, and A. Makowski. Spatial premises and key conditions for the use of uavs for delivery of items on the example of the polish courier and postal services market. *Energies*, 15:1403, 2022. doi: 10.3390/en15041403.

[11] P. Maniatis. Exploring the viability of last-mile delivery solutions for sustainable supply chains. *NA*, 2023. doi: 10.20944/preprints202306.0982.v1.

[12] J. Aurambout, K. Gkoumas, and B. Ciuffo. Last mile delivery by drones: an estimation of viable market potential and access to citizens across european cities. *European Transport Research Review*, 11, 2019. doi: 10.1186/s12544-019-0368-2.

[13] H. Li and F. Wang. Branch-price-and-cut for the truck–drone routing problem with time windows. *Naval Research Logistics (Nrl)*, 70:184–204, 2022. doi: 10.1002/nav.22087.

[14] H. Min. Leveraging drone technology for last-mile deliveries in the e-tailing ecosystem. *Sustainability*, 15:11588, 2023. doi: 10.3390/su151511588.

[15] Y. Chow, W. Yat, and Y. Lau. Preliminary study of drone delivery systems in hong kong. *NA*, pages 305–317, 2022. doi: 10.52458/978-81-955020-5-9-30.

[16] T. Jones. International commercial drone regulation and drone delivery services. *NA*, 2017. doi: 10.7249/rr1718.3.

[17] D. Resnik and K. Elliott. Using drones to study human beings: ethical and regulatory issues. *Science and Engineering Ethics*, 25:707–718, 2018. doi: 10.1007/s11948-018-0032-6.

[18] F. Stephan, N. Reinsperger, M. Grünthal, D. Paulicke, and P. Jahn. Human drone interaction in delivery of medical supplies: a scoping review of experimental studies. *Plos One*, 17:e0267664, 2022. doi: 10.1371/journal.pone.0267664.

[19] A. Sanjab, W. Saad, and T. Başar. Prospect theory for enhanced cyber-physical security of drone delivery systems: a network interdiction game. *NA*, 2017. doi: 10.1109/icc.2017.7996862.

[20] F. J. Mateen, K. Leung, A. C. Vogel, A. Cissé, and T. C. Y. Chan. A drone delivery network for antiepileptic drugs: a framework and modelling case study in a low-income country. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 114: 308–314, 2020. doi: 10.1093/trstmh/trz131.

[21] Ashesh Sinha Ali Tolooie and Aida Khosh Raftar Nouri. Heuristic approach for optimising reliable supply chain network using drones in last-mile delivery under uncertainty. *International Journal of Systems Science: Operations & Logistics*, 11(1): 2301610, 2024. doi: 10.1080/23302674.2023.2301610. URL https://doi.org/10.1080/23302674.2023.2301610.

[22] Susan Lund, James Manyika, Lola Woetzel, Jacques Bughin, Mekala Krishnan, Jeongmin Seong, and Mac Muir. Globalization in transition: The future of trade and value chains, Jan 2019. URL https://www.mckinsey.com/featured-insights/innovation-and-growth/globalization-in-transition-the-future-of-trade-and-value-chains.

[23] D. Coll, N. Acharya, and H. Acharya. Creating a strong supply chain network design for success, 2023. URL https://fulfillmentiq.com/supply-chain-network-design-for-success/.

[24] T. P. Raptis, A. Passarella, and M. Conti. Data management in industry 4.0: State of the art and open challenges. *IEEE Access*, 7:97052–97093, 2019.

[25] J. Yang, Y. Liu, and Y. Jia. Influence of trust relationships with suppliers on manufacturer resilience in covid-19 era. *Sustainability*, 14(15):9235, 2021.

[26] M. Fargnoli, A. Lleshaj, M. Lombardi, N. Sciarretta, and G. Di Gravio. A bim-based pss approach for the management of maintenance operations of building equipment. *Buildings*, 9(6):139, 2019.

[27] K. Govindan, D. Kannan, T. B. Jørgensen, and T. S. Nielsen. Supply chain 4.0 performance measurement: A systematic literature review, framework development, and empirical evidence. *Transportation Research Part E: Logistics and Transportation Review*, 164:102725, 2022.

[28] J. Sada. Unveiling the power of optimization models in supply chain management: Basics and advantages, 2024. URL https://www.linkedin.com/pulse/unveiling-power-optimization-models-supply-chain-management-sada-f6ikc/.

[29] I. Jackson, D. Ivanov, A. Dolgui, and J. Namdar. Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. *International Journal of Production Research*, 62(17):6120–6145, 2024.

[30] F. Habibi, R. K. Chakrabortty, and A. Abbasi. Evaluating supply chain network resilience considering disruption propagation. *Computers & Industrial Engineering*, 183:109531, 2023.

[31] S. Joshi. A review on sustainable supply chain network design: Dimensions, paradigms,

concepts, framework and future directions. *Sustainable Operations and Computers*, 3: 136–148, 2021.

[32] S. Schmidt and A. Saraceni. Consumer acceptance of drone-based technology for last-mile delivery. *Research in Transportation Economics*, 103:101404, 2024.

[33] R. Kellermann, T. Biehle, and L. Fischer. Drones for parcel and passenger transportation: A literature review. *Transportation Research Interdisciplinary Perspectives*, 4: 100088, 2020.

[34] M. Gao, C. H. Hugenholtz, T. A. Fox, M. Kucharczyk, T. E. Barchyn, and P. R. Nesbit. Weather constraints on global drone flyability. *Scientific Reports*, 11, 2021.

[35] Y. Shi, Y. Lin, B. Li, and R. Y. M. Li. A bi-objective optimization model for the simultaneous pickup and delivery of medical supplies with drones. *Computers & Industrial Engineering*, 171:108389, 2022.

[36] M. Mhlanga, T. Cimini, M. Amaechi, C. Nwaogwugwu, and A. McGahan. From a to o-positive: Blood delivery via drones in rwanda, 2021. URL https://reachalliance.org/wp-content/uploads/2021/03/Zipline-Rwanda-Final-April19.pdf.

[37] S. S. Bacanli, E. Elgeldawi, B. Turgut, and D. Turgut. Uav charging station placement in opportunistic networks. *Drones*, 6(10):293, 2022.

[38] S. A. Hassnain Mohsan, N. Q. Hamood Othman, M. A. Khan, H. Amjad, and J. .Zywiołek. A comprehensive review of micro uav charging techniques. *Micromachines*, 13 (6), 2022.

[39] C. Bian, H. Li, F. Wallin, A. Avelin, L. Lin, and Z. Yu. Finding the optimal location for public charging stations–a gis-based milp approach. *Energy Procedia*, 158:6582–6588, 2019.

[40] LinkedIn. How can stochastic optimization help you optimize your

supply chain?, 2024. URL https://www.linkedin.com/advice/0/how-can-stochastic-optimization-help-you-optimize-beb5f.

[41] M. Ghorbani, M. Nourelfath, and M. Gendreau. A two-stage stochastic programming model for selective maintenance optimization. *Reliability Engineering & System Safety*, 223:108480, 2022.

[42] F. Stranieri, E. Fadda, and F. Stella. Combining deep reinforcement learning and multi-stage stochastic programming to address the supply chain inventory management problem. *International Journal of Production Economics*, 268:109099, 2024.

[43] A. Ghasemi, F. Farajzadeh, C. Heavey, J. Fowler, and C. T. Papadopoulos. Simulation optimization applied to production scheduling in the era of industry 4.0: A review and future roadmap. *Journal of Industrial Information Integration*, 39:100599, 2024.

[44] V. Kanade. What is the markov decision process? definition, working, and examples, 2024. URL https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-markov-decision-process/.

[45] D. J. Barrientos R., M. C. C. Medina, B. J. T. Fernandes, and P. V. A. Barros. The use of reinforcement learning algorithms in object tracking: A systematic literature review. *Neurocomputing*, 596:127954, 2024.