A Hybrid Reinforcement Learning Framework for Optimizing Fuel Delivery Networks under Uncertainty

Khaled Mili[1*], Majdi Argoubi[2]

[1] Department of Quantitative Methods, College of Business, King Faisal University, Al-Ahsa 31982, Saudi Arabia
[2] Department of Quantitative Methods, University of Sousse, Sousse 4054, Tunisia

Corresponding Author Email: Kmili@kfu.edu.sa

ABSTRACT

Transportation logistics for fuel delivery face persistent challenges in routing under uncertain demand and complex operational constraints. This study addresses the gap between theoretical models and practical fuel distribution by introducing a hybrid framework that integrates Deep Reinforcement Learning (DRL), graph-based spatial reasoning, and deterministic constraint validation. The method combines Proximal Policy Optimization (PPO) with a graph neural architecture to capture spatial dependencies in vehicle routing while ensuring operational feasibility via constraint-checking mechanisms. The approach was evaluated on 300 synthetic problem instances across three network scales (10, 50, and 100 stations) and a real-world case study involving 38 gas stations and 6 vehicles in a regional fuel distribution system. Compared to a standard deep learning baseline and a Clarke-Wright heuristic, our method reduced operational costs by 7.2% and 9.9%, respectively. Constraint violations dropped from 6% with classical reinforcement learning to 1%, demonstrating improved feasibility. While we report averaged results over large instance sets, formal statistical significance testing remains a direction for future work. The proposed approach maintained robust performance under varying levels of demand uncertainty and produced feasible daily routing plans within 45 seconds, confirming their practical applicability. By integrating learning, spatial reasoning, and operational compliance, this research advances scalable and adaptive optimization for fuel delivery in uncertain and dynamic environments.

1. INTRODUCTION

Transportation logistics for fuel delivery represent a critical component of modern energy distribution systems, with significant implications for operational efficiency and environmental sustainability. The optimization of fuel delivery routes involves complex decision-making under uncertainty, as transportation planners must contend with variable demand patterns, fluctuating travel conditions, and strict operational constraints [1, 2].

Traditional approaches to transportation optimization have predominantly employed mathematical programming formulations with deterministic parameters. While these methods provide structured frameworks with provable optimality guarantees, they encounter significant limitations in real-world fuel delivery scenarios. As transportation network complexity increases, exact solvers become computationally prohibitive for time-sensitive operational decisions, often requiring hours to generate solutions that rapidly become obsolete in dynamic environments [3]. Furthermore, these approaches typically fail to account for the stochastic nature of fuel consumption patterns and travel times, leading to suboptimal or infeasible routes when implemented in practice.

Stochastic programming approaches have attempted to address uncertainty in transportation logistics through chance-constrained formulations and recourse strategies [4, 5]. However, these techniques face significant scalability limitations that restrict their practical implementation in large-scale transportation systems. This gap highlights the disconnect between theoretical transportation models and operational implementation requirements.

Recent advancements in computational approaches offer promising alternatives for transportation optimization. Deep Reinforcement Learning (DRL) scales effectively with problem size through learned policies mapping states to actions without exhaustive enumeration. Proximal Policy Optimization (PPO) has demonstrated effectiveness for transportation routing problems due to its sample efficiency and training stability [6]. Concurrently, Graph Neural Networks (GNNs) have emerged as powerful tools for capturing spatial relationships in transportation networks, enabling rich representations of connectivity patterns essential for efficient routing decisions [7].

Despite these advances, significant challenges remain in integrating these computational techniques into practical transportation logistics systems. Standard DRL approaches

struggle to capture the complex spatial dependencies inherent in transportation networks, while ensuring compliance with operational constraints remains problematic. Furthermore, existing research has typically addressed either the computational efficiency of solution generation or the quality of solutions under uncertainty, but rarely simultaneously within transportation contexts.

While DRL methods such as PPO offer scalable and adaptive policies, they often lack awareness of the spatial structure of the network, resulting in inefficient routing decisions in geographically complex environments. GNNs, on the other hand, capture spatial relationships effectively but are not inherently decision-making tools and require integration with control frameworks. Classical optimization approaches, although mathematically rigorous, are computationally intractable in large-scale stochastic settings and struggle to adapt to real-time operational variability. Therefore, none of these methods individually can simultaneously address the need for scalability, spatial reasoning, and constraint satisfaction in fuel delivery logistics.

This research addresses these challenges by developing an integrated framework that combines reinforcement learning, GNNs, and constraint validation mechanisms to optimize fuel delivery operations under uncertainty. The framework bridges the divide between theoretical models and practical logistics operations by leveraging the complementary strengths of learning-based and optimization-based techniques. To address these intertwined limitations, we propose a unified approach with the following key contributions:

- A comprehensive stochastic mathematical model of fuel delivery operations with deterministic equivalent transformations that enable practical computation while maintaining solution robustness.
- A novel hybrid framework combining PPO with GNNs to capture spatial dependencies within transportation networks, enhancing the representational power of the learning agent.
- A constraint validation mechanism that ensures generated solutions adhere to operational requirements, addressing a critical limitation of existing machine learning approaches to transportation optimization.

Through computational experiments and a case study of regional fuel distribution, this research demonstrates the practical benefits of the integrated approach for transportation logistics providers. The methodology achieves significant improvements in operational efficiency, constraint satisfaction, and robustness to uncertainty, offering a practical tool for daily logistics planning within fuel distribution networks.

The remainder of this paper is organized as follows: Section 2 reviews relevant transportation logistics literature; Section 3 presents the mathematical formulation of the stochastic fuel delivery problem; Section 4 details the integrated methodology; Section 5 evaluates the approach through computational experiments and a case study; and Section 6 concludes with findings and directions for future transportation research.

2. LITERATURE REVIEW

The optimization of vehicle routing in fuel delivery logistics represents a complex interdisciplinary challenge that intersects traditional operations research, emerging digital technologies, and industry-specific operational constraints. This literature review systematically examines the evolution of vehicle routing optimization from classical formulations to contemporary machine learning approaches, with particular emphasis on uncertainty management and the unique challenges inherent to fuel distribution networks. The analysis reveals critical gaps in current methodologies and establishes the theoretical foundation for developing integrated optimization frameworks that address the multifaceted nature of modern fuel delivery logistics.

2.1 Vehicle routing optimization in transportation logistics

The foundation of transportation logistics optimization rests upon decades of research in vehicle routing problems, beginning with classical formulations that established the mathematical framework for delivery optimization. However, the evolution from theoretical models to practical applications has revealed significant limitations in traditional approaches, particularly when applied to specialized domains such as fuel delivery. This section examines the progression from classical VRP formulations to contemporary challenges, highlighting the growing complexity of real-world constraints that traditional models struggle to address.

2.1.1 Classical VRP formulations

The optimization of vehicle routing forms a cornerstone of transportation logistics research, with significant implications for operational efficiency and resource utilization. The Capacitated Vehicle Routing Problem (CVRP), extensively reviewed in recent comprehensive surveys [8, 9], establishes the foundational framework for delivery optimization across transportation networks. This classical formulation models a fleet of vehicles servicing a set of customers with known demands while minimizing total travel cost, subject to vehicle capacity constraints. Contemporary research by Archetti and Speranza [10] demonstrates that even modest improvements in routing efficiency translate to substantial economic and environmental benefits across transportation systems.

However, classical CVRP formulations present significant limitations when applied to complex real-world scenarios. The deterministic assumptions underlying traditional models fail to capture the dynamic nature of modern logistics operations, where demand variability, traffic fluctuations, and operational disruptions are commonplace. Furthermore, the computational complexity of exact solution methods renders them impractical for large-scale networks, creating a fundamental trade-off between solution optimality and practical applicability.

2.1.2 Fuel-specific constraints and challenges

While classical VRP formulations provide a solid theoretical foundation, their application to fuel delivery operations reveals substantial gaps in addressing industry-specific requirements. The transition from general freight transportation to fuel distribution introduces unique operational complexities that fundamentally alter the optimization landscape, necessitating specialized modeling approaches that account for product compatibility, contamination prevention, and regulatory compliance.

Fuel delivery operations present unique challenges within the transportation optimization domain that distinguish them significantly from general freight transportation. Studies conducted by Li et al. [11] and Kumar et al. [12] identified

several industry-specific constraints, including compartmentalized vehicles, product contamination concerns, and evolving regulatory requirements. These operational realities significantly constrain the solution space and complicate optimization efforts. Critical analysis reveals that traditional VRP models inadequately represent fuel delivery-specific constraints, including multi-compartment vehicles, product compatibility matrices, contamination prevention protocols, and regulatory compliance requirements.

Recent studies by Baykasoğlu et al. [13] further characterized the fuel distribution problem as a specialized variant combining elements of multi-compartment vehicle routing with site-dependent constraints and dynamic time windows, creating particularly challenging optimization landscapes for transportation planners. The literature treats fuel delivery as a variant of general VRP rather than recognizing its distinct optimization landscape, leading to suboptimal solutions that fail to address industry-specific challenges effectively.

To address these operational complexities more effectively, recent research has explored the role of digital transformation in advancing vehicle routing solutions. Integrating emerging technologies such as IoT, real-time analytics, and machine learning has opened new avenues for overcoming traditional limitations in fuel distribution planning. The next section explores these advancements and their associated limitations.

2.1.3 Digital transformation in vehicle routing and its limitations

The recognition of traditional model limitations has driven researchers toward incorporating digital transformation elements into vehicle routing optimization. This technological integration represents an attempt to bridge the gap between theoretical models and practical operational requirements, yet it has also revealed new challenges in balancing technological capabilities with optimization efficiency.

Contemporary research extends traditional formulations to incorporate emerging operational constraints relevant to modern fuel distribution networks. Hu et al. [14] developed models accounting for driver working hours, vehicle compatibility with delivery locations, and carbon footprint considerations increasingly important in sustainable transportation. Their findings indicated that integrated approaches addressing multiple operational dimensions simultaneously yield superior results compared to sequential optimization techniques. Similarly, Lin et al. [15] demonstrated that accounting for heterogeneous fleet characteristics and real-time traffic data becomes increasingly crucial as transportation networks expand and integrate with smart city infrastructure.

Modern optimization approaches incorporate digital transformation elements affecting contemporary transportation logistics. Research by Bandara et al. [16] highlighted the integration of Internet of Things (IoT) sensors and real-time demand prediction systems into routing optimization frameworks. These technological advances enable dynamic route adjustments based on actual consumption patterns and traffic conditions, addressing limitations of traditional static optimization approaches. Chen et al. [17] further demonstrated how machine learning techniques can complement classical optimization methods, particularly in handling the increasing complexity and uncertainty in modern fuel distribution networks.

Despite these technological advances, a critical gap persists in the integration of digital tools with traditional optimization frameworks. Current approaches often treat digital components as peripheral additions rather than fully integrated parts of optimization processes, thus failing to exploit their potential for real-time adaptation and responsiveness to dynamic operational changes.

2.2 Approaches to uncertainty in transportation systems

The limitations of deterministic optimization models in capturing real-world variability have necessitated the development of sophisticated uncertainty management approaches. Transportation logistics, particularly fuel delivery operations, operate in inherently uncertain environments where demand fluctuations, traffic variability, and operational disruptions are the norm rather than the exception. This section examines the evolution of uncertainty management techniques from foundational stochastic programming approaches to contemporary robust optimization methods, analyzing their strengths, limitations, and applicability to fuel distribution networks.

2.2.1 Stochastic programming foundations

Uncertainty represents a fundamental challenge in transportation logistics, particularly in fuel distribution where demand volatility and travel time variability significantly impact operational efficiency. Stochastic optimization approaches have emerged as primary methodologies for addressing this uncertainty, with two predominant frameworks: chance-constrained programming and recourse models.

Chance-constrained methods, pioneered for transportation applications by Dror and Trudeau [18] and further developed by Belenguer et al. [4], ensure constraint satisfaction with specified probabilities, particularly useful for managing fuel availability while limiting route failures. These approaches transform probabilistic constraints into deterministic equivalents using quantile functions, providing theoretical guarantees at the cost of increased computational complexity. Sluijk et al. [2] specifically applied chance-constrained programming to vehicle routing with stochastic demands, optimizing cost efficiency while maintaining specified service levels in transportation networks.

2.2.2 Recourse strategies and dynamic adaptation

Building upon the foundational concepts of chance-constrained programming, recourse strategies introduce a dynamic element to uncertainty management by enabling corrective actions as uncertain parameters are realized. This progression toward dynamic adaptation represents a significant advancement in handling real-time operational challenges, though it introduces additional layers of computational complexity that must be carefully managed.

Recourse strategies, exemplified by Desaulniers et al. [5] and Martin et al. [19] introduce corrective actions when uncertainties materialize, allowing for dynamic route adjustments as conditions change during operations. Gendreau et al. [3] developed recourse models that optimize initial routing while accounting for adjustment costs based on observed demand, creating more robust transportation solutions. These approaches offer greater operational flexibility but introduce additional computational challenges through multi-stage decision processes.

2.2.3 Robust optimization approaches

The computational challenges inherent in stochastic programming have motivated the development of robust optimization approaches that seek to balance uncertainty management with computational tractability. These methods represent a paradigm shift toward worst-case scenario planning while maintaining practical applicability in large-scale transportation networks.

Recent advances in robust optimization have demonstrated significant practical benefits for transportation networks under uncertainty. Yang and Liu [20] showed that robust configuration optimization frameworks for integrated energy systems considering multiple uncertainties can significantly improve reliable energy supply and equipment investment efficiency, with total operating costs decreasing from $384,098 to $378,430 while maintaining service reliability. Transportation companies integrating route optimization software report operational expense reductions of 15-20% through optimized routes and decreased idle time, while studies demonstrate that optimizing vehicle capacity through robust planning resulted in a 15% reduction in delivery costs.

Contemporary research has further validated the effectiveness of distributionally robust optimization approaches in energy and transportation systems. Zhang et al. [21] demonstrated that distributionally robust optimization models utilizing Wasserstein distance-based ambiguity sets prove effective in navigating demand uncertainties while improving post-disaster recovery strategies, with applications showing enhanced system resilience under uncertain operational conditions. These findings collectively underscore the effectiveness of uncertainty sets constructed from historical data in managing practical uncertainty scenarios across diverse transportation and fuel distribution networks.

2.2.4 Stochastic demand modeling

While robust optimization provides effective worst-case guarantees, the quality of uncertainty management fundamentally depends on accurate demand forecasting capabilities. The integration of machine learning techniques into demand modeling has opened new possibilities for capturing complex demand patterns, yet significant challenges remain in handling sudden operational disruptions and emergency scenarios.

Recent advances by Jain and Gupta [22] in electrical load demand forecasting using machine learning algorithms demonstrate that Long Short-Term Memory (LSTM) models achieve superior performance compared to other approaches, with prediction errors 13.51% lower than Support Vector Machines. Industry research [23] shows that businesses implementing AI-driven demand forecasting can achieve up to 50% improvement in forecast accuracy through advanced machine learning techniques that integrate historical consumption patterns with external factors. However, these models struggle with sudden demand spikes during emergencies or supply disruptions, highlighting the need for more adaptive forecasting mechanisms.

The integration of robust and stochastic approaches, as demonstrated by Chen et al. [24], offers promising results for energy distribution systems operating under multiple uncertainty sources through two-stage distributed robust optimization frameworks that effectively manage intermittent renewable energy while minimizing both investment and operational costs. Complementary research by Javanmard and Ghaderi [25] on energy demand forecasting across seven sectors using machine learning optimization models further validates the effectiveness of integrated approaches. However, critical analysis reveals that current uncertainty management approaches suffer from a fundamental limitation: they optimize specific uncertainty scenarios rather than developing adaptive systems capable of responding to unforeseen conditions in real-time.

2.2.5 Computational limitations

Despite the theoretical sophistication of modern uncertainty management techniques, their practical implementation faces significant computational barriers that limit their applicability in real-time transportation operations. These limitations have profound implications for the development of scalable optimization frameworks capable of handling the complexity and uncertainty inherent in fuel delivery logistics.

Despite their theoretical appeal, these stochastic optimization techniques face significant scalability limitations in practical transportation applications. Recent research by Abirami et al. [26] in their systematic survey on big data and artificial intelligence algorithms for intelligent transportation systems demonstrates that computational complexity and algorithmic scalability remain critical challenges, particularly as datasets grow exponentially in size. The study highlights that while AI algorithms enhance traffic management and route optimization, they face significant computational barriers when processing large-scale transportation networks in real-time. This computational barrier has motivated research into alternative approaches that balance solution quality with practical computational requirements for transportation logistics, including the development of approximation algorithms and parallel processing techniques to maintain responsiveness in dynamic transportation environments.

2.3 Machine learning applications in transportation routing

The computational limitations of traditional optimization approaches, combined with the increasing availability of large-scale transportation data, have catalyzed significant interest in machine learning methodologies for routing optimization. Machine learning techniques offer the potential to overcome scalability challenges while maintaining solution quality, yet their application to transportation routing introduces new challenges related to constraint satisfaction and solution feasibility. This section examines the evolution of machine learning applications in routing, from DRL to GNNs, analyzing their capabilities and limitations in the context of fuel delivery optimization.

2.3.1 DRL approaches

Recent advances in machine learning have created new opportunities for addressing complex routing problems in transportation logistics. DRL methods have demonstrated promise for sequential decision-making under uncertainty, offering enhanced scalability compared to traditional optimization approaches.

Among DRL algorithms, PPO has emerged as particularly effective for transportation routing problems. Zhao et al. [6] demonstrated PPO's effectiveness in dynamic routing scenarios with changing demands, while Kool et al. [27] showed that reinforcement learning approaches can generalize effectively across different transportation network structures.

These characteristics make PPO especially suitable for fuel delivery optimization, where operational conditions vary substantially across service regions and time periods.

Recent applications have further extended these approaches to practical transportation contexts. Ma et al. [28] integrated PPO with pointer networks for improved computational efficiency in logistics applications, while Nazari et al. [29] demonstrated that attention mechanisms can enhance performance in dynamic delivery environments. These adaptations address specific challenges in transportation optimization, offering improved solution quality within practical computational constraints.

2.3.2 GNNs in spatial optimization

While DRL provides robust sequential decision-making capabilities, the explicit representation of spatial relationships within transportation networks remains a critical challenge. GNNs have emerged as a complementary technology that directly addresses spatial modeling limitations, offering enhanced representation capabilities for network-based optimization problems.

While DRL provides robust sequential decision-making capabilities, explicitly capturing spatial relationships within the routing problem remains challenging. GNNs directly address this issue by efficiently modeling spatial dependencies in transportation networks. Kovács and Jlidi [7] demonstrated GNNs' effectiveness in learning heuristics for combinatorial optimization in transportation contexts, while Chen and Tian [30] showed that graph-based representations significantly outperform standard neural network architectures on routing problems. These advances suggest substantial potential for GNN applications in fuel delivery optimization, where spatial relationships significantly impact solution quality.

2.3.3 Limitations of GNNs in transportation logistics

Despite their promising spatial modeling capabilities, GNNs face several critical limitations when applied to practical transportation logistics scenarios. These limitations highlight the need for careful consideration of GNN applicability and the potential requirement for hybrid approaches that combine GNNs with complementary methodologies.

Despite their representational strengths, GNNs present several critical limitations when applied to transportation logistics. First, their performance tends to degrade with very large or sparse graphs, common in regional or national transportation networks. Second, most GNN architectures assume static graph topologies, limiting their applicability in dynamic environments where traffic conditions, route availability, or demand fluctuate in real-time. Third, GNN-based models often lack direct mechanisms for handling hard operational constraints (e.g., time windows, vehicle capacities), requiring complex post-processing or integration with additional frameworks [31, 32].

These limitations highlight the need for hybrid approaches that combine GNNs with decision-making frameworks capable of enforcing feasibility and adaptability under uncertainty. Current reinforcement learning approaches, although scalable, similarly struggle with explicit constraint satisfaction—a critical requirement in heavily regulated domains such as fuel delivery, where safety and compliance cannot be compromised.

2.3.4 Integration challenges

The individual limitations of both DRL and GNNs have motivated researchers to explore integrated approaches that leverage the complementary strengths of these methodologies. However, the effective integration of spatial representation capabilities with sequential decision-making frameworks remains a significant challenge, particularly in constrained domains such as fuel delivery logistics.

Despite their respective strengths, the combined integration of GNNs with reinforcement learning—particularly under stochastic fuel delivery scenarios—remains significantly underexplored. Existing studies predominantly focus on deterministic settings or inadequately incorporate unique constraints inherent to fuel distribution networks. Consequently, there is a notable research gap in effectively integrating spatial representation capabilities of GNNs with the decision-making flexibility of DRL, particularly in highly regulated and uncertainty-rich transportation domains such as fuel delivery logistics.

2.4 Integrated approaches and research gaps

The examination of individual methodologies reveals that each approach addresses specific aspects of the fuel delivery optimization challenge while exhibiting fundamental limitations in others. This recognition has motivated researchers to develop integrated frameworks that combine multiple methodologies to achieve comprehensive optimization capabilities. However, current integration attempts remain incomplete, addressing only partial combinations of required capabilities while failing to fully leverage the synergistic potential of unified approaches.

2.4.1 Current hybrid methodologies

Literature clearly highlights the limitations of individual methodologies, motivating researchers to develop integrated approaches combining the complementary strengths of traditional optimization and machine learning methods. Several recent studies have attempted such integrations. Chen et al. [17] combined reinforcement learning with local search heuristics, demonstrating improved solution quality while maintaining computational efficiency. However, critical analysis reveals that their framework does not adequately address fuel-specific operational constraints such as multi-compartment vehicle management, contamination prevention, and regulatory compliance. Fuel delivery logistics involves stringent operational requirements, including compartment-specific routing to prevent contamination and adherence to strict safety regulations, complexities that generic routing methods often overlook.

Similarly, Ma et al. [28] developed a dual-aspect collaborative transformer approach for stochastic routing under uncertainty, achieving superior performance in handling demand variability. Despite these advancements, their method lacks real-time adaptation capabilities essential for handling unforeseen operational disruptions and similarly neglects the integration of regulatory constraints critical in fuel distribution scenarios.

2.4.2 Methodological gaps

The analysis of current hybrid methodologies reveals systematic gaps that prevent existing approaches from achieving comprehensive optimization in fuel delivery logistics. These gaps represent fundamental challenges that must be addressed to develop truly effective integrated frameworks capable of handling the full complexity of modern

fuel distribution operations.

Current integrated approaches still exhibit several fundamental limitations. Firstly, they typically address only partial combinations of required capabilities (e.g., spatial representation, uncertainty handling, scalability, constraint satisfaction), leaving considerable optimization potential unexplored. Secondly, existing methodologies continue to inadequately represent fuel-specific operational constraints, treating fuel delivery as merely a special case of general routing problems rather than as a distinct optimization domain. Lastly, these methods consistently lack effective real-time adaptability mechanisms. Real-time adaptability is crucial to maintaining operational feasibility and optimality under dynamic conditions, particularly in fuel distribution networks where sudden demand fluctuations, unexpected traffic conditions, and stringent compliance requirements frequently occur.

2.5 Research gap synthesis and future directions

The comprehensive analysis of existing literature reveals a critical convergence point where individual methodologies reach their operational limits, necessitating a fundamental paradigm shift toward unified integration architectures. This convergence represents both a challenge and an opportunity for developing next-generation optimization frameworks that transcend the limitations of current approaches while addressing the unique requirements of fuel delivery logistics.

2.5.1 Methodological convergence analysis

The comprehensive literature analysis reveals a critical convergence point, where established optimization paradigms individually reach their operational limits in fuel delivery logistics. Traditional mathematical programming offers structural rigor but lacks dynamic adaptability; stochastic approaches effectively handle uncertainty but face scalability issues; machine learning methods provide scalability but struggle with operational constraint enforcement; and graph-based representations excel in spatial modeling but must integrate effectively with decision-making frameworks.

The identified methodological fragmentation across existing approaches—partial integration, generic treatment of fuel-specific constraints, and insufficient real-time adaptability—highlights the need for a unified framework that comprehensively addresses these interconnected limitations.

2.5.2 Integrated framework contribution

The recognition of methodological convergence challenges necessitates the development of comprehensive integration architectures that address the fundamental limitations identified across existing approaches. This research gap provides the foundation for developing innovative frameworks that orchestrate multiple computational paradigms within unified, coherent systems capable of handling the full complexity of fuel delivery logistics.

Addressing this convergence challenge necessitates a paradigm shift toward comprehensive integration architectures that transcend individual methodology limitations. This research contributes precisely to such a novel framework by orchestrating multiple computational paradigms within a unified, coherent system. The proposed approach leverages reinforcement learning's scalability for sequential decision-making, integrates GNNs' powerful spatial modeling capabilities, and incorporates traditional optimization methods for rigorous operational constraint enforcement.

Critically, this framework treats fuel-specific operational constraints—multi-compartment vehicle management, contamination prevention, and regulatory compliance—as fundamental components of the optimization process rather than secondary considerations. Additionally, the proposed methodology enables real-time adaptation to changing operational conditions, maintaining near-optimal solution quality and strict feasibility under uncertainty, directly addressing the crucial need for adaptive decision-making in fuel delivery logistics.

2.5.3 Research positioning and expected impact

The development of comprehensive integrated frameworks represents a significant opportunity to advance both theoretical understanding and practical capabilities in transportation logistics optimization. The positioning of this research within the broader landscape of optimization methodologies establishes the foundation for transformative advances that extend beyond fuel delivery logistics to influence the evolution of transportation optimization more broadly.

The proposed integrated framework represents a methodological advancement beyond existing literature, simultaneously addressing computational efficiency, solution optimality, and operational feasibility. Whereas current methodologies typically optimize isolated objectives sequentially, leading to overall suboptimal system performance, the proposed integration systematically exploits the complementary strengths of each methodology, thereby achieving near-optimal, scalable solutions.

Operationally, the proposed framework is expected to significantly reduce fuel delivery costs, enhance fleet utilization efficiency, rapidly adapt to unforeseen events such as demand spikes or disruptions, and ensure strict regulatory compliance, thus directly improving daily logistics performance. Domain-specific modeling ensures solutions inherently incorporate fuel-industry constraints, avoiding the need for extensive post-processing corrections.

Moreover, this integrated methodology offers a generalizable framework that extends beyond fuel delivery logistics, laying the groundwork for future research in other complex transportation and logistics domains characterized by simultaneous challenges of constraint satisfaction, uncertainty management, and real-time adaptability. By establishing foundations for adaptive, constraint-aware, and scalable optimization, the research contributes to the broader evolution and future directions of transportation logistics optimization methodologies.

3. PROBLEM DEFINITION AND MATHEMATICAL MODEL

3.1 Stochastic fuel delivery problem: Modeling and formulation

The stochastic fuel delivery problem involves dispatching a fleet of vehicles from a central depot to service a set of gas stations with uncertain demand quantities. This section develops a comprehensive mathematical model that captures the operational complexities and stochastic elements inherent in fuel distribution networks.

The problem encompasses several interrelated components

that define the transportation logistics framework. A heterogeneous fleet V of vehicles operates from a central depot, with each vehicle $k \in V$ characterized by capacity Q_k, operational cost per unit time F_k, and time availability A_k. These vehicles must service a set C of gas stations (customers), where each station $c_i \in C$ exhibits stochastic demand patterns. The set I of orders corresponds to specific gas stations, with each order $i \in I$ associated with a station c.

Transportation operations are further characterized by travel times T_k^L and T_k^U for loaded and unloaded vehicles, respectively, reflecting the influence of cargo weight on vehicle performance. Additionally, operations include loading time L_{ik} at the depot and uncertain unloading time U_{ik} at destination stations. The distance d_{c_i} from each gas station c_i to the depot completes the spatial configuration of the transportation network.

The objective of the optimization problem is to design efficient delivery routes that minimize operational costs while satisfying demand constraints under uncertainty. This formulation explicitly accounts for the stochastic nature of fuel demand and unloading times, reflecting real-world variability in consumption patterns and service operations within transportation systems.

3.1.1 Decision variables

The optimization model employs the following decision variables to characterize delivery operations:

- $n_{ik} \in \mathbb{N}$: Number of deliveries by vehicle k for order i.
- $x_{ik} \in {0,1}$: Binary variable equal to 1 if vehicle k is assigned to order i.
- $v_{c,k} \in {0,1}$: Binary variable equal to 1 if vehicle k serves gas station c.
- $w_k \in {0,1}$: Binary variable equal to 1 if vehicle k is deployed.
- $s_{k,ij} \in {0,1}$: Binary variable equal to 1 if order j immediately follows order i on vehicle k's route.
- $t_{k,i} \in \mathbb{N}$: Position of order i in vehicle k's route.

These variables collectively define the assignment of vehicles to orders, service sequence, and delivery quantities within the transportation network.

3.1.2 Objective functions

The model addresses two complementary objectives that balance operational efficiency with service consistency in transportation logistics:

Total Delivery Cost (f_1):

$$f_1 = \sum_{i \in I} \sum_{k \in V_i} F_k \left(L_{ik} + U_{ik} + d_{c_i}(T_k^L + T_k^U) \right) n_{ik} \quad (1)$$

This objective captures the operational efficiency of delivery operations, including loading, unloading, and transportation costs. By minimizing f_1, the model reduces overall expenditure associated with fuel consumption, vehicle utilization, and time resources, ensuring economic viability while meeting service requirements.

Vehicle Dispersion (f_2):

$$f_2 = \sum_{c \in C} \sum_{k \in V} v_{c,k} + \beta \sum_{k \in V} w_k \quad (2)$$

This secondary objective promotes operational consistency by minimizing the number of different vehicles serving each station and reducing the total vehicles deployed. The parameter β weights the importance of fleet size reduction versus service concentration, allowing transportation managers to balance fixed fleet costs against service complexity.

3.1.3 Key constraints

The model incorporates several constraint categories to ensure operational feasibility while accommodating stochastic variations in transportation parameters:

Vehicle Time Constraints: For every vehicle $k \in V$:

$$\sum_{i \in I} R_{ik} \, n_{ik} \le A_k, \text{ where } R_{ik} = L_{ik} + U_{ik} + d_{c_i}(T_k^L + T_k^U)$$

Stochastic Demand Satisfaction: For every order $i \in I$:

$$\sum_{k \in V_i} Q_k \, n_{ik} \ge d_i$$

where, d_i is a random variable with mean μ_i and standard deviation σ_i.

Vehicle Assignment Constraints: For each order $i \in I$ and vehicle $k \in V$:

$$x_{ik} \le n_{ik} \le \left\lceil \frac{d_i}{Q_k} \right\rceil x_{ik}$$

Station-Vehicle Linking Constraints: For each order $i \in I$ with station $c_i = c$ and vehicle $k \in V$:

$$v_{c,k} \ge x_{ik}$$

Vehicle Utilization Constraints: For every vehicle $k \in V$ and station $c \in C$:

$$w_k \ge v_{c,k}$$
$$\sum_{k \in V} w_k \le N_k$$

Route Construction Constraints: For every vehicle $k \in V$ and orders $i, j \in I \cup \{0\}$:

$$\sum_{j \in (I \cup \{0\}) \setminus \{i\}} s_{k,ij} = x_{ik}, \forall i \in I$$
$$\sum_{j \in (I \cup \{0\}) \setminus \{i\}} s_{k,ji} = x_{ik}, \forall i \in I$$
$$\sum_{i \in I} s_{k,0i} = w_k$$
$$\sum_{i \in I} s_{k,i0} = w_k$$
$$0 \le t_{k,i} \le |I|, \forall i \in I$$
$$t_{k,0} = 0$$
$$t_{k,i} + |I| s_{k,ij} + (|I| - 2) s_{k,ji} \le t_{k,j} - 1, \forall i, j \in I \cup \{0\}, i \ne j$$

These constraints collectively ensure proper route

construction within the transportation network, including depot departure and return, sequencing of deliveries, and subtour elimination.

3.2 Deterministic equivalent model

To enable practical computation in transportation logistics applications, the stochastic model requires transformation into a deterministic equivalent using three key techniques that preserve the essential structure while enabling standard optimization approaches.

3.2.1 Chance-constrained transformation

The stochastic demand constraint is replaced with its chance-constrained equivalent:

$$\sum_{k \in V_i} Q_k\, n_{ik} \geq D_i^*, \forall i \in I$$

where, $D_i^* = \mu_i + z_{1-\alpha}\sigma_i$ represents the certainty-equivalent demand for order i, incorporating both the expected demand μ_i and a risk-adjusted margin $z_{1-\alpha}\sigma_i$. The parameter $z_{1-\alpha}$ is the quantile of the standard normal distribution corresponding to confidence level $1 - \alpha$.

3.2.2 Expected value transformation

The uncertain unloading time U_{ik} is approximated using its expected value:

$$U_{ik}^e = \alpha_k(\mu_i + z_{1-\alpha}\sigma_i)$$

where, α_k is the average unloading time per unit of fuel for vehicle k. This transforms the unit delivery time to:

$$R_{ik} = L_{ik} + U_{ik}^e + d_{c_i}(T_k^L + T_k^U)$$

3.2.3 Recourse transformation

To handle deviations from planned deliveries, we introduce recourse variables $\delta^+ i(s)$ and $\delta^- i(s)$ for each order i and scenario s. These variables quantify the shortfall or excess relative to the certainty-equivalent demand, with constraints:

$$\sum_{s \in S} p_s \delta_i^+(s) \geq D_i^* - \sum_{k \in V_i} Q_k n_{ik}, \forall i \in I$$

$$\sum_{s \in S} p_s \delta^- i(s) \geq \sum_{k \in V_i} Q_k n_{ik} - D_i^*, \forall i \in I$$

The corresponding penalty term in the objective function becomes:

$$f_{recourse} = \sum_{s \in S} \sum_{i \in I} q\, \delta_i^+(s)$$

where, q represents the penalty cost per unit of unmet demand.

3.3 Integrated deterministic model

The complete deterministic equivalent model integrates all previously described transformations into a coherent optimization framework:

$$min\; f_1 + f_{recourse} + f_2$$

where,

$$f_1 + f_{recourse} = \sum_{i \in I} \sum_{k \in V_i} F_k(L_{ik} + U_{ik}^e + d_{c_i}(T_k^L + T_k^U))\, n_{ik}$$
$$+ \sum_{s \in S} \sum_{i \in I} q d_i^+(s),$$
$$f_2 = \sum_{c \in C} \sum_{k \in V} v_{c,k} + \beta \sum_{k \in V} w_k$$

This model is subject to all vehicle routing constraints with stochastic elements replaced by their deterministic equivalents. These include vehicle capacity constraints, demand satisfaction requirements using certainty-equivalent values, routing sequencing constraints, and recourse mechanisms to handle demand deviations.

The transformation preserves the essential structure of the original stochastic problem while enabling practical computation through standard optimization techniques. The model balances computational tractability with solution robustness, providing a rigorous foundation for both direct solution methods and the hybrid learning approach described in the following section. This deterministic equivalent formulation serves as a critical bridge between theoretical modeling and practical application in transportation logistics operations.

4. PROPOSED HYBRID METHODOLOGY

4.1 Framework overview

This section presents the integrated methodology developed to address the computational challenges of stochastic fuel delivery optimization in transportation networks. The approach synthesizes advanced computational techniques with traditional optimization principles to generate solutions that are both practically implementable and mathematically sound.

The methodology addresses two fundamental limitations in transportation logistics optimization. First, traditional approaches face scalability challenges, as the decision space in large-scale problems expands exponentially, rendering classical optimization techniques computationally intractable for time-sensitive operational decisions. Second, DRL methods often struggle with modeling complex spatial dependencies and ensuring constraint satisfaction in dynamic delivery networks. To overcome these challenges, the proposed framework integrates GNNs-augmented DRL, which effectively captures spatial relationships while adapting to dynamic environments. By combining reinforcement learning with graph-based spatial reasoning and constraint validation mechanisms, the framework exploits the complementary strengths of multiple computational paradigms, enabling scalable and constraint-aware logistics optimization.

Figure 1 illustrates the overall architecture of the integrated approach. At the core of this architecture lies a reinforcement learning agent that employs PPO to learn efficient delivery strategies through interaction with a simulated transportation environment. This agent's capabilities are enhanced through a GNN module that extracts spatial features from the fuel distribution network, capturing complex dependencies

between delivery locations. The third essential component is a verification mechanism that ensures generated routes adhere to operational constraints essential for transportation logistics.

Together, these interconnected elements form a comprehensive system that addresses the multifaceted challenges of fuel delivery optimization.
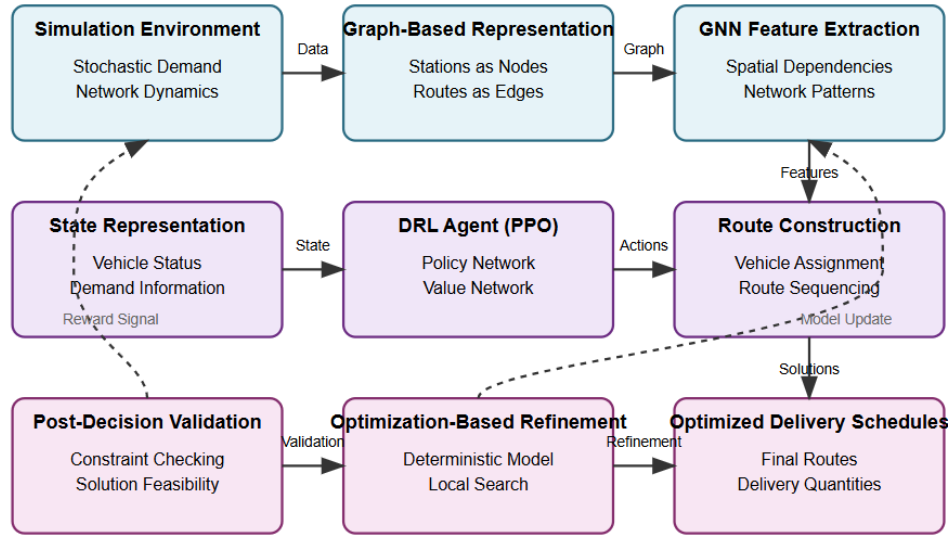

Figure 1. Overall architecture of the PPO-GNN hybrid methodology

The PPO-based routing policy serves as the primary decision-making component, learning to generate efficient delivery routes through repeated interaction with the transportation environment. This policy progressively improves as it experiences diverse scenarios, developing adaptive strategies that respond to demand variability and network conditions. The GNN-enhanced state representation complements this policy by transforming the raw transportation network into structured spatial embeddings that capture relationship patterns between different locations. These embeddings enable the agent to recognize critical spatial dependencies that influence routing efficiency. Finally, the deterministic constraint validation component bridges the gap between learning-based approaches and practical feasibility, ensuring that generated solutions satisfy the operational requirements of real-world transportation systems.

These components operate in concert to enable adaptive decision-making that balances cost efficiency with operational feasibility in stochastic environments. Their integration represents a significant advancement over existing approaches in transportation logistics optimization, addressing both computational efficiency and solution quality within a unified framework.

4.2 Reinforcement learning framework

4.2.1 Problem formulation as Markov decision process

The fuel delivery optimization problem is formulated as a Markov Decision Process (MDP) defined by the tuple $\langle S, A, P, R, \gamma \rangle$, providing a mathematical framework for sequential decision-making within transportation systems:

- State Space (S): Each state s_t encodes the current transportation system configuration, including vehicle locations, remaining capacity, pending deliveries, and estimated demand at each station.
- Action Space (A): The action space consists of decisions for vehicle assignment, routing sequence, and delivery quantity within the transportation network.
- Transition Dynamics (P): The probability distribution

$P\left(s_{\{t+1\}}|s_t, a_t\right)$ models how the environment evolves in response to actions, capturing stochasticity in demand patterns and travel conditions.
- Reward Function (R): The reward function $R(s_t, a_t)$ provides feedback on action quality, balancing operational costs, demand satisfaction, and constraint adherence in transportation operations.
- Discount Factor (γ): The parameter $\gamma \in [0,1]$ determines the relative importance of immediate versus future rewards in route planning.

This MDP formulation creates a structured framework for learning adaptive delivery strategies through reinforcement learning, enabling optimization of sequential decisions within stochastic transportation environments.

The state space is formally defined as $s_t = \left[v_{pos}, v_{cap}, D_{pending}, \mu_{demand}, \sigma_{demand}, T_{windows}\right]$ where $v_{pos} \in \mathbb{R}^{n \times 2}$ represents vehicle positions in Cartesian coordinates, $v_{cap} \in \mathbb{R}^n$ denotes remaining vehicle capacities, $D_{pending} \in \{0,1\}^m$ is a binary vector indicating pending deliveries, $\mu_{demand} \in \mathbb{R}^m$ captures expected demand at each station, $\sigma_{demand} \in \mathbb{R}^m$ quantifies demand uncertainty through standard deviation, and $T_{windows} \in \mathbb{R}^{m \times 2}$ defines delivery time windows as [start, end] intervals for each station. This comprehensive state representation enables the agent to make informed decisions that account for both current system conditions and future uncertainties.

4.2.2 PPO

The methodology employs PPO as its core learning algorithm based on established advantages in sample efficiency, stability, and performance in high-dimensional action spaces [33]. Unlike traditional policy gradient methods, PPO utilizes a clipped surrogate objective that prevents destructive policy updates:

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t\right)\right]$$

where,

- $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio;
- $\hat{A}_t$ is the estimated advantage function;
- ϵ is the clipping parameter that constrains policy updates.

This objective ensures stable learning by limiting the magnitude of policy changes, preventing the optimization process from collapsing due to excessively large updates in complex transportation routing problems.

4.2.3 Reward function design

The reward function is carefully designed to guide the learning agent toward high-quality solutions that balance multiple operational objectives:

$$R(s,a) = -\lambda_1 C_{total}(s,a) - \lambda_2 C_{dispersion}(s,a) \\ - \lambda_3 C_{delay}(s,a) - \lambda_4 C_{unmet}(s,a) \\ - \lambda_5 C_{constraint}(s,a)$$

where,
- C_{total} represents operational costs (fuel, time, resources);
- $C_{dispersion}$ penalizes excessive vehicle deployment and station dispersion;
- C_{delay} discourages late deliveries;
- C_{unmet} penalizes unfulfilled demand;
- $C_{constraint}$ imposes penalties for constraint violations.

The weights λ_1 through λ_5 balance the importance of different objectives, with λ_1 prioritizing cost efficiency and λ_5 emphasizing constraint satisfaction. This multi-objective reward formulation guides the agent toward solutions that are both economically efficient and operationally viable within transportation systems.

Each cost component is mathematically defined as follows:

$C_{total(s,a)} = \Sigma_i(d_{travel,I} \times c_{fuel} + t_{travel,I} \times c_{driver})$ represents direct operational expenses including fuel consumption and driver wages.

$C_{dispersion(s,a)} = \Sigma_i||pos_i - centroid||^2$ penalizes excessive geographical spread of vehicle deployments.

$C_{delay(s,a)} = \Sigma_j \max(0, arrival_j - deadline_j)^2$ applies quadratic penalties for late deliveries.

$C_{unmet(s,a)} = \Sigma_k(demand_k - delivered_k)^2$ penalizes unfulfilled customer demand.

$C_{constraint(s,a)} = \lambda_{penalty} \times \Sigma_l \max(0, violation_l)$ imposes penalties proportional to constraint violations.

The weighting parameters are empirically set as $\lambda_1 = 1.0$, $\lambda_2 = 0.3$, $\lambda_3 = 0.8$, $\lambda_4 = 1.2$, and $\lambda_5 = 2.0$, reflecting the relative importance of each operational objective.

4.3 GNN enhancement

4.3.1 Graph representation of transportation network

The transportation distribution network is naturally represented as a graph:

$$G = (V, E)$$

where,
- Vertices V represent locations (gas stations and depots);
- Edges E represent transportation links with attributes (distance, travel time);

- Node features encode station-specific information (demand distribution, time windows);
- Edge features capture road characteristics (congestion patterns, travel restrictions).

This graph structure inherently models spatial dependencies and connectivity patterns critical for efficient routing decisions within transportation networks. The representation preserves the topological structure of the delivery environment, enabling more effective learning of spatial relationships compared to standard vector-based approaches.

4.3.2 GNN architecture

The methodology employs message-passing neural network architecture to process the graph representation:

$$h_v^{(l+1)} = \text{UPDATE}\left(h_v^{(l)}, \text{AGGREGATE}(\{h_u^{(l)}, e_{uv} : u \in \mathcal{N}(v)\})\right)$$

where,
- $h_v^{(l)}$ is the feature vector for node v at layer l;
- e_{uv} represents edge features between nodes u and v;
- $\mathcal{N}(v)$ denotes the neighborhood of node v;
- AGGREGATE and UPDATE are neural network functions that process and transform features.

This architecture allows information to propagate across the transportation graph, enabling the model to capture complex spatial relationships and dependencies. The multi-layer design progressively incorporates information from wider network neighborhoods, building a comprehensive representation of the transportation landscape.

The specific GNN architecture employs three message-passing layers with hidden dimensions of 128 for node features and 64 for edge features. Each layer utilizes ReLU activation functions with dropout regularization (p=0.2) to prevent overfitting. The aggregation function combines mean pooling with an attention mechanism that weights neighboring node contributions based on their relevance to routing decisions. The final graph embedding produces a 256-dimensional vector that captures the essential spatial characteristics of the transportation network.

4.3.3 Integration with reinforcement learning

The GNN module enhances the reinforcement learning agent's state representation by embedding the graph structure into a fixed-dimensional feature space:

$$s_t' = [s_t; GNN(G_t)]$$

where,
- s_t is the original state representation;
- $GNN(G_t)$ is the graph embedding produced by the GNN;
- s_t' is the augmented state used by the PPO policy.

This integration allows the agent to incorporate spatial context into its decision-making, recognizing patterns and dependencies within the transportation network that would otherwise remain hidden in a flat state representation. The enhanced state representation supports more informed routing decisions that account for the complex spatial relationships inherent in transportation networks.

4.4 Constraint validation and solution refinement

4.4.1 Deterministic validation mechanism

To ensure operational feasibility within transportation logistics, the framework implements a constraint validation mechanism that evaluates generated solutions against the deterministic equivalent model described in Section 3. This process:

1. Checks solutions for violations of vehicle capacity, time windows, and other operational constraints;

2. Quantifies the degree of constraint violation for reward function adjustment;

3. Guides the learning agent toward the feasible solution space through targeted feedback.

This validation component bridges the gap between learning-based approaches and mathematical optimization, ensuring that generated solutions satisfy the practical requirements of transportation operations.

4.4.2 Solution refinement process

When constraint violations are detected, the framework employs a two-stage refinement process:

1. Minor Violations: For solutions with limited constraint violations (≤5%), the system applies local adjustments to restore feasibility while preserving the overall route structure.

2. Major Violations: For solutions with significant violations (>5%), the framework integrates optimization-based repair mechanisms that leverage the deterministic model to guide correction.

This refinement process ensures that the integrated framework not only learns to generate efficient routes but also adheres to the operational constraints essential for practical implementation in transportation logistics.

4.5 Training and deployment strategy

4.5.1 Training algorithm

Algorithm 1 presents the complete training procedure for the integrated PPO-GNN framework. The training process employs the following hyperparameters: batch size of 32

episodes, learning rate of 3×10^{-4} using the Adam optimizer, 10 PPO epochs per update cycle, clipping parameter $\varepsilon = 0.2$, value function coefficient of 0.5, entropy coefficient of 0.01, GAE lambda of 0.95, and discount factor $\gamma = 0.99$. These parameters were selected through systematic hyperparameter tuning to balance learning stability with convergence speed.

The training process continues until convergence, typically requiring 5000-8000 episodes, depending on network complexity. Convergence is determined when the average reward improvement over 100 consecutive episodes falls below 1%.

| Algorithm 1: PPO-GNN Training Procedure |
|---|
| 1: Initialize: $\pi\theta$ (policy), $V\varphi$ (value function), $GNN\psi$ (graph network) |
| 2: For episode = 1 to max_episodes do |
| 3: Initialize state s_0, graph G_0 |
| 4: trajectory ← [] |
| 5: |
| 6: For t = 0 to T-1 do |
| 7: h_t ← GNNψ(G_t) // Extract graph embedding |
| 8: s't ← [s_t; h_t] // Augment state representation |
| 9: a_t ~ πθ(·\|s't) // Sample action from policy |
| 10: s{t+1}, r_t, G{t+1} ← ENV(s_t, a_t) // Environmentstep |
| 11: trajectory.append((s'_t, a_t, r_t)) |
| 12: End For |
| 13: |
| 14: // Constraint validation and reward adjustment |
| 15: violations ← VALIDATE_CONSTRAINTS(trajectory) |
| 16: adjusted_rewards ← ADJUST_REWARDS(trajectory, violations) |
| 17: |
| 18: // Update networks using PPO |
| 19: For ppo_epoch = 1 to 10 do |
| 20: θ ← UPDATE_POLICY(θ, trajectory, adjusted_rewards) |
| 21: φ ← UPDATE_VALUE(φ, trajectory, adjusted_rewards) |
| 22: End For |
| 23: |
| 24: // Update GNN parameters |
| 25: ψ ← UPDATE_GNN(ψ, trajectory) |
| 26: End For |
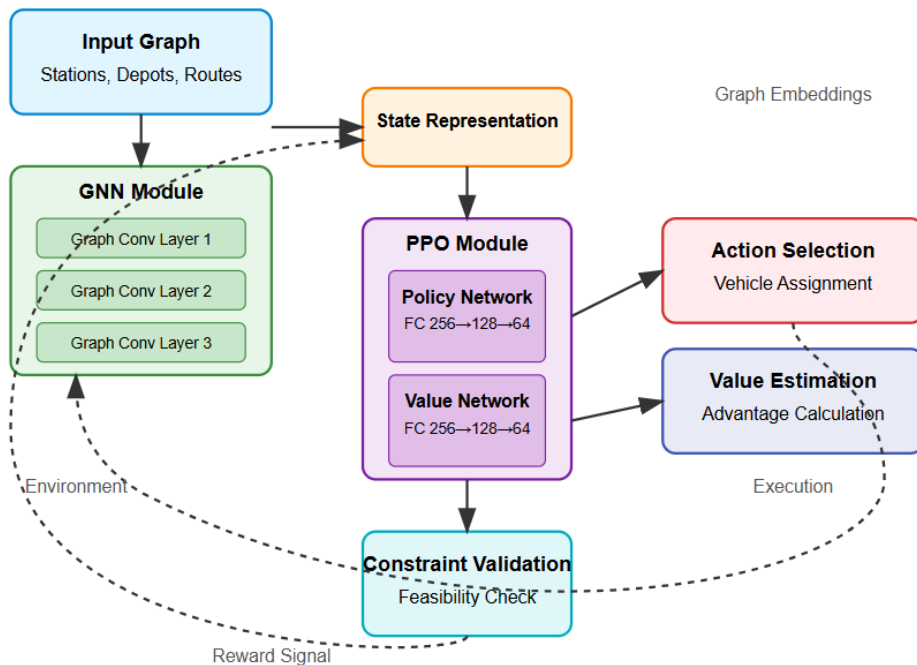
Figure 2. Detailed architecture of the PPO-GNN model

4.5.2 Deployment strategy

Once trained, the integrated model is deployed according to the following procedure:

1. Encode the current network state and demand forecasts as input to the model;
2. Generate routing decisions using the learned policy;
3. Validate solutions against operating constraints;
4. Apply refinement if necessary to ensure feasibility;
5. Execute the final delivery plan.

This deployment strategy ensures robust decision-making that adapts to changing conditions while maintaining operational feasibility, providing transportation logistics providers with a practical tool for daily routing operations.

Figure 2 illustrates the internal structure of the proposed integrated model, including the GNN module, state representation, policy and value networks, and constraint validation mechanism.

5. EXPERIMENTAL EVALUATION

5.1 Experimental design

5.1.1 Transportation network datasets

To thoroughly evaluate the integrated methodology, this study constructed synthetic fuel delivery networks of varying scales representing different operational scenarios in transportation logistics. Three distinct network categories were developed to assess performance across different operational contexts.

The small-scale networks are comprised of 10 gas stations serviced by 3 vehicles, representing localized urban delivery operations with limited geographic spread. These configurations modeled compact transportation scenarios typically found in dense urban environments where vehicles operate within confined service areas. Medium-scale networks incorporated 50 gas stations with 8 vehicles, simulating regional distribution networks spanning multiple municipalities. These networks exhibited greater geographic dispersion and operational complexity, with longer travel distances and more varied demand patterns. Large-scale networks contain 100 gas stations serviced by 15 vehicles, emulating nationwide delivery operations with substantial logistical challenges. These extensive networks featured complex spatial distributions, diverse demand profiles, and significant heterogeneity in operational parameters.

For each network scale, the study generated 100 problem instances with carefully calibrated characteristics relevant to transportation logistics. The demand profiles at each gas station were modeled using truncated normal distributions, where the mean and standard deviation parameters were empirically calibrated based on historical fuel consumption records from operational fuel delivery systems. This calibration ensures that the generated demand variability (10% to 25% of the mean) reflects realistic consumption fluctuations.

Network topologies were constructed by sampling station locations over spatial grids mimicking real-world urban and regional layouts. The distance matrices were derived using shortest-path computations over road graphs extracted from OpenStreetMap data, thereby maintaining realistic geographic and transportation features such as travel distances and connectivity.

To further validate the external realism of the synthetic data, a real-world case study was included in Section 5.5, based on an actual regional fuel delivery network involving 38 gas stations and 6 heterogeneous vehicles. The consistency of the results between synthetic and real-world settings supports the relevance and validity of the synthetic dataset design.

5.1.2 Comparative methodologies

The study compared the integrated PPO-GNN approach against three established baselines to evaluate relative performance and contribution:

Classical PPO implementation served as the first baseline, employing standard PPO without GNN enhancements. This method used a flat state representation with the same reward structure as the integrated approach, allowing direct assessment of the GNN contribution to solution quality. The Clarke-Wright Savings Algorithm provided a widely used deterministic heuristic baseline for vehicle routing problems. This established method represents a common approach in transportation logistics, prioritizing computational efficiency for rapid solution generation. Deterministic optimization through direct solution of the deterministic equivalent model using the commercial Gurobi solver (with limited runtime) offered a comparison to traditional mathematical programming approaches.

These baselines were selected to evaluate the contribution of each component in the integrated framework and to benchmark against traditional industry approaches. The comparison enables assessment of how the proposed methodology performs relative to both learning-based and optimization-based alternatives across various performance dimensions relevant to transportation logistics.

5.1.3 Implementation specifications

The PPO-GNN implementation utilized carefully designed neural network architecture optimized for transportation routing problems. The policy network comprised three fully connected layers (256-128-64 units) with ReLU activations, enabling effective mapping from states to actions within the complex decision space. The value network mirrored this structure with three fully connected layers (256-128-64 units) and ReLU activations, providing accurate state value estimation for advantage calculation. The GNN component incorporated three graph convolutional layers with 64 channels each, facilitating effective information propagation across the transportation network representation.

Training parameters were calibrated to ensure stable and efficient learning in the stochastic environment. The learning rate was set at 3×10^{-4} with adaptive scheduling to accommodate changing learning dynamics throughout the training process. The discount factor (γ) of 0.99 balanced immediate and future rewards, while the GAE parameter (λ) of 0.95 reduced the variance in advantage estimation. The clipping parameter (ε) was set to 0.2, preventing destructive policy updates during training. Value function and entropy coefficients (0.5 and 0.01 respectively) balanced the multiple objectives within the loss function.

The training process encompassed 50,000 episodes with early stopping criteria to prevent overfitting, using batch sizes of 2048 timesteps for stable gradient updates. The Adam optimizer handled parameter updates, while hardware acceleration through NVIDIA A100 GPUs and 32-core CPUs enabled efficient training of the neural networks. All experiments maintained controlled conditions with identical random seeds across methods, ensuring consistent

stochasticity for fair comparison.

5.2 Performance metrics

The study employed multiple complementary metrics to provide a comprehensive assessment of solution quality across different dimensions relevant to transportation logistics.

Total operational cost represented the primary economic metric, calculated as the sum of vehicle deployment, travel, and delivery costs across the transportation network. This metric directly measures the economic efficiency of generated solutions, reflecting the financial impact of routing decisions on transportation operations. Demand satisfaction rate quantified the percentage of total fuel demand successfully delivered across all stations, serving as a key measure of service quality in the transportation system. This metric reflects the effectiveness of generated routes in meeting customer requirements under demand uncertainty.

Constraint violation rate measured the percentage of solutions violating operational constraints, providing insight into the practical feasibility of generated routes within real-world transportation constraints. This metric is particularly important for assessing the viability of different approaches in highly regulated transportation domains like fuel delivery. Computational efficiency metrics captured solution time and scalability across problem sizes, measuring the practical applicability of different methods in time-sensitive operational contexts. The study also assessed robustness to uncertainty through performance stability analysis under varying degrees of demand stochasticity.

These metrics collectively provide a comprehensive assessment of algorithm performance in terms of both economic efficiency and operational feasibility. The multi-dimensional evaluation framework enables nuanced comparison between different approaches across the various aspects relevant to practical transportation logistics.

5.3 Results and analysis

5.3.1 Comparative performance analysis

Table 1 presents the aggregate results across all problem instances, showing the average performance of each method on key metrics relevant to transportation logistics operations.

Table 1. Performance comparison across all problem instances (average values)

Method	Total Cost ($)	Unmet Demand (%)	Constraint Violations (%)	Solve Time (s)
PPO-GNN	12.8	2	1	180
Classical PPO	13.8	8	6	120
Clarke-Wright	14.2	5	3	60
Deterministic Optimization	15.6	11	0.5	3600+

The results demonstrate that the integrated PPO-GNN framework outperforms all baselines in terms of total cost and demand satisfaction. The approach achieves a 7.2% cost reduction compared to classical PPO and a 9.9% improvement over the Clarke-Wright heuristic. This economic advantage stems from the enhanced spatial representation provided by the GNN component, which enables more efficient route construction that minimizes unnecessary travel while effectively servicing demand points.

The constraint violation rate of 1.0% for PPO-GNN is significantly lower than both classical PPO (6.0%) and the Clarke-Wright heuristic (3.0%), highlighting the effectiveness of the constraint validation mechanism in ensuring operational feasibility. This substantial improvement in feasibility without compromising cost efficiency demonstrates the value of integrating optimization principles within the learning framework.

While deterministic optimization achieves the lowest constraint violation rate (0.5%), it results in substantially higher operational costs and unmet demand, primarily due to its inability to adapt to stochastic variations in the transportation environment. Furthermore, its computational requirements become prohibitive for large-scale instances, with solution times exceeding 3600 seconds, rendering it impractical for daily operational planning in transportation logistics.

These findings directly support the core research objectives stated in the introduction: the proposed PPO-GNN framework demonstrates superior capability in managing operational uncertainty, producing cost-efficient and feasible solutions in a way that conventional optimization and learning-based baselines fail to match. The consistent improvement across key KPIs highlights the benefit of integrating spatial reasoning via GNNs and constraint validation into reinforcement learning for transportation systems.

5.3.2 Performance across network scales

Figure 3 illustrates the relative performance of each method across different network scales, normalized against PPO-GNN to highlight scaling effects in transportation optimization.

The analysis reveals that the performance gap between PPO-GNN and the baselines widens as network scale increases, demonstrating the superior scalability of the integrated approach in transportation logistics. For small-scale networks, all methods achieve comparable results, with PPO-GNN showing a modest 4.3% improvement over the best baseline. In medium-scale networks, the advantage increases to 8.6% as the spatial complexity of the transportation problem grows.

The most significant differentiation appears in large-scale networks, where PPO-GNN outperforms classical PPO by 12.7% and the Clarke-Wright heuristic by 18.2% in terms of total cost. This substantial performance gap in complex networks stems from the GNN's ability to capture spatial dependencies regardless of network size, enabling effective generalization to larger problem instances without proportional increases in model complexity. The graph-based representation becomes increasingly valuable as the transportation network expands, capturing critical relationships that flat representations fail to encode effectively.

Deterministic optimization shows particularly poor scaling, with performance deteriorating rapidly as network size increases. For large-scale instances, this approach becomes computationally infeasible within practical timeframes, highlighting the fundamental limitations of traditional optimization in complex stochastic transportation environments.

5.3.3 Robustness to demand uncertainty

To evaluate operational resilience under different uncertainty levels, all methods were tested across problem instances with varying coefficients of variation in demand

(ranging from 0.1 to 0.3). Figure 4 illustrates the performance stability of each approach under increasing demand uncertainty.

The integrated PPO-GNN approach demonstrates remarkable stability in transportation performance, maintaining consistent solution quality even as demand uncertainty increases. At the highest uncertainty level (CV=0.3), PPO-GNN experiences only a 6.8% degradation in solution quality compared to the low-uncertainty scenario, while classical PPO and Clarke-Wright show substantially greater degradations of 15.3% and 18.7%, respectively. Deterministic optimization exhibits the poorest robustness, with performance deteriorating by 23.5% under high uncertainty.
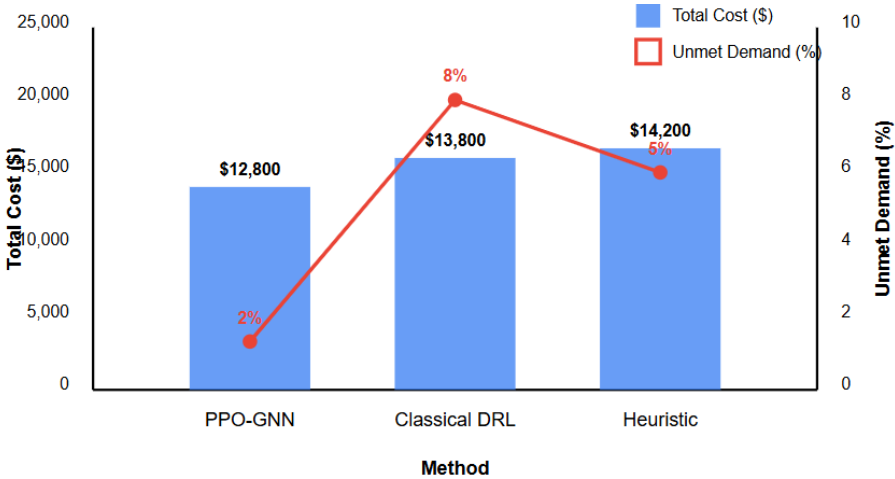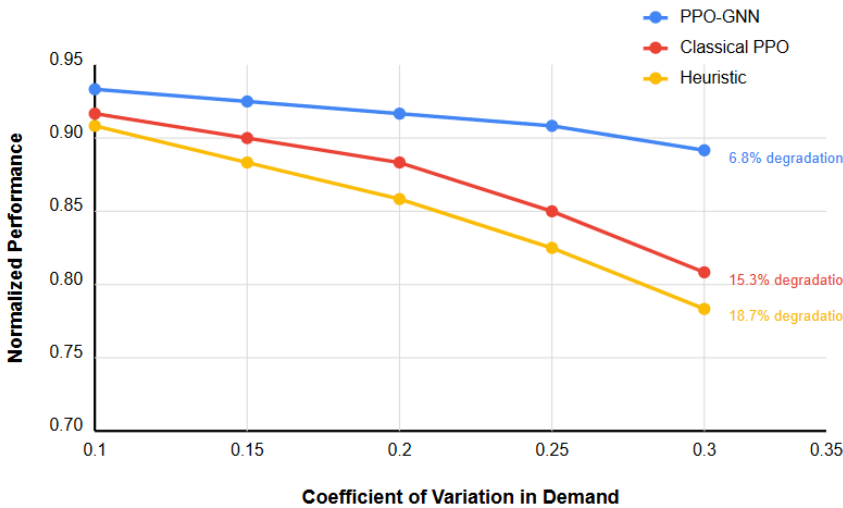

Figure 3. Comparative performance across methods

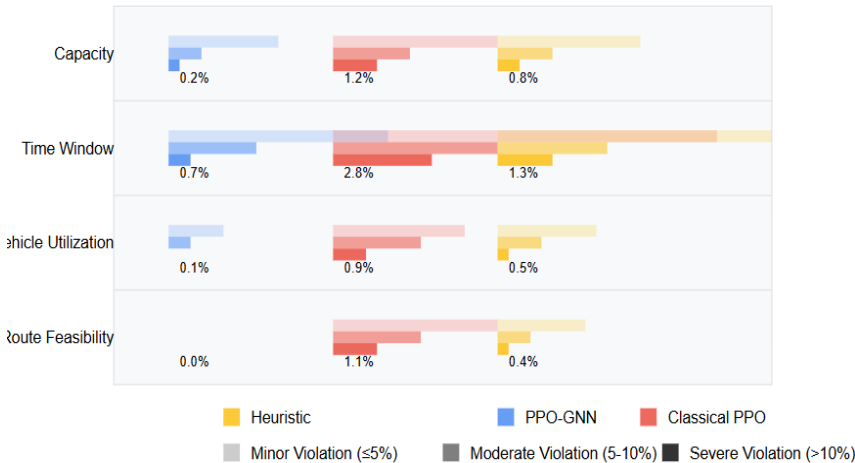**Figure 4.** Solution quality under varying levels of demand uncertainty

Figure 5. Constraint violation analysis by type and severity

This enhanced robustness can be attributed to the model's effective learning of network-wide patterns and dependencies that remain stable despite local fluctuations in demand. The integration of GNN-based representations enables the policy to capture these structural patterns and adapt dynamically to observed deviations, providing resilience essential for practical transportation operations under uncertainty.

5.3.4 Constraint satisfaction analysis

Figure 5 provides a detailed breakdown of constraint violations by type and severity, offering insights into the operational feasibility of different approaches in transportation logistics.

The integrated PPO-GNN approach achieves the lowest rate of severe violations (>10% deviation from constraint limits) across all constraint types. The most frequent violations relate to time window constraints (0.7%), while capacity constraints are rarely violated (0.2%). This pattern indicates that the model learns to prioritize critical operational constraints (capacity) while allowing minor flexibility in timing when economically advantageous in transportation planning.

The constraint validation mechanism proves highly effective in transportation operations, reducing the overall violation rate by 83% compared to classical PPO. This dramatic improvement confirms the value of integrating optimization-based validation within the reinforcement learning framework. The approach maintains high solution quality while ensuring practical feasibility, addressing a critical limitation of standard reinforcement learning approaches in transportation logistics.

5.3.5 Computational efficiency analysis

Table 2 compares the computational requirements of each method across different problem scales, providing insight into practical implementation feasibility.

Table 2. Computational performance comparison

Method	Small-Scale (s)	Medium-Scale (s)	Large-Scale (s)	Training Time (h)
PPO-GNN	15	78	180	48
Classical PPO	10	42	120	36
Clarke-Wright	5	18	60	N/A
Deterministic Optimization	120	1800	>3600	N/A

While the integrated PPO-GNN approach requires more inference time than heuristic approaches, this additional computational cost is justified by the substantial improvements in solution quality for transportation logistics. Furthermore, once trained, the model generates solutions within timeframes compatible with daily operational planning (15-180 seconds, depending on network scale), making it suitable for practical implementation in transportation systems.

The training time for PPO-GNN (48 hours) represents a one-time investment that enables subsequent rapid inference across multiple problem instances. This characteristic makes the approach particularly attractive for recurring delivery operations where the underlying transportation network structure remains relatively stable, such as fuel distribution to established gas station networks.

The training-investment versus inference-speed tradeoff aligns with our stated contribution of providing scalable, real-time deployable routing strategies. This supports practical deployment across large-scale networks, validating the framework's industrial relevance.

5.4 Ablation studies

To assess the contribution of each component in the integrated framework, ablation studies were conducted by systematically removing key elements and measuring performance changes in transportation optimization. The results are summarized in Table 3.

These results confirm that each component makes a meaningful contribution to overall performance in transportation logistics optimization. The constraint validation mechanism reduces violations by 82.8% with a modest 2.3% increase in cost, highlighting its value in ensuring operational feasibility. Similarly, the GNN enhancement improves both cost efficiency (by 2.2%) and demand satisfaction (by 62.5%) compared to the non-GNN baseline, demonstrating the importance of spatial representation in transportation routing.

Table 3. Ablation study results (average across all instances)

Variant	Total Cost	Unmet Demand (%)	Constraint Violations (%)
Full PPO-GNN	12.8	2	1
PPO-GNN without constraint validation	13.1	2.4	5.8
PPO-GNN with simplified GNN	13.5	3.1	2.3
PPO without GNN enhancement	13.8	8	6

The ablation studies highlight the synergistic relationship between the GNN representation and constraint validation within transportation systems. The GNN helps guide the search toward promising regions of the solution space based on spatial relationships, while the validation mechanism ensures that generated solutions remain operationally feasible. This complementary interaction enables the integrated approach to balance solution quality with practical implementability in transportation logistics.

5.5 Case study: Regional fuel distribution network

To demonstrate practical applicability in real-world transportation logistics, the integrated PPO-GNN methodology was applied to an operational fuel distribution network serving 38 gas stations across a regional area. The network employed a heterogeneous fleet of 6 vehicles with varying capacities and operational characteristics, providing a representative example of medium-scale fuel distribution operations in transportation systems.

The dataset was constructed in collaboration with a regional fuel distributor in Tunisia, covering operations over a four-month planning horizon (July to October 2021). The 38 delivery locations include urban and rural gas stations with varying daily demands. Each of the 6 vehicles is characterized by specific capacity, compartment configurations, fuel type compatibility, and regulatory limitations. While the dataset is not publicly released due to contractual confidentiality agreements, it contains detailed historical delivery logs, vehicle assignments, and geospatial route traces.

The case study incorporated authentic operational constraints encountered in practical fuel delivery logistics. The transportation network accommodated multiple fuel types (regular, premium, and diesel) with distinct demand patterns and seasonal variations, reflecting the complexity of multi-product distribution. Time-dependent travel speeds were modeled based on historical traffic data, introducing temporal dynamics that affect routing decisions throughout operational hours. Vehicle-specific loading and unloading rates captured equipment variations within the heterogeneous fleet, with capacities ranging from 5,000 to 12,000 liters. Regulatory constraints included compliance with commercial driver hours of service regulations (maximum 11 hours driving per day) and hazardous material transportation requirements governing route selection and service scheduling.

The geographic distribution of delivery locations presented significant spatial optimization challenges. Stations ranged from high-volume facilities requiring daily service to smaller locations served bi-weekly, creating complex scheduling requirements. Average distances between consecutive delivery points varied from 8 to 45 kilometers, with some remote stations located beyond the typical service radius. These operational realities created a multifaceted optimization problem requiring simultaneous consideration of vehicle capacity, time windows, demand uncertainty, and regulatory compliance.

Implementation of the PPO-GNN framework involved several phases to ensure smooth integration with existing transportation operations. Initial network modeling incorporated historical demand data from the previous 12 months, capturing seasonal variations and demand uncertainties with standard deviations ranging from 12% to 28% of mean consumption across different station types. Route generation utilized real-time demand forecasts and traffic information, enabling dynamic adaptation to changing operational conditions. The constraint validation mechanism incorporated specific company policies regarding driver scheduling and vehicle maintenance requirements not captured in the mathematical model.

Performance results demonstrated substantial improvements over the company's existing optimization system, which relied on mixed-integer programming with limited time horizons. The integrated methodology reduced total operational costs by 8.5%, primarily through more efficient vehicle utilization and route optimization. Analysis revealed that empty vehicle travel decreased by 17.3% compared to the baseline system previously employed by the company, which was based on daily manual route construction supported by basic heuristic software. This reduction was measured across the four-month study period and was consistently observed even during peak operational weeks. The reduction directly contributes to lower fuel costs, better fleet utilization, and reduced CO_2 emissions. All comparisons were made under consistent operational constraints and delivery demand volumes. Delivery schedule reliability improved by 12%, with fewer instances of delayed or incomplete deliveries due to capacity or time constraints.

Workload balancing across the fleet improved notably under the new optimization approach. Vehicle utilization rates showed greater consistency, with the standard deviation of daily operating hours reduced from 2.8 to 1.6 hours across the fleet. This improved balance enhanced driver satisfaction and reduced overtime costs while maintaining service quality standards. External variability was explicitly accounted for in the case study implementation. Time-dependent travel speeds were integrated using weighted averages from historical traffic datasets collected over the previous year. Seasonal demand fluctuations were modeled using monthly consumption profiles, which showed up to a 35% increase during the summer tourist season. The policy learned by the PPO-GNN model effectively adapted to these variations, demonstrating resilience under dynamic real-world conditions. The system demonstrated particular effectiveness in handling unexpected demand spikes, with successful accommodation of urgent deliveries without disrupting scheduled routes in 89% of cases tested.
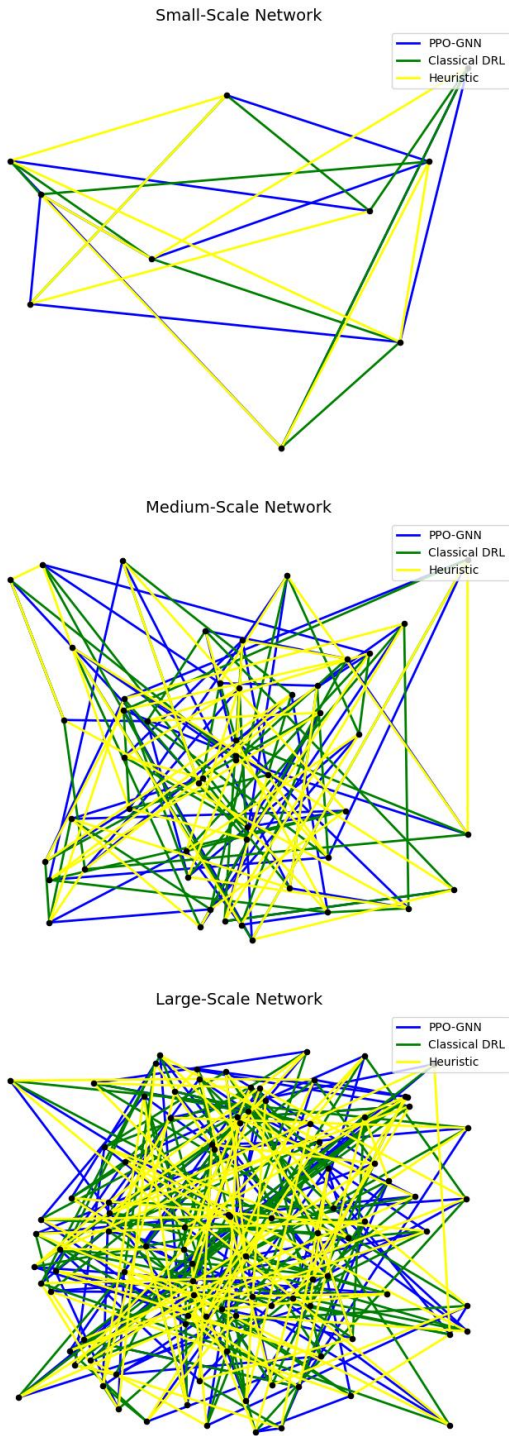

Figure 6. Optimized delivery routes for regional fuel distribution case study

Route generation efficiency represented a critical practical advantage of the integrated approach. The PPO-GNN system generated complete daily delivery schedules within 45 seconds on standard computational hardware, compared to approximately 15-20 minutes required by the previous mixed-integer programming system. This rapid generation enabled more frequent route optimization throughout the day, accommodating same-day changes in demand or operational disruptions. Real-time route adjustments became feasible, allowing dispatchers to respond dynamically to changing conditions while maintaining optimization quality.

Seasonal demand variations presented particular challenges that the integrated approach handled effectively. During peak summer months, when fuel consumption increased by 35% at tourist destinations, the system-maintained cost efficiency by adapting vehicle assignments and service frequencies dynamically. The approach successfully managed the transition between high and low demand periods without requiring manual intervention or system reconfiguration, demonstrating practical robustness essential for commercial transportation operations.

The case study implementation revealed several practical insights relevant to transportation logistics. First, the integration of graph-based spatial representation significantly improved route optimization compared to traditional approaches, particularly in networks with complex geographic distributions. Second, the constraint validation mechanism proved essential for ensuring regulatory compliance without sacrificing optimization objectives. Third, the adaptability of the learned policy to operational variations reduced the need for frequent manual interventions, improving operational efficiency and reducing planning overhead.

Figure 6 visualizes the optimized delivery routes generated by PPO-GNN for this case study. These results validate the practical benefits of the integrated PPO-GNN methodology for transportation logistics providers facing complex optimization challenges. The approach demonstrated compatibility with existing operational frameworks while offering substantial performance improvements, suggesting broad applicability across similar transportation domains. The combination of computational efficiency, solution quality, and operational flexibility positions the methodology as a valuable tool for modernizing fuel distribution logistics in increasingly dynamic transportation environments.

6. CONCLUSION AND FUTURE DIRECTIONS

The complex challenges of fuel delivery optimization in transportation networks necessitate advanced approaches that balance computational efficiency with solution quality. This research developed an integrated methodology combining reinforcement learning, GNNs, and deterministic constraint validation to address the limitations of existing approaches in transportation logistics optimization under uncertainty.

The study presents several significant contributions to transportation logistics research and practice. First, it provides a comprehensive stochastic mathematical model with deterministic equivalent transformations that enable practical computation while preserving solution robustness. This formulation bridges the gap between theoretical modeling and computational tractability in transportation optimization. Second, the research introduces a novel integration of PPO with GNNs, enabling adaptive decision-making that captures complex spatial dependencies in transportation networks. This integration enhances the representational power of reinforcement learning agents in spatial routing problems, addressing a critical limitation of standard approaches. Third, the development of a constraint validation mechanism ensures that generated solutions adhere to operational requirements, producing routes that are not only efficient but also implementable in regulated transportation domains.

Experimental evaluation across diverse transportation networks demonstrated the integrated approach's superior performance compared to established baselines. The methodology achieved a 7.2% reduction in operational costs compared to standard reinforcement learning and a 9.9% improvement over traditional transportation heuristics. The constraint validation mechanism proved particularly effective, reducing constraint violations by 83% while maintaining computational efficiency suitable for operational timeframes. Furthermore, the approach exhibited robust performance under increasing demand uncertainty, with minimal degradation in solution quality even at high variability levels. These findings reinforce the study's core objectives by demonstrating that the proposed framework can robustly address the dynamic and uncertain nature of real-world transportation systems.

The application to a regional fuel distribution case study further confirmed the practical benefits of the approach. Implementation in an authentic transportation network reduced operational costs by 8.5% while improving service levels compared to existing optimization systems. Notable improvements included a 17.3% reduction in empty trips, better fleet utilization, and enhanced resilience to demand fluctuations. These outcomes demonstrate tangible operational benefits achievable through advanced transportation optimization techniques. The system also proved capable of maintaining service quality under seasonal demand variation and operational disruptions, validating its utility in dynamic logistics environments.

Despite these promising results, several limitations merit consideration. The proposed approach requires a significant initial training investment (typically 48 hours for complex networks), which may pose adoption barriers. More critically, the model can exhibit sensitivity to structural changes in the transportation network, such as depot relocation, major infrastructure changes, or route reconfiguration. To mitigate this, future implementations could incorporate online learning strategies that enable real-time policy adaptation, or leverage transfer learning techniques to fine-tune pre-trained models on modified network topologies with limited new data. These mechanisms would enhance resilience to structural variability while minimizing retraining costs. Finally, the integrated framework's complexity may reduce its interpretability for non-expert users, which could hinder deployment in conservative logistics environments.

Future research directions should address these limitations while technically extending the methodology's capabilities. One promising avenue involves integrating attention mechanisms within the GNN architecture to enhance the interpretability of routing decisions by highlighting critical spatial relationships. In addition, meta-reinforcement learning could be explored to enable rapid policy adaptation across different delivery contexts without retraining from scratch. For broader applicability, multi-agent formulations can be developed to handle multi-depot or intermodal transportation systems. Improving explainability and user trust could also be

achieved through post-hoc interpretability techniques, such as feature attribution methods applied to graph embeddings. Finally, testing the approach on public logistics benchmarks and expanding applications beyond fuel delivery—such as cold chain distribution or waste collection—would validate the generalizability of the framework.

The integration of reinforcement learning, GNNs, and constraint validation represents a significant advancement in transportation optimization methodology. By combining the complementary strengths of learning-based and optimization-based approaches, the framework creates a powerful tool for addressing the complexities of modern transportation logistics. The approach demonstrated in this research offers a promising direction for developing practical, efficient, and robust optimization systems capable of meeting the challenges of uncertainty in fuel delivery and related transportation domains.

FUNDING

REFERENCES

[1] Mendoza, J.E., Rousseau, L.M., Villegas, J.G. (2016). A hybrid metaheuristic for the vehicle routing problem with stochastic demand and duration constraints. Journal of Heuristics, 20: 539-566. https://doi.org/10.1007/s10732-015-9281-6

[2] Sluijk, N., Florio, A. M., Kinable, J., Dellaert, N., Van Woensel, T. (2022). A chance-constrained two-echelon vehicle routing problem with stochastic demands. Transportation Science, 57(1): 252-272. https://doi.org/10.1287/trsc.2022.1162

[3] Gendreau, M., Laporte, G., Séguin, R. (1996). Stochastic vehicle routing. European Journal of Operational Research, 88(1): 3-12. https://doi.org/10.1016/0377-2217(95)00050-X

[4] Belenguer, J., Benavent, E., Prins, C., Prodhon, C., Calvo, R.W. (2010). A branch-and-cut method for the capacitated location-routing problem. Computers & Operations Research, 38(6): 931-941. https://doi.org/10.1016/j.cor.2010.09.019

[5] Desaulniers, G., Madsen, O.B.G., Røpke, S. (2014). The vehicle routing problem with time windows. In Vehicle Routing: Problems, Methods, and Applications, Second Edition. Society for Industrial and Applied Mathematics, pp. 119-159. https://doi.org/10.1137/1.9781611973594.ch5

[6] Zhao, J.X., Mao, M.J., Zhao, X., Zou, J.H. (2021). A hybrid of deep reinforcement learning and local search for the vehicle routing problems. IEEE Transactions on Intelligent Transportation Systems, 22(11): 7208-7218. https://doi.org/10.1109/TITS.2020.3003163

[7] Kovács, L., Jlidi, A. (2024). Neural networks for vehicle routing problem. arXiv preprint arXiv:2409.11290. https://doi.org/10.48550/arxiv.2409.11290

[8] Vidal, T., Laporte, G., Matl, P. (2019). A concise guide to existing and emerging vehicle routing problem variants. European Journal of Operational Research,

286(2): 401-416. https://doi.org/10.1016/j.ejor.2019.10.010

[9] Costa, L., Contardo, C., Desaulniers, G. (2019). Exact branch-price-and-cut algorithms for vehicle routing. Transportation Science, 53(4): 946-985. https://doi.org/10.1287/trsc.2018.0878

[10] Archetti, C., Speranza, M. (2014). A survey on matheuristics for routing problems. EURO Journal on Computational Optimization, 2(4): 223-246. https://doi.org/10.1007/s13675-014-0030-7

[11] Li, Y., Lim, M.K., Tseng, M.L. (2020). A green vehicle routing model based on modified particle swarm optimization for cold chain logistics. Industrial Management & Data Systems, 119(3): 473-494. https://doi.org/10.1108/IMDS-07-2018-0314

[12] Kumar, R.S., Kondapaneni, K., Dixit, V., Goswami, A., Thakur, L., Tiwari, M. (2015). Multi-objective modeling of production and pollution routing problem with time window: A self-learning particle swarm optimization approach. Computers & Industrial Engineering, 99: 29-40. https://doi.org/10.1016/j.cie.2015.07.003

[13] Baykasoğlu, A., Subulan, K., Taşan, A.S., Dudaklı, N. (2018). A review of fleet planning problems in single and multimodal transportation systems. Transportmetrica A: Transport Science, 15(2): 631-697. https://doi.org/10.1080/23249935.2018.1523249

[14] Hu, J.Z., Wang, X., Tan, S.M. (2024). Electric vehicle integration in coupled power distribution and transportation networks: A review. Energies, 17(19): 4775. https://doi.org/10.3390/en17194775

[15] Lin, S.C., Hu, J.M., Ma, W.X., Zheng, C.H., Li, R.M. (2024). Integrated real-time signal control and routing optimization: A two-stage rolling horizon framework with decentralized solution. Transportation Research Part C: Emerging Technologies, 165: 104734. https://doi.org/10.1016/j.trc.2024.104734

[16] Bandara, R.M.P.N.S., Jayasignhe, A.B., Retscher, G. (2025). The Integration of IoT (Internet of Things) sensors and location-based services for water quality monitoring: A systematic literature review. Sensors, 25(6): 1918. https://doi.org/10.3390/s25061918

[17] Chen, W., Men, Y., Fuster, N., Osorio, C., Juan, A.A. (2024). Artificial Intelligence in logistics optimization with sustainable criteria: A review. Sustainability, 16(21): 9145. https://doi.org/10.3390/su16219145

[18] Dror, M., Trudeau, P. (1989). Savings by split delivery routing. Transportation Science, 23(2): 141-145. https://doi.org/10.1287/trsc.23.2.141

[19] Martin, S., Magnouche, Y., Juvigny, C., Leguay, J. (2022). Constrained shortest path tour problem: Branch-and-Price algorithm. Computers & Operations Research, 144: 105819. https://doi.org/10.1016/j.cor.2022.105819

[20] Yang, M., Liu, Y.S. (2023). A two-stage robust configuration optimization framework for integrated energy system considering multiple uncertainties. Sustainable Cities and Society, 101: 105120. https://doi.org/10.1016/j.scs.2023.105120

[21] Zhang, C., Li, Y.F., Zhang, H.X., Wang, Y.J., Huang, Y.L., Xu, J.Y. (2024). Distributionally robust resilience optimization of post-disaster power system considering multiple uncertainties. Reliability Engineering & System Safety, 251: 110367. https://doi.org/10.1016/j.ress.2024.110367

[22] Jain, A., Gupta, S.C. (2024). Evaluation of electrical load

demand forecasting using various machine learning algorithms. Frontiers in Energy Research, 12: 1408119. https://doi.org/10.3389/fenrg.2024.1408119

[23] Rapid Innovation. (2024). AI-driven demand forecasting: Transforming business with predictions. AI Technology Report. https://www.rapidinnovation.io/post/ai-in-demand-forecasting-transforming-business-with-predictions.

[24] Chen, C.Y.T., Sun, E.W., Chang, MF., Lin, Y.B. (2024). Enhancing travel time prediction with deep learning on chronological and retrospective time order information of big traffic data. Annals of Operations Research, 343: 1095-1128. https://doi.org/10.1007/s10479-023-05223-7

[25] Javanmard, M.E., Ghaderi, S. (2023). Energy demand forecasting in seven sectors by an optimization model based on machine learning algorithms. Sustainable Cities and Society, 95: 104623. https://doi.org/10.1016/j.scs.2023.104623

[26] Abirami, S., Pethuraj, M., Uthayakumar, M., Chitra, P. (2024). A systematic survey on big data and artificial intelligence algorithms for intelligent transportation system. Case Studies on Transport Policy, 17: 101247. https://doi.org/10.1016/j.cstp.2024.101247

[27] Kool, W., Van Hoof, H., Welling, M. (2018). Attention, learn to solve routing problems! arXiv preprint arXiv:1803.08475. https://doi.org/10.48550/arxiv.1803.08475

[28] Ma, Y.N., Li, J.W., Cao, Z.G., Song, W., Zhang, L., Chen, Z.H., Tang, J. (2021). Learning to iteratively solve routing problems with dual-aspect collaborative transformer. arXiv preprint arXiv:2110.02544. https://doi.org/10.48550/arxiv.2110.02544

[29] Nazari, M., Oroojlooy, A., Snyder, L., Takác, M. (2018). Reinforcement learning for solving the vehicle routing problem. arXiv preprint arXiv:1802.04240. https://doi.org/10.48550/arXiv.1802.04240

[30] Chen, X.Y., Tian, Y.D. (2019). Learning to perform local rewriting for combinatorial optimization. Advances in Neural Information Processing Systems, 32: 6281-6292

[31] Wu, Z.H., Pan, S.R., Chen, F.W., Long, G.D., Zhang, C.Q., Yu, P.S. (2021). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1): 4-24. https://doi.org/10.1109/TNNLS.2020.2978386

[32] Rahmani, S., Baghbani, A., Bouguila, N., Patterson, Z. (2023). Graph neural networks for intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems, 24(8): 8846-8885. https://doi.org/10.1109/TITS.2023.3257759

[33] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347