

Intelligent IoT Based Supply Chain For Fresh Produce: A Hybrid Reinforcement Learning And Optimization Approach

by

Chirag Seth

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2025

© Chirag Seth 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Fruits and vegetables form a vital component of the global economy; however, their distribution poses complex logistical challenges due to high perishability, supply fluctuations, strict quality and safety standards, and environmental sensitivity. In this thesis, we propose an adaptive optimization model that accounts for delays, travel time, and temperature variations impacting produce shelf life, and compare its performance against traditional methods such as Robust Optimization (RO), Distributionally Robust Optimization (DRO), and Stochastic Programming (SP).

Our adaptive model significantly outperforms traditional methods by enabling real-time route and temperature adjustments during transit. Empirical results using synthetic and IoT-derived data inspired by a realistic last mile delivery in Toronto show that the adaptive model improves average product shelf life by over 18% and reduces freshness deviation by 80%, with only a marginal increase in travel time. Furthermore, we introduce a Hybrid Model that combines pre-optimized static routes with real-time RL-based corrections. This approach mitigates the limitations of both static and reactive planning by following optimal routes under normal conditions and dynamically overriding them in response to disruptions such as traffic delays or temperature excursions.

Our results demonstrate that the proposed framework retains global efficiency while providing localized adaptability, making it a robust and practical solution for cold-chain logistics. This thesis thus offers a comprehensive, data-driven strategy to enhance sustainability, minimize spoilage, and increase responsiveness in fresh produce supply chains.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Professor Mehrdad Pirnia and Professor James H. Bookbinder, for their unwavering guidance, support, and encouragement throughout the course of my graduate studies. Their expertise, insightful feedback, and patient mentorship have been instrumental in shaping this thesis. I am deeply grateful to my parents back in India, whose sacrifices and unconditional love have been the foundation of all my achievements. In particular, I would like to thank my father, whose constant motivation and belief in my abilities have inspired me to persevere and strive for excellence at every stage. Despite coming from an environment where I had no direct mentors in this academic path, my parents stood as my strongest pillars of strength. Their resilience and optimism instilled in me the determination to overcome every challenge.

I would also like to acknowledge my friends, who have been a source of encouragement, this journey would not have been possible without their presence and belief in me. A very special thanks goes to Ananya Kappal whose constant encouragement, understanding, and emotional support have helped me navigate the highs and lows of this research journey. Her belief in me has been a driving force in my pursuit of this goal.

I am sincerely grateful to my thesis readers, Professor Safa Ernay and Professor Hossein Abouee Mehrizi, for their time to review my work.

Dedication

This is dedicated to my wonderful parents, Pawan and Sonia Seth.

Table of Contents

Author’s Declaration	ii
Abstract	iii
Acknowledgements	iv
Dedication	v
List of Figures	x
List of Tables	xi
List of Abbreviations	xii
1 Introduction	1
1.1 Literature Review	3
1.1.1 Shelf-Life Modeling for Fresh Produce	4

1.1.2	IoT Sensors	5
1.1.3	Existing Approaches in Supply Chain Optimization For Fresh Produce	8
1.1.4	Reinforcement Learning in Supply Chain Management	9
2	Fresh Food Supply Chain Problem	15
2.1	Supply Chain Network and Assumptions	17
2.2	Shelf-Life Calculation	19
2.3	Optimization Models	22
2.3.1	Deterministic Model	22
2.3.2	Robust Optimization Model	25
2.3.3	Stochastic Programming Model	28
2.3.4	Distributionally Robust Model	30
2.3.5	The Proposed Adaptive Optimization Model	31
3	Reinforcement Learning and Hybrid Model	37
3.1	Introduction	37
3.2	RL for Dynamic Logistics Decision Making	40
3.3	Model Formulation	41
3.3.1	State Space ($\mathcal{S}$)	41
3.3.2	Action Space ($\mathcal{A}$)	42

3.3.3	Transition Function	43
3.3.4	Reward Function	45
3.3.5	Learning Process	47
3.4	Hybrid Model Formulation	47
3.4.1	Hybrid Learning Mechanism	49
3.4.2	Hybrid Algorithmic Formulation	50
4	Implementation, Experiments, and Results	53
4.1	IoT-based Sensors for Real-Time Monitoring	53
4.1.1	Synthetic Data Generation for Model Evaluation	55
4.1.2	Experimental Results	57
4.2	Scenario Generation for the proposed Optimization Model	60
4.2.1	Scenario based comparative analysis between a deterministic model and the proposed adaptive model	63
4.2.2	Shelf life comparative analysis between a deterministic model and the proposed adaptive model	65
4.3	Parameter Tuning for the proposed adaptive optimization Model	66
5	Conclusions and Future Work	70
5.1	Limitations of the Study	71
5.2	Directions for Future Research	72

References	74
APPENDICES	81
.1 Appendix A: Shelf-Life Model Derivations	81
.1.1 A.1 Arrhenius-Based Shelf Life Model	81
.1.2 A.2 Q10-Based Shelf Life Model	84
.2 Appendix B: IoT Device Specifications	85

List of Figures

1.1	Impact of temperature on the appearance of red raspberries during a storage period of 7 days [14].	7
4.1	Left: Tive Tag affixed to a shipment for temperature monitoring. Right: Flowchart illustrating the data movement from the Tive Tag to the cloud and mobile application.	54
4.2	Network structure with N delivery stops and a warehouse.	56
4.3	Scenario-wise comparison of the adaptive optimization model's performance in terms of total trip time versus total freshness deviation.	62
4.4	Total Travel Hours (Deterministic vs Adaptive)	64
4.5	Total Freshness Deviation (Deterministic vs Adaptive)	65
4.6	Sensitivity analysis of correction strength β : (a) Total temperature deviation and (b) final shelf life.	68
4.7	Sensitivity analysis of fluctuation severity τ_{ik} : (a) Total temperature deviation and (b) final shelf life.	69

List of Tables

2.1	Key sets, decision variables, and parameters used in shelf-life and transportation modeling.	18
3.1	Notations for RLModel.	41
4.1	Total transportation time (in hours) for each model variant under normal and disrupted conditions	57
1	Tive Tag Specifications	86

List of Abbreviations

DQN	Deep Q-Network
DRO	Distributionally Robust Optimization
IoT	Internet of Things
RL	Reinforcement Learning
RO	Robust Optimization
SC	Supply Chain
SP	Stochastic Programming
VRP	Vehicle Routing Problem

Chapter 1

Introduction

The global food supply chain is under mounting pressure, with one-third of all food produced approximately 1.3 billion tons is lost or wasted annually. Perishable items such as fresh fruits, vegetables, dairy, and meat account for a disproportionately large share of this waste, with spoilage rates ranging from 20% to 60% depending on the region and infrastructure conditions [27]. Beyond the economic burden, estimated at nearly \$940 billion per year, the environmental toll of wasted perishables—through squandered water, energy, and labor—is staggering. Given the fragile shelf life of these goods, minimizing loss during transportation has become not just an operational concern but a global imperative [23].

Perishable logistics are categorically more complex than those for non-perishables. Unlike dry goods, perishables are highly sensitive to external conditions such as temperature, humidity, vibration, and transit delays. For example, the shelf life of semi-dried persimmons fluctuates drastically under minor thermal variations, rendering traditional

routing and storage models inadequate [38]. Perishables deteriorate progressively during each stage—storage, transit, and last mile delivery—necessitating a continuous, responsive approach to logistical planning. This is where the Internet of Things (IoT) has transformative potential [31] by collecting real-time data from temperature sensors, humidity monitors, and GPS trackers allowing logistics operators to detect condition breaches and re-optimize delivery routes on the fly. As demonstrated in the dairy logistics domain, incorporating sensor feedback significantly reduces spoilage while lowering operational costs [22], indicating a significant opportunity for similar approaches in broader perishable supply chains.

The challenge therefore, lies in coupling robust, model-driven planning with sensor-driven adaptivity. While deterministic and robust optimization frameworks offer structured solutions for logistics under uncertainty, they remain limited in responsiveness in real time. On the other hand, learning approaches like reinforcement learning (RL) can adapt on the fly but often lack global consistency. Integrating these paradigms into a hybrid framework could enable both optimal planning and real-time correction—addressing a critical gap in current literature.

This thesis proposes a comprehensive framework for intelligent transportation of fresh produce by combining optimization-based routing with reinforcement learning for real-time adaptation. First, we propose an adaptive optimization model in response to sensor data and compare multiple traditional models, including deterministic, robust, distributionally robust, and stochastic approaches, considering freshness, temperature deviation, and delivery delays. Next, we design a reinforcement learning model that uses IoT-derived sensor inputs to make dynamic route adjustments during transit. Finally, we introduce a hybrid

model that initializes the trip with an optimized route and activates RL-based corrections when disruptions occur.

The remainder of this chapter is focussed on reviewing the literature on shelf-life modeling for fresh produce, IoT based systems, supply chain optimization, and learning-based logistics.

1.1 Literature Review

Perishable food supply chains are characterized by high levels of complexity due to the biological sensitivity of products, the need for stringent environmental control, and frequent disruptions across logistics networks. These challenges are especially pronounced in developing countries and small-scale enterprises, where infrastructure limitations and lack of cold chain integration amplify spoilage risks and operational inefficiencies [5]. To better understand these challenges and identify potential solutions, this section reviews relevant literature across multiple dimensions. We begin by exploring the importance of shelf-life estimation and its modeling through classical approaches such as the Arrhenius and Q_{10} equations. Next, we discuss the role of IoT-enabled sensor systems in capturing real-time environmental data critical to managing perishability. We then examine a range of optimization approaches—including deterministic, stochastic, robust, and distributionally robust optimization—that have been applied to supply chain decision-making under uncertainty. Finally, we review recent advances in reinforcement learning (RL) and its emerging role in enabling adaptive, data-driven logistics strategies. By synthesizing insights from

these diverse areas, we identify existing gaps and motivate the need for integrated frameworks that combine sensor data, empirical models, and algorithmic intelligence to enhance freshness preservation and operational efficiency.

1.1.1 Shelf-Life Modeling for Fresh Produce

Shelf life estimation is critical in the perishable food supply chain due to its direct impact on food quality, safety, and economic efficiency. Traditional models such as the Arrhenius equation and Q_{10} models are commonly employed to describe the effect of temperature on the degradation rate of fresh produce. The Arrhenius model assumes that reaction rates accelerate exponentially with temperature and requires precise estimation of activation energy and reference rate—parameters that are often difficult to determine for a wide variety of fruits and vegetables in heterogeneous supply chains [40]. The Q_{10} model, a simplified exponential expression based on temperature differential effects, is thus often used in practice due to its lower data requirements, despite being less physiologically grounded [25]. However, both models typically rely on assumptions of uniform ambient conditions and ignore real-time variations that significantly affect spoilage rates. Ambient parameters such as temperature, relative humidity, gas composition (O_2 and CO_2), and vibration are all known to influence respiration rate, microbial activity, and nutrient degradation in fruits and vegetables [32]. Various studies advocate the integration of IoT-enabled sensor systems to track the preceding parameters dynamically across the supply chain. For example, the real-time temperature gradients within the pallets can vary by more than $10^{\circ}C$, affecting the predicted shelf life by more than 15% depending on the position of the

package [32]. IoT systems can provide fine-grained ambient profiles, and when integrated with decay models can enable dynamic shelf-life prediction and more informed logistics decisions. Thus, while classical models laid the foundation, their limitations necessitate hybrid approaches combining empirical modeling with sensor-driven contextual data.

Recent empirical research further highlight these modeling challenges. [11] found that although the Arrhenius model can effectively estimate shelf life under controlled environments, its reliance on accurate activation energy values and consistent temperature data makes it impractical for variable cold chains [11]. Their Q_{10} values for semi-dried persimmons varied drastically across temperature bands, reinforcing the need for dynamic adaptation. Moreover, Aung and Chang (2023) emphasized that real-world cold chains suffer from temperature and inconsistencies, which static models cannot account for, thereby underlining the significance of integrating IoT for granular monitoring and traceability [4]. These findings reinforce that robust shelf-life prediction must account for dynamic, sensor-driven environmental variability across the logistics chain.

1.1.2 IoT Sensors

The Internet of Things (IoT) has emerged as a transformative technology in supply chain management, particularly in the management of perishable goods, due to its ability to provide real-time monitoring and decision-making insights. IoT integrates physical objects via sensors and actuators, allowing continuous tracking of critical parameters like temperature, humidity, and geographic position, crucial for maintaining the quality and safety of perishable products [43]. Recent advances in IoT, including sensor-driven shelf-life predic-

tion, real-time logistics re-planning, and automated quality control, have shown significant potential to reduce food spoilage and optimize supply chain efficiency [33]. In addition, IoT technologies address common challenges faced by traditional supply chains, such as information asymmetry and fragmented logistics, thus improving traceability, reliability, and ultimately consumer trust [28]. Hence, the implementation of IoT-based systems in the perishable supply chain not only improves operational efficiency but also offers strategic advantages through increased transparency and responsiveness to dynamic market conditions.

IoT-enabled traceability frameworks can provide pallet-level visibility of environmental exposure, allowing for proactive route adjustments or dynamic pricing. Such systems not only minimize losses, but also help automate regulatory compliance, enhance inventory accuracy, and build consumer trust through transparent reporting.

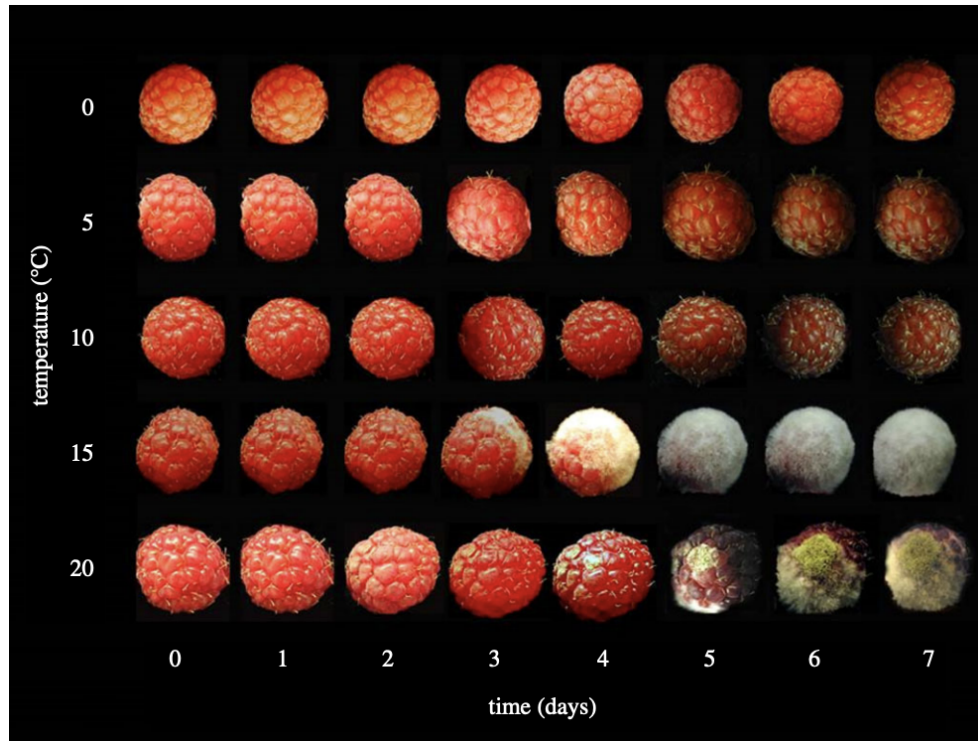


Figure 1.1: Impact of temperature on the appearance of red raspberries during a storage period of 7 days [14].

Figure 1.1 visually demonstrates the rapid deterioration of red raspberries stored under varying temperatures over a 7-day period. As temperature increases, spoilage accelerates significantly—visible by fungal growth and discoloration from Day 4 onward, particularly at 15°C and 20°C. This clearly illustrates the sensitivity of perishable goods to thermal conditions and highlights the importance of maintaining strict temperature control in cold chain logistics. Such temperature variations often arise in real-world supply chains due to the co-transportation of different produce with incompatible storage requirements, frequent door openings during multi-stop deliveries, or unexpected transit delays. These disruptions

can lead to local thermal excursions that accelerate deterioration and waste. IoT-enabled sensing technologies offer a viable solution by providing real-time tracking of environmental parameters, enabling early interventions that can significantly reduce spoilage and improve decision-making throughout the logistics process.

1.1.3 Existing Approaches in Supply Chain Optimization For Fresh Produce

Optimization methods can play a central role in mitigating spoilage and inefficiencies across perishable supply chains. Traditional deterministic approaches often fail under real-world variability, prompting researchers to adopt stochastic models that incorporate uncertain demands and travel times [12]. However, stochastic models typically assume known probability distributions, which may not reflect reality, hence the need for robust optimization (RO) and distributionally robust optimization (DRO) approaches. These models protect against worst-case or ambiguous distribution scenarios, which was shown in Govindan et al. (2014), who used metaheuristics to solve a multi-objective location-routing problem under sustainability constraints [15]. Further, cross-docking-based vehicle scheduling models have been proposed to minimize early/tardy delivery penalties in time-sensitive systems [1], but these often address single-stage or myopic decisions.

Other publications further emphasize the increasing sophistication of optimization techniques in perishable supply chains. A robust multi-objective MILP model was proposed to incorporate perishability, food waste valorization, and supply uncertainty for sustainable

food networks [24]; however, their approach remains limited to static decisions and lacks adaptive responsiveness to real-time data. An exact branch-and-cut algorithm was developed for a perishable production-inventory-routing problem to achieve high precision, but it faces scalability limits and lacks integration with IoT-based freshness tracking [36]. A genetic algorithm was presented for the integrated location-inventory-routing problem with perishables, yet it did not address re-optimization based on sensor feedback, highlighting a need for dynamic, data-driven models [16]. A multi-objective hub location and scheduling model was developed to integrate perishability constraints with sustainability goals by minimizing transportation costs, carbon emissions, and loss of freshness [30]. Their use of a specific metaheuristic provides effective Pareto solutions, though the model assumes fixed fleet capacity and lacks real-time

1.1.4 Reinforcement Learning in Supply Chain Management

The application of Reinforcement Learning (RL) in supply chain management has gained substantial momentum in recent years, fueled by advancements in sensor technologies, the Internet of Things (IoT), and the growing complexity of logistics networks. RL offers a sequential decision-making framework that enables adaptive learning from real-time feedback, making it particularly well-suited to environments characterized by uncertainty, perishability, and multi-objective trade-offs.

A comprehensive survey on the intersection of RL and supply chain management was provided, emphasizing how emerging technologies such as blockchain, 5G, and IoT can support real-time data collection, visibility, and responsive control [37]. The authors argue

that the integration of RL with sensor networks enables intelligent policy adaptation to external disruptions and internal state changes. Similarly, RL’s increasing role in logistics and supply chain operations was analyzed, highlighting its ability to enhance decisions under uncertainty when combined with IoT-generated data streams [[46]]. These studies establish RL not only as a robust learning paradigm but also as a bridge between operational research and data-driven intelligence.

A vaccine supply chain framework that integrates blockchain, IoT, and machine learning was presented as a prototype for RL applications in perishable goods transport [17]. The centrality of IoT sensors in enabling effective policy learning was reinforced, with real-time, high-frequency sensor data described as the backbone of state awareness in RL-based models [39]. The use of RL to manage blockchain-based IoT networks was further illustrated, ensuring efficient sensor data submission and communication integrity [[29]]. These perspectives lay the groundwork for deploying RL in dynamic, sensor-rich environments where logistics decisions must be made continuously.

RL was employed to optimize physical internet-based logistics, enabling scalable and modular transport strategies [35]. The enhancement of supply chain visibility using neural networks and RL was explored, particularly for interpreting unstructured, real-time sensor data [18]. Global supply chains were modeled using RL in foundational work, marking an early but significant step toward intelligent logistics [34]. An AIoT-based RL framework for energy-efficient logistics was recently introduced, stressing that RL can achieve both operational and environmental goals—a key concern in cold chain systems where energy-intensive refrigeration must be balanced with quality preservation [19].

Despite this progress, the application of RL in fresh food transportation remains limited in scope. A hybrid deep reinforcement learning approach was proposed to support proactive transshipment decisions in dual-channel systems [26]. While the model addresses demand volatility between online and offline platforms, it is constrained to inventory reallocation and does not extend to in-transit routing decisions or freshness-aware scheduling. A DRL-based routing strategy was developed to minimize carbon emissions and enforce delivery time windows, although perishability was simplified to fixed deadlines, neglecting continuous degradation due to fluctuating temperatures [48].

DRL was applied for adaptive logistics scheduling, dynamically adjusting routes to match demand patterns [47]. Yet, the model assumes uniform product behavior and does not incorporate item-specific thermal sensitivity or freshness decay. RL was also merged with blockchain to enhance traceability in agri-food supply chains, improving transparency and long-term planning, though it lacks responsiveness to transient temperature shifts and fails to optimize real-time delivery based on freshness metrics [10].

RL has been applied in warehouse-level inventory and pricing optimization. RL was used to improve stock turnover for agri-food items [2], while DRL was employed to optimize dynamic pricing strategies [21]. These models show strong performance in static settings but do not address real-time disruptions or multi-product perishability during transport. An attempt to bridge this gap was made by integrating RL with IoT-based sensing for cold chain logistics, enabling adaptive adjustments to transportation based on environmental feedback [20]. However, this approach lacks a hybrid formulation that combines RL’s adaptability with structured optimization techniques such as Mixed Integer Linear

Programming (MILP) or Distributionally Robust Optimization (DRO), which are crucial for ensuring strict route feasibility, stop timing, and freshness constraints.

These studies collectively reveal the strengths and shortcomings of current RL approaches in perishable logistics. While they leverage real-time data and learning-based control, they often fail to model perishability as a continuous function of time and temperature. Additionally, few frameworks incorporate hybrid architectures that blend the reactive adaptability of RL with the constraint satisfaction and rigor of optimization models.

Despite the promising potential of IoT-based technologies and learning algorithms, the current literature remains fragmented, with most studies addressing shelf-life estimation, sensor integration, or optimization in isolation. There is a lack of end-to-end integration between real-time sensor feedback and decision-making algorithms that can adapt dynamically to disruptions during transportation. Many models either oversimplify perishability using fixed deadlines or ignore the cumulative impact of fluctuating environmental conditions. Furthermore, while some optimization models achieve high precision, they are often computationally intensive and not scalable in real-time scenarios. These limitations underscore the need for unified, adaptive frameworks that can ingest continuous environmental data, model spoilage progression realistically, and generate feasible, responsive logistics strategies that balance freshness preservation with operational efficiency.

Given the inherent complexities and significant economic impacts associated with perishable goods supply chains, as well as the promising potential of IoT-based technologies to address these challenges, it is evident that a systematic and integrated approach is cru-

cial. While the existing literature emphasizes individual methods and IoT solutions, there remains a clear gap in the development of comprehensive models that combine optimization techniques with advanced real-time IoT tracking systems and real-time algorithmic approaches such as reinforcement learning to solve this problem.

Our work advances the state of the art by developing an Adaptive Optimization framework, enabling dynamic re-optimization as new IoT sensor data becomes available, this bridges a critical gap in temporal decision-making and practical deployability as well as an RL framework. This thesis addresses these gaps and Therefore, the main contributions of this paper are as follows:

- Proposing an adaptive optimization model that dynamically adjusts temperature deviations based on real-time feedback from IoT sensors, enabling efficient cold-chain delivery under uncertainty.
- Integrating IoT-enabled shelf-life estimation using both Arrhenius and Q_{10} models to support continuous, sensor-driven spoilage prediction for multiple product types with varying thermal sensitivities.
- Designing a hybrid reinforcement learning framework based on Deep Q-Networks (DQN) that complements optimization models by providing real-time corrections in the presence of disruptions, effectively combining global route optimality with local adaptability.
- Conducting extensive computational experiments using synthetic IoT-based datasets to benchmark the proposed models against classical approaches—Deterministic, Ro-

bust Optimization (RO), Stochastic Programming (SP), and Distributionally Robust Optimization (DRO)—highlighting improvements in freshness retention and shelf-life preservation.

The remainder of this thesis is structured as follows. Chapter 2 defines the fresh food supply chain problem and presents the underlying assumptions, temperature-based degradation models, and the various optimization models, including our proposed adaptive approach. Chapter 3 describes the learning-based models developed to complement the optimization frameworks. Chapter 4 details the implementation, experimental setup, and performance comparisons under various disruption scenarios. Finally, Chapter 5 concludes the thesis by discussing limitations, practical implications, and directions for future research

Chapter 2

Fresh Food Supply Chain Problem

Due to the perishable nature of fresh food products, the fresh food supply chain requires robust monitoring and adaptive techniques to maintain quality and minimize waste. The perishability of these goods introduces complexities such as temperature fluctuations, shelf-life degradation, and unpredictable delays. Typically, distributors pre-plan delivery routes based on the expected requirements for the following day. While this approach simplifies logistics, it also makes the route rigid, leading to inefficiencies in terms of both food wastage and total travel time. Since most spoilage of fresh produce occurs during transportation, it is crucial to implement adaptive tracking systems that monitor ambient parameters such as temperature in real time. This would enable informed decision-making during transit. During transport and distribution, products are frequently loaded and unloaded, often encountering temperature variations and other environmental changes that can accelerate shelf-life deterioration. Additionally, unexpected factors such as traffic congestion, order

cancellations, road closures, and weather fluctuations can significantly impact travel time and overall delivery efficiency. Addressing these challenges with real-time adaptability is essential to optimizing the supply chain and reducing waste. To effectively address these challenges, we need an adaptive, real-time model capable of handling uncertainties in fresh produce transportation. However, to build a reliable real-time system, establishing a benchmark as a starting point is equally important. Distributing companies typically pre-plan delivery routes based on early requirements and historical route data. Here, optimization techniques can play a crucial role by leveraging historical data to generate the most efficient delivery plan—minimizing both total travel time and the shelf-life degradation of each product. For real-time monitoring, reinforcement learning models can be highly beneficial. These models enable an agent to learn optimal routes through both offline and online learning, allowing it to make informed, real-time decisions based on changing conditions during transit. This chapter will explore the foundation of online learning, its background, and the principles of shelf-life calculations, which will be crucial when integrating them into advanced models. In later chapters, we will discuss optimization approaches along with reinforcement learning (RL) approach and finally we will see the hybrid model that combine both optimization techniques and RL.

Online learning refers to a sequential decision-making framework wherein the model is updated incrementally as new information becomes available. Unlike traditional batch optimization, online learning enables real-time adaptability to dynamic environments by continuously refining decisions based on feedback. In the context of fresh food logistics, online learning facilitates responsive route adjustments under real-time disturbances such

as delays, temperature fluctuations, and cancellations.

2.1 Supply Chain Network and Assumptions

This research proposes reoptimization for perishable products that need to be delivered to various stops while ensuring minimal spoilage and efficient transportation given the availability of sensor data. Let $\mathcal{N} = \{0, 1, \dots, N-1\}$ represent the set of all locations, where 0 denotes the warehouse, and each $i \in \mathcal{N} \setminus \{0\}$ corresponds to a delivery stop. Each product belongs to a category $k \in \mathcal{K} = \{1, \dots, K\}$, where K is the total number of perishable product types. Each product k has a defined shelf life, which represents the duration for which it remains fresh under optimal conditions. The remaining shelf life at a given stop i is denoted as $L_{ik} \geq 0$, which depletes over time and is influenced by temperature variations and delays. The rate of decay is dependent on both internal product properties and external environmental conditions, making real-time tracking essential.

Each route between locations has an associated distance $d_{ij} \geq 0$ and travel time $T_{ij} \geq 0$, both of which can be affected by external factors such as traffic congestion, road closures, and weather conditions. Additionally, unexpected delays $\delta_i \geq 0$ can occur at locations due to loading, unloading, or unforeseen circumstances. Temperature control during transport is critical, as deviations can lead to spoilage. Each product type k requires a storage temperature θ_k , but the actual transportation temperature $\hat{\theta}_k$ may fluctuate due to ambient conditions, vehicle refrigeration efficiency, and frequent loading/unloading events. The absolute deviation from the required temperature at location i for product k is represented

Table 2.1: Key sets, decision variables, and parameters used in shelf-life and transportation modeling.

Symbol	Description	Units
<i>Sets</i>		
$\mathcal{N}$	Set of all nodes (warehouse + delivery stops)	–
$\mathcal{K}$	Set of perishable product types	–
$\mathcal{S}$	Set of scenarios (stochastic model)	–
Ξ	Uncertainty set (robust model)	–
$\mathcal{P}$	Ambiguity set of probability distributions (DRO)	–
<i>Decision Variables</i>		
x_{ij}	1 if vehicle travels from node i to j , 0 otherwise	binary
u_i	Sequencing variable for node i (MTZ formulation)	integer
t_{ik}	Temperature of product k at node i	°C
D_{ik}	Absolute deviation from ideal temperature θ_k	°C
S_{ik}	Slack variable for temperature bounds	°C
I_{tk}	Inventory of product k at time t	units (e.g., kg)
<i>Parameters</i>		
t	Actual temperature	K
t_0	Reference temperature	K
L_T	Shelf life at temperature T	hours
L_{T_0}	Shelf life at reference temperature T_0	hours
E_a	Activation energy	J/mol
R	Universal gas constant (≈ 8.314)	J/mol·K
k, k_0	Reaction rate constant and pre-exponential factor	1/hour
Q_{10}	Q_{10} temperature coefficient	dimensionless
T_{ij}	Travel time between node i and j	hours
d_{ij}	Distance between node i and j	km
δ_i	Delay at node i	hours
τ_{ik}	Ambient temperature shift at node i for product k	°C
θ_k	Ideal storage temperature for product k	°C
$\theta_k^{\min}, \theta_k^{\max}$	Min/max allowable temperatures for product k	°C
L_k	Initial shelf life of product k	hours
q_k	Demand for product k	units (e.g., kg)
$\mu_{T_{ij}}, \sigma_{T_{ij}}^2$	Mean and variance of travel time from i to j	hours, hours ²
$\mu_{\delta_i}, \sigma_{\delta_i}^2$	Mean and variance of delay at node i	hours, hours ²
β	Temperature correction factor	[0,1]
λ_1, λ_2	Penalty weights for temperature deviation and slack	cost units

by $D_{ik} \geq 0$. To account for flexibility in real-world operations, a slack variable $S_{ik} \geq 0$ is introduced, which allows minor temperature deviations within an acceptable range. Products also have specific demand quantities $q_k \geq 0$ that need to be fulfilled while minimizing spoilage and optimizing transportation efficiency. Due to natural environmental shifts, the temperature at a given location may vary, represented by a temperature shift parameter $\tau_{ik} \in \mathbb{R}$. To consider these shifts, a correction factor $\beta \in [0, 1]$ is applied, ensuring controlled temperature adjustments. The total inventory level at stop i for product k is represented as $I_{ik} \geq 0$, which is dynamically updated as deliveries progress.

Decisions during transportation include determining whether a vehicle should travel from location i to j , represented by a binary decision variable $x_{ij} \in \{0, 1\}$, where $x_{ij} = 1$ indicates that the vehicle moves from i to j , and 0 otherwise. The order in which locations are visited is tracked using a sequencing variable $u_i \geq 0$, which ensures that delivery routes maintain logical consistency.

Table 2.1 summarizes the key decision variables and parameters related to temperature dynamics, product perishability, routing decisions, and logistical parameters. For clarity and ease of understanding, each of these notations will be introduced and explained in detail as they appear in the respective sections of the methodology.

2.2 Shelf-Life Calculation

Since degradation rates depend significantly on temperature and environmental factors, a dynamic shelf-life estimation model would allow real-time decision-making at key points

such as delivery stops or after door openings. By incorporating a predictive model, we can ensure that deliveries are prioritized based on the freshness of each item, leading to reduced waste and improved efficiency. To quantify the impact of temperature variations on shelf life, we will use two models: the Arrhenius model and the Q_{10} model. The Arrhenius model [3] is a mechanistic approach that explains how reaction rates, including food spoilage processes, are influenced by temperature. It is based on the Arrhenius equation:

$$k = k_0 \exp \left(-\frac{E_a}{RT} \right), \quad (2.1)$$

where k is the temperature-dependent degradation rate constant of the food quality index, k_0 is the pre-exponential (frequency) factor, E_a is the activation energy (in J/mol) required to initiate the degradation process, R is the universal gas constant (8.314 J/mol·K), and T is the absolute temperature (in Kelvin) at which spoilage begins.

Using this relationship, we can estimate the remaining shelf life at a given temperature by applying:

$$L_T = L_{T_0} \exp \left[\frac{E_a}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \right], \quad (2.2)$$

where L_T denotes the estimated shelf life of the product at temperature T , and L_{T_0} is the known shelf life under a reference storage temperature T_0 . This exponential relationship models the accelerated degradation of perishable products under elevated temperatures.

The Arrhenius model is highly accurate for predicting spoilage behavior when the chemical kinetics are well characterized for that product. However, its practical application can be challenging as it requires product-specific activation energy values, which may not always be available for a wide range of fresh produce.

As an alternative, the Q_{10} model provides a simpler empirical approach to estimating shelf life. It assumes that for every 10°C rise in temperature, the rate of deterioration increases by a factor of Q_{10} , typically ranging between 1.5 and 3 for food products. The shelf-life relationship is given by:

$$L_T = L_{T_0} Q_{10}^{\frac{T_0 - T}{10}}, \quad (2.3)$$

where Q_{10} is the temperature coefficient for spoilage acceleration, and T , T_0 , L_T , and L_{T_0} are defined as above. Unlike the Arrhenius model, the Q_{10} model does not require detailed knowledge of activation energies, making it more practical for applications where specific chemical properties of the produce are unknown. One can also modify this to Q_5 or Q_3 based on the application domain, especially when the temperature variation is relatively small or constrained. However, the Q_{10} model assumes a fixed exponential relationship between temperature and spoilage, which may not always be accurate for certain produce that exhibit non-linear degradation behavior under fluctuating environmental conditions. In this work, the Arrhenius model is used when sufficient product-specific data is available to provide precise degradation estimates, while the Q_{10} model is employed as

a more adaptable alternative when dealing with a variety of produce types where exact kinetic parameters are unknown.

2.3 Optimization Models

In the following sections, we explore a sequence of mathematical optimization models—including *Deterministic*, *Stochastic*, *Robust Optimization (RO)*, and *Distributionally Robust Optimization (DRO)*, which are each suited to different operational contexts and data availability scenarios. These models incorporate perishability constraints either explicitly through shelf-life degradation estimates or implicitly via temperature-based thresholds to assess the quality and viability of delivery schedules. Shelf-life calculations derived from the Arrhenius or Q_{10} models serve as critical inputs for quantifying freshness and prioritizing deliveries and are used to benchmark the performance of the proposed adaptive model in Section 2.3.5.

2.3.1 Deterministic Model

Assuming perfect knowledge of all parameters, a deterministic model provides a baseline formulation for the transportation planning problem. In this case, the objective is to minimize the total travel cost while satisfying operational and perishability constraints. The generic structure of a deterministic optimization problem is given by:

$$[\mathcal{D}] : \min_{x \in X} f(x, \bar{\xi}) \quad (2.4)$$

$$\text{s.t. } f_i(x, \bar{\xi}) \leq 0 \quad \forall i \in M, \quad (2.5)$$

where x represents the decision variables (such as routing choices), $\bar{\xi}$ is the known vector of input parameters (including travel times, delays, and shelf-life values), and M is the index set of constraints. Based on this structure, the specific formulation of the transportation problem is as follows:

$$\min \sum_{i \in N} \sum_{j \in N} (T_{ij} + \delta_i) x_{ij} \quad (2.6)$$

subject to

$$\sum_{j \in N \setminus \{0\}} x_{0j} = 1 \quad (2.7)$$

$$\sum_{i \in N \setminus \{0\}} x_{i0} = 1 \quad (2.8)$$

$$\sum_{j \in N \setminus \{i\}} x_{ij} = 1 \quad \forall i \in N \setminus \{0\} \quad (2.9)$$

$$\sum_{j \in N \setminus \{i\}} x_{ji} = 1 \quad \forall i \in N \setminus \{0\} \quad (2.10)$$

$$x_{ii} = 0 \quad \forall i \in N \quad (2.11)$$

$$u_i - u_j + nx_{ij} \leq n - 1 \quad \forall i \neq j, i, j \in N \setminus \{0\} \quad (2.12)$$

$$\sum_{i \in N} \sum_{j \in N} (\delta_i + t_{ij}) x_{ij} \leq L_k - R_k \quad \forall k \in K \quad (2.13)$$

The objective function in Equation (2.6) minimizes the total transportation cost, which includes the travel time t_{ij} between nodes i and j , and the service delay δ_i incurred at node i . The binary decision variable $x_{ij} \in \{0, 1\}$ indicates whether the vehicle moves directly from node i to node j .

Constraints (2.7) and (2.8) ensure that the delivery route originates and terminates at the depot, denoted by node 0. Constraints (2.9) and (2.10) guarantee that each node $i \in N \setminus \{0\}$ is visited exactly once, with a unique incoming and outgoing arc.

Constraint (2.11) eliminates self-loops by enforcing $x_{ii} = 0$ for all $i \in N$, ensuring that the vehicle does not return to the same node without progressing. Constraint (2.12) implements the subtour elimination rule using the Miller–Tucker–Zemlin (MTZ) formulation. Here, $u_i \in \mathbb{R}^+$ is a sequencing variable that defines the order of visiting node i , and n is the total number of the nodes. Constraint (2.13) imposes a perishability limit for which product type $k \in \mathcal{K}$, the total time spent in transit, computed as the sum of all t_{ij} and δ_i values along the selected route, must not exceed the available shelf life window $L_k - R_k$, where L_k is the initial shelf life at dispatch and R_k is the minimum required remaining shelf life upon delivery, assuming $L_k - R_k > 0$.

This deterministic model serves as a foundational formulation for routing perishable goods under idealized conditions, where all parameters—travel times t_{ij} , delays δ_i , and shelf-life values L_k, R_k —are known in advance and remain fixed throughout the planning horizon.

2.3.2 Robust Optimization Model

This section introduces the concept of robust optimization, detailing its principles and methodologies [42]. Initially proposed by Soyster [41] and further advanced by Ben-Tal and Nemirovski [6, 7, 8], robust optimization has become a vital approach for tackling problems under uncertainty. By accounting for uncertainties and variations in input parameters, robust optimization offers a framework for optimization under uncertainty, ensuring system performance and reliability despite unpredictable factors. Consider the general form of a

deterministic optimization problem as defined in (2.4)–(2.5), where x represents the decision variables, $X \subset \mathbb{R}^n$ is the feasible set, $\bar{\xi}$ is the vector of nominal values of the uncertain parameters ξ , and $M = \{1, \dots, m\}$ is the index set of constraints. This problem seeks the optimal x to minimize the objective function $f(x, \bar{\xi})$ under constraints $f_i(x, \bar{\xi}) \leq 0, \forall i \in M$.

In the optimization problem (2.7)–(2.12), some parameters such as travel and handling times are inherently non-deterministic. When $\xi \neq \bar{\xi}$, the optimal solution obtained may not be feasible for some constraints $f_i(x, \xi) \leq 0$, or even if all constraints are satisfied, the objective function value may vary for different ξ . This limitation leads to the adoption of robust optimization, which aims to minimize the worst-case objective function and ensure solution feasibility across all possible realizations of uncertain parameters within a predefined uncertainty set. Therefore, the robust counterpart of the preceding problem is formulated as follows:

$$[\mathcal{RO}] : \min_{x \in X} \sup_{\xi \in \Xi} f(x, \xi) \tag{2.14}$$

$$\text{s.t.} \quad f_i(x, \xi) \leq 0 \quad \forall \xi \in \Xi, \forall i \in M, \tag{2.15}$$

where Ξ is the uncertainty set of the parameter ξ . That is, Ξ represents all possible values that may be assumed by the random parameter ξ . This framework minimizes the worst-case objective in (2.14) under the constraints (2.15), ensuring that a solution x is

feasible for all realizations of the uncertain parameter ξ within the uncertainty set Ξ . The tractable reformulation of the problem depends on both the functions f and the structure of the uncertainty set Ξ . We define the robust perishable delivery optimization problem considering the worst-case uncertainties in travel time, delays, and shelf life variations. The goal is to minimize the transportation cost while ensuring that the freshness of the products is preserved during transportation, even under uncertain conditions. The uncertainty set used in this study is a box uncertainty set, where the travel times and delays are bounded within given minimum and maximum values.

The mathematical model is given as:

$$\min \sum_{i \in N} \sum_{j \in N} (T_{\max,ij} + \delta_{\max,i}) x_{ij} \quad (2.16)$$

In Equation (2.16), $T_{\max,ij}$ denotes the maximum possible travel time between nodes i and j , and $\delta_{\max,i}$ denotes the maximum possible delay at location i . The variable $x_{ij} \in \{0, 1\}$ is the binary routing decision variable. Subject to constraints (2.7)–(2.12), enforcing routing feasibility, and:

$$\sum_{i \in N} \sum_{j \in N} (\delta_{\max,i} + T_{\max,ij}) x_{ij} \leq L_k - R_k \quad \forall k \in K \quad (2.17)$$

Constraint (2.17) ensures that, for each product $k \in \mathcal{K}$, the total transit time in the worst-case scenario (including worst-case delays $\delta_{\max,i}$ and travel times $T_{\max,ij}$) does not exceed the freshness window $L_k - R_k$, where L_k is the initial shelf life and R_k is the mini-

mum required remaining shelf life upon delivery.

The uncertainty set for travel times and delays is defined as:

$$T_{ij} \in [T_{ij}^{\min}, T_{ij}^{\max}], \quad \forall i \in N, \forall j \in N \quad (2.18)$$

$$\delta_i \in [\delta_i^{\min}, \delta_i^{\max}], \quad \forall i \in N \quad (2.19)$$

Here, $T_{ij}^{\min}$ and $T_{ij}^{\max}$ represent the minimum and maximum possible travel times between nodes i and j , respectively. Similarly, $\delta_i^{\min}$ and $\delta_i^{\max}$ define the minimum and maximum possible delays at location i .

2.3.3 Stochastic Programming Model

Real-world transportation involves uncertainty in travel times, demand, and handling delays. An alternative method to handle randomness is Stochastic Programming (SP), introduced by [13]. SP considers discrete scenarios for each parameter and optimizes the expected value based on their probability distributions, derived from historical data, leading to the following formulation:

$$\min_x \mathbb{E}_\xi [f(x, \xi)] \quad (2.20)$$

In this formulation, x represents the vector of decision variables, ξ is the random vector

representing uncertainty in system parameters, and $f(x, \xi)$ is the cost function that depends on both decision variables and realizations of uncertainty. The operator $\mathbb{E}_\xi[\cdot]$ denotes the expectation over all realizations of ξ , thereby accounting for the expected transportation cost across possible scenarios.

In Stochastic Programming, uncertainty in travel times t_{ij} and delays δ_i is modeled by a finite set of discrete scenarios. Each scenario $s \in S$ represents a particular realization of these uncertain parameters, with associated probability $p_s \in [0, 1]$, where $\sum_{s \in S} p_s = 1$. The travel time and delay values for scenario s are denoted by t_{ij}^s and δ_i^s , respectively. The model optimizes the expected value of the total transportation cost while respecting perishability constraints. The base model uses the objective function given by (2.6) and routing constraints (2.7)–(2.12), but replaces the in-transit shelf-life constraint with the following expected constraint:

$$\sum_{s \in S} p_s \sum_{i \in N} \sum_{j \in N} (\delta_i^s + t_{ij}^s) x_{ij} \leq L_k - R_k \quad \forall k \in K \quad (2.21)$$

The constraint (2.21) ensures that, on average across all defined scenarios $s \in S$, the total time spent in transit—including scenario-specific delays δ_i^s and travel times t_{ij}^s —does not reduce the usable shelf life of any product k below the required level. This formulation balances expected performance with perishability constraints by incorporating probabilistic modeling of operational uncertainty. The risk aversion of the solution can be tuned by introducing a robustness parameter z , which can modify the constraint threshold or enter

the objective in extended formulations.

2.3.4 Distributionally Robust Model

Distributionally Robust Optimization (DRO) extends stochastic programming by considering a set of probability distributions rather than a single fixed distribution [9]. The goal is to minimize the worst-case expected cost:

$$\min_x \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [f(x, \xi)], \quad (2.22)$$

where $\mathcal{P}$ is an ambiguity set of probability distributions. DRO is useful in SC logistics where disruptions such as weather, strikes, or demand surges lead to distributional uncertainty. The tractable reformulation of the problem depends on the structure of the ambiguity set. In this study, a moment-based ambiguity set is used, which considers the mean and variance of the uncertain parameters, such as travel times T_{ij} and delays δ_i . The ambiguity set is defined as follows:

$$\mathcal{P} = \left\{ P \mid \mathbb{E}_P[T_{ij}] = \mu_{T_{ij}}, \mathbb{E}_P[\delta_i] = \mu_{\delta_i}, \text{Var}_P[T_{ij}] = \sigma_{T_{ij}}^2, \text{Var}_P[\delta_i] = \sigma_{\delta_i}^2 \right\}, \quad (2.23)$$

In this ambiguity set, $\mu_{T_{ij}}$ and $\sigma_{T_{ij}}^2$ represent the mean and variance of the travel time between locations i and j , while μ_{δ_i} and $\sigma_{\delta_i}^2$ represent the mean and variance of the delay

time at location i . The mathematical model is given by (2.6), subject to constraints (2.7)–(2.12) and:

$$\sum_{i \in N} \sum_{j \in N} (\mu_{\delta_i} + \mu_{T_{ij}}) x_{ij} + z \cdot \sum_{i \in N} \sum_{j \in N} (\sigma_{\delta_i}^2 + \sigma_{T_{ij}}^2) x_{ij}^2 \leq Q_k \cdot L_k - R_k \quad \forall k \in K \quad (2.24)$$

In this model μ_{δ_i} and $\mu_{T_{ij}}$ are the mean delay times and travel times, $\sigma_{\delta_i}^2$ and $\sigma_{T_{ij}}^2$ are the variances of the delay times and travel times, $Q_k \geq 0$ represents the demand for product k , $L_k \geq 0$ represents the available shelf life for product k , $R_k \geq 0$ represents the required shelf life for product k . The moment-based DRO in-transit time constraint (2.24) ensures that the total time spent in transit does not exceed the available shelf life L_k minus the required shelf life R_k for each product, considering both the mean and variance of the travel times and delays.

2.3.5 The Proposed Adaptive Optimization Model

In real-world fresh-food logistics, environmental conditions, such as ambient temperature and handling delays, introduce dynamic and location-specific effects on product quality. Unlike static formulations above, this model incorporates temperature feedback and adjustment mechanisms across sequential stops to minimize spoilage and energy costs while preserving product freshness. The model minimizes total cost considering travel time, temperature deviations, and operational flexibility. In practice, this optimization is carried out iteratively, where at each visited node i , a restricted MILP is solved to determine the next

stop j , and the temperature state is updated using feedback from current conditions. At each iteration t , we define two dynamic subsets of nodes: the visited set $\mathcal{V}_t \subseteq \mathcal{N}$ and the remaining set $\mathcal{R} = \mathcal{N} \setminus \mathcal{V}$, where $\mathcal{N}$ denotes the set of all delivery locations including the depot. The set $\mathcal{V}$ is updated after each optimization step by including the node selected in the previous hop, and $\mathcal{R}$ is updated accordingly. The decision model then selects the best next node $j \in \mathcal{R}$ to visit from the current node $i \in \mathcal{V}$, minimizing the combined travel and freshness penalty locally at each stage. The objective function and constraints presented below apply to each such sub-problem in the adaptive optimization sequence.

The mathematical formulation for each decision step is given as follows:

$$\min_{j \in \mathcal{R}} (T_{ij} + \delta_i) x_{ij} + \lambda_1 \sum_{k \in \mathcal{K}} D_{jk} + \lambda_2 \sum_{k \in \mathcal{K}} S_{jk} \quad (2.25)$$

subject to constraints (2.7)–(2.12) and:

$$t_{0k} = \theta_k \quad \forall k \in \mathcal{K} \quad (2.26)$$

$$t_{jk} \geq t_{ik} + \tau_{jk} + \beta(\theta_k - t_{ik}) - M(1 - x_{ij}) - S_{jk} \quad \forall i, j \in \mathcal{N}, i \neq j, k \in \mathcal{K} \quad (2.27)$$

$$t_{jk} \leq t_{ik} + \tau_{jk} + \beta(\theta_k - t_{ik}) + M(1 - x_{ij}) + S_{jk} \quad \forall i, j \in \mathcal{N}, i \neq j, k \in \mathcal{K} \quad (2.28)$$

$$\theta_k^{\min} \leq t_{ik} \leq \theta_k^{\max} \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (2.29)$$

$$D_{ik} \geq t_{ik} - \theta_k \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (2.30)$$

$$D_{ik} \geq \theta_k - t_{ik} \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \quad (2.31)$$

This model includes a dynamic temperature update mechanism that accounts for ambient shifts τ_{jk} , moderated by the temperature correction factor β , along each traveled arc (i, j) . In this formulation: t_{0k} is the initial temperature of product k at the warehouse (node 0), t_{jk} is the temperature of product k at location j , τ_{jk} denotes the ambient temperature shift along the arc from node i to j , $\beta \in [0, 1]$ is the temperature correction factor moderating adjustment toward the required temperature θ_k , M is a large positive constant used for constraint relaxation when the arc (i, j) is not selected (i.e., when $x_{ij} = 0$), $S_{jk} \geq 0$ is a slack variable allowing temperature deviations to maintain feasibility, θ_k is the required storage temperature for product k , $\theta_k^{\min}$ and $\theta_k^{\max}$ are the minimum and maximum allowable temperatures for product k , D_{ik} represents the absolute temperature deviation of product k at node i , $\lambda_1, \lambda_2 \geq 0$ are penalty coefficients on temperature deviations and slack variables, respectively, $x_{ij} \in \{0, 1\}$ is the route decision variable indicating if arc (i, j) is traversed, T_{ij} is the travel time between nodes i and j , δ_i is the delay at node i , $\mathcal{N}$ is the set of all nodes (including warehouse and delivery stops), $\mathcal{K}$ is the set of all product types.

The objective in Equation (2.25) minimizes local travel time and delay costs at each hop, penalizes temperature deviations weighted by λ_1 , and penalizes slack usage weighted by λ_2 , over all candidate destinations $j \in \mathcal{R}$. The total route cost is accumulated incrementally across all stages of the rolling-horizon execution. Routing feasibility is enforced by constraints (2.7) to (2.12), which respectively ensure route start/end at the warehouse, single visits to each delivery stop, elimination of self-loops, and prevention of subtours via the MTZ formulation.

Constraint (2.26) sets the initial temperature of each product at the warehouse. In each subsequent MILP re-optimization, the current temperature t_{ik} serves as the initial state for that subproblem. Constraints (2.27) and (2.28) impose lower and upper bounds on temperature evolution between successive nodes, with the big- M term deactivating constraints on arcs not selected in the route. Constraint (2.29) restricts product temperatures within allowable limits. Constraints (2.30) and (2.31) compute the absolute deviation D_{ik} from the target temperature θ_k .

To better illustrate the workings of the proposed model, we present a stepwise algorithm 1 that reflects the actual rolling-horizon implementation. At each stage of the delivery process, the model evaluates only the remaining candidate nodes that are yet to be visited. This subset is referred to as $\mathcal{R}$, and is updated after each decision.

The optimization at each step solves a restricted MILP based on the current node i , and selects the best next node $j \in \mathcal{R}$ based on travel time, delay, temperature drift, and expected freshness penalties. The temperature state is propagated forward using the dynamic correction mechanism described in Equations (2.27)–(2.28). The process continues until all stops have been visited, and the route concludes with a return to the warehouse.

Algorithm 1 Rolling-Horizon Adaptive Optimization with Temperature Feedback

Require: Node set $\mathcal{N} = \{0, 1, \dots, n-1\}$ (warehouse 0 and customers),

- 1: Product set $\mathcal{K} = \{1, \dots, K\}$, travel times T_{ij} , delays δ_i ,
- 2: Ambient temperature shifts τ_{jk} , temperature bounds $[\theta_k^{\min}, \theta_k^{\max}]$,
- 3: Ideal temperatures θ_k , correction factor β , penalties λ_1, λ_2 , big- M

Ensure: Route $\{x_{ij} = 1\}$, temperature profile $\{t_{ik}\}$, total cost

Phase 0: Initialization

- 4: Set $t_{0k} \leftarrow \theta_k$ for all $k \in \mathcal{K}$ *(initial temperature)*
- 5: $\mathcal{V} \leftarrow \{0\}$, $\mathcal{R} \leftarrow \mathcal{N} \setminus \{0\}$
- 6: Initialize cost: $TotalCost \leftarrow 0$, route: $Route \leftarrow \emptyset$

Phase 1– n : Iterative Re-Optimization and Feedback

- 7: **while** $\mathcal{R} \neq \emptyset$ **do**
- 8: Let $i \leftarrow$ most recently visited node
- 9: Solve MILP defined by constraints (2.7)–(2.12), (2.26)–(2.31), restricted to nodes $\mathcal{V} \cup \mathcal{R}$
- 10: Select next node $j \in \mathcal{R}$ such that $x_{ij} = 1$
- 11: **for all** $k \in \mathcal{K}$ **do**
- 12: Compute drift τ_{jk} and correction $\beta(\theta_k - t_{ik})$
- 13: Predict temperature at j :

$$t_{jk} \leftarrow t_{ik} + \tau_{jk} + \beta(\theta_k - t_{ik})$$

- 14: Impose constraints:

$$\begin{aligned} t_{jk} &\geq t_{ik} + \tau_{jk} + \beta(\theta_k - t_{ik}) - M(1 - x_{ij}) - S_{jk} \\ t_{jk} &\leq t_{ik} + \tau_{jk} + \beta(\theta_k - t_{ik}) + M(1 - x_{ij}) + S_{jk} \\ \theta_k^{\min} &\leq t_{jk} \leq \theta_k^{\max} \\ D_{jk} &\geq t_{jk} - \theta_k \\ D_{jk} &\geq \theta_k - t_{jk} \end{aligned}$$

- 15: **end for**
- 16: Add to route: $Route \leftarrow Route \cup \{(i, j)\}$
- 17: Update cost:

$$TotalCost += (T_{ij} + \delta_i)x_{ij} + \lambda_1 \sum_k D_{jk} + \lambda_2 \sum_k S_{jk}$$

- 18: $\mathcal{V} \leftarrow \mathcal{V} \cup \{j\}$, $\mathcal{R} \leftarrow \mathcal{R} \setminus \{j\}$
- 19: **end while**

Phase $n+1$: Return to Warehouse 35

- 20: Let $i \leftarrow$ last visited customer
 - 21: Solve final leg $i \rightarrow 0$ and update:
 - 22: $Route \leftarrow Route \cup \{(i, 0)\}$, $TotalCost += (T_{i0} + \delta_i)x_{i0}$
 - Return:** adaptive route $\{x_{ij} = 1\}$, temperatures $\{t_{ik}\}$, and final cost
-

This chapter examined the fresh food supply chain problem, highlighting perishability, temperature fluctuations, and operational uncertainties. We defined the network structure, assumptions, and notation, and presented shelf-life estimation methods via the Arrhenius and Q_{10} models. Four optimization frameworks were formulated: deterministic Model for ideal conditions, RO for worst-case bounded uncertainties, SP for scenario-based analysis, and DRO for ambiguity in probability distributions. Finally, we proposed the Adaptive Optimization Model, which integrates real-time temperature feedback, ambient condition adjustments, and rolling-horizon reoptimization to dynamically preserve product freshness and improve routing efficiency.

Chapter 3

Reinforcement Learning and Hybrid Model

3.1 Introduction

Reinforcement Learning (RL) is a branch of machine learning focused on sequential decision-making, where an agent interacts with an environment to learn an optimal strategy by maximizing long-term rewards [45]. RL has found significant applications in robotics, autonomous systems, gaming, and logistics, owing to its ability to adapt to dynamic and uncertain environments. Unlike supervised learning, which relies on labeled data, RL operates on the principle of an agent improving its behavior based on reward signals obtained through a balance of exploration and exploitation.

The formal framework of RL is modeled as a Markov Decision Process (MDP), repre-

sented by the tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma)$, where:

- $\mathcal{S}$ is the set of states, each representing a different configuration of the environment.
- $\mathcal{A}$ is the set of actions the agent can take.
- $P(s'|s, a)$ is the transition probability, denoting the likelihood of reaching state s' after taking action a in state s .
- $r(s, a) : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function that assigns a numerical reward to each state-action pair.
- $\gamma \in [0, 1]$ is the discount factor, which controls the importance of future rewards.

The agent follows a policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$, which maps states to actions. The goal of the agent is to learn the optimal policy π^* , which maximizes the expected cumulative reward:

$$V^\pi(s) = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \middle| s_0 = s, \pi \right].$$

Additionally, the action-value function (Q-function) is defined as:

$$Q^\pi(s, a) = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \middle| s_0 = s, a_0 = a, \pi \right].$$

Here, R denotes the reward received after taking action a in state s . Q-learning is one of the most widely used RL techniques for discrete action spaces. The Q-learning update rule is:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R + \gamma \max_{a'} Q(s', a') - Q(s, a) \right].$$

In large state-action spaces, Deep Q-Networks (DQN) use a deep neural network $Q_\theta(s, a)$ parameterized by θ to approximate the Q-function. The network is trained using the loss function:

$$L(\theta) = \mathbb{E}_{(s,a,R,s') \sim \mathcal{D}} \left[(y - Q_\theta(s, a))^2 \right],$$

where the target value y is:

$$y = R + \gamma \max_{a'} Q_{\theta^-}(s', a'),$$

and θ^- represents the parameters of a separate target network, which stabilizes the training process.

In real-time logistics, RL can be leveraged to make adaptive decisions in dynamic environments. In the context of our problem, RL will play a crucial role in real-time route optimization and temperature control during transportation. The state space includes variables such as vehicle location, remaining shelf life of products, ambient temperature, and real-time traffic conditions. The action space consists of selecting the next delivery stop and adjusting the storage temperature of the product. Rewards will be calculated based on timely deliveries, minimization of spoilage, and energy efficiency of the transportation process.

3.2 RL for Dynamic Logistics Decision Making

In general, optimization models excel at providing optimal decisions based on predefined constraints and parameters. However, in real-time dynamic environments, such as delivery logistics, where traffic, weather, and demand can fluctuate, traditional optimization models may struggle to adapt to unforeseen changes. For example, a food delivery company like Uber Eats or FreshDirect may experience sudden traffic jams or severe weather conditions during peak hours, delaying orders. Similarly, a wholesale produce supplier may receive an urgent order increase from a key grocery chain while in-transit, requiring a reallocation of inventory. These scenarios highlight the need for a model that can continuously learn and adapt to dynamic changes, instead of relying solely on static optimization methods.

RL is an ideal solution for these challenges. Unlike traditional optimization models, RL can be trained on historical data and then deployed to make real-time decisions. The RL model learns from interactions with its environment, constantly updating its policy based on the rewards received for actions taken. This allows it to adjust decisions dynamically, ensuring efficient delivery while minimizing delays, spoilage, and excess travel.

Below, we summarize the key notations specific to the RL model. For additional variables related to shelf-life, temperature modeling, and transportation dynamics, please refer to Table [2.1](#).

In the following section, we will provide a detailed formulation of the RL model, including the state and action spaces, reward functions, and the learning process the agent undergoes to optimize delivery decisions.

Table 3.1: Notations for RLModel.

Notation	Description	Units
$\mathcal{S}$	Set of all states (e.g., locations, inventory, temperature)	-
$\mathcal{A}$	Set of all actions the agent can take	-
$P(s' s, a)$	Transition probability from state s to s' given action a	-
$r(s, a)$	Reward function for a state-action pair based on delivery	arbitrary units
γ	Discount factor for future rewards	[0,1] (dimensionless)
π	Policy mapping states to actions	-
$V^\pi(s)$	Value function for state s under policy π	units (e.g., reward)
$Q^\pi(s, a)$	Q-function for state-action pair (s, a) under policy π	units (e.g., reward)
$Q(s, a)$	Action-value function in Q-learning	units (e.g., reward)
α	Learning rate for Q-learning update	[0,1] (dimensionless)
ϵ	Exploration factor (epsilon-greedy)	[0,1] (dimensionless)

3.3 Model Formulation

In this section, we present the formal mathematical formulation of our RL framework for adaptive routing and temperature control in perishable goods delivery. The model dynamically interacts with uncertain transportation conditions such as varying delays, fluctuating temperatures, and stochastic demand. The environment is modeled as a Markov Decision Process (MDP) where the agent learns to optimize delivery schedules by minimizing spoilage, delays, and maximizing product freshness based on observed state information at each delivery stop.

3.3.1 State Space ($\mathcal{S}$)

The state space $\mathcal{S}$ represents the complete system configuration at each decision stage during the delivery process. Each state $s \in \mathcal{S}$ consists of multiple variables capturing

the physical location of the vehicle, the current inventory levels, the remaining shelf life of products, the cumulative elapsed time, and the thermal deviation from ideal storage conditions. Formally, the state can be represented as:

$$s = (i, \{I_k\}_{k=1}^K, \{L_{ik}\}_{k=1}^K, \Delta, \{D_{ik}\}_{k=1}^K),$$

where $i \in \{0, 1, \dots, N - 1\}$ denotes the current delivery stop index, with $i = 0$ corresponding to the warehouse. The term I_k represents the inventory level of product $k \in \{1, \dots, K\}$ currently available on the vehicle. The remaining shelf life of product k at the current stop i is denoted by L_{ik} , which dynamically evolves as the vehicle progresses through the delivery route and is directly affected by the temperature fluctuations encountered during transportation. The cumulative travel time since departure is denoted by Δ , and D_{ik} represents the absolute deviation between the observed temperature $\hat{\theta}_k$ and the ideal storage temperature θ_k for product k , computed as $D_{ik} = |\theta_k - \hat{\theta}_k|$.

Shelf life L_{ik} is updated at each visited stop based on temperature-sensitive degradation models. Specifically, either the Arrhenius or the Q10 model is employed depending on product-specific characteristics as explained in Section 2.2. These formulations enable dynamic shelf life estimation at each stop as the transportation environment evolves.

3.3.2 Action Space ($\mathcal{A}$)

At each decision epoch, the RL agent selects an action $a \in \mathcal{A}$, which simultaneously determines the next delivery stop and the thermal control adjustment required for preserving

product freshness. The action is represented as a tuple:

$$a = \left(j, \{ \Delta \hat{\theta}_k \}_{k=1}^K \right),$$

where $j \in \{0, 1, \dots, N-1\}$ corresponds to the next delivery stop to visit, and $\Delta \hat{\theta}_k$ is the temperature correction applied for product k .

Simultaneously, the agent adjusts storage temperature using a correction mechanism defined by:

$$\Delta \hat{\theta}_k = \beta \cdot (\hat{\theta}_k - \theta_k), \quad \beta \in [0, 1],$$

where β is a tunable correction factor controlling the aggressiveness of temperature adjustment for product k . This allows the agent to dynamically fine-tune temperature deviations and mitigate accelerated spoilage due to suboptimal thermal exposure during transportation.

Through this joint control policy, the agent adaptively balances routing efficiency and freshness preservation while responding to real-time environmental fluctuations.

3.3.3 Transition Function

The transition function $P(s'|s, a)$ models the environment's dynamics by describing how the system evolves from the current state s to the next state s' after executing action a . In the context of our perishable goods delivery problem, the transition occurs upon

completion of each delivery leg, where both routing and thermal factors jointly affect state evolution.

At each transition, the vehicle moves from its current location i to the selected next stop j . The travel time associated with this move consists of distance-dependent time and a random delay time component. The total travel time τ_{ij} incurred in the transition is computed as:

$$\tau_{ij} = t_{ij} + \delta_j,$$

where t_{ij} denotes the base travel time from stop i to j , and δ_j represents the random delay encountered at stop j . Both t_{ij} and δ_j are drawn from empirical distributions generated via historical datasets in the simulation model.

Inventory levels are updated by subtracting the delivered amount at stop j . If RI_k^j represents the quantity delivered for product k at stop j , then the inventory update is defined as:

$$I'_k = I_k - RI_k^j.$$

For each k , the remaining shelf life after transition is recalculated using the degradation model (Q10 or Arrhenius) described previously, where the actual temperature $\hat{\theta}_k$ is modified based on the applied temperature adjustment $\Delta\hat{\theta}_k$. Finally, the accumulated elapsed time is incremented as:

$$\Delta' = \Delta + \tau_{ij}.$$

The temperature deviation in the new state is also updated based on sensor readings and

correction applied:

$$D'_{ik} = |\theta_k - (\hat{\theta}_k + \Delta\hat{\theta}_k)|.$$

Thus, the complete state s' is deterministically obtained from state s and action a , conditional on the stochastic realization of delay δ_j . Therefore, $P(s'|s, a)$ remains stochastic due to the random delay component but fully tractable.

3.3.4 Reward Function

The reward function is designed to jointly capture three core objectives: maximizing delivery quantity, minimizing freshness loss, and reducing delivery delays and travel distance. At each decision step, the cumulative reward $r(s, a)$ is defined as:

$$r(s, a) = R_{\text{del}} - R_{\text{tr}} - R_{\text{sp}},$$

where R_{del} , R_{tr} , and R_{sp} correspond to delivery reward, travel penalty, and spoilage penalty, respectively.

The delivery reward R_{del} directly promotes serving customers by rewarding delivered quantities. It is expressed as:

$$R_{\text{del}} = \gamma \cdot \sum_{k \in K} Del_{jk},$$

where Del_{jk} is the amount of product k delivered at stop j , and γ is a global reward coefficient uniformly applied across all product types. Assigning higher γ values

encourages the RL agent to prioritize delivery stops that can fulfill higher demands earlier in the route, ensuring that maximum demand is satisfied before shelf life degrades.

The travel penalty R_{tr} penalizes the agent for incurring longer travel times and delays. This term ensures that shorter or more efficient routes are preferred. It is defined as:

$$R_{\text{tr}} = \lambda_2 \cdot \tau_{ij},$$

where $\tau_{ij} = t_{ij} + \delta_j$ combines base travel time t_{ij} and stochastic delay δ_j for the transition from stop i to stop j , and λ_2 is the weight assigned to penalize longer travel durations.

Finally, the spoilage penalty R_{sp} quantifies freshness loss due to temperature deviations. This term penalizes the system for exposing inventory to suboptimal thermal conditions during transit:

$$R_{\text{sp}} = \lambda_1 \sum_{k \in K} D'_{ik}$$

where D'_{ik} is the temperature deviation for product k at stop j after applying thermal control.

This weighted structure enables the agent to learn a policy that simultaneously prioritizes early high-volume deliveries, chooses shorter efficient routes, and minimizes product degradation due to environmental deviations. Such reward shaping reflects realistic trade-offs in perishable supply chain operations, and can be tuned in practice using cost data, expert inputs, or hyperparameter search to adapt to real-world business constraints.

3.3.5 Learning Process

The reinforcement learning agent employs a Deep Q-Network (DQN) to approximate the optimal policy π^* for sequential decision-making in dynamic perishable goods transportation. Given the large dimensionality of the state space and the stochastic nature of real-world transportation delays and shelf-life degradation, the DQN framework is well-suited to approximate the action-value function $Q(s, a)$.

At each step, the agent observes the current state s , selects an action a according to an ϵ -greedy policy to balance exploration and exploitation, receives an immediate reward $r(s, a)$ as defined in the reward function, and transitions to a new state s' . The experiences (s, a, r, s') are stored in a replay buffer $\mathcal{M}$ to decorrelate samples. The complete training process is outlined in Algorithm 2.

3.4 Hybrid Model Formulation

Although reinforcement learning demonstrates strong adaptability to real-time variations in delivery logistics, it can occasionally suffer from local greediness when no disruptions such as delays, temperature deviations, or demand anomalies occur. In such stable scenarios, the RL model may choose sub-optimal routes purely based on short-term reward signals, deviating from globally optimal routes computed under full information. This tradeoff highlights the need for a hybrid approach that combines the pre-computed global optimality of classical optimization models with the adaptive flexibility of RL-based real-time decision making.

Algorithm 2 Deep Q-Learning for Perishable Goods Delivery Optimization

- 1: Initialize replay buffer $\mathcal{M}$
- 2: Initialize Q-network parameters θ and target network $\theta^- \leftarrow \theta$
- 3: **for** each episode **do**
- 4: Reset environment: set vehicle at warehouse, initialize inventory I_k , shelf life L_{ik} , and elapsed time Δ
- 5: Observe initial state s
- 6: **while** delivery route not completed **do**
- 7: Select action a using ϵ -greedy policy:

$$a = \begin{cases} \text{random action} & \text{with probability } \epsilon \\ \arg \max_{a'} Q_{\theta}(s, a') & \text{otherwise} \end{cases}$$

- 8: Execute action a , observe next stop j , travel time t_{ij} , delay δ_j , updated shelf life L_{jk} , updated inventory I'_k , and compute reward r
- 9: Transition to new state s'
- 10: Store transition (s, a, r, s') into $\mathcal{M}$
- 11: Sample minibatch from $\mathcal{M}$ and update network parameters θ by minimizing:

$$L(\theta) = \frac{1}{2} \sum_i (y_i - Q_{\theta}(s_i, a_i))^2$$

where $y_i = r_i + \gamma \max_{a'} Q_{\theta^-}(s'_i, a')$

- 12: Every C steps, update target network: $\theta^- \leftarrow \theta$
 - 13: Set $s \leftarrow s'$
 - 14: **end while**
 - 15: **end for**
-

To address these challenges, we design a hybrid optimization-RLframework that leverages the strengths of both methods. The approach operates in two stages:

- **Pre-optimization stage:** Before dispatch, a deterministic route plan is computed offline using an optimization model, which minimizes total travel distance while satisfying delivery demands across all stops. This optimization module utilizes classical

linear programming methods to determine an initial globally optimal route, assuming full knowledge of demand and network topology at dispatch time.

- **Real-time adaptive stage:** During the trip, the RL agent operates in parallel with the optimized plan. If no unusual events occur, such as unexpected delays, shelf-life deterioration, or temperature excursions, the vehicle continues to follow the pre-computed optimized route. However, when stochastic disturbances occur in real time (e.g., road closures, sudden delays, or sharp temperature deviations), the RL agent dynamically overrides the planned route and selects alternative stops that better balance delivery success, spoilage avoidance, and travel penalties under the evolving conditions.

To further stabilize learning, reward shaping is introduced where higher rewards are provided to the RL agent when its decisions remain aligned with the optimized route under normal conditions. This dual-mode learning structure allows the system to prioritize global optimality while retaining adaptability to real-world uncertainties.

3.4.1 Hybrid Learning Mechanism

At the start of each delivery episode, the system computes an optimized initial delivery route R^* using a traditional optimization solver based on total travel distance minimization, delivery demands, and initial shelf-life values. This optimized path serves as the primary routing plan.

During execution, the vehicle follows R^* sequentially. At each stop j , the environment observes the realized travel time t_{ij} , stochastic delay δ_j , and any product temperature deviation D_{jk} . As long as these remain within their respective acceptable thresholds, i.e.,

$$\delta_j \leq \Delta_\delta, \quad \text{and} \quad D_{jk} \leq \Delta_D \quad \forall k,$$

the predefined route R^* is maintained. This allows the system to exploit the global optimality of the offline solution when conditions are stable.

If either of the thresholds is violated, the system transitions into RL-controlled adaptive mode. Here, a Deep Q-Network (DQN) agent actively selects the next delivery stop based on current state s , considering updated inventory levels I_k , remaining shelf-life L_k , elapsed time τ , and delivery demand fulfillment. The RL agent receives reward signals that balance multiple objectives:

$$r(s, a) = R_{\text{del}} - R_{\text{tr}} - R_{\text{sp}},$$

where R_{del} is the delivery reward, R_{tr} penalizes travel and delay times, and R_{sp} penalizes freshness loss.

3.4.2 Hybrid Algorithmic Formulation

The overall hybrid learning process is summarized in Algorithm 3.

Algorithm 3 Hybrid Optimization-Guided Adaptive Deep Q-Learning

```
1: Initialize: Replay buffer  $\mathcal{M}$ , DQN parameters  $\theta$ , target network  $\theta^- \leftarrow \theta$ , predefined
   optimal path  $R^*$ , episode counter episode  $\leftarrow 1$ 

2: while episode  $\leq$  MaxEpisodes do

3:   Reset environment: inventory  $I_k$ , shelf-life  $L_k$ , warehouse location  $i = 0$ , time  $\tau = 0$ ,
   path index  $p = 1$ , visited set  $\mathcal{V} \leftarrow \{0\}$ .

4:   Set done  $\leftarrow$  False.

5:   while not done do

6:     if  $\delta_j \leq \Delta_\delta$  and  $D_{jk} \leq \Delta_D$  for all  $k$  (nominal condition) then

7:       Move to next stop  $a = R^*[p]$ , update state variables, set  $p \leftarrow p + 1$ .

8:     else

9:       Select  $a = \arg \max_{a'} Q_\theta(s, a')$  (with  $\epsilon$ -greedy).

10:      Execute  $a$ : observe  $t_{ij}$ ,  $\delta_j$ ,  $D_{jk}$ , update inventory, shelf-life, time  $\tau$ .

11:      Compute reward  $r(s, a)$ ; store transition  $(s, a, r, s')$  in  $\mathcal{M}$ .

12:      Sample minibatch, compute targets  $y_i$ , minimize loss  $L(\theta)$ , periodically up-
      date  $\theta^- \leftarrow \theta$ .

13:    end if

14:    if termination conditions met then

15:      done  $\leftarrow$  True

16:    end if

17:  end while

18:  episode  $\leftarrow$  episode + 1

19: end while
```

This chapter introduced the role of RL in dynamic logistics decision-making for perishable goods delivery, highlighting its adaptability compared to static optimization approaches. We formulated the RL model with a framework, defining its state space, action space, transition dynamics, and reward structure, and implemented it using a deep Q-Network (DQN) to handle large, stochastic state-action spaces. The chapter also presented the Hybrid Model (HM), which integrates an offline optimization-based route plan with an online RL agent for real-time adaptability. Under nominal conditions, the system follows the pre-optimized route, while disruptions trigger the RL-based adaptive control to balance delivery efficiency, freshness preservation, and travel cost minimization. This combined approach leverages the global optimality of classical optimization with the real-time responsiveness of RL, ensuring robust performance under uncertainty.

Chapter 4

Implementation, Experiments, and Results

The goal of the implemented RL-based, optimization-based, and hybrid approaches is to evaluate their performance in optimizing delivery routes, maintaining product freshness, and minimizing delays and costs during the transportation of perishable products. The synthetic data generated serves as a benchmark for both RL-based and optimization-based models, enabling a comprehensive comparison of these approaches.

4.1 IoT-based Sensors for Real-Time Monitoring

To facilitate data collection, Tive Tags, a specialized type of IoT-enabled data logger designed to track temperature fluctuations throughout the transportation process, are uti-

lized [44]. These Tags offer an efficient and non-intrusive solution for monitoring perishable goods without requiring complex installation procedures. Tive Tags, which are thin, flexible labels supporting end-to-end temperature monitoring of sensitive products, are suited to the data collection needs due to their ease of use and cost-effectiveness. The process begins with activation, which involves affixing the Tive Tag to the target location or shipment and initiating it by tapping with a Near-Field Communication (NFC)-enabled smartphone. Once activated, the data logging phase is started, during which temperature data is recorded at regular intervals which was set for every 5 mins throughout the monitoring period. Finally, data retrieval is conducted by tapping the Tag again to upload the recorded data to the Tive cloud platform via the mobile application. A comprehensive and secure audit trail of temperature data is ensured by this approach, which is crucial for analyzing the environmental conditions experienced during transit.

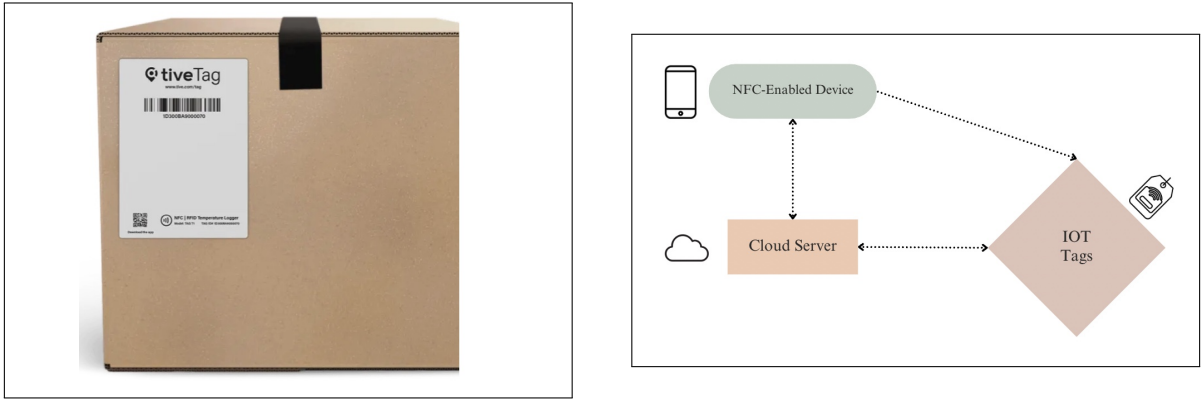


Figure 4.1: Left: Tive Tag affixed to a shipment for temperature monitoring. Right: Flowchart illustrating the data movement from the Tive Tag to the cloud and mobile application.

4.1.1 Synthetic Data Generation for Model Evaluation

To assess the robustness and performance of RL, optimization-based, and hybrid models, a synthetic dataset was generated, simulating real-world operational scenarios involving perishable goods. This dataset is based on IoT sensor data as well as assumptions about delivery operations and uncertainties commonly encountered in logistics.

The synthetic delivery network comprises a central warehouse, represented by node 0, and a set of $N = 10$ delivery stops inspired by a realistic last mile delivery in Toronto, Ontario. Each stop $i \in \mathcal{N} \setminus \{0\}$ is associated with a base travel distance d_{0i} from the warehouse, where:

$$d_{0i} \sim \mathcal{U}(d_{\min}, d_{\max})$$

Additionally, delays are introduced at each stop $i \in \mathcal{N}$, which simulate operational disruptions like traffic congestion, roadblocks, or handling issues. These delays are modeled as random variables:

$$\delta_i \sim \mathcal{U}(\delta_{\min}, \delta_{\max})$$

A set of perishable products, denoted as $\mathcal{K} = \{1, 2, \dots, K\}$, is included in the synthetic data. Each product $k \in \mathcal{K}$ has an initial shelf life L_k^0 , representing the maximum time the product remains fresh under ideal temperature conditions. The shelf life is perturbed by a random factor:

$$\epsilon_k \sim \mathcal{U}(-\alpha, \alpha)$$

which accounts for variations in freshness and spoilage rates:

$$L_k = L_k^0(1 + \epsilon_k)$$

Each product also has a required minimum shelf life R_k , ensuring its viability upon delivery. These parameters help enforce product integrity during transportation and are considered in the optimization and RL models.

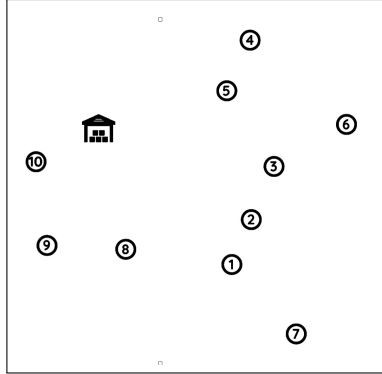


Figure 4.2: Network structure with N delivery stops and a warehouse.

This synthetic dataset ensures that both RL and optimization models are tested under comparable, real-world-inspired conditions, including delays, travel distances, and product freshness decay. In the next section, we present the results of applying these models to the synthetic data, followed by scenario testing and sensitivity analysis on optimization models. These results will allow us to evaluate how each approach performs under various opera-

tional uncertainties, and provide insights into the effectiveness of hybrid RL-optimization models for perishable goods logistics.

4.1.2 Experimental Results

The results from the models analyzed reveal several important insights into the performance of various logistics optimization strategies. The total transportation time for each model variant under normal and disrupted conditions is presented in Table 4.1.

Table 4.1: Total transportation time (in hours) for each model variant under normal and disrupted conditions

Model	Total Time (hours)
Deterministic	9.31
Stochastic Programming (SP)	9.42
Adaptive	9.80
Hybrid Model	10.45
Distributionally Robust Optimization (DRO)	10.50
Reinforcement Learning	11.0
Robust Optimization (RO)	15.50

The deterministic model performs with a total transportation time of 9.31 hours, achieving the most optimal delivery time. It assumes a perfect scenario without any disruptions, and the delivery route is predetermined. This model works well under stable, disruption-free conditions, ensuring the most efficient use of time. However, real-world logistics often

involve uncertainties like roadblocks or delays, which this model does not consider. The stochastic programming (SP) model performs slightly worse, with a total time of 9.42 hours. This model accounts for some uncertainty in the form of potential delays and varying travel times. However, despite incorporating some level of uncertainty, it still performs quite similarly to the deterministic model. This slight increase in time reflects the model's attempt to factor in randomness but still adheres to a largely static approach, meaning it doesn't adapt to major disruptions dynamically unless there is prior knowledge of the probability distribution functions of uncertain variables. The adaptive model incorporates dynamic decision-making, which allows it to respond to environmental factors such as temperature fluctuations and delays. It performs with a total time of 9.80 hours, which is slightly worse than the deterministic and SP models. This indicates the model's ability to balance adaptability with efficiency. The increase in time compared to the baseline models reflects the trade-off of maintaining flexibility in responding to real-time disruptions while still prioritizing efficiency in ideal conditions.

The hybrid model combines the optimal route generation from the deterministic model with reinforcement learning (RL) to adapt dynamically in response to disruptions. Under normal conditions, it matches the performance of the deterministic model with a total time of 9.31 hours, showing that it can follow the optimal path when no disruptions occur. However, when a disruption such as a roadblock or temperature deviation increase after a certain limit is introduced, the hybrid model's performance increases to 10.45 hours. This increase indicates that while the model is capable of adjusting its route dynamically, the real-time decision-making comes with a slight penalty in terms of time and distance.

The distributionally robust optimization (DRO) model, which is designed to handle worst-case uncertainty, results in a total time of 10.50 hours. This model takes a conservative approach by adding more time to the delivery process in order to account for all possible uncertainties. Although this ensures reliability in the face of potential disruptions, it sacrifices efficiency compared to models like the adaptive and hybrid models, which balance efficiency and flexibility more effectively. The reinforcement learning (RL) model, while capable of adapting to disruptions, performs with a total time of 11.0 hours. This outcome highlights the RL model’s tendency to prioritize short-term rewards over long-term efficiency. The RL model is highly flexible and adapts well to real-time disruptions, but this flexibility comes at the cost of potentially suboptimal decision-making when there are no significant disruptions. The RL model often chooses routes that appear optimal in the short term but may lead to inefficiencies when there are no major disruptions.

Finally, the robust optimization (RO) model, which is the most conservative approach, results in a total transportation time of 15.50 hours. The RO model explicitly considers the worst-case scenario by adding significant buffers to account for uncertainty. While this ensures the model’s reliability in highly uncertain conditions, it leads to a substantial increase in time and distance. This approach is highly reliable but inefficient, making it less suitable for dynamic environments where both efficiency and adaptability are required.

4.2 Scenario Generation for the proposed Optimization Model

To systematically evaluate the robustness and practical effectiveness of the proposed adaptive optimization model under real-world uncertainties, we conducted an extensive scenario-based testing framework. The purpose was to simulate diverse operational conditions that fresh food logistics systems routinely encounter, such as random handling delays and ambient temperature fluctuations during transportation. Our approach involved mathematically generating controlled perturbations around key parameters, ensuring that the variability introduced remained statistically justified and representative of actual logistics challenges. Specifically, scenarios were constructed by introducing stochastic variations in two critical parameters: travel delays and ambient temperature shifts. The parameter ranges and probability distributions were selected based on operational data from logistics studies, thermal stability characteristics of common fresh produce, and standard deviation estimates reflective of real-world variability. Gaussian distributions were chosen due to their well-known properties under the central limit theorem, enabling a realistic aggregation of small environmental factors, whereas truncated Gaussians for delay magnitudes ensured non-negativity while preserving natural spread. Each delivery location $i \in \mathcal{N}$ independently triggered a potential delay occurrence modeled as a Bernoulli random variable:

$$\mathbb{P}(\text{Delay at stop } i) = p_d,$$

where $p_d = 0.2$ represents a 20% chance of encountering a delay. If a delay was realized, its magnitude δ_i was drawn from a truncated Gaussian distribution:

$$\delta_i \sim \max\{0, \mathcal{N}(\mu_d, \sigma_d^2)\},$$

with mean $\mu_d = 0.5$ hours and standard deviation $\sigma_d = 0.2$ hours. The truncation at zero ensures that physically infeasible negative delays are avoided, maintaining the realism of operational disturbances. In parallel, ambient temperature shifts were modeled as independent continuous random variables at each delivery stop j and for each product $k \in \mathcal{K}$. The temperature change τ_{jk} was sampled from a standard Gaussian distribution:

$$\tau_{jk} \sim \mathcal{N}(0, \sigma_\tau^2),$$

with $\sigma_\tau = 1.0^\circ\text{C}$. Fifty independent scenarios were generated under this framework, each representing a distinct realization of travel delays and ambient temperature shifts. For each scenario, the adaptive optimization model was re-solved, and the key outcomes — total travel time, cumulative temperature deviations, and slack usage — were recorded. This repeated solving approach provided empirical insight into the distribution of model performance across random environmental perturbations, allowing us to assess robustness in both average and worst-case senses.

We visualized the trade-off between total trip time and cumulative freshness deviation across all tested scenarios. The resulting scatter plot, shown in Figure 4.3, illustrates the

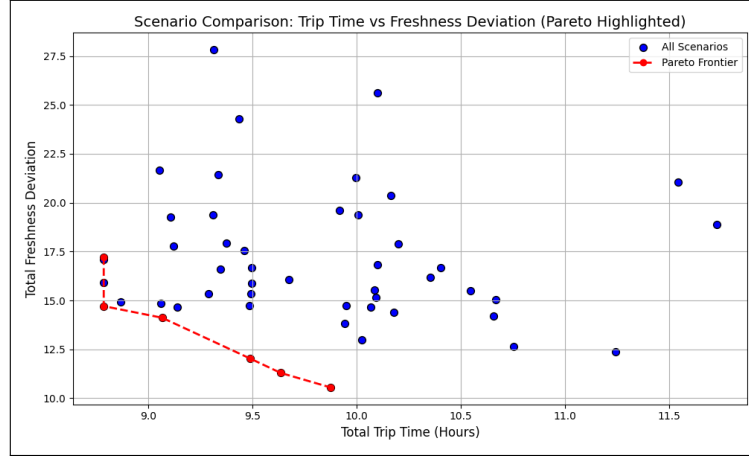


Figure 4.3: Scenario-wise comparison of the adaptive optimization model’s performance in terms of total trip time versus total freshness deviation.

bi-objective performance of the adaptive optimization model, where each blue dot represents a distinct scenario outcome generated from randomized delay and temperature perturbations. The x-axis denotes the total trip time (in hours), while the y-axis indicates the cumulative absolute temperature deviation across all stops and product types.

A Pareto frontier is overlaid in red, connecting the set of non-dominated solutions—i.e., those for which no other scenario yields both lower trip time and lower freshness deviation simultaneously. These Pareto-optimal solutions reveal the best-achievable trade-offs under the model’s adaptive control framework. Notably, the frontier is concave and downward-sloping, reflecting the classical tension between rapid delivery and tight temperature control. Scenarios with shorter trip durations (left side of the plot) tend to exhibit higher freshness deviation, suggesting more aggressive routing with less thermal stabilization. In contrast, points near the lower-right corner demonstrate better temperature adherence but

at the expense of longer routes, possibly due to conservative pacing or more temperature-conducive paths.

This distribution confirms the model’s capacity to balance competing objectives dynamically. The dispersion of the blue points around the frontier also highlights the inherent variability in outcomes under uncertain conditions—validating the need for an adaptive formulation. The frontier itself serves as a decision-support tool, enabling logistics planners to select a delivery policy that aligns with operational priorities (e.g., prioritizing freshness over speed in sensitive SCs). Furthermore, the tight clustering near the lower-left segment of the frontier suggests that moderate increases in delivery time can yield substantial gains in temperature stability, an insight that could inform risk-aware transportation planning for perishable goods.

4.2.1 Scenario based comparative analysis between a deterministic model and the proposed adaptive model

We executed 50 randomized scenarios using Gaussian-distributed travel delays and ambient temperature changes, applying both models to each scenario. Two key performance metrics were assessed: total travel time and total freshness deviation. Figure 4.4 presents the travel time (in hours) across 50 scenarios for both models. On the x -axis, we plot the scenario index, while the y -axis indicates the total route time. The adaptive model

consistently incurs slightly higher travel times, ranging approximately from 12.5 to 15.8 hours across scenarios, compared to 10.5 to 13.5 hours in the deterministic model. This increase reflects the routing flexibility required to maintain product temperature closer to ideal values, suggesting a deliberate trade-off for improved freshness.

Figure 4.5 illustrates the cumulative temperature deviation from the ideal across all stops

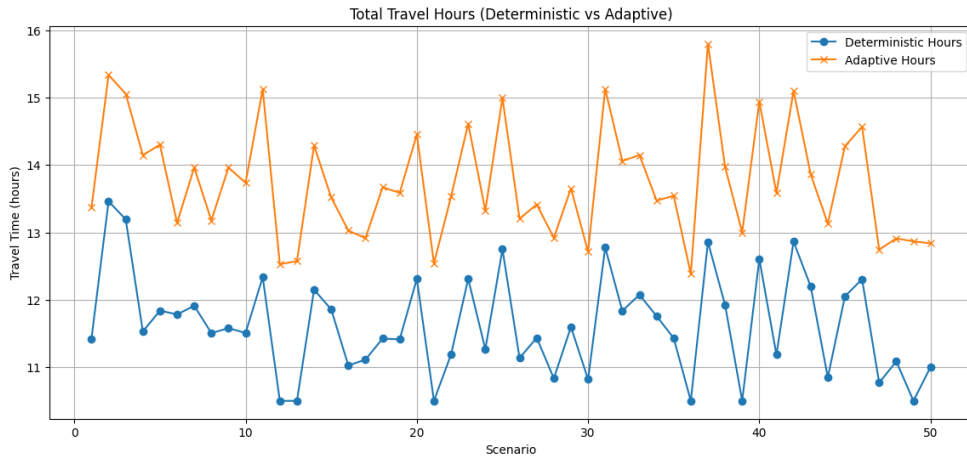


Figure 4.4: Total Travel Hours (Deterministic vs Adaptive)

and products. The deterministic model exhibits a constant total deviation of approximately 121°C across all scenarios over 10 stops for four produce, reflecting the lack of adaptive correction. In contrast, the adaptive model significantly reduces the deviation (to between 13 and 30°C), demonstrating its ability to dynamically regulate internal temperature despite external fluctuations.

The deterministic model optimizes solely for travel efficiency, resulting in minimal route time but large deviations from temperature requirements, risking spoilage and product degradation. On the other hand, the adaptive model integrates a temperature control

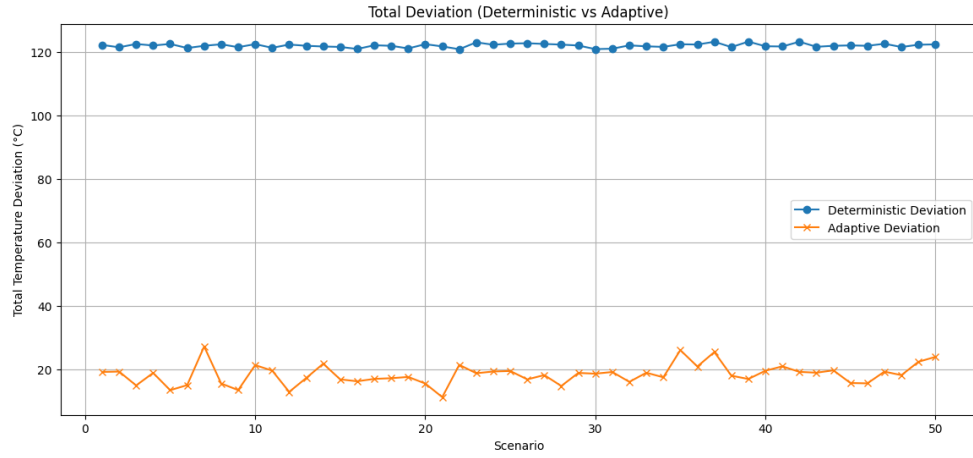


Figure 4.5: Total Freshness Deviation (Deterministic vs Adaptive)

mechanism into the optimization, accepting slightly longer travel routes to minimize the temperature deviation penalty. This dual-objective approach is beneficial for logistics involving perishable items, as it ensures compliance with freshness standards while maintaining operational feasibility. Notably, the adaptive model achieves up to an 80% reduction in freshness deviation on average, validating its practical advantage in cold-chain logistics and food supply networks.

4.2.2 Shelf life comparative analysis between a deterministic model and the proposed adaptive model

In addition, to evaluate the performance of the proposed adaptive and deterministic models with respect to product shelf life, we also conducted an experiment using a single representative product item, apple, chosen for its extensive study in post-harvest research and

well-documented shelf life characteristics. Apples typically exhibit an average shelf life of approximately 30 days when stored at an ideal temperature of 5°C, which we assumed as the initial product temperature at the start of the delivery route. The modeled delivery trip consisted of 10 stops, including the warehouse, simulating a realistic logistic scenario for the distribution of fresh produce. In this example, temperature variations during transit were allowed within a range of $\pm 3^{\circ}\text{C}$ from the ideal, reflecting typical fluctuations encountered in cold-chain transportation. Shelf life was estimated at the completion of the route using a Q_{10} temperature coefficient of 2, which implies that the degradation rate of the product doubles for every 10 ° C increase in temperature above the ideal. The adaptive model maintained an average product temperature of 6.23°C during transit, resulting in an estimated shelf life of 27.54 days. Conversely, the deterministic model, with a higher average temperature of 8.68°C, yielded a reduced shelf life of 23.25 days. These findings highlight the superior ability of the adaptive model to preserve freshness of the produce by dynamically mitigating temperature deviations, thus improving shelf life relative to the deterministic baseline.

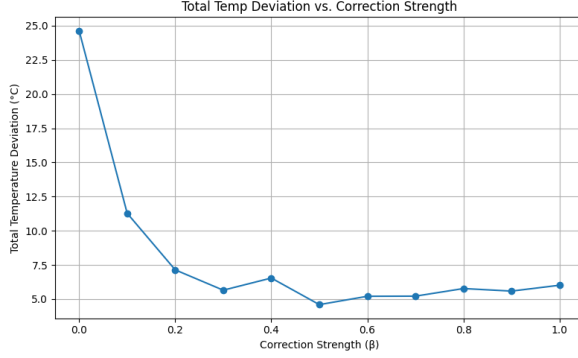
4.3 Parameter Tuning for the proposed adaptive optimization Model

To evaluate the performance of the adaptive temperature control model in cold-chain delivery, we also conducted sensitivity analyses on two key parameters: correction strength

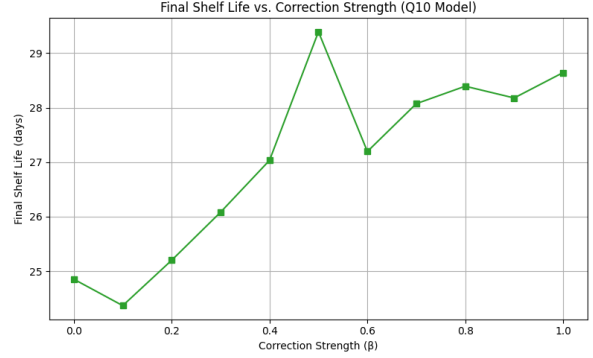
(β) and ambient temperature shifts (τ_{ik}). These govern how actively the system mitigates ambient disturbances and directly influence product freshness across the delivery route.

The model assumes that perishable goods must be maintained near 5°C, within an allowable range of 2°C to 8°C, with shelf life degradation modeled using the Q10 function ($Q_{10} = 2$) and a reference shelf life of 30 days. Ambient disturbances are introduced through a uniform distribution $\mathcal{U}[0, 1]$ for the correction factor β , and via Gaussian noise with varying standard deviation σ_τ for ambient temperature shifts τ_{ik} . Figure 4.6 illustrates the impact of increasing correction strength. As shown in Figure 4.6a, the total temperature deviation decreases sharply with increasing β , falling from approximately 25.5°C at $\beta = 0.0$ to around 7.5°C at $\beta = 0.4$, beyond which it stabilizes around 5–6°C. This indicates that even modest correction significantly reduces cumulative thermal deviation. The diminishing returns beyond $\beta = 0.4$ reflect the system’s limited ability to counteract random, bidirectional fluctuations through stronger correction—especially when successive stops experience opposing temperature shifts.

Figure 4.6b shows that final shelf life increases with β , peaking near 29.5 days around $\beta = 0.5$, then plateauing. The non-monotonic behavior suggests that overly aggressive correction can lead to minor overcompensations, reducing thermal stability. This aligns with practical cold-chain systems, where energy-efficient controllers or passive cooling systems avoid abrupt changes that may cause oscillations or stress refrigeration units. Thus, a moderate correction strength ($\beta \in [0.3, 0.5]$) offers a practical trade-off between energy cost, control effort, and freshness preservation.



(a) Total temperature deviation vs. correction strength (β).



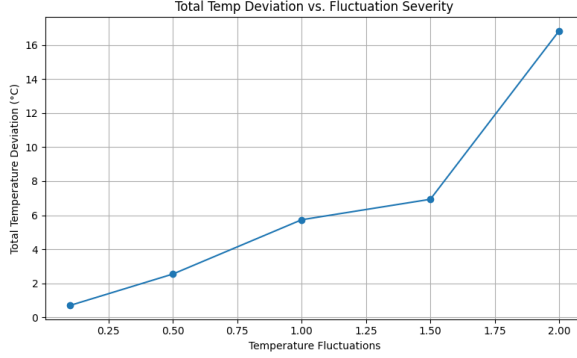
(b) Final shelf life vs. correction strength (β).

Figure 4.6: Sensitivity analysis of correction strength β : (a) Total temperature deviation and (b) final shelf life.

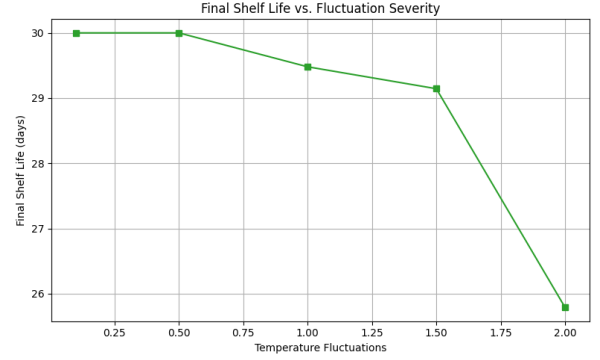
Figure 4.7 presents the model's response to increasing ambient temperature shifts (τ_{ik}), with β fixed. In Figure 4.7a, the total temperature deviation rises monotonically from 1°C at $\tau_{ik} = 0.25$ to 16°C at $\tau_{ik} = 2.0$, indicating the system's inability to fully correct large, rapid disturbances even at maximum correction strength. In real-world delivery conditions, this may correspond to frequent door openings, poor insulation, or transit through varying climate zones—factors that cause sharp and compounding thermal stress.

Figure 4.7b shows the shelf life degradation resulting from these deviations. At low τ_{ik} , shelf life remains near the reference of 30 days, but it drops sharply at higher τ_{ik} , reaching 26 days at $\tau_{ik} = 2.0$. Given that shelf life decay under the Q10 model is exponential in temperature, even small thermal violations have amplified effects. This mirrors practical scenarios where highly sensitive goods (e.g., vaccines, seafood, dairy) require more than

algorithmic control—such as insulated containers, real-time temperature tracking, or priority delivery routing—to avoid spoilage.



(a) Total temperature deviation vs. fluctuation severity (τ_{ik}).



(b) Final shelf life vs. fluctuation severity (τ_{ik}).

Figure 4.7: Sensitivity analysis of fluctuation severity τ_{ik} : (a) Total temperature deviation and (b) final shelf life.

Together, these results emphasize that moderate correction strengths are generally sufficient in stable environments, allowing for energy-efficient and effective temperature control. However, in high-volatility scenarios (e.g., long-haul deliveries with varying ambient exposure or cross-border logistics), controller effectiveness saturates. In such cases, passive insulation, proactive rerouting, and IoT-based real-time monitoring become necessary to preserve quality. These findings suggest that operational strategies should jointly optimize controller settings (β) and environmental resilience (mitigating τ_{ik}) to ensure freshness and safety in cold-chain logistics.

Chapter 5

Conclusions and Future Work

This research has systematically investigated the application of advanced optimization and reinforcement learning models to address the critical challenges of inefficiency and spoilage within fresh produce supply chains. The primary objective was to demonstrate that, by integrating real-time data from IoT sensors with intelligent decision-making frameworks, it is possible to significantly improve operational efficiency and preserve the quality of perishable goods during transit. The core contribution of this work lies in the formulation and evaluation of a novel adaptive optimization model and a hybrid reinforcement learning framework. Through extensive simulation and comparative analysis, this thesis provides compelling evidence supporting this hypothesis.

The findings confirm that the ideal model is contingent on the operational environment and priorities. The deterministic model established an optimal baseline under ideal conditions, achieving a total travel time of 9.31 hours but failing to mitigate freshness loss,

resulting in a product shelf life of only 23.25 days. In contrast, the proposed adaptive optimization model, by leveraging real-time feedback, achieved a remarkable 80% average reduction in freshness deviation. This came at a justifiable increase in travel time to 9.80 hours but substantially extended product shelf life to 27.54 days, proving its effectiveness in dynamic environments. For scenarios characterized by stability punctuated by severe disruptions, the hybrid model offers the most pragmatic solution by leveraging deterministic optimality in normal conditions and the reactive agility of reinforcement learning to navigate unforeseen events. A purely reinforcement learning approach, while highly adaptive, is most appropriate in highly stochastic and unpredictable environments where pre-computed plans have little value. In conclusion, this research confirms that a shift from static planning to dynamic, data-driven strategies can create a more efficient, resilient, and sustainable fresh produce supply chain.

5.1 Limitations of the Study

While this research provides valuable insights, it is important to acknowledge its limitations, which offer a foundation for subsequent work. The models were primarily evaluated using a comprehensive synthetic dataset, which was inspired by the real operation of the last mile delivery company in Toronto, Ontario. While designed to reflect realistic operational parameters, it was supplemented by only a limited set of data from live IoT sensors. This approach, though enabling controlled and repeatable experiments, does not fully capture the spectrum of complexity and noise inherent in large-scale, real-world logistics data.

A more extensive data collection effort using live IoT sensors would be required for full validation. Furthermore, the study’s scenario analysis, though covering 50 distinct simulations, represents a finite subset of possible real-world events and was not tested against more complex or cascading disruptions, such as simultaneous vehicle breakdowns or correlated delays. The investigation of RL was also primarily centered on a Deep Q-Network (DQN) architecture, and a more exhaustive exploration of alternative RL algorithms and reward-shaping functions could yield policies with improved performance. Finally, the supply chain was modeled as a single-depot, single-vehicle routing problem, a common simplification that does not encompass the complexities of multi-depot, multi-vehicle fleet coordination standard in large-scale distribution networks.

5.2 Directions for Future Research

Building upon the findings and limitations of this thesis, several promising directions for future research emerge. The most critical next step is to achieve Scalability and Deployment by transitioning from simulation to a pilot program with a logistics partner. Testing the models on real logistics networks, such as regional food hubs, would enable validation against real-world operational data and performance evaluation at scale under diverse demand patterns.

Future work must also focus on deeper Integration with Real-Time Data Streams. This involves incorporating live feeds from GPS, traffic APIs, and advanced temperature sensors to create a dynamic digital twin of the logistics network, allowing for more accurate,

real-time updates to travel times, delays, and temperature deviations. The current single-vehicle model should be extended to address Multi-Vehicle Coordination. This involves evolving the problem into a multi-agent system, where a fleet of vehicles coordinates deliveries while sharing shelf-life and route constraints. This is a natural fit for Multi-Agent Reinforcement Learning (MARL), which could optimize system-wide efficiency. Concurrently, Cold Chain Compliance Automation should be pursued by enabling autonomous temperature control decisions using sensor feedback to minimize spoilage without manual intervention. A final, vital avenue for future research is the explicit inclusion of Sustainability Objectives, such as minimizing carbon emissions by incorporating detailed fuel consumption models, to align the logistical optimization with pressing environmental and corporate sustainability goals.

References

- [1] Dwi Agustina, C.K.M. Lee, and Rajesh Piplani. Vehicle scheduling and routing at a cross docking center for food supply chains. *International Journal of Production Economics*, 152:29–41, 2014.
- [2] R. Akila. A deep reinforcement learning approach for optimizing inventory management in the agri-food supply chain. *Semantic Scholar Preprint*, 2024.
- [3] Svante Arrhenius. Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren. *Zeitschrift für Physikalische Chemie*, 4:226–248, 1889.
- [4] Myo Min Aung and Yoon Seok Chang. *Cold Chain Management*. Springer Series in Advanced Manufacturing. Springer Nature Switzerland, 2023.
- [5] M. Balaji, K. Arshinder, et al. Modeling the causes of food wastage in indian perishable food supply chain. *Resources, Conservation and Recycling*, 114:153–167, August 2016.
- [6] Aharon Ben-Tal and Arkadi Nemirovski. Robust convex optimization. *Mathematics of operations research*, 23(4):769–805, 1998.

- [7] Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of uncertain linear programs. *Operations research letters*, 25(1):1–13, 1999.
- [8] Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88:411–424, 2000.
- [9] Dimitris Bertsimas, Melvyn Sim, and Meilin Zhang. Adaptive distributionally robust optimization. *Management Science*, 65(2):604–618, 2019.
- [10] H. Chen, Z. Chen, F. Lin, and P. Zhuang. Effective management for blockchain-based agri-food supply chains using deep reinforcement learning. *IEEE Transactions on Industrial Informatics*, 17(9):1–10, 2021.
- [11] Ji-Young Choi, Hyeon-Jeong Lee, Jeong-Seok Cho, Yeong-Min Lee, Jin-Ho Woo, and Kwang-Deog Moon. Prediction of shelf-life and changes in the quality characteristics of semidried persimmons stored at different temperatures. *Food Science and Biotechnology*, 26(5):1255–1262, 2017.
- [12] Fabrizio Dabbene, Paolo Gay, and Nicola Sacco. Optimisation of fresh-food supply chains in uncertain environments, part i: Background and methodology. *Biosystems Engineering*, 99(3):348–359, 2008.
- [13] George B. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3–4):197–206, 1955.

- [14] Maria Cristina do Nascimento Nunes, Michela Nicometo, Jean-Pierre Emond, Roberta Bianca Melis, and Ilias Uysal. Improvement in fresh fruit and vegetable logistics quality: berry logistics field studies. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2017):20130307, 2014.
- [15] Kannan Govindan, Ahmad Jafarian, Reza Khodaverdi, and Kannan Devika. Two-echelon multiple-vehicle location–routing problem with time windows for optimization of sustainable supply chain network of perishable food. *International Journal of Production Economics*, 152:9–28, 2014.
- [16] Abdelhalim Hiassat, Ali Diabat, and Iyad Rahwan. A genetic algorithm approach for location-inventory-routing problem with perishable products. *Journal of Manufacturing Systems*, 42:93–103, 2017.
- [17] H. Hu, J. Xu, M. Liu, and MK. Lim. Vaccine supply chain management: An intelligent system utilizing blockchain, iot and machine learning. *Journal of Business Research*, 2023.
- [18] AK. Kalusivalingam and A. Sharma. Enhancing supply chain visibility through ai: Implementing neural networks and reinforcement learning algorithms. *International Journal of Artificial Intelligence and Machine Learning*, 2020.
- [19] A. Kannagi and PK. Goswami. Reinforcement learning to manage energy efficient supply chains. *IEEE Xplore*, 2024.

- [20] V. Kavididevi and S. K. Monikapreethi. Iot-enabled reinforcement learning for enhanced cold chain logistics performance in refrigerated transport. *IEEE Access*, 2024.
- [21] A. Kavooosi, R. Tavakkoli-Moghaddam, and H. Sajedi. Dynamic pricing and inventory control of perishable products by a deep reinforcement learning algorithm. *Expert Systems with Applications*, 2025.
- [22] N Khan, S Ahmad, S Raza, and A Khan. Cost effective route optimization for dairy product delivery. *KJMR*, 2025.
- [23] Mustafa Kirci, Ola Isaksson, and Roland Seifert. Managing perishability in the fruit and vegetable supply chains. *Sustainability*, 14(9):5378, 2022.
- [24] Ramesh Krishnan, K. Arshinder, and Renu Agarwal. Robust optimization of sustainable food supply chain network considering food waste valorization and supply uncertainty. *Computers & Industrial Engineering*, 171:108499, 2022.
- [25] Anna Lamberty and Judith Kreyenschmidt. Ambient parameter monitoring in fresh fruit and vegetable supply chains using internet of things-enabled sensor and communication technology. *Foods*, 11(6):1777, 2022.
- [26] J. Lee, Y. Shin, and I. Moon. A hybrid deep reinforcement learning approach for a proactive transshipment of fresh food in the online–offline channel system. *Computers & Industrial Engineering*, 2024.
- [27] P Liu, A Hendalianpour, and J Razmi. A solution algorithm for integrated production-

- inventory-routing of perishable goods with transshipment and uncertain demand. *Complex Intelligent Systems*, 2021.
- [28] Heng Luo, Minjie Zhu, Sengang Ye, Hanping Hou, Yong Chen, and Larissa Bulysheva. An intelligent tracking system based on internet of things for the cold chain. *Internet Research*, 26(2):435–445, 2016.
- [29] N. Mhaisen, N. Fetais, A. Erbad, and A. Mohamed. To chain or not to chain: A reinforcement learning approach for blockchain-enabled iot monitoring applications. *Future Generation Computer Systems*, 2020.
- [30] MirMohammad Musavi and Ali Bozorgi-Amiri. A multi-objective sustainable hub location-scheduling problem for perishable food supply chain. *Computers & Industrial Engineering*, 113:766–778, 2017.
- [31] Abhijit Pal and Krishna Kant. Iot-based sensing and communications infrastructure for the fresh food supply chain. *IEEE Computer*, 51(2):76–80, 2018.
- [32] Amitangshu Pal and Krishna Kant. Smart sensing, communication, and control in perishable food supply chain. *ACM Transactions on Sensor Networks (TOSN)*, 16(1):1–41, 2020.
- [33] Zhibo Pang, Qiang Chen, Weili Han, and Lirong Zheng. Value-centric design of the internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion. *Information Systems Frontiers*, 17(2):289–319, 2015.

- [34] P. Pontrandolfo, A. Gosavi, and OG. Okogbaa. Global supply chain management: a reinforcement learning approach. *International Journal of Production Research*, 2002.
- [35] E. Puskás, Á. Budai, and G. Bohács. Optimization of a physical internet based supply chain using reinforcement learning. *European Transport Research Review*, 2020.
- [36] Yuzhuo Qiu, Jun Qiao, and Panos M. Pardalos. Optimal production, replenishment, delivery, routing and inventory management policies for products with perishable inventory. *Omega*, 82:193–204, 2019.
- [37] B. Rolf, I. Jackson, M. Müller, and S. Lang. A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, 2023.
- [38] M Selukar, P Jain, and T Kumar. Inventory control of multiple perishable goods using deep reinforcement learning for sustainable environment. *Sustainable Computing: Informatics and Systems*, 2022.
- [39] R. Singh, P. Mandal, and S. Nayak. Reinforcement learning approach in supply chain management: A review. In *AI-Driven Technologies in Supply Chain Management*. Wiley, 2024.
- [40] M.J. Sousa-Gallagher, A. Tank, and R. Sousa. Emerging technologies to extend the shelf life and stability of fruits and vegetables. In *The Stability and Shelf Life of Food*, pages 393–419. Elsevier, 2016.

- [41] Allen L Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5):1154–1157, 1973.
- [42] X Andy Sun and Álvaro Lorca. Robust optimization in electric power systems operations. *Integration of Large-Scale Renewable Energy into Bulk Power Systems: From Planning to Operation*, pages 227–258, 2017.
- [43] S. Taj, A. S. Imran, Z. Kastrati, S. M. Daudpota, R. A. Memon, and J. Ahmed. Iot-based supply chain management: A systematic literature review. *Internet of Things*, 24:100982, 2023.
- [44] Tive, Inc. Tive Tag, 2025. Accessed: 2025-06-02.
- [45] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.
- [46] Y. Yan, AHF. Chow, CP. Ho, YH. Kuo, and Q. Wu. Reinforcement learning for logistics and supply chain management: Methodologies, state of the art, and future opportunities. *Computers & Industrial Engineering*, 2022.
- [47] H. Yu. Research on fresh product logistics transportation scheduling based on deep reinforcement learning. *Journal of Food Quality*, 2022:1–12, 2022.
- [48] Y. Zou, Q. Gao, H. Wu, and N. Liu. Carbon-efficient scheduling in fresh food supply chains with a time-window-constrained deep reinforcement learning model. *Sensors*, 24(23):7461, 2024.

APPENDICES

.1 Appendix A: Shelf-Life Model Derivations

.1.1 A.1 Arrhenius-Based Shelf Life Model

The Arrhenius equation describes the temperature dependence of the rate constant k for a chemical or biological degradation process as:

$$k = k_0 \exp\left(\frac{-E_a}{RT}\right), \quad (1)$$

where:

- k is the rate constant at absolute temperature T [K],
- k_0 is the pre-exponential (frequency) factor [1/time],
- E_a is the activation energy [J/mol],
- R is the universal gas constant [8.314 J/mol · K],

- T is the absolute temperature [K].

Assuming spoilage follows first-order kinetics, the quality index Q (e.g., firmness, color, microbial count) of a perishable product decreases exponentially with time:

$$Q(t) = Q_0 \exp(-kt), \quad (2)$$

where:

- Q_0 is the initial quality index at time $t = 0$,
- $Q(t)$ is the quality at time t ,
- k is the degradation rate constant.

Let Q_{crit} be the minimum acceptable quality index below which the product is considered spoiled. The shelf life L_T at temperature T is the time when $Q(t) = Q_{\text{crit}}$:

$$Q_{\text{crit}} = Q_0 \exp(-kL_T) \quad (3)$$

$$\Rightarrow \ln \left(\frac{Q_{\text{crit}}}{Q_0} \right) = -kL_T \quad (4)$$

$$\Rightarrow L_T = \frac{1}{k} \ln \left(\frac{Q_0}{Q_{\text{crit}}} \right) \quad (5)$$

Note that $\ln(Q_0/Q_{\text{crit}})$ is a product-specific constant. Hence, shelf life is inversely proportional to k :

$$L_T \propto \frac{1}{k} \quad (6)$$

Substituting the Arrhenius expression for k :

$$L_T \propto \frac{1}{k_0 \exp(-E_a/RT)} \quad (7)$$

$$= \frac{1}{k_0} \exp\left(\frac{E_a}{RT}\right) \quad (8)$$

Now, define L_{T_0} as the shelf life at a reference temperature T_0 :

$$L_{T_0} = \frac{1}{k_0} \exp\left(\frac{E_a}{RT_0}\right) \quad (9)$$

Taking the ratio of shelf lives at two different temperatures T and T_0 :

$$\frac{L_T}{L_{T_0}} = \frac{\frac{1}{k_0} \exp\left(\frac{E_a}{RT}\right)}{\frac{1}{k_0} \exp\left(\frac{E_a}{RT_0}\right)} \quad (10)$$

$$= \exp\left(\frac{E_a}{R} \left(\frac{1}{T} - \frac{1}{T_0}\right)\right) \quad (11)$$

Therefore, the shelf life at temperature T can be written in terms of the known shelf life at reference temperature T_0 :

$$L_T = L_{T_0} \exp \left(\frac{E_a}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \right) \quad (12)$$

This formulation allows precise estimation of residual shelf life under varying temperatures using known values of activation energy E_a and reference shelf life L_{T_0} . It is widely used in cold chain logistics and food safety monitoring for perishable goods.

.1.2 A.2 Q10-Based Shelf Life Model

The Q10 model provides an empirical estimation of the change in reaction rate (and thus shelf life) per 10°C rise in temperature. It is defined as:

$$Q_{10} = \frac{k_{T+10}}{k_T}, \quad (13)$$

where k_T is the degradation rate at temperature T [°C]. Assuming first-order kinetics, shelf life is inversely proportional to the rate constant:

$$\frac{L_T}{L_{T+10}} = Q_{10} \quad \Rightarrow \quad L_T = L_{T_0} \cdot Q_{10}^{(T_0-T)/10}. \quad (14)$$

Thus, the shelf life at any temperature T is:

$$L_T = L_{T_0} \cdot Q_{10}^{(T_0-T)/10}, \quad (15)$$

where:

- L_{T_0} is the shelf life at reference temperature T_0 ,
- T and T_0 are in degrees Celsius [$^{\circ}\text{C}$],
- Q_{10} typically ranges from 1.5 to 3 for fresh produce.

This formulation is easier to implement in real-time logistics systems as it requires fewer product-specific parameters compared to the Arrhenius model. However, it assumes exponential spoilage behavior, which may not hold under rapid or fluctuating temperature conditions.

.2 Appendix B: IoT Device Specifications

The Tive Tag is a single-use, non-lithium NFC-enabled temperature logger designed for cold chain logistics. Its ultra-thin form factor and robust waterproof design make it particularly suitable for perishable food transport. Below are the technical specifications sourced from the official Tive documentation.

The Tive Tag’s ability to log temperature data in a secure, tamper-evident format without lithium content makes it compliant with international air cargo regulations. Data can be accessed using standard NFC-capable mobile devices, making it ideal for last-mile visibility and quality assurance.

Table 1: Tive Tag Specifications

Specification	Details
Operating Temperature Range	−30°C to +50°C
Temperature Accuracy	±0.5°C (−10°C to +50°C) ±1.0°C outside this range
Measurement Interval	Configurable (1 to 720 minutes)
Recording Capacity	4,864 readings (up to 50 days at 15-minute intervals)
Interface	NFC (ISO 14443 Type 2), optional UHF RFID (EPC Gen2)
Dimensions	95 mm × 125 mm × 1 mm
Water Resistance	IP67 (dustproof and waterproof)
Battery Type	Non-lithium, alkaline (approx. 40 mAh)
Battery Life	More than 1 year (storage and operation)
Certifications	FCC, CE, RoHS, EN 12830 (transport compliance)