

Article

Optimizing Large-Scale Sugarcane Transportation under Capacity and Queue Constraints Using a Hybrid Evolutionary Framework

Putis Wittayasin^a, Nuchsa Kriengkorakot^{b,*}, and Preecha Kriengkorakot^c

Department of Industrial Engineering, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand
E-mail: ^aputis.wi.67@ubu.ac.th, ^{b,*}nuchsa.k@ubu.ac.th (Corresponding author), ^cpreecha.k@ubu.ac.th

Abstract. Large-scale sugarcane transportation requires balancing transportation efficiency with congestion at processing facilities under capacity constraints. This study proposes a queue-aware bi-objective optimization framework integrating Non-dominated Sorting Genetic Algorithm II (NSGA-II), Adaptive Large Neighborhood Search (ALNS), and reinforcement learning (RL) for state-dependent search-mode control. The objectives are to minimize total transportation distance and average queue waiting time. The framework was evaluated using a real-world case comprising 199 subdistricts and four sugar mills in northeastern Thailand. Exhaustive enumeration on a reduced instance was first used to validate the ability of the proposed method to approximate a known reference Pareto front. Large-scale experiments compared the RL-guided framework with standalone NSGA-II, Random ALNS, tuned MOEA/D, and a greedy baseline. The RL-guided method achieved the best mean performance, with a transportation distance of 144,392.49 km and an average waiting time of 1.383 h/trip, together with the best mean HV and IGD values. Friedman and Holm-adjusted Wilcoxon tests confirmed significant improvements over NSGA-II for distance, HV, and IGD and over MOEA/D for all evaluated measures, while differences from Random ALNS were not statistically significant. Relative to the non-optimized baseline, the proposed method reduced transportation distance by 9.93% and fuel consumption and CO₂ emissions by 9.05%. The results demonstrate the potential of queue-aware RL-guided optimization to support efficient and sustainable large-scale agricultural transportation planning.

Keywords: Sugarcane transportation, Multi-objective optimization, Reinforcement learning, Adaptive large neighborhood search, Queue-aware logistics

ENGINEERING JOURNAL Volume 30 Issue 9

Received 11 May 2026

Accepted 1 September 2026

Published 30 September 2026

Online at <https://engj.org/>

DOI:10.4186/ej.2026.30.9.43

1. Introduction

Agricultural logistics is a critical component of agro-industrial supply chains because the timely and efficient movement of harvested products from geographically dispersed production areas to processing facilities directly affects transportation efficiency, resource utilization, processing continuity, and overall supply-chain performance. Sugarcane logistics is particularly challenging because large quantities of harvested cane must be transported within a limited harvesting period while transportation resources and mill processing capacities are constrained. Mathematical optimization has therefore received considerable attention across different stages of the sugarcane supply chain, including harvesting, transportation, processing, and integrated planning. Nevertheless, a recent review of optimization models for the sugarcane supply chain indicates that research remains concentrated on harvesting-related decisions, while greater integration among transportation, processing, and operational interactions is still needed [1]. This limitation is particularly relevant in large production regions where numerous geographically dispersed supply areas must be coordinated with a relatively small number of processing facilities.

Transportation and processing decisions are strongly interdependent in sugarcane logistics. A transportation plan that minimizes geographical distance alone may allocate large quantities of cane to the nearest mills and consequently concentrate truck arrivals at a limited number of facilities. Because mills have finite unloading and processing capacities, such assignments can increase congestion and truck waiting times even when transportation distance is reduced. Higgins [2] demonstrated this interaction in a real sugarcane road-transport system, showing that improved coordination of vehicle arrivals can reduce queueing and mill idle time. Similarly, Aliano Filho et al. [3] developed a bi-objective model integrating sugarcane harvesting and transportation and demonstrated that meaningful trade-offs emerge when multiple operational criteria are optimized simultaneously. These findings indicate that efficient sugarcane transportation cannot be evaluated solely through transportation distance or cost. Congestion and service performance at processing facilities must also be considered because geographically efficient assignments do not necessarily produce operationally efficient arrival patterns.

Both exact optimization and metaheuristic approaches have been applied to sugarcane logistics. Jena and Poggi [4] formulated tactical and operational sugarcane harvest-planning problems using mixed-integer programming and introduced computational strategies to improve solution efficiency. Exact approaches are valuable for obtaining optimal solutions on tractable instances; however, their computational burden can increase substantially when a problem contains hundreds of supply locations, several processing facilities, capacity restrictions, and operational interactions. Metaheuristics

therefore remain attractive for large and operationally detailed problems. Pitakaso and Sethanan [5], for example, developed an Adaptive Large Neighborhood Search (ALNS) for large-scale sugarcane inbound logistics involving harvester assignment, routing, time windows, and shared resources. Their results demonstrate the potential of neighborhood-based search for complex agricultural logistics problems in which the solution structure cannot be efficiently explored using simple allocation or routing rules alone.

Multi-objective evolutionary optimization provides an appropriate framework when logistics decisions involve conflicting performance criteria. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) proposed by Deb et al. [6] uses elitist non-dominated sorting and diversity preservation to approximate multiple trade-off solutions without requiring predetermined objective weights. This characteristic is suitable for the present study because reducing transportation distance and reducing queue waiting time can require different allocation decisions. Nevertheless, population-based evolutionary search alone may require substantial computational effort to refine solutions in large combinatorial spaces. Large neighborhood search provides a complementary mechanism by modifying significant portions of a candidate solution and exploring alternative neighborhoods during local improvement [7]. The combination of NSGA-II and neighborhood-based search therefore has a clear methodological rationale: NSGA-II supports global Pareto-based exploration and population diversity, whereas local search provides focused intensification around promising solutions.

Recent developments in machine-learning-assisted optimization provide additional opportunities to improve search control. Bengio et al. [8] highlighted the growing role of machine learning in combinatorial optimization, while Karimi-Mamaghan et al. [9] showed that learning mechanisms can support several components of metaheuristics, including parameter control, operator selection, initialization, and cooperation among search strategies. Reinforcement learning (RL) is particularly relevant because it can adapt decisions according to the state of the optimization process and feedback from previous actions. Recent studies have integrated RL directly with ALNS. Reijnen et al. [10] employed deep reinforcement learning to control ALNS components according to the evolving search state, whereas Johnn et al. [11] formulated operator selection as a Markov decision process and developed a graph reinforcement learning framework for adaptive large neighborhood search. RL-assisted optimization has also recently been introduced into sugarcane logistics for heterogeneous fleet scheduling, routing, and machine-breakdown conditions [12]. However, these studies do not specifically address large-scale sugarcane allocation in which transportation decisions are explicitly evaluated together with mill queue congestion.

Related engineering logistics studies further demonstrate the importance of hybrid and multi-

objective optimization in practical transportation systems. Previous studies published in *Engineering Journal* have examined dynamic multi-depot vehicle routing [13], capacitated vehicle routing with multiple objectives [14], and sustainable multi-objective production and distribution planning [15]. More recent studies have further investigated robust fuzzy multi-objective aggregate production planning under uncertainty [16] and hybrid optimization for supply chain network design under demand uncertainty [17]. These studies collectively demonstrate the relevance of multi-objective and hybrid optimization for complex engineering and supply-chain decision problems. In addition, decomposition-based multi-objective algorithms such as MOEA/D provide an alternative to dominance-based approaches by decomposing a multi-objective problem into a set of scalar subproblems [18]. Such methodological diversity is important when evaluating a newly proposed framework because comparison only with closely related variants may not provide sufficient evidence of competitive performance.

The present study also extends our previous research on large-scale Thai sugarcane transportation [19]. The earlier study considered a capacitated transportation system involving 199 subdistricts and four factories and applied an RL-guided NSGA-II–ALNS framework to minimize transportation distance and the number of truck trips. Although that framework demonstrated the feasibility of learning-guided hybrid optimization for a large agricultural transportation network, congestion and truck waiting at processing facilities were not explicitly represented. The present study addresses this limitation by replacing trip minimization with a queue-aware service objective and explicitly evaluating truck arrivals, limited unloading capacity, service processes, and waiting times at the mills. The principal research gap therefore lies at the intersection of large-scale capacitated transportation, queue-aware operational evaluation, multi-objective evolutionary optimization, and learning-guided local search. Existing sugarcane studies have addressed harvesting, routing, scheduling, or transportation efficiency, while RL-assisted metaheuristics have primarily focused on generic routing or direct operator-control mechanisms. Evidence remains limited regarding state-dependent hybrid search for balancing transportation efficiency and processing congestion in a real-world agricultural network with hundreds of supply nodes.

To address this gap, this study proposes a queue-aware RL-guided hybrid NSGA-II–ALNS framework that simultaneously minimizes total transportation distance and average queue waiting time per trip. Candidate solutions are decoded into feasible subdistrict-to-mill allocations under supply and processing-capacity requirements, after which truck trips, arrival times, and unloading queues are evaluated through a queue-simulation mechanism. NSGA-II performs global Pareto-based exploration, while selective ALNS provides local improvement. Unlike approaches in which RL

directly selects an individual neighborhood operator, the proposed controller uses information describing the current optimization state—including search progress, temperature, recent improvement, congestion condition, utilization level, and distance condition—to select a higher-level search mode, after which an operator is sampled from the corresponding mode-specific distribution. This hierarchical search-mode mechanism constitutes the main algorithmic distinction of the framework. The method is evaluated on a real-world 199-subdistrict, four-mill case and compared with a greedy baseline, standalone NSGA-II, random-ALNS hybridization, and tuned MOEA/D. Its reliability is further examined through exhaustive Pareto validation on a tractable instance, normalized Pareto-quality indicators, parameter-sensitivity analysis, repeated independent runs, and paired non-parametric statistical tests.

2. Materials and methods

2.1. Problem Description and Case Study

The study considers a daily sugarcane transportation system in northeastern Thailand in which harvested sugarcane from geographically dispersed subdistricts must be allocated to multiple processing mills under finite processing capacities. The system is represented as a capacitated multi-source, multi-destination transportation network. Each subdistrict has a known daily sugarcane supply, whereas each mill has a maximum daily processing capacity. The planning problem therefore determines feasible subdistrict-to-mill allocations while balancing transportation efficiency and congestion at the receiving facilities.

The main real-world instance consists of 199 sugarcane-producing subdistricts and four processing mills. Total daily sugarcane supply is 98,687 tons, whereas the aggregate mill capacity is 118,000 tons/day, corresponding to an overall utilization of approximately 83.6% and a capacity margin of 19,313 tons/day. Although the system is feasible in aggregate, mill capacities are unevenly distributed and the supply locations are spatially dispersed. Consequently, allocating cane solely to the geographically nearest mills may reduce transportation distance but can concentrate truck arrivals and increase unloading congestion. Redistributing supply to less congested mills may reduce waiting time at the expense of longer transportation distances.

The input data include subdistrict-level sugarcane supply, mill processing capacities, and pairwise origin-to-mill distances. Individual mill capacities range from approximately 12,000 to 50,000 tons/day, and origin-to-mill distances range from approximately 10 to 150 km. Geographic information was processed using GIS/QGIS and Python to prepare the distance matrix, while demand and capacity data were checked for consistency before optimization. The underlying production and industry

data were obtained from the Office of the Cane and Sugar Board, Thailand [20].

Transportation demand is converted into truck trips during solution decoding. Two payload classes, 20 and 40 tons per trip, are represented in the operational evaluation. The model does not impose an explicit finite-fleet availability constraint; instead, sufficient contracted trucking capacity is assumed to execute the generated trips within the daily planning horizon. The number of trips is determined from the allocated sugarcane quantity and the corresponding truck payload capacity. This formulation focuses on sugarcane-to-mill allocation and the resulting transportation and congestion effects rather than detailed fixed-fleet vehicle scheduling.

Accordingly, the problem is formulated as a bi-objective capacitated transportation problem that minimizes total transportation distance and average queue waiting time per trip. The mathematical formulation and feasibility constraints are presented in Section 2.2, while the temporal queue-evaluation mechanism is described in Section 2.3. The principal characteristics of the case study are summarized in Table 1.

Table 1. Characteristics of the large-scale case study

Characteristic	Value
Sugarcane-producing subdistricts	199
Processing mills	4
Total daily supply	98,687 tons/day
Total mill capacity	118,000 tons/day
Overall capacity utilization	83.6%
Capacity margin	19,313 tons/day
Individual mill capacity	12,000–50,000 tons/day
Origin-to-mill distance range	10–150 km
Truck payload classes	20 and 40 tons/trip
Planning horizon	Daily / single-period
Objectives	Distance and average queue waiting time

2.2. Mathematical Formulation

Let S denote the set of sugarcane-producing subdistricts and F the set of processing mills. For each subdistrict $s \in S$, D_s denotes the available sugarcane supply (tons/day), and C_f denotes the daily processing capacity of mill $f \in F$. Let d_{sf} be the base distance between subdistrict s and mill f , and let $\bar{d}_{sf}$ denote the effective transportation distance used in the operational evaluation. Each subdistrict is associated with a predetermined truck payload capacity $q_s \in \{20, 40\}$ tons/trip. The principal allocation variable is $x_{sf} \geq 0$, representing the quantity of sugarcane transported from subdistrict s to mill f . The corresponding number of trips is denoted by $t_{sf} \in \mathbb{Z}_+$.

The problem simultaneously minimizes total transportation distance and average queue waiting time per trip. The first objective is

$$\min Z_1 = \sum_{s \in S} \sum_{f \in F} \bar{d}_{sf} t_{sf}, \quad (1)$$

where transportation distance is accumulated over all generated truck trips. The second objective is

$$\min Z_2 = \frac{1}{N} \sum_{s \in S} \sum_{f \in F} \sum_{r=1}^{t_{sf}} W_{sfr}, \quad (2)$$

where W_{sfr} is the queue waiting time of trip r transporting sugarcane from s to f , and

$$N = \sum_{s \in S} \sum_{f \in F} t_{sf}$$

is the total number of truck trips. The waiting-time values are generated by the queue-evaluation procedure described in Section 2.3.

All harvested sugarcane must be allocated within the planning horizon

$$\sum_{f \in F} x_{sf} = D_s, \forall s \in S \quad (3)$$

The total quantity assigned to each mill cannot exceed its daily processing capacity

$$\sum_{s \in S} x_{sf} \leq C_f, \forall f \in F \quad (4)$$

The number of truck trips is determined from the allocated quantity and the payload capacity assigned to each subdistrict

$$t_{sf} = \frac{x_{sf}}{q_s}, \forall s \in S, f \in F, \quad (5)$$

with $t_{sf} = 0$ when $x_{sf} = 0$. This relationship allows the final trip on an origin–destination allocation to carry less than the nominal truck capacity, consistent with the implemented trip-generation procedure. The decision domains are

$$x_{sf} \geq 0, \quad t_{sf} \in \mathbb{Z}_+, \forall s \in S, f \in F. \quad (6)$$

Constraints (3)–(6) are enforced through the capacity-aware decoding procedure rather than treated as penalty constraints within the evolutionary algorithm. Subdistricts are processed according to the encoded priority sequence and allocated first to their nearest feasible mill; if the remaining mill capacity is insufficient, the residual quantity is transferred to the next feasible mill until the complete supply is assigned. This mechanism ensures supply conservation and prevents mill-capacity violations for every evaluated solution.

The model does not impose a separate finite-fleet availability constraint because sufficient contracted trucking capacity is assumed for the daily planning horizon. Instead, vehicle requirements are represented through the integer trip variables in (5), with truck payload capacities fixed at either 20 or 40 tons per trip for each subdistrict. This assumption separates the strategic allocation and congestion problem considered here from a detailed fixed-fleet vehicle-scheduling problem. In the implementation, the assigned load is explicitly split into successive truck trips according to the corresponding payload capacity.

2.3. Queue-Aware Evaluation Model

A queue-aware evaluation model is embedded in the solution evaluation process to capture congestion and truck waiting at the processing mills. After a feasible subdistrict-to-mill allocation is obtained, the allocated sugarcane is divided into truck trips according to the assigned truck payload capacity. Each trip is then characterized by its dispatch time, travel time, arrival time, service time, and queue waiting time.

For each subdistrict s , the base dispatch time is distributed across the 24-h planning horizon according to its position p_s in the decoded priority sequence

$$T_s^{base} = \frac{p_s}{|S|-1} H \quad (7)$$

where $H = 24$ h. To avoid artificially synchronized departures, a small reproducible jitter ε_s within ± 0.15 h is added. For repeated trips from the same subdistrict, consecutive departures are separated by $g = 0.10$ h. The dispatch time of trip r is therefore

$$T_{sr}^{dep} = \max(0, T_s^{base} + \varepsilon_s) + (r - 1)g. \quad (8)$$

The final implementation uses a fixed planning horizon of 24 h, a dispatch jitter of ± 0.15 h, and a 0.10-h interval between successive trips from the same subdistrict.

Travel time is estimated using an average truck speed of $v = 35$ km/h. Because the spatial distance matrix does not explicitly represent all road-network deviations, the base distance d_{sf} is adjusted using a detour factor $\delta = 1.3$

$$\bar{d}_{sf} = \delta d_{sf} \quad (9)$$

and the corresponding travel and arrival times are

$$T_{sfr}^{travel} = \frac{\bar{d}_{sf}}{v}, \quad (10)$$

$$T_{sfr}^{arr} = T_{sr}^{dep} + T_{sfr}^{travel}. \quad (11)$$

These relationships correspond directly to the implemented queue evaluation, in which the adjusted

distance is divided by the average speed and added to the dispatch time to determine mill arrival time.

Each processing mill is modeled with six parallel unloading bays. The total service rate of mill f is derived from its daily processing capacity C_f over the 24-h operating period

$$\mu_f = \frac{C_f}{24}, \quad (12)$$

and the service rate available at each unloading bay is

$$\mu_f^{bay} = \frac{\mu_f}{6}. \quad (13)$$

For a truck carrying load L_{sfr} , the service time includes a fixed setup time of 0.08 h and the load-dependent unloading time

$$T_{sfr}^{ser} = 0.08 + \frac{L_{sfr}}{\mu_f^{bay}}. \quad (14)$$

Thus, mills with greater daily processing capacity have proportionally higher service rates while retaining the same number of parallel unloading bays.

At each mill, trucks are processed according to ascending arrival time. Priority position and trip sequence are used as deterministic tie-breakers when arrival times coincide. Each arriving truck is assigned to the unloading bay with the earliest available time. If A_{fb} denotes the next available time of the selected bay b , the service starting time and queue waiting time are calculated as

$$T_{sfr}^{start} = \max(T_{sfr}^{arr}, A_{fb}), \quad (15)$$

$$W_{sfr} = T_{sfr}^{start} - T_{sfr}^{arr}. \quad (16)$$

After service begins, the bay remains occupied for T_{sfr}^{ser} , and its availability time is updated to $T_{sfr}^{start} + T_{sfr}^{ser}$. No artificial upper limit is imposed on waiting time. The resulting waiting times for all trips are aggregated to obtain the average queue waiting time used as objective Z_2 in Eq. (2). The implemented procedure therefore represents a multi-server queue with finite parallel unloading capacity and explicitly links transportation allocation, arrival patterns, mill capacity, and congestion.

The queue evaluation is reproducible for a given candidate solution. Although dispatch jitter is introduced to generate realistic temporal variation, it is generated deterministically from the priority position using a fixed random seed rather than being resampled at every objective evaluation. Consequently, identical candidate solutions produce identical queue-performance values, allowing fair comparisons among optimization algorithms while avoiding unrealistically synchronized truck arrivals.

2.4. Solution Representation and Capacity-Aware Decoding

Each candidate solution is represented by a continuous priority vector

$$\mathbf{r} = (r_1, r_2, \dots, r_{|S|}), \quad 0 \leq r_s \leq 1, \quad (17)$$

where each element corresponds to one sugarcane-producing subdistrict. The vector does not directly encode factory assignments or truck routes. Instead, the subdistrict processing order is obtained by sorting the priority values in ascending order

$$\pi = \text{argsort}(\mathbf{r}), \quad (18)$$

where π defines the sequence in which subdistricts are processed by the decoder.

For each subdistrict in the priority sequence, candidate mills are examined in ascending order of transportation distance. The decoder assigns as much sugarcane as possible to the nearest mill with available processing capacity. If the remaining capacity of that mill is insufficient, the unallocated quantity is transferred to the next nearest feasible mill, and this process continues until the entire supply of the subdistrict has been assigned. Thus, a subdistrict may be split among more than one mill only when the capacity of a closer mill is insufficient. This decoding procedure directly enforces the supply-balance and mill-capacity requirements defined in Eqs. (3) and (4).

After allocation, the assigned quantities are converted into truck trips according to the predetermined 20- or 40-ton payload capacity associated with each subdistrict. The resulting trips are then passed to the queue-aware evaluation model described in Section 2.3 to calculate transportation distance and average waiting time. Therefore, each priority vector is deterministically transformed into a feasible transportation plan and its two objective values.

This indirect representation is used throughout NSGA-II and the subsequent local-search procedures. Modifying the priority sequence can change both mill allocation and truck arrival patterns while feasibility is restored automatically by the decoder. Consequently, the search algorithms can explore alternative allocation patterns without requiring explicit repair of supply or mill-capacity violations after every search operation.

2.5. Proposed RL-Guided Hybrid NSGA-II-ALNS Framework

The proposed framework integrates NSGA-II [6] with ALNS-based local search [7] and reinforcement learning (RL) to balance global exploration, local improvement, and adaptive search control. NSGA-II serves as the main multi-objective optimization backbone, maintaining a population of candidate priority vectors and evolving them through Pareto-based selection. Each

candidate is decoded using the capacity-aware procedure in Section 2.4 and evaluated using total transportation distance and average queue waiting time.

To improve promising regions of the search space without applying computationally expensive local search to the entire population, ALNS is selectively activated at predefined intervals. Approximately 25% of the population is targeted for local improvement, while a fraction of high-quality elite solutions is preserved from perturbation. The selected priority sequences are modified using a pool of neighborhood operators and subsequently re-decoded and re-evaluated through the same capacity and queue models. This selective mechanism reduces unnecessary evaluations while retaining local-search intensification.

The RL component does not directly select an individual neighborhood operator. Instead, it performs a higher-level search-control decision by selecting one of five search modes: DISTANCE, QUEUE_RELIEF, BALANCED, INTENSIFY, or SKIP. Each active mode is associated with a predefined probability distribution over the available neighborhood operators. After the RL controller selects a mode according to the current search state, an operator is sampled from the corresponding mode-specific distribution and applied to the selected priority sequence. This hierarchical structure allows the search behavior to respond to current distance, congestion, utilization, and search-progress conditions without learning a separate action for every individual operator.

The resulting candidate is decoded, evaluated, and considered for acceptance using a simulated-annealing-based mechanism that favors Pareto or normalized scalar improvement while allowing controlled diversification. Queue-relief moves are additionally restricted from causing excessive deterioration in transportation distance. Accepted solutions are returned to the NSGA-II population, and the process continues until the termination criterion is reached. The detailed RL state representation, search-mode actions, learning procedure, and policy parameters are described in Section 2.6.

2.6. RL Search-Mode Policy

The RL controller is formulated as a discrete state-action policy that determines the search mode applied during selective local improvement. Rather than treating individual neighborhood operators as RL actions, the controller selects a higher-level search mode, thereby reducing the action space and allowing multiple operators with similar search purposes to be grouped under a common policy.

The state representation summarizes both search progress and current solution characteristics. Each state contains nine discretized components: generation progress, current search temperature, previously selected mode, recent normalized solution change, mean queue waiting time, maximum mill utilization, proportion of flexible subdistricts, transportation-distance stretch

relative to nearest-mill allocation, and the number of subdistricts experiencing high waiting times. The available actions are the five search modes defined in Section 2.5: DISTANCE, QUEUE_RELIEF, BALANCED, INTENSIFY, and SKIP. For the four active modes, an individual neighborhood operator is subsequently sampled from the corresponding predefined mode-specific probability distribution; the SKIP action performs no local perturbation.

The reward reflects improvements in both optimization objectives. A Pareto-dominating move receives the highest positive reward, whereas other improving moves are evaluated using normalized changes in transportation distance and waiting time with phase-dependent weights. Non-improving moves receive a negative reward, while additional mode-specific bonuses or penalties encourage behavior consistent with the current search condition. For example, distance-reduction moves are favored during the later distance-focused phase, whereas queue-relief moves receive additional credit when they reduce waiting time.

During policy training, the state-action value is updated using an advantage-adjusted Q-learning rule

$$B(s) \leftarrow B(s) + \beta[R - B(s)], \quad (19)$$

$$Q(s, a) \leftarrow Q(s, a) + \eta [R - B(s) + \gamma \max_{a'} Q(s', a') - Q(s, a)], \quad (20)$$

where R is the observed reward, $B(s)$ is a state-dependent reward baseline, η is the learning rate, and γ is the discount factor. The baseline reduces the influence of state-specific reward magnitude and allows the Q-values to represent the relative advantage of alternative search modes. The implemented training procedure uses multiple independent seeds, and the resulting Q-tables are averaged and stored for subsequent optimization runs.

For the final optimization experiments, the pretrained Q-table is loaded and held fixed rather than retrained for each benchmark run. The stored Q-values are adjusted according to the current search phase and operational conditions, after which one of the highest-valued candidate modes is selected using temperature-controlled probabilistic sampling with a small exploration probability. This separation between offline policy training and inference prevents information leakage between benchmark runs while retaining limited stochastic exploration during optimization. The detailed training and inference parameters are reported in Section 2.8.

2.7. Algorithm Workflow

The optimization process begins with a greedy-informed initial population of priority vectors. The overall workflow of the proposed queue-aware RL-guided hybrid NSGA-II–ALNS framework is illustrated in Fig. 1. Each individual is decoded into a feasible subdistrict-to-mill allocation using the procedure

described in Section 2.4 and subsequently evaluated using the transportation-distance and queue-waiting objectives. The mean objective values of the initial population are also used to construct an auxiliary normalized scalar score for local-search targeting and acceptance decisions. NSGA-II itself remains Pareto-based and does not use this scalar score for non-dominated sorting or population survival.

At each generation, NSGA-II performs its standard evolutionary update to maintain global exploration and solution diversity. At predefined intervals, a subset of approximately 25% of the population is selected for local improvement, while high-quality elite individuals are protected from perturbation. For each selected individual, the current search state is constructed and passed to the pretrained RL policy, which selects one of the five search modes described in Section 2.6. If an active mode is selected, a neighborhood operator is sampled from its mode-specific probability distribution and applied to the priority sequence.

The modified sequence is then converted back into a priority vector, decoded into a capacity-feasible transportation plan, and evaluated using the same queue-aware model. The candidate solution is accepted if it provides a Pareto improvement or an improvement in the auxiliary normalized scalar score; otherwise, controlled non-improving moves may be accepted through the simulated-annealing mechanism. Additional safeguards restrict queue-relief moves that cause excessive increases in transportation distance. Accepted solutions replace their corresponding individuals in the current population before the evolutionary search continues.

The procedure is repeated until the prescribed number of generations is reached. The final population is re-evaluated, and non-dominated sorting is applied to extract the final Pareto approximation representing the trade-off between transportation distance and average queue waiting time. This workflow combines Pareto-based global exploration, selective local improvement, queue-aware evaluation, and state-dependent RL search control within a unified optimization framework.

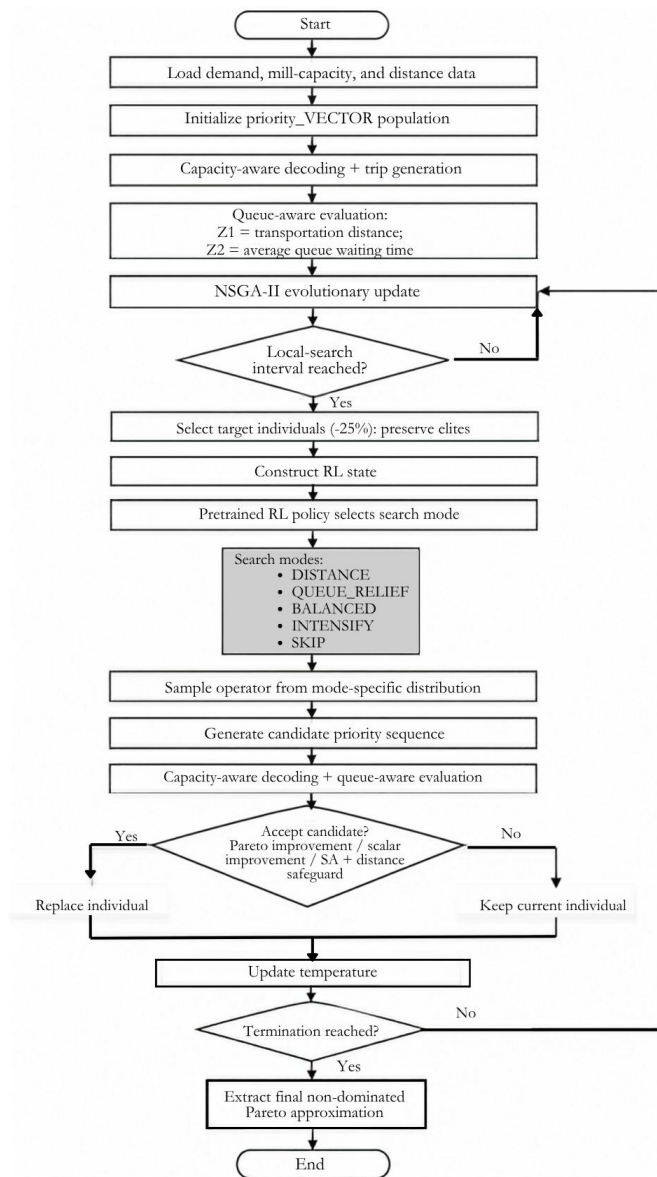


Fig. 1. Overall workflow of the proposed queue-aware RL-guided hybrid NSGA-II-ALNS framework

2.8. Experimental Design and Parameter Settings

The computational experiments were conducted under a common experimental setting to ensure a fair comparison among the stochastic optimization methods. The main experiments used a population size of 30 and a termination limit of 100 generations. Ten matched random seeds (11, 22, 33, 44, 55, 66, 77, 88, 99, and 100) were used for repeated runs. For each seed, the evolutionary methods started from the same greedy-informed initial population in order to reduce variation caused only by initialization. Because the hybrid methods perform additional local-search evaluations, the runtime of each method is also reported.

The RL search-mode policy was pretrained independently before the final benchmark experiments. The pretraining phase used five seeds (101, 202, 303, 404, and 505), a population size of 20, and 100 generations. The Q-learning parameters were set as follows: learning rate $\eta = 0.20$, discount factor $\gamma = 0.90$, exploration

probability $\epsilon = 0.08$, initial softmax temperature = 1.00, minimum temperature = 0.25, and reward-baseline learning rate = 0.10. During the final optimization runs, the pretrained Q-table was fixed and used only for inference.

For the final RL-guided optimization, the inference policy used an initial softmax temperature of 0.90, a minimum value of 0.35, and a small exploration probability $\epsilon = 0.02$. The simulated-annealing component used an initial temperature $T_0 = 200$ and a cooling factor $\alpha = 0.98$. Selective ALNS was activated every 5 generations. The main parameter settings are summarized in Table 2.

To evaluate robustness, a one-factor-at-a-time sensitivity analysis was additionally conducted for the proposed RL-guided framework using five representative seeds (11, 33, 55, 77, and 99). The tested settings included $T_0 \in \{100, 200, 400\}$, $\alpha \in \{0.95, 0.98, 0.995\}$, ALNS frequency $\in \{3, 5, 10\}$, initial inference softmax temperature $\in \{0.60, 0.90, 1.20\}$, and inference exploration probability $\epsilon \in \{0, 0.02, 0.05\}$. The baseline configuration was retained for the final benchmark comparison, and the sensitivity analysis was used only to assess robustness rather than to retrospectively choose the best parameter combination.

Table 2. Main experimental and algorithmic parameter settings

Parameter	Setting
Population size	30
Number of generations	100
Number of benchmark runs	10
Benchmark seeds	11, 22, 33, 44, 55, 66, 77, 88, 99, 100
Selective ALNS target fraction	25% of population
ALNS invocation frequency	Every 5 generations
Initial SA temperature, T_0	200
Cooling factor, α	0.98
RL pretraining seeds	101, 202, 303, 404, 505
RL pretraining population size	20
RL pretraining generations	100
Q-learning rate, η	0.20
Discount factor, γ	0.90
Training exploration probability, ϵ	0.08
Training softmax temperature	1.00 to 0.25
Inference exploration probability, ϵ	0.02
Inference softmax temperature	0.90 to 0.35

2.9. Validation and Benchmark Methods

To validate the complete queue-aware bi-objective evaluation on a tractable problem, an **8-subdistrict, two-mill instance** was constructed from the real case-study data. Because the adopted solution representation is priority-based, all $8! = 40,320$ possible priority sequences were exhaustively enumerated. Each sequence was decoded and evaluated using the same capacity-aware allocation and queue-simulation procedures as the large-scale problem. The non-dominated solutions obtained from this enumeration were used as an exhaustively enumerated reference Pareto front within the adopted priority-encoding space.

The small-instance validation compared standalone NSGA-II, NSGA-II + ALNS with random search control, and the proposed RL-guided NSGA-II–ALNS framework over 10 matched runs. Their approximation sets were evaluated against the enumerated reference front to assess how closely each method recovered the known trade-off structure. This experiment complements the large-scale analysis by providing a reference-based validation of the complete distance–queue formulation rather than validating transportation distance alone.

For the full 199-subdistrict, four-mill instance, five methods were compared: Greedy nearest assignment, standalone NSGA-II, NSGA-II + ALNS (Random), the proposed RL-guided NSGA-II–ALNS, and MOEA/D [18]. The greedy method provides a deterministic operational baseline, standalone NSGA-II evaluates the contribution of local search, and Random ALNS isolates the benefit of RL-guided search control. MOEA/D was included as an established decomposition-based multi-objective baseline and its principal neighborhood settings were tuned before the final comparison. The stochastic algorithms were evaluated using matched random seeds and consistent population and generation settings where applicable; runtime was reported separately because the hybrid methods perform additional local-search evaluations. The performance indicators and statistical comparison procedures are described in Section 2.10.

2.10. Performance Indicators and Statistical Analysis

Multi-objective solution quality was evaluated using Hypervolume (HV) and Inverted Generational Distance (IGD) [21], [22]. Because transportation distance and queue waiting time have substantially different numerical scales, both objectives were normalized using common min–max bounds derived from the pooled approximation sets of all stochastic multi-objective methods and runs

$$Z_{ik}^{norm} = \frac{Z_{ik} - Z_k^{min}}{Z_k^{max} - Z_k^{min}} \quad (21)$$

where Z_{ik} is objective k of solution i , and Z_k^{min} and Z_k^{max} are the common minimum and maximum values, respectively. The same normalization bounds were applied to every method to ensure direct comparability.

For the large-scale instance, the empirical reference Pareto front was constructed from the non-dominated solutions in the pooled approximation sets of NSGA-II, Random ALNS, RL-guided NSGA-II–ALNS, and tuned MOEA/D across all 10 runs. HV was calculated in the normalized objective space using the fixed reference point $r = (1.10, 1.10)$, where larger values indicate better convergence and coverage. IGD was computed using the same normalized empirical reference front, with smaller values indicating closer approximation to the reference trade-off surface. The deterministic Greedy nearest solution was retained as an operational baseline but excluded from construction of the reference front and from HV/IGD comparisons because it produces only a single deterministic solution rather than a stochastic Pareto approximation.

Statistical comparisons were conducted using the matched random seeds described in Section 2.8. For transportation distance, average queue waiting time, HV, and IGD, differences among the four stochastic algorithms were first assessed using the Friedman test at a significance level of $\alpha = 0.05$. When the omnibus test indicated significant differences, pairwise comparisons were performed using two-sided Wilcoxon signed-rank tests, with Holm correction applied separately within each performance measure to control the family-wise error rate.

Potential outliers were examined using the conventional $1.5 \times$ interquartile range (IQR) criterion. All runs were retained in the final analysis, including identified outliers, because they represent genuine stochastic variation in algorithm performance rather than invalid observations. Results are therefore reported using both mean $\pm$ standard deviation and distributional comparisons where appropriate.

2.11. Environmental and Economic Evaluation

In addition to the two optimization objectives, fuel consumption, fuel cost, and CO₂ emissions were calculated as derived performance indicators to evaluate the economic and environmental implications of each transportation plan. Two truck payload classes were considered: 20-ton and 40-ton trucks, with representative fuel efficiencies of 5.0 km/L and 3.45 km/L, respectively, based on empirical truck-fuel data under Thai operating conditions [23].

For each generated trip i , fuel consumption was estimated from the traveled distance d_i and the fuel efficiency e_i of the corresponding truck class

$$Fuel = \sum_{i=1}^N \frac{d_i}{e_i} \quad (22)$$

Total fuel cost was then calculated as

$$Fuel\ Cost = Fuel \times P_d \quad (23)$$

where P_d is the diesel price in THB/L. A reference diesel price of 44.4 THB/L [24] was used consistently for comparative economic evaluation and did not influence the optimization process.

Carbon dioxide emissions were estimated from total diesel consumption using a diesel-combustion emission factor of 2.70 kg CO₂/L [25].

$$CO_2 = Fuel \times 2.70 \quad (24)$$

Fuel consumption, fuel cost, and CO₂ emissions were treated as post-optimization performance indicators rather than additional optimization objectives. Under the adopted fixed fuel-efficiency and emission-factor assumptions, these indicators are largely derived from transportation activity and would therefore provide limited independent information as an additional objective. The optimization consequently focuses on the operational trade-off between transportation distance and queue waiting time, while the derived indicators are used to quantify the associated sustainability benefits.

3. Results

3.1. Exhaustive Validation on a Small Instance

The proposed framework was first validated on the 8-subdistrict, two-mill instance described in Section 2.9. Exhaustive enumeration of all 40,320 priority sequences produced 39,729 unique objective vectors, from which five non-dominated solutions formed the reference Pareto front within the adopted priority-encoding space. The reference solutions covered transportation distances from 117,716.09 to 121,077.76 km and average queue waiting times from 0.164 to 0.611 h/trip, confirming a clear trade-off between transportation efficiency and mill congestion.

The three stochastic algorithms were subsequently evaluated over 10 matched runs against this reference front. As shown in Table 3, the proposed RL-guided NSGA-II–ALNS achieved the highest mean normalized HV (1.1845), corresponding to 98.97% of the reference hypervolume, and the lowest mean IGD (0.1210). Standalone NSGA-II achieved 97.99% HV recovery with an IGD of 0.1273, whereas Random ALNS achieved 95.81% HV recovery with an IGD of 0.1571. The RL-guided method also recovered an average of 2.4 of the five reference Pareto points (48%), compared with 2.0 points (40%) for NSGA-II and 1.1 points (22%) for Random ALNS.

The reference hypervolume was 1.1969, and all algorithmic HV values remained below this value, indicating internally consistent reference-based evaluation. Although none of the heuristics recovered the complete five-point reference front in every run, all three closely approximated the known trade-off surface. The RL-guided framework provided the closest overall approximation, particularly in terms of HV recovery and exact reference-point recovery, at the expense of a longer runtime resulting from its additional local-search and policy-evaluation operations. Overall, the results demonstrate that the proposed framework closely approximates the exhaustively enumerated reference

trade-off surface, providing supporting validation before application to the full-scale case.

Table 3. Performance against the exhaustively enumerated reference Pareto front

Method	HV (mean ± SD)	HV recovery (%)	IGD (mean ± SD)	Exact PF recovery (%)	Run time (s)
RL-guided NSGA-II– ALNS	1.1845 ± 0.0054	98.97	0.1210 ± 0.0502	48	33.61 ± 1.47
NSGA-II	1.1728 ± 0.0197	97.99	0.1273 ± 0.0454	40	14.74 ± 0.54
NSGA-II + ALNS (Random)	1.1467 ± 0.0282	95.81	0.1571 ± 0.0384	22	20.41 ± 0.98

3.2. Large-Scale Comparative Performance

The five optimization methods were evaluated on the full-scale instance comprising 199 subdistricts and four sugar mills. Table 4 summarizes the representative-solution performance obtained over 10 matched runs. The proposed RL-guided NSGA-II–ALNS achieved the best mean performance for both primary objectives, with a transportation distance of 144,392.49 ± 1,352.61 km and an average queue waiting time of 1.383 ± 0.136 h/trip.

Table 4. Large-scale comparative performance over 10 matched runs

Method	Distance (km), mean ± SD	Avg. waiting (h/trip), mean ± SD	Runtime (s), mean ± SD
RL-guided NSGA-II– ALNS	144,392.49 ± 1,352.61	1.383 ± 0.136	148.06 ± 1.13
NSGA-II + ALNS (Random)	145,061.80 ± 1,123.67	1.519 ± 0.145	90.93 ± 2.74
NSGA-II	146,588.25 ± 1,616.68	1.531 ± 0.168	64.88 ± 0.37
MOEA/D	149,053.39 ± 3,228.11	5.187 ± 0.380	75.68 ± 1.49
Greedy nearest	165,945.96 ± 0.00	10.354 ± 0.000	0.02 ± 0.01

The Random ALNS variant produced results close to those of the RL-guided method and exhibited slightly lower run-to-run variability in transportation distance. Standalone NSGA-II showed higher mean distance and waiting time, whereas MOEA/D produced substantially higher queue waiting despite remaining competitive in transportation distance. The deterministic Greedy nearest method was the fastest but performed worst for both operational objectives.

The proposed RL-guided method required the longest computational time because it combines NSGA-II evolution with selective local search, queue-aware re-evaluation, and state-dependent search-mode control. Nevertheless, its runtime remained within a few minutes per planning run, while providing the best mean distance–waiting performance among the evaluated methods. Statistical significance of these differences is examined separately in Section 3.4.

3.3. Pareto-Front Quality

The quality of the Pareto approximations was evaluated using the normalized Hypervolume (HV) and Inverted Generational Distance (IGD) indicators described in Section 2.10. The empirical reference front was constructed from the pooled non-dominated solutions generated by the four stochastic multi-objective methods, with the fixed normalized HV reference point $r = (1.10, 1.10)$. Figure 2 illustrates the resulting trade-off between transportation distance and average queue waiting time.

The proposed RL-guided NSGA-II–ALNS achieved the best mean Pareto-front quality, obtaining the highest HV (1.0676 ± 0.0614) and the lowest IGD (0.0782 ± 0.0272). Random ALNS ranked second, followed by standalone NSGA-II, whereas MOEA/D showed substantially poorer convergence to the empirical reference front. The RL-guided method generated an average of approximately **11 non-dominated solutions per run**, providing a useful range of alternatives for balancing transportation efficiency and congestion.

Table 5. Pareto-front quality over 10 matched runs

Method	HV (mean $\pm$ SD)	IGD (mean $\pm$ SD)	Avg. non-dominated solutions
RL-guided NSGA-II–ALNS	1.0676 ± 0.0614	0.0782 ± 0.0272	11.0 ± 4.97
NSGA-II + ALNS (Random)	1.0516 ± 0.0446	0.0865 ± 0.0206	11.4 ± 2.76
NSGA-II	0.9771 ± 0.0543	0.1249 ± 0.0249	12.1 ± 5.51
MOEA/D	0.2304 ± 0.0574	0.7667 ± 0.0670	1.4 ± 0.52

Although Random ALNS exhibited lower run-to-run variability in both HV and IGD, the RL-guided framework achieved better mean values for both indicators. These results indicate that state-dependent search-mode control improved the average convergence and coverage of the Pareto approximation, while the statistical significance of the observed differences is examined in Section 3.4.

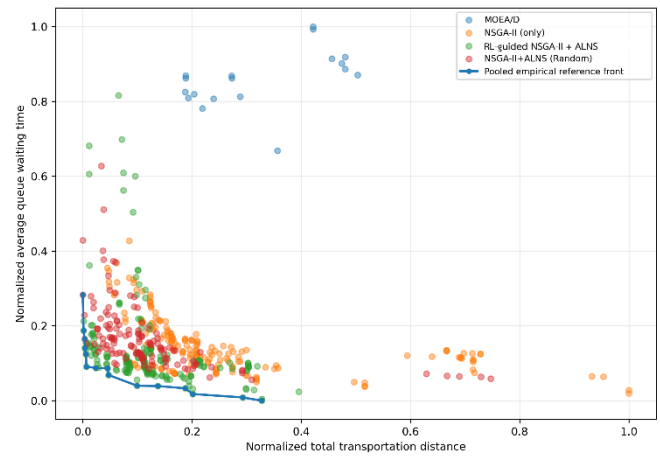


Fig. 2. Normalized Pareto approximations and pooled empirical reference front for the large-scale 199×4 instance.

3.4. Statistical Analysis

The Friedman test indicated significant differences among the four stochastic optimization methods for all evaluated performance measures: transportation distance ($\chi^2 = 20.28$, $p < 0.001$), average queue waiting time ($\chi^2 = 21.36$, $p < 0.001$), HV ($\chi^2 = 25.32$, $p < 0.001$), and IGD ($\chi^2 = 22.44$, $p < 0.001$). Pairwise Wilcoxon signed-rank tests with Holm correction were subsequently used to compare the proposed RL-guided framework with the competing methods.

Table 6. Pairwise statistical comparison of the proposed RL-guided method

Performance measure	Comparator	Holm-adjusted p-value	Significance
Distance	Random ALNS	0.193	No
	NSGA-II	0.023	Yes
	MOEA/D	0.012	Yes
Queue waiting time	Random ALNS	0.129	No
	NSGA-II	0.111	No
	MOEA/D	0.012	Yes
HV	Random ALNS	0.625	No
	NSGA-II	0.012	Yes
	MOEA/D	0.012	Yes
IGD	Random ALNS	0.625	No
	NSGA-II	0.018	Yes
	MOEA/D	0.012	Yes

The proposed RL-guided framework therefore achieved statistically significant improvements over standalone NSGA-II in transportation distance, HV, and IGD, and over tuned MOEA/D in all four performance measures. Although the RL-guided method also achieved better mean values than Random ALNS for distance,

waiting time, HV, and IGD, these differences were not statistically significant after Holm correction. Thus, the results support an advantage of RL-guided search over the standalone evolutionary and decomposition-based baselines, while the evidence does not establish statistical superiority over Random ALNS.

Potential outliers identified using the $1.5 \times \text{IQR}$ criterion were retained because they represented valid stochastic runs. These included one high-distance MOEA/D run, two distance outliers for Random ALNS, and one favorable low-IGD RL-guided run. Retaining these observations avoids selectively removing legitimate run-to-run variability.

3.5. Parameter Sensitivity and Robustness

A post-hoc one-factor-at-a-time sensitivity analysis was conducted to examine the robustness of the proposed RL-guided framework. Five matched seeds were used while varying the simulated-annealing temperature T_0 , cooling factor α , ALNS invocation frequency, initial inference temperature, and exploration probability ϵ . The baseline configuration ($T_0 = 200$, $\alpha = 0.98$, ALNS frequency = 5, initial inference temperature = 0.90, and $\epsilon = 0.02$) was fixed before this analysis and was not selected retrospectively from the sensitivity results.

Table 7. Selected results of the parameter sensitivity analysis

Setting	Distance (km)	Waiting (h/trip)	HV	IGD
Baseline	144,198.18	1.398	1.0641	0.0807
$\alpha=0.95$	144,922.34	1.317	1.0803	0.0713
$\alpha=0.995$	144,938.16	1.365	1.0632	0.0812
ALNS frequency = 3	144,681.77	1.360	1.0487	0.0887
ALNS frequency = 10	145,129.29	1.508	1.0149	0.1092
Initial temperature = 0.60	145,040.26	1.363	1.0643	0.0880
Initial temperature = 1.20	145,350.17	1.357	1.0530	0.0973
$\epsilon=0$	144,759.08	1.381	1.0617	0.0803
$\epsilon=0.05$	146,694.63	1.453	1.0180	0.1174

Changes in T_0 from 100 to 400 produced essentially unchanged solution quality, indicating low sensitivity to the initial simulated-annealing temperature within the examined range. Overall, the maximum observed changes were approximately 1.73% in transportation distance, 7.89% in waiting time, and 4.63% in HV, indicating relatively stable performance for these

measures. IGD was more sensitive, with a maximum change of approximately 45.49%. The most noticeable deterioration occurred at $\epsilon = 0.05$ and with an ALNS interval of 10 generations, suggesting that excessive exploration or less frequent local improvement can reduce Pareto-front quality.

Although some alternative settings improved individual measures—for example, $\alpha = 0.95$ reduced waiting time and improved HV and IGD at the expense of a small increase in transportation distance—no single alternative consistently improved all criteria. The baseline configuration was therefore retained for the final benchmark experiments, supporting the conclusion that the reported performance is not dependent on a narrowly tuned parameter combination.

3.6. Environmental and Economic Impact

The environmental and economic effects of transportation optimization were assessed using fuel consumption, CO₂ emissions, and fuel cost. The non-optimized Before configuration was used as the operational reference. Tables 8 and 9 summarize the corresponding environmental and economic outcomes.

Table 8. Transportation distance and fuel consumption

Method	Distance (km)	Fuel (L)
Before	160,307.58	36,571.38
RL-guided	144,392.49 (−9.93%)	33,261.72 (−9.05%)
Random ALNS	145,061.80 (−9.51%)	33,354.15 (−8.80%)
NSGA-II	146,588.25 (−8.56%)	33,722.36 (−7.79%)
Greedy	165,945.96 (+3.52%)	38,322.21 (+4.79%)

Table 9. Environmental and economic impacts

Method	CO ₂ (kg)	Fuel cost (THB)
Before	98,742.73	1,623,769.30
RL-guided	89,806.65 (−9.05%)	1,476,820.44 (−146,948.87)
Random ALNS	90,056.22 (−8.80%)	1,480,924.43 (−142,844.87)
NSGA-II	91,050.37 (−7.79%)	1,497,272.81 (−126,496.49)
Greedy	103,469.96 (+4.79%)	1,701,506.04 (+77,736.73)

Values in parentheses indicate changes relative to the Before configuration.

Relative to the Before configuration, the proposed RL-guided framework reduced transportation distance by 9.93% and fuel consumption and CO₂ emissions by 9.05%. This corresponds to an estimated fuel-cost saving

of approximately 146,949 THB per planning cycle. Random ALNS and standalone NSGA-II also produced substantial reductions, whereas the Greedy strategy increased all three derived impact indicators relative to the baseline.

Figure 3 compares the spatial transportation patterns before and after optimization using the proposed RL-guided framework. The optimized allocation shows a more consolidated transportation structure with fewer long-distance assignments, consistent with the reductions in total transportation distance, fuel consumption, and CO₂ emissions reported in Tables 8 and 9.

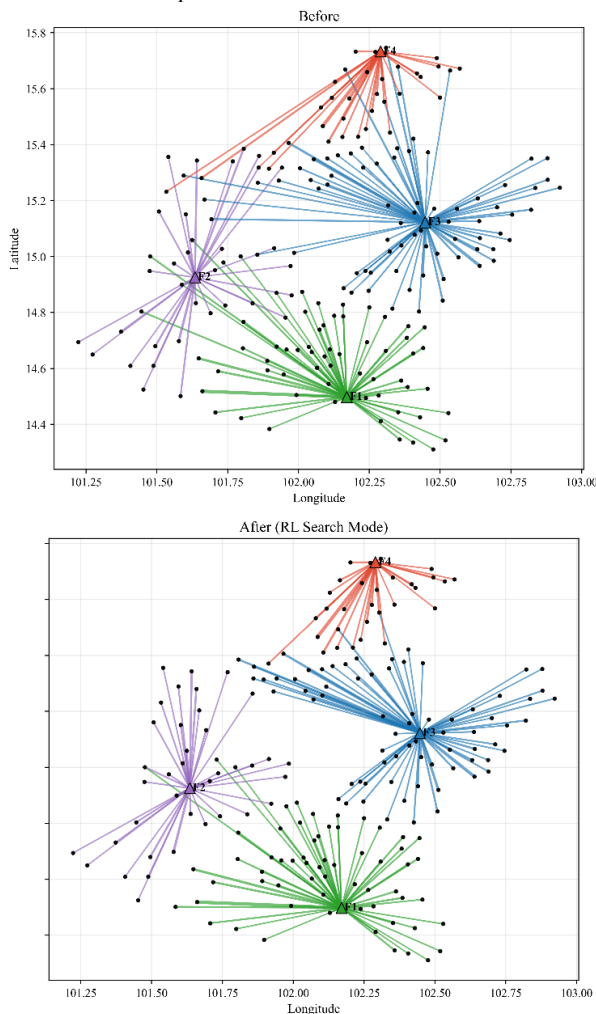


Fig. 3. Sugarcane transportation flows before and after optimization using the proposed RL-guided framework.

4. Discussion

4.1. Interpretation and Comparison with Previous Studies

The results indicate that integrating ALNS with NSGA-II improves solution quality for large-scale sugarcane transportation planning. The RL-guided framework achieved the best mean transportation distance, queue waiting time, HV, and IGD, while statistically significant improvements over standalone NSGA-II were observed for distance, HV, and IGD. These findings are consistent with previous studies

showing that neighborhood-based search can enhance evolutionary optimization in large-scale routing and transportation problems [5], [7].

The contribution of reinforcement learning should, however, be interpreted cautiously. Although the RL-guided method achieved better mean values than Random ALNS across all evaluated measures, the differences were not statistically significant after Holm correction, and Random ALNS showed lower variability for several indicators. Thus, the principal benefit of RL in this study is its state-dependent guidance of search behavior, rather than universal statistical superiority over random neighborhood selection. This interpretation is consistent with recent RL-assisted metaheuristic studies emphasizing adaptive search decisions based on optimization feedback [8]–[12].

Compared with previous sugarcane transportation studies that mainly emphasize transportation distance, scheduling, or harvesting–transport coordination [2]–[5], this study explicitly incorporates average mill queue waiting time as a second objective. The resulting Pareto solutions therefore represent a practical trade-off between transportation efficiency and processing congestion. Together with hierarchical RL-guided search and evaluation on a 199-subdistrict, four-mill system, this queue-aware formulation represents the main extension of the proposed framework.

4.2. Practical and Managerial Implications

The proposed framework provides a decision-support approach for coordinating sugarcane allocation and transportation across multiple mills. Rather than assigning cane solely to the nearest facility, planners can consider both transportation distance and expected queue waiting time. This is particularly relevant when nearby mills approach their processing capacity, because locally attractive assignments may create congestion and increase truck waiting.

The generated Pareto solutions also allow managers to select transportation plans according to operational priorities. During periods of high fuel cost, greater emphasis may be placed on distance-efficient solutions, whereas under severe mill congestion, solutions with lower waiting time may be preferred. A balanced Pareto solution can therefore support coordinated decisions among transportation planners, mill operators, and cane-supply managers without requiring a single fixed weighting between the two objectives.

In addition, the reductions in transportation distance obtained by the optimized plans translate into lower fuel consumption, CO₂ emissions, and fuel expenditure. These derived benefits indicate that improving allocation and congestion management can simultaneously support operational efficiency and sustainability. The framework may therefore serve as a practical planning tool for daily or seasonal sugarcane transportation, provided that local capacity, travel, and queue parameters are updated to reflect actual operating conditions.

4.3. Limitations and Future Research

Several limitations should be acknowledged. First, the queue model assumes predefined travel speeds, unloading capacity, service characteristics, and reproducible dispatch-time perturbations. Although these assumptions enable consistent comparison among algorithms, actual sugarcane transportation may involve stochastic travel times, variable unloading rates, equipment failures, and irregular truck arrivals. Second, the model assumes sufficient contracted trucking capacity and does not explicitly represent a finite fleet, vehicle reuse, or detailed driver scheduling.

The empirical evaluation is also based on a single large-scale case comprising 199 subdistricts and four sugar mills. Although exhaustive validation on a reduced instance and repeated stochastic experiments support the reliability of the computational results, further testing on additional regions, network structures, demand levels, and congestion conditions would strengthen the generalizability of the findings. In addition, the RL policy was pretrained and fixed during the final benchmark experiments; therefore, its transferability to substantially different operating environments remains to be examined.

Future research should incorporate stochastic travel and service times, explicit fleet availability, vehicle scheduling, and dynamic or rolling-horizon replanning. The RL component could also be extended toward online or transfer-learning mechanisms that adapt to changing congestion and capacity conditions. Finally, validation using operational data from additional sugar mills and real-time transportation systems would provide stronger evidence of practical applicability and scalability.

5. Conclusion

This study developed a queue-aware bi-objective sugarcane transportation framework that integrates NSGA-II, selective ALNS, and reinforcement learning for state-dependent search-mode control. The model simultaneously considers transportation distance and average mill queue waiting time under supply, processing-capacity, and heterogeneous truck-payload constraints. Exhaustive validation on a reduced instance showed that the proposed framework closely approximated the known reference Pareto trade-off surface.

For the large-scale 199-subdistrict, four-mill case, the RL-guided framework achieved the best mean performance in transportation distance, queue waiting time, HV, and IGD. Statistical analysis confirmed significant improvements over standalone NSGA-II for distance, HV, and IGD and over tuned MOEA/D for all evaluated measures. However, the differences between RL-guided and Random ALNS were not statistically significant after Holm correction, indicating that the benefit of reinforcement learning should primarily be interpreted as improved state-dependent search guidance

rather than universal superiority over random local search.

Beyond optimization performance, the proposed approach reduced transportation distance by approximately 9.93%, with corresponding reductions of 9.05% in fuel consumption and CO₂ emissions relative to the non-optimized baseline. These findings demonstrate that jointly considering transportation efficiency and mill congestion can support more balanced, sustainable, and operationally meaningful sugarcane logistics planning. Future extensions should incorporate stochastic travel and service conditions, explicit fleet scheduling, and dynamic replanning for broader real-world deployment.

Acknowledgements

This research was supported by the Department of Industrial Engineering, Faculty of Engineering, Ubon Ratchathani University. The authors would also like to thank the Office of the Cane and Sugar Board (OCSB), Thailand, for providing real-world data used in this research.

References

- [1] E. S. Teixeira, S. Rangel, H. O. Florentino, and S. A. de Araujo, "A review of mathematical optimization models applied to the sugarcane supply chain," *International Transactions in Operational Research*, vol. 30, no. 4, pp. 1755–1788, 2023, doi: 10.1111/itor.13056.
- [2] A. Higgins, "Scheduling of road vehicles in sugarcane transport: A case study at an Australian sugar mill," *European Journal of Operational Research*, vol. 170, no. 3, pp. 987–1000, 2006, doi: 10.1016/j.ejor.2004.07.055.
- [3] A. Aliano Filho, T. Melo, and M. V. Pato, "A bi-objective mathematical model for integrated planning of sugarcane harvesting and transport operations," *Computers & Operations Research*, vol. 134, Art. no. 105419, 2021, doi: 10.1016/j.cor.2021.105419.
- [4] S. D. Jena and M. Poggi, "Harvest planning in the Brazilian sugar cane industry via mixed integer programming," *European Journal of Operational Research*, vol. 230, no. 2, pp. 374–384, 2013, doi: 10.1016/j.ejor.2013.04.011.
- [5] R. Pitakaso and K. Sethanan, "Adaptive large neighborhood search for scheduling sugarcane inbound logistics equipment and machinery under a sharing infield resource system," *Computers and Electronics in Agriculture*, vol. 158, pp. 313–325, 2019, doi: 10.1016/j.compag.2019.02.001.
- [6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002, doi: 10.1109/4235.996017.

- [7] S. Ropke and D. Pisinger, "An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows," *Transportation Science*, vol. 40, no. 4, pp. 455–472, 2006, doi: 10.1287/trsc.1050.0135.
- [8] Y. Bengio, A. Lodi, and A. Prouvost, "Machine learning for combinatorial optimization: A methodological tour d'horizon," *European Journal of Operational Research*, vol. 290, no. 2, pp. 405–421, 2021, doi: 10.1016/j.ejor.2020.07.063.
- [9] M. Karimi-Mamaghan, M. Mohammadi, P. Meyer, A. M. Karimi-Mamaghan, and E.-G. Talbi, "Machine learning at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art," *European Journal of Operational Research*, vol. 296, no. 2, pp. 393–422, 2022, doi: 10.1016/j.ejor.2021.04.032.
- [10] R. Reijnen, Y. Zhang, H. C. Lau, and Z. Bukhsh, "Online control of adaptive large neighborhood search using deep reinforcement learning," *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 34, no. 1, pp. 475–483, 2024, doi: 10.1609/icaps.v34i1.31507.
- [11] S.-N. Johnn, V.-A. Darvari, J. Handl, and J. Kalcsics, "A graph reinforcement learning framework for neural adaptive large neighbourhood search," *Computers & Operations Research*, vol. 172, Art. no. 106791, 2024, doi: 10.1016/j.cor.2024.106791.
- [12] R. Pitakaso, K. Sethanan, C. Chamnanlor, C.-F. Chien, S. Gonwirat, K. Worasan, and M. K. Lim, "Optimizing sugarcane bale logistics operations: Leveraging reinforcement learning and artificial multiple intelligence for dynamic multi-fleet management and multi-period scheduling under machine breakdown constraints," *Computers and Electronics in Agriculture*, vol. 236, Art. no. 110431, 2025, doi: 10.1016/j.compag.2025.110431.
- [13] K. Meesuptaweekoon and P. Chaovaitwongse, "Dynamic vehicle routing problem with multiple depots," *Engineering Journal*, vol. 18, no. 4, pp. 135–149, 2014, doi: 10.4186/ej.2014.18.4.135.
- [14] S. Horng and P. Yenradee, "Performance comparison of two-phase LP-based heuristic methods for capacitated vehicle routing problem with three objectives," *Engineering Journal*, vol. 24, no. 5, pp. 145–159, 2020, doi: 10.4186/ej.2020.24.5.145.
- [15] K. Nupet and P. Yenradee, "New sustainable approach for multi-objective production and distribution planning in supply chain," *Engineering Journal*, vol. 27, no. 7, pp. 1–26, 2023, doi: 10.4186/ej.2023.27.7.1.
- [16] N. Sutthibutr, N. Chiadamrong, K. Hiraishi, and S. Thajchayapong, "A unified optimization model with proportional fairness and robustness of fuzzy multi-objective aggregate production planning in supply chain under uncertain environments," *Engineering Journal*, vol. 28, no. 5, pp. 25–52, 2024, doi: 10.4186/ej.2024.28.5.25.
- [17] U. L. T. Tran and N. Chiadamrong, "Maximizing EVA in a cocoa supply chain network design with a hybrid optimization approach: A case study in Vietnam," *Engineering Journal*, vol. 29, no. 7, pp. 61–83, 2025, doi: 10.4186/ej.2025.29.7.61.
- [18] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007, doi: 10.1109/TEVC.2007.892759.
- [19] P. Wittayasin, N. Kriengkarakot, and P. Kriengkarakot, "A reinforcement learning-guided hybrid NSGA-II + ALNS framework for large-scale capacitated transportation optimization in Thai sugarcane logistics," *Journal of Current Science and Technology*, vol. 16, no. 3, Art. no. 189, 2026, doi: 10.59796/jcst.V16N3.2026.189.
- [20] Office of the Cane and Sugar Board, *Sugarcane Production and Industry Data*, Ministry of Industry, Thailand, 2023.
- [21] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms—A comparative case study," in *Parallel Problem Solving from Nature—PPSN V*, A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, Eds. Berlin, Germany: Springer, 1998, pp. 292–301, doi: 10.1007/BFb0056872.
- [22] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000, doi: 10.1162/106365600568202.
- [23] T. Naparswad, "A study of fuel consumption of trucks," M.Eng. thesis, School of Civil Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand, 2013.
- [24] Energy Policy and Planning Office, Ministry of Energy, Thailand, "Retail oil price structure, Apr. 15, 2026," Apr. 15, 2026..
- [25] U.S. Environmental Protection Agency, 2025 *GHG Emission Factors Hub*, Jan. 2025. [Online]. Available: <https://www.epa.gov/system/files/documents/2025-01/ghg-emission-factors-hub-2025.pdf>. Accessed: Sep. 3, 2026.



Putis Wittayasin is a PhD student in the Industrial Engineering Department, Faculty of Engineering, Ubon Ratchathani University, Thailand.



Nuchsa Kriengkorakot is an Associate Professor in the Industrial Engineering Department, Faculty of Engineering, Ubon Ratchathani University, Thailand.



Preecha Kriengkorakot is an Associate Professor in the Industrial Engineering Department, Faculty of Engineering, Ubon Ratchathani University, Thailand.