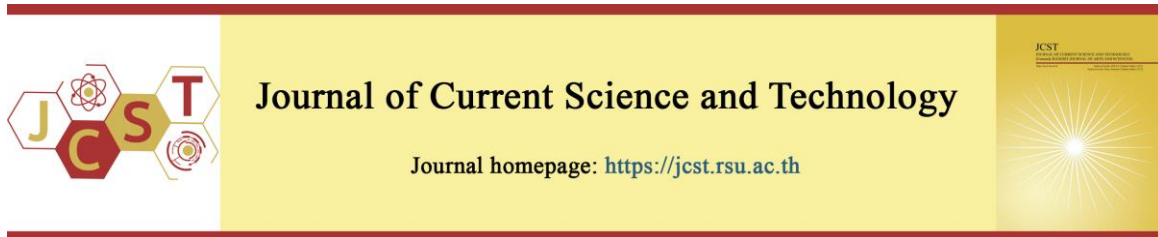


Cite this article: Wittayasin, P., Kriengkorakot, N., & Kriengkorakot, P. (2026). A reinforcement learning-guided hybrid NSGA-II + ALNS framework for large-scale capacitated transportation optimization in Thai sugarcane logistics. *Journal of Current Science and Technology*, 16(3), Article 189. <https://doi.org/10.59796/jcst.V16N3.2026.189>



A Reinforcement Learning-Guided Hybrid NSGA-II + ALNS Framework for Large-Scale Capacitated Transportation Optimization in Thai Sugarcane Logistics

Putis Wittayasin, ORCID: 0009-0008-5118-5494
Nuchara Kriengkorakot, ORCID: 0009-0003-6300-6733
Preecha Kriengkorakot*, ORCID: 0009-0009-1722-7678

Department of Industrial Engineering, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand

*Corresponding author; E-mail: preecha.k@ubu.ac.th

Received 16 November 2025; Revised 13 December 2025; Accepted 14 February 2026; Published 25 June 2026

Abstract

Efficient planning of large-scale agricultural transportation requires balancing travel distance, fleet utilization, and factory capacity constraints. While mixed-integer linear programming (MILP) becomes computationally intractable for large-scale instances and conventional metaheuristics rely on static operator-selection mechanisms, adaptive learning-guided approaches for multi-objective capacitated transportation remain limited. This study proposes a reinforcement learning-guided hybrid NSGA-II + ALNS framework to minimize total transportation distance and truck trips in sugarcane logistics. A real-world case involving 199 subdistricts and four processing plants (796 origin–destination pairs) in northeastern Thailand is examined. Compared with a greedy nearest-assignment baseline, the proposed method reduces total transportation distance from 123,313.52 km to 109,245.22 km (by 11.41%), fuel consumption from 28,131.83 L to 25,108.96 L (by 10.75%), and CO₂ emissions from 75,955.94 kg to 67,794.20 kg (by 10.75%), resulting in an estimated fuel cost saving of approximately 96,550 Thai Baht per cycle. Statistical validation using ANOVA and Tukey's HSD confirms that performance differences are significant at the 95% confidence level. The results demonstrate that reinforcement learning-guided operator adaptation improves convergence stability, Pareto-front quality, and environmental performance in large-scale bi-objective agricultural transportation systems.

Keywords: reinforcement learning, hybrid NSGA-II, adaptive large neighborhood search, ALNS; Sugarcane transportation, agricultural logistics, capacitated transportation problem

1. Introduction

Agricultural logistics plays a central role in supporting large-scale production systems, particularly in countries where agriculture contributes substantially to national economic output. Sugarcane is one of the most significant agricultural commodities worldwide, accounting for 78.9% of global sugar production in 2022 (Krungsri Research, 2023). Thailand supplies 5.6% of global output and exports nearly 70% of its production annually. Beyond sugar, the industry generates high-value by-products, such as molasses for ethanol and bagasse for renewable energy, further

emphasizing the importance of sustainable logistics management.

The northeastern region of Thailand, especially Nakhon Ratchasima, forms a dense production and processing hub in which harvested sugarcane must be transported daily from numerous subdistricts to multiple sugar mills (Office of the Cane and Sugar Board, 2023). This inbound logistics system is operationally complex due to spatially heterogeneous supply origins, variable daily factory capacities, queuing delays, and the need to coordinate a large fleet of trucks over long distances. These challenges are

further aggravated by the rapid post-harvest deterioration of sugarcane quality. Physiological respiration, enzymatic inversion, and microbial activity progressively reduce sucrose content after harvest, leading to deterioration in commercial cane sugar (CCS) and lower factory recovery rates, particularly under tropical conditions (Rein, 2016). Empirical evidence shows that transportation delays and prolonged waiting times significantly diminish sugar recovery and economic returns (Solomon et al., 2006). Thus, reducing travel distance and ensuring timely delivery are essential for both growers and mill operators.

At the strategic planning level, the daily assignment of supplied sugarcane from multiple origins to mills with strict processing capacities can be modeled as a capacitated transportation problem (CTP), a classical multi-source, multi-destination allocation problem that seeks to minimize total distribution cost or distance subject to supply and capacity constraints. While mixed-integer linear programming (MILP) can yield optimal solutions for small instances, it becomes computationally intractable for large-scale systems involving hundreds of supply nodes and multiple factories (Teixeira et al., 2023; Aliano Filho et al., 2020).

Despite recent advances in metaheuristics and machine learning, three critical gaps remain in large-scale agricultural transportation optimization. First, scalability limitations prevent exact optimization approaches from handling real-world instances with hundreds of origins. Second, many traditional metaheuristics rely on static or manually tuned operator-selection probabilities that cannot adapt to problem-specific search dynamics. Third, although agricultural transportation inherently involves conflicting objectives, such as minimizing travel distance and fleet usage, most reinforcement-learning-enhanced heuristics focus primarily on single-objective formulations. These limitations restrict adaptability, robustness, and multi-objective decision support in practical agricultural logistics systems.

Recent studies highlight the benefits of integrating metaheuristics with data-driven learning mechanisms. Algorithms such as genetic algorithms, differential evolution, and adaptive large neighborhood search (ALNS) have demonstrated strong performance in large-scale logistics planning (Jena & Poggi, 2013; Hoff et al., 2010). Reinforcement learning (RL), in particular, provides a mechanism for adaptive operator selection based on real-time search feedback, enhancing convergence stability and robustness

(Bengio et al., 2021; Karimi-Mamaghan et al., 2022; Reijnen et al., 2024; Johnn et al., 2024).

To address the identified gaps, this study develops a reinforcement learning-guided hybrid NSGA-II + ALNS framework for large-scale sugarcane transportation planning. The problem is formulated as a bi-objective capacitated transportation problem that simultaneously minimizes total transportation distance and the number of truck trips under factory capacity constraints.

This study makes four primary contributions: (1) Methodological: the first integration of RL-guided operator selection within an NSGA-II + ALNS framework for bi-objective capacitated transportation problems; (2) Algorithmic: the design of a Q-learning based adaptive destroy repair operator mechanism that improves upon static (Random) and supervised (RF, XGBoost) selection strategies; (3) Empirical: comprehensive benchmarking on a real-world Thai sugarcane network involving 199 subdistricts and four processing plants, with MILP validation on small- and medium-scale instances; and (4) Practical: demonstration of significant transportation-distance reduction and fuel-cost savings, providing actionable decision support for sustainable agricultural logistics planning.

Comprehensive computational experiments, statistical validation, and disruption scenarios are conducted to evaluate solution quality, robustness, and practical applicability in large-scale agricultural logistics systems.

2. Objectives

The objectives of this study are as follows:

- To formulate a realistic large-scale sugarcane transportation problem as a bi-objective capacitated transportation problem (CTP), aiming to simultaneously minimize total transportation distance and the number of truck trips under factory capacity constraints.
- To develop a hybrid multi-objective optimization framework that integrates NSGA-II with reinforcement learning-guided adaptive large neighborhood search (ALNS), where reinforcement learning dynamically selects improvement operators based on search performance.
- To evaluate the effectiveness of the proposed hybrid framework by benchmarking it against conventional non-learning heuristics and alternative optimization methods using identical datasets and experimental settings.
- To assess the robustness of the proposed approach under factory disruption scenarios, examining its

ability to reallocate transportation flows while maintaining feasibility under varying supply and capacity conditions.

3. Materials and Methods

3.1 Problem Formulation (Bi-objective Capacitated Transportation Problem)

The daily sugarcane–mill assignment in this study is modeled as a single-period capacitated transportation problem (CTP). A set of subdistricts S represents supply nodes and a set of sugar mills, F represents demand (processing) nodes. For each subdistrict, $s \in S$, a known amount of harvested cane, S_s (tons/day) is available for delivery within the planning horizon. Each mill $f \in F$ has a finite processing capacity, C_f (tons/day). A homogeneous fleet of trucks with payload capacity Q (tons/trip) is used to transport cane from subdistricts to mills. Let d_{sf} denote the road distance (km) between subdistrict s and mill f .

Two sets of decision variables are defined as follows:

$x_{sf} \geq 0$: tons of sugarcane shipped from subdistrict s to mill f ,
 $t_{sf} \in \mathbb{Z}_+$: number of truck trips assigned from s to f .

The bi-objective capacitated transportation problem considered in this study aims to simultaneously optimize transportation efficiency and fleet utilization. The first objective, denoted as Z_1 , minimizes the total transportation distance (km), calculated as the sum of the travel distance on each transportation lane multiplied by the corresponding number of truck trips. The second objective, denoted as Z_2 , minimizes the total number of truck trips, which serves as a proxy for fleet utilization, driver workload, and fixed operating costs. The resulting bi-objective optimization problem is formulated as follows:

$$\text{Minimize } \begin{cases} Z_1 = \sum_{s \in S} \sum_{f \in F} d_{sf} t_{sf}, & (1) \\ Z_2 = \sum_{s \in S} \sum_{f \in F} t_{sf}. & (2) \end{cases}$$

these objectives are subject to the following constraints:

Supply (harvest) conservation: all harvested cane at each subdistrict must be assigned to mills within the planning horizon:

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

mill capacity constraint: the total inflow to each mill cannot exceed its daily processing capacity:

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

truck-capacity linking constraint: the assigned flow on each lane must be compatible with the number of truck trips and capacity Q :

$$x_{sf} \leq Q t_{sf}, \quad \forall s \in S, \forall f \in F. \quad (5)$$

Although Constraint (5) is expressed using an aggregated payload parameter Q for modeling clarity, heterogeneous truck capacities (20 and 40 tons per trip) are explicitly considered during the decoding stage (Algorithm 1a). In the evaluation phase, actual truck assignments are determined based on subdistrict demand volume and fleet configuration, ensuring that the operational implementation reflects the mixed-fleet setting.

Non-negativity and integrality: the quantity of sugarcane shipped on each lane must be non-negative and continuous, whereas the number of truck trips assigned to each lane must take non-negative integer values:

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

The truck payload capacities considered in this study are $Q_{\text{small}} = 20$ tons per trip and $Q_{\text{large}} = 40$ tons per trip. For modeling simplicity, the mathematical formulation employs an aggregated representation Q , while heterogeneous truck assignment is implemented during the decoding stage (Algorithm 1a), where actual truck types are allocated according to subdistrict demand and fleet configuration.

This study assumes that the total harvested supply does not exceed the aggregate processing capacity, i.e., $\sum_{s \in S} S_s \leq \sum_{f \in F} C_f$, which holds for the empirical dataset used in this study; therefore, all harvested cane can be allocated within the day without backlogging.

The pair (Z_1, Z_2) defines a bi-objective capacitated transportation problem. In this sugarcane transportation setting, the two objectives exhibit a generally positive correlation: reducing total transportation distance (Z_1) often leads to fewer truck trips (Z_2) through more efficient allocation of flows to nearby mills. However, trade-offs may arise when nearby mills approach their capacity limits, requiring partial loads or reassignment to more distant facilities.

Importantly, Z_2 is subject to a structural lower bound determined by the ratio between total harvested volume and truck payload capacity. Given fixed daily demand and fleet configuration, the minimum feasible

number of trips is largely constrained by this theoretical bound. Consequently, most competitive solutions converge to a similar minimum trip level. Nevertheless, retaining Z_2 as an explicit objective is essential to prevent artificial load fragmentation, in which distance could be reduced by increasing the number of trips through smaller shipment allocations. Including Z_2 therefore imposes optimization pressure that ensures distance reductions are achieved without inflating fleet usage, therefore preserving operational feasibility.

This interaction creates a sparse but meaningful Pareto structure, which is analyzed in Section 4.4. A feasible solution is said to be non-dominated (Pareto-optimal) if no other feasible solution can reduce one objective without worsening the other (Deb et al., 2002). Because exact multi-objective MILP approaches become computationally expensive for the large-scale instances considered, this work approximates the Pareto set using the hybrid metaheuristic framework described in Section 3.3, while the CTP formulation above provides a transparent, operations-oriented representation of the sugarcane–mill assignment problem. Similar bi-objective transportation formulations have been applied in logistics and routing planning to balance multiple cost-related criteria under capacity constraints (Sun & Lang, 2015).

3.1.1 Derived Performance Indicators (Fuel Cost and Emissions)

Although the primary optimization objectives focus on minimizing total transportation distance and the number of truck trips, additional performance indicators are computed to support economic and environmental interpretation. Total fuel cost is estimated using a standard distance-based formulation:

$$\text{Fuel cost (THB)} = \text{Distance (km)} \times r_f \times \text{diesel price (THB/L)} \quad (7)$$

where r_f denotes a representative fuel consumption coefficient (L/km) derived from the operating characteristics of the truck fleet considered in this study. Two truck classes are modeled, corresponding to medium-duty trucks (approximately 20 tons) and heavy-duty trucks (approximately 40 tons). Based on empirical field measurements of heavy-duty trucks operating under Thai road conditions (Naparswad, 2013), the fuel efficiencies are set to 5.0 km/L for 20-ton trucks and 3.45 km/L for 40-ton trucks. The appropriate fuel consumption rate is applied according to the truck type used on each transportation lane, and

total fuel consumption is calculated using the optimized travel distances.

The diesel price used for cost estimation is 31.94 THB/L, based on the official nationwide retail diesel pump price in Thailand on September 24, 2025, as announced by the Energy Policy and Planning Office (EPPO) (2025), Ministry of Energy. This standardized price is used for comparative economic evaluation and does not influence the optimization process.

Carbon dioxide (CO_2) emissions are estimated using fuel-based emission factors. According to the U.S. Environmental Protection Agency (EPA) GHG Emission Factors Hub (2025), diesel fuel combustion emits 2.70 kg CO_2 per liter. These derived indicators do not directly influence the optimization process but provide additional insights into the economic and environmental implications of the optimized transportation plans.

3.2 Datasets and Study Instances

The datasets used in this study were derived from real-world sugarcane transportation operations in the northeastern region of Thailand, representing a capacitated multi-source, multi-destination allocation system between sugarcane-producing subdistricts and sugar factories. Three benchmark instances were developed to reflect different levels of problem scale and computational complexity: a small instance (10 subdistricts and 2 factories), a medium instance (20 subdistricts and 3 factories), and a large real-world instance (199 subdistricts and 4 factories). Each dataset preserves realistic spatial distribution, heterogeneous demand patterns, and factory capacity limitations to ensure operational fidelity.

The distance matrix was constructed using the geographic coordinates of subdistricts and factories obtained from QGIS and exported in CSV format. Pairwise distances (in kilometers) between each subdistrict and each factory were computed using Python-based spatial calculations. These static distances represent consistent spatial relationships and are independent of real-time traffic conditions. The demand dataset contains the daily harvested sugarcane volume (tons/day) for each subdistrict, whereas the factory dataset specifies the daily processing capacity (tons/day) of each mill (Table 1). All datasets were preprocessed and validated using Microsoft Excel and Python (Pandas and NumPy) to ensure consistency across problem scales.

Table 1 Summary of dataset characteristics

Instance	Subdistricts	Factories	Total Demand (tons/day)	Distance Range (km)	Factory Capacity (tons/day per factory)
Small	10	2	9,000 (sum across 10 subdistricts)	20 – 140	12,000 – 32,000
Medium	20	3	15,000 (sum across 20 subdistricts)	20 – 140	12,000 – 50,000
Large	199	4	98,687 (sum across 199 subdistricts)	10 – 150	12,000 – 50,000

To support both validation and scalability analysis, the small and medium instances were solved using a MILP model in LINGO, providing exact or near-exact baseline allocations. The large instance, representing an entire province, served as the main benchmark for evaluating the proposed metaheuristic and RL-guided hybrid methods. This hierarchical design enables a consistent assessment of algorithmic performance across increasing levels of complexity.

The dataset structure follows established practices in agricultural supply-chain modeling, in which spatially distributed farms must be allocated to multiple processing centers under capacity and distance constraints (Higgins, 2006). Similar dataset configurations have been widely used in sugarcane-transportation studies, particularly those focusing on transportation efficiency, multi-factory allocation, and balanced resource utilization (Chetthamrongchai et al., 2001).

The integration of real geospatial distances and empirical demand–capacity records ensures that the benchmark instances realistically reflect the operational complexity of sugarcane transportation in Thailand. This high-fidelity dataset foundation allows for a rigorous evaluation of the hybrid optimization frameworks introduced in subsequent sections.

3.2.1 Demand–Capacity Balance and Feasibility Margin

For the large-scale instance, the total daily demand across 199 subdistricts is 98,687 tons, while the aggregate processing capacity of the four factories is 118,000 tons per day. This corresponds to a demand–capacity ratio of approximately 83.6%, leaving system slack of 19,313 tons (16.4%). Although overall feasibility is guaranteed, the relatively high utilization level creates localized capacity pressure, leading to meaningful trade-offs between transportation distance and factory allocation decisions within the bi-objective framework.

3.3 Algorithmic Framework

The proposed optimization framework integrates non-dominated sorting genetic algorithm II (NSGA-

II) with ALNS, enhanced by an RL mechanism for adaptive operator selection.

This hybridization is designed to exploit NSGA-II’s strong global exploration and ALNS’s efficient local improvement, while the RL agent dynamically learns which neighborhood operators yield the best progress toward Pareto-optimal solutions.

3.3.1 NSGA-II Backbone

NSGA-II serves as the primary multi-objective optimization engine in the proposed framework. Each individual in the population represents a complete allocation vector, mapping every subdistrict to a feasible factory while satisfying demand and capacity constraints. The encoding contains no routing sequence; instead, it models a capacitated assignment structure consistent with the transportation formulation described in Section 3.1.

The algorithm employs tournament selection, simulated binary crossover (SBX), and polynomial mutation to generate new candidate allocations across successive generations. Non-dominated sorting ranks individuals according to Pareto dominance, while the crowding-distance metric preserves solution diversity along the Pareto front (Deb et al., 2002). The population size and termination settings were selected to balance solution quality and computational cost for the large-scale instance (199 subdistricts $\times$ 4 factories). In a preliminary calibration (seed = 101), increasing the population from $N = 30$ to $N = 50$ yielded only marginal reductions in final distance ($\approx 0.26 - 1.04\%$ across methods), while increasing runtime by approximately 70 – 74%. Therefore, this study set $N = 30$ and $G = 100$ as a cost-effective configuration enabling fair benchmarking over 20 independent runs and five methods within practical computation time.

At predefined intervals, NSGA-II candidate solutions are refined using the ALNS procedure (Section 3.3.2), which operates on the decoded priority-induced ordering of subdistricts, $ids = argsort(x)$, rather than directly modifying a factory-assignment vector. Specifically, ALNS applies simple neighborhood moves (swap, shift, and reverse) to ids to perturb the assignment sequence, and the modified

sequence is then re-decoded using the capacity-feasible nearest-first rule to reconstruct a neighboring feasible allocation under the CTP constraints. This hybrid interaction enables NSGA-II to perform broad global exploration while ALNS provides focused local improvement by efficiently exploring alternative feasible allocation sequences while maintaining feasibility through the decoder.

3.3.2 Adaptive Large Neighborhood Search (ALNS)

The ALNS framework is incorporated into the NSGA-II backbone as a local-improvement mechanism that perturbs and reconstructs subdistrict-factory allocation decisions rather than vehicle routes. This design ensures full consistency with the capacitated transportation structure of the problem, in which the decision variable is the assignment of each supply node to a feasible processing center subject to capacity constraints.

At each ALNS invocation, a destroy operator removes a selected subset of subdistricts from the current allocation vector, creating partial solutions with unassigned supply nodes. A repair operator then reinserts these subdistricts by selecting feasible factories based on insertion cost, distance, capacity slack, or other allocation heuristics. This destroy-repair paradigm is widely used in large-scale assignment and distribution problems in which neighborhoods are defined over allocation decisions rather than route sequences (Ropke & Pisinger, 2006).

The operator portfolio implemented in this study includes random removal, demand-weighted removal, and worst-cost removal, all adapted to the CTP structure. These operators do not involve route construction or sequencing; instead, they perturb the allocation vector to explore alternative distribution patterns. A simulated annealing (SA) acceptance mechanism is employed to control search diversification.

At iteration t , a newly generated solution is accepted either if it improves at least one objective or with a probability, $\text{Pr}(\text{accept})$, defined as:

$$\text{Pr}(\text{accept}) = \exp\left(-\frac{\Delta s}{T_t}\right) \quad (8)$$

where $\text{Pr}(\text{accept})$ denotes the probability of accepting a non-improving solution, Δs represents the normalized scalar deterioration of the objective value, and T_t is the temperature at iteration t , updated according to a geometric cooling schedule, $T_t = \alpha^t T_0$. Such acceptance mechanisms are widely used to enhance search diversification and prevent premature

convergence in assignment-based metaheuristics (Arabani & Farahani, 2012).

To avoid manual operator tuning, ALNS maintains adaptive operator weights based on their historical contributions to objective improvement. After each iteration, operators that successfully improve feasibility or objective values receive score rewards, while ineffective operators are gradually down-weighted. This adaptive mechanism has proven effective in heterogeneous logistics and assignment applications (Pisinger & Ropke, 2018). The updated weights influence operator selection probabilities until the next reinforcement-learning update (see Section 3.3.3).

Overall, the ALNS module provides flexible, large-scale neighborhood exploration tailored specifically to the capacitated transportation structure of sugarcane allocation, allowing the hybrid algorithm to efficiently refine NSGA-II candidate solutions under complex demand-capacity constraints.

3.3.3 RL-Guided Operator Selection (Proposed)

To enhance adaptability beyond heuristic weighting, an RL-based decision module is incorporated to guide operator selection within the ALNS framework. At each iteration, the algorithm defines a state s_t , an action a_t corresponding to the choice of destroy-repair operators, and a reward r_t derived from objective improvement or Pareto-dominance gain.

The agent employs an ϵ -greedy policy and updates Q-values according to:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \eta [r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (9)$$

where η is the learning rate and γ is the discount factor, controlling the relative importance of future rewards. In this study, γ is set to 0.95, a commonly used value in reinforcement learning that balances short-term performance improvement with longer-term learning stability. This setting encourages the selection of operators that yield consistent benefits across successive iterations rather than isolated immediate gains.

3.3.4 Algorithm Workflow

The proposed workflow integrates three components into a single procedure: (i) data-driven instance construction and objective evaluation, (ii) a population-based global search using NSGA-II with periodic ALNS local improvement, and (iii) an RL policy that adaptively selects destroy-repair operators in ALNS. The unified pseudocode below summarizes the complete pipeline.

Algorithm notation. Z_1 = total distance (km),
 Z_2 = total truck trips; Scalarize ($\cdot$) = normalized
scalar combination used only for SA acceptance; “PF”
= non-dominated set (Pareto front); operators $\in$

{swap, shift, reverse, ...}. RL state $s = (\text{bin}_g, \text{bin}_T, \text{last_op}, \text{bin}_{\Delta s})$ reward uses Pareto gain / $-\Delta\text{scalar}$ (Pareto-aware).

Algorithm 1(a). Instance Construction and Evaluatio

```

Algorithm 1(a). Instance Construction and Objective Evaluation for Bi-objective Capacitated
Transportation Problem
Input : Distance matrix D, demand DEMANDs, factory capacity FCAPs,
        truck types {small=20t, large=40t}
Output: Evaluate(x)  $\rightarrow$  ( $Z_1, Z_2$ )

A1 Build min_fac_lookup[s]  $\leftarrow$  factories sorted by D[s,  $\cdot$ ] for each subdistrict s
A2 PROCEDURE DecodeCluster(ids):
A3   Initialize fact_infos[f] with CAP=0, LOADS={}, IDS=[]
A4   for each index in ids (in order):
A5     s  $\leftarrow$  subdistrict_of(index); remaining  $\leftarrow$  DEMANDs[s]
A6     for f in min_fac_lookup[s]:
A7       free  $\leftarrow$  FCAPs[f] - fact_infos[f].CAP
A8       alloc  $\leftarrow$  min(remaining, max(free, 0))
A9       if alloc > 0: update CAP/LOADS/IDS; remaining  $\leftarrow$  remaining - alloc
A10      if remaining  $\leq$  0: break
A11   return fact_infos

A12 PROCEDURE AssignTrucks(fact_infos):
A13   for each (f, LOADS[s]):
A14     d  $\leftarrow$  LOADS[s]
A15     while d > 0:
A16       truck  $\leftarrow$  truck_type(s)  $\in$  {small, large}; cap  $\leftarrow$  capacity(truck)
A17       load  $\leftarrow$  min(d, cap)
A18       append trip(sub=s, factory=f, load, truck); d  $\leftarrow$  d - load

A19 PROCEDURE Evaluate(x):
A20   ids  $\leftarrow$  argsort(x); info  $\leftarrow$  DecodeCluster(ids); AssignTrucks(info)
A21    $Z_1 \leftarrow \sum_{\text{trip}} D[\text{trip.subdistrict}, \text{trip.factory}]$ 
A22    $Z_2 \leftarrow \text{number\_of\_trips}(\text{info})$ 
A23   return ( $Z_1, Z_2$ ), info, ids

```

Algorithm 1(b). RL-guided NSGA-II + ALNS Optimization

```

Algorithm 1(b). RL-guided NSGA-II + ALNS for Bi-objective Optimization
Input : Evaluate( $\cdot$ ), population size N, generations G,
        ALNS interval f, SA parameters ( $T_0, \alpha$ ),
        RL parameters ( $\epsilon, \tau, \eta, \gamma$ )
Output: Approximate Pareto set A and final population P

B1 P  $\leftarrow$  GreedyNearestInitialization(N) with small jitters
B2 A  $\leftarrow \emptyset$ ; T  $\leftarrow T_0$ 
B3 for g = 1 ... G do
B4   Evaluate all individuals in P
B5   A  $\leftarrow$  UpdateParetoArchive(A, P) // non-dominated sorting + crowding distance
B6   P  $\leftarrow$  NSGAII_Operators(P) // selection, SBX crossover, mutation
B7   if (g mod f == 0) then // periodic ALNS improvement
B8     for each solution s  $\in$  P do
B9       (F1, info, ids)  $\leftarrow$  Evaluate(s.X); s1  $\leftarrow$  Scalarize(F1)
B10      a  $\leftarrow$  SelectOperator(policy=RL/ML/Random, state={g,T,last_op,last_Δs})
B11      x'  $\leftarrow$  ApplyOperator(ids, a)  $\rightarrow$  decision vector
B12      (F2, _, _)  $\leftarrow$  Evaluate(x'); s2  $\leftarrow$  Scalarize(F2); Δs  $\leftarrow$  s2 - s1
B13      if (F2 < F1) OR (Δs < 0) OR (rand < exp(-Δs / max(T, ε))):
B14        s.X  $\leftarrow$  x'; s.F  $\leftarrow$  F2; last_op  $\leftarrow$  a; last_Δs  $\leftarrow$  Δs
B15      else:
B16        last_Δs  $\leftarrow$  0
B17      end for
B18      T  $\leftarrow \alpha \cdot T$ 
B19    end if
B20  end for
B21  return A, P

```

Algorithm 1(c). RL Operator Selection Policy

```

Algorithm 1(c). Reinforcement Learning Policy for Operator Selection
R1 Maintain  $Q(s, a)$  over discretized bins of  $g$ ,  $T$ ,  $last\_op$ , and  $\Delta s$ 
R2 FUNCTION SelectOperator_RL(state  $s$ ):
R3   if  $Q(s, \cdot)$  undefined or  $rand < \epsilon$ : choose  $a$  uniformly at random
R4   else: pick top-2 actions by  $Q(s, \cdot)$ ; sample  $a$  via softmax with temperature  $\tau$ 
R5   return  $a$ 
R6 After acceptance decision, observe new state  $s'$  and reward  $r$  (Pareto-gain or  $-\Delta scalar$ )
R7 Update rule:  $Q(s, a) \leftarrow Q(s, a) + \eta [r + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a)]$ 
R8 Decay  $\tau$  and  $\epsilon$  gradually per schedule; update bins for next iteration

```

Remarks.

- (i) *Pareto-first acceptance* (Step B13) ensures that non-dominated solutions are always preserved, while simulated annealing allows controlled uphill exploration for diversification.
- (ii) The scalarization score (Steps B9–B12) is used only for stochastic acceptance, not for ranking.
- (iii) The RL policy (Algorithm 1c) adopts tabular Q-learning with discretized state bins for lightweight adaptability.
- (iv) Additional ALNS operators (e.g., insertion, 2-opt) can be seamlessly added to the same interface without retraining the policy.

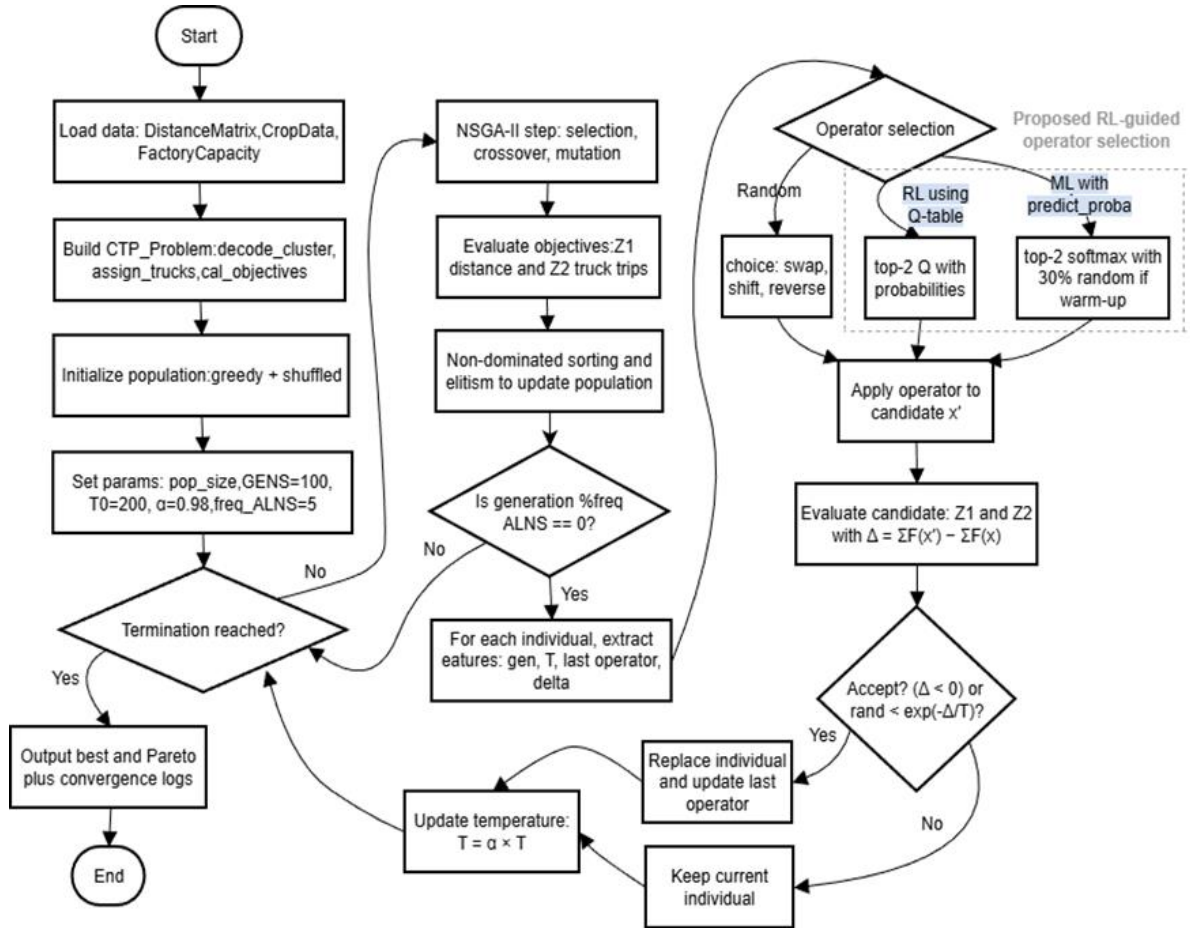


Figure 1 illustrates the overall workflow of the proposed Hybrid NSGA-II + RL-guided ALNS framework

The algorithm integrates NSGA-II for global multi-objective exploration with an ALNS-style neighborhood search for local adaptive refinement. Reinforcement learning is embedded as a hyper-heuristic mechanism to dynamically select neighborhood perturbation operators (swap, shift, and reverse) based on observed search performance. The acceptance of candidate solutions follows a simulated annealing criterion to balance exploration and exploitation. This design enables context-aware operator adaptation while preserving the Pareto-based search structure of NSGA-II, as illustrated in Figure 1.

3.4 Experimental Design and Parameterization

To ensure consistent and reproducible comparisons, all algorithms were executed under identical computational settings. Five optimization variants were evaluated — (1) NSGA-II (only), (2) NSGA-II + ALNS (Random), (3) Hybrid NSGA-II + RL-ALNS, (4) Hybrid NSGA-II + RF-ALNS, and (5) Hybrid NSGA-II + XGBoost-ALNS.

All methods were initialized from the same greedy-based population and identical random seeds to eliminate initialization bias. Experiments were performed on a workstation equipped with an AMD Ryzen 5 4500U CPU and 8 GB of RAM, implemented in Python 3.12.4 using the Pymoo 0.6.0 framework. Each method was executed for 20 independent runs and results were reported as mean $\pm$ standard deviation.

3.4.1 Parameter Configuration

The parameter settings of NSGA-II, ALNS, and the RL module were determined through preliminary calibration experiments, following

standard practices in the evolutionary and large neighborhood search literature (Deb et al., 2002; Ropke & Pisinger, 2006). The complete parameter configuration is summarized in Table 2.

In the ALNS component, the initial temperature T_0 is set to 200. This parameter is dimensionless and controls the probability of accepting non-improving solutions in the simulated annealing mechanism. The value was selected based on pilot experiments to ensure sufficient exploration during the early search phase of the large-scale transportation problem. Since the scalar deterioration measure is normalized and the temperature decreases geometrically with $\alpha = 0.98$, this setting provides a balanced transition from exploration to exploitation.

3.4.2 Evaluation Metrics

The two optimization objectives total transportation distance (Z_1) and total number of truck trips (Z_2) were jointly minimized in accordance with the MILP formulation in Section 3.1. Performance was assessed based on convergence behavior, the quality and spread of obtained Pareto fronts and computational time per run. Comparative statistical significance among the five algorithms was subsequently tested using one-way ANOVA followed by Tukey's HSD post-hoc analysis at a significance level of $\alpha = 0.05$ (details in Section 3.7).

3.5 Multi-objective Evaluation Indicators

To comprehensively assess the quality of the obtained Pareto-optimal solutions, two widely used multi-objective indicators were employed: hypervolume (HV) and inverted generational distance (IGD).

Table 2 Algorithm parameter summary

Component	Parameter	Symbol / Description	Value / Range
NSGA-II	Population size	N	30
	Number of generations	G	100
	Crossover probability	P_c	0.9
	Mutation probability	P_m	0.1
	Distribution index (crossover / mutation)	η_c, η_m	20 / 20
ALNS	Initial temperature	T_0	200
	Cooling factor	α	0.98
	Invocation frequency	f	every 5 generations
RL module	Learning rate	η	0.20
	Discount factor	γ	0.95
	Exploration rate	ϵ	0.05 (fixed)
	Softmax temperature	τ	0.9 $\rightarrow$ 0.4 (decay)
	Reward design		+1 for Pareto gain; $-\Delta s$ (scalarized deterioration)

The HV metric quantifies the volume of the objective space dominated by the approximated Pareto front and bounded by a predefined reference point, r (Zitzler & Thiele, 1998). A larger HV indicates superior performance, reflecting both convergence toward the optimal trade-off surface and solution diversity. In this study, the reference point was defined as

$$\mathbf{r} = (Z_1^{\max}, Z_2^{\max}) \quad (10)$$

where $Z_1^{\max}$ and $Z_2^{\max}$ represent the maximum observed values of distance and truck trips across all algorithms and runs. This conservative setting follows common recommendations for HV analysis in multi-objective optimization.

IGD evaluates the average distance from each point on a reference Pareto set P^* to its nearest point on the obtained front, P (Zitzler et al., 2000). It is computed as:

$$\text{IGD}(P^*, P) = \frac{1}{|P^*|} \sum_{x \in P^*} \min_{y \in P} \|x - y\| \quad (11)$$

since the true Pareto front is unknown for the large-scale instance, P^* was constructed as the union of the best non-dominated solutions obtained across all algorithms and runs, following standard practice in empirical evolutionary multi-objective studies.

Lower IGD values indicate better convergence accuracy and distribution uniformity. Before computing HV and IGD, all objectives were normalized using min-max scaling to avoid distortion due to differences in magnitude. Indicator values were averaged over five independent runs for each algorithm to ensure statistical robustness and comparability.

3.6 Statistical Analysis

To verify the statistical significance of performance differences among the compared optimization methods, a one-way analysis of variance (ANOVA) was conducted on the final objective values (i.e., total distance and total truck trips). The ANOVA test evaluated whether the mean performance across the five algorithms differed significantly under identical experimental conditions. When the null hypothesis of equal means was rejected, Tukey's honestly significant difference (HSD) post-hoc

test was applied at a 95% confidence level ($\alpha = 0.05$) to identify specific pairwise differences. The reported statistics include the F ratio, degrees of freedom (df_1 , df_2), and corresponding p -values, highlighting which hybrid configurations achieved statistically significant improvements. This rigorous analysis ensured that the observed performance differences were not due to random sampling variations but reflected genuine algorithmic improvements.

3.7 Reproducibility

For reproducibility, all random seeds and configuration parameters (population size, number of generations, ALNS frequency, cooling rate, and learning hyperparameters) were fixed and documented. The input datasets DistanceMatrix.xlsx, CropData.xlsx, and FactoryCapacity.xlsx along with result logs (Pareto fronts, convergence curves, and summary tables) are available for replication. The implementation conforms exactly to the pseudocode presented in Algorithm 1, ensuring transparent evaluation and consistent benchmarking across all tested methods.

4. Results

4.1 Validation with Exact MILP (LINGO)

Two small-scale benchmark instances 10 subdistricts with 2 factories (10×2) and 20 subdistricts with 3 factories (20×3) (Table 4) were solved using an exact MILP model implemented in LINGO (branch-and-bound). These exact solutions served as ground-truth references for validating the performance of the metaheuristic algorithms. The comparison focused on two objectives: total transportation distance and total number of truck trips.

Across both datasets, all metaheuristic variants including NSGA-II alone and the four hybrid ALNS-based extensions (random, RL-guided, RF-guided, and XGBoost-guided) achieved objective values within approximately 1.5% of the exact MILP optimum. This small optimality gap confirms the correctness of the problem encoding and the reliability of the hybrid design prior to scaling up to the large 199×4 instance.

Table 3 Validation results for small-scale instance (10×2)

Method	Distance (km) $\pm$ SD	Trips $\pm$ SD	Δ distance (km)	Gap (%) relative to LINGO	Runtime (s) $\pm$ SD
LINGO (MILP)	10,589.80 (–)	292.4 (–)	0.00	0.00	0.10
NSGA-II (Only)	10,753.54 ($\pm$ 0.00)	297.0 ($\pm$ 0.0)	163.74	1.546	7.35 $\pm$ 0.03
NSGA-II + ALNS (Random)	10,753.54 ($\pm$ 0.00)	297.0 ($\pm$ 0.0)	163.74	1.546	10.22 $\pm$ 0.04
NSGA-II + ALNS (RL)	10,753.54 ($\pm$ 0.00)	297.0 ($\pm$ 0.0)	163.74	1.546	10.17 $\pm$ 0.02
NSGA-II + ALNS (RF)	10,753.54 ($\pm$ 0.00)	297.0 ($\pm$ 0.0)	163.74	1.546	10.34 $\pm$ 0.01
NSGA-II + ALNS (XGB)	10,753.54 ($\pm$ 0.00)	297.0 ($\pm$ 0.0)	163.74	1.546	10.90 $\pm$ 0.10

Table 4 Validation results for medium-scale instance (20×3)

Method	Distance (km) $\pm$ SD	Trips $\pm$ SD	Δ distance (km)	Gap (%) relative to LINGO	Runtime (s) $\pm$ SD
LINGO (MILP)	24,843.50 (–)	629.8 (–)	0.00	0.00	0.03
NSGA-II (only)	25,243.40 $\pm$ 80.32	640.0 $\pm$ 0.0	399.9	1.610 $\pm$ 0.32	15.51 $\pm$ 0.11
NSGA-II + ALNS (random)	25,203.88 $\pm$ 14.49	639.0 $\pm$ 0.0	360.38	1.451 $\pm$ 0.06	21.57 $\pm$ 0.03
NSGA-II + ALNS (RL)	25,203.88 $\pm$ 14.49	639.0 $\pm$ 0.0	360.38	1.451 $\pm$ 0.06	21.59 $\pm$ 0.06
NSGA-II + ALNS (RF)	25,198.59 $\pm$ 14.49	639.0 $\pm$ 0.0	355.09	1.429 $\pm$ 0.06	21.92 $\pm$ 0.39
NSGA-II + ALNS (XGB)	25,232.82 $\pm$ 86.18	640.0 $\pm$ 0.0	389.32	1.567 $\pm$ 0.35	22.24 $\pm$ 0.03

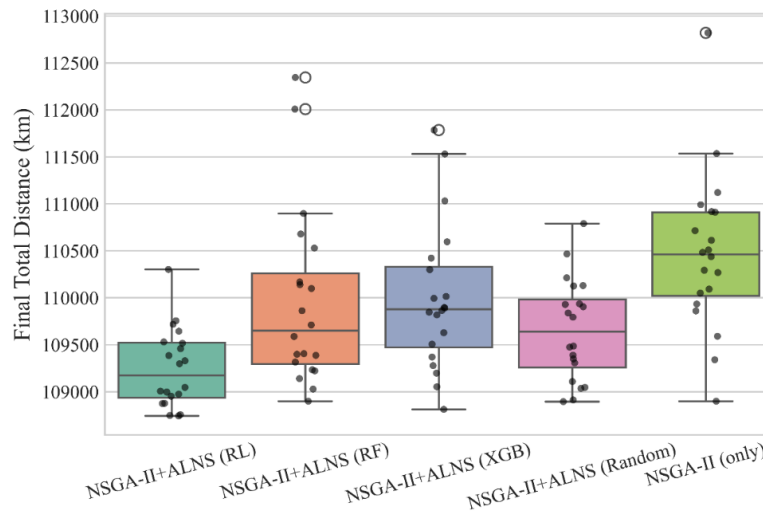


Figure 2 Distribution of final solutions (20 independent runs)

The ALNS-enhanced methods consistently outperformed the standalone NSGA-II, reducing the optimality gap from 1.61% to approximately 1.43% while maintaining computation times below 23 seconds. The RL-guided variant achieved nearly identical performance to the RF-based method while exhibiting a slightly faster average runtime, demonstrating reliability and generalization across different instance sizes.

4.2 Optimization Performance

To evaluate large-scale performance, 20 independent runs were conducted on the 199×4 capacitated transportation problem. Figure 2 and Table 5 compare the results of five methods: NSGA-II (Only), NSGA-II + ALNS (random), NSGA-II + ALNS (RF), NSGA-II + ALNS (XGBoost), and the proposed NSGA-II + ALNS (RL). For comparison, the MILP model was also implemented in LINGO. While LINGO successfully obtained optimal solutions for small and medium-sized instances, it failed to converge to a feasible or optimal solution for the large-scale instance within a reasonable computational time. This

observation confirms the computational intractability of exact MILP solvers for large-scale sugarcane transportation planning and justifies the adoption of hybrid metaheuristic approaches.

As shown in Figure 2, NSGA-II + ALNS (RL) consistently achieved the lowest final transportation distance with the smallest dispersion, indicating superior convergence stability. In contrast, the random ALNS and standalone NSGA-II methods exhibited wider variability, reflecting less reliable search behavior without learning-guided operator control.

Quantitatively, Table 5 indicates that NSGA-II + ALNS (RL) achieves the lowest mean total distance (109,245.22 km) with a relatively small standard deviation (417.87 km), demonstrating both superior solution quality and strong robustness across runs. The RL-guided variant consistently outperforms the RF-guided, XGBoost-guided, and random ALNS approaches. Notably, the average number of truck trips remains almost identical across all methods, confirming that the observed improvements stem from more efficient allocation decisions rather than increased fleet usage.

Table 5 Performance summary across 20 independent runs (199×4 instance)

Method	Final Distance (km) $\pm$ SD	Trips (mean)	Runtime (s)
NSGA-II + ALNS (RL)	109,245.22 $\pm$ 417.87	3,881.15	229.11
NSGA-II + ALNS (RF)	109,953.19 $\pm$ 946.57	3,881.20	181.89
NSGA-II + ALNS (XGB)	109,991.99 $\pm$ 780.38	3,881.35	188.71
NSGA-II + ALNS (Random)	109,656.71 $\pm$ 538.34	3,881.20	181.94
NSGA-II (Only)	110,469.27 $\pm$ 839.93	3,881.45	130.39

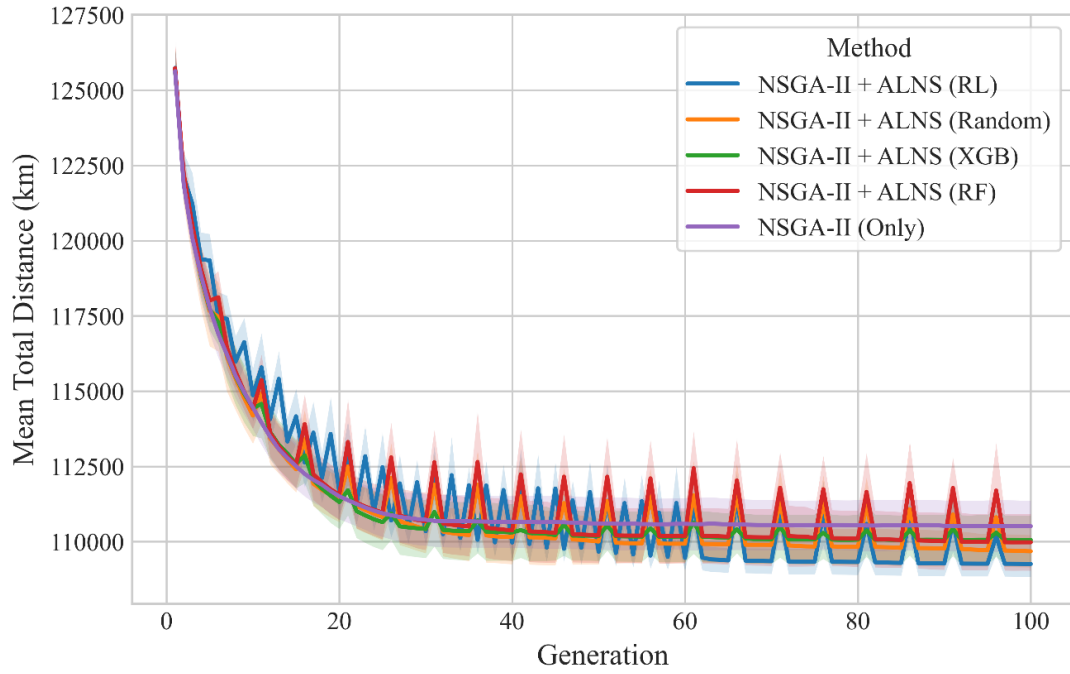


Figure 3 Convergence curves of optimization methods with error bands (mean $\pm$ SD)

The mean number of truck trips is nearly identical across all methods because trip counts are primarily determined by fixed demand volumes and truck capacity constraints. Since all algorithms operate under the same demand, fleet configuration, and trip assignment rules, improvements are achieved mainly through more efficient allocation of flows to factories, resulting in shorter travel distances rather than changes in the number of trips.

4.3 Convergence Analysis

Figure 3 shows the convergence behavior of the five optimization methods over 100 generations in terms of the mean total transportation distance. All methods improve rapidly during the early stage (generations 0 – 20), indicating effective initialization and global exploration.

After this phase, differences in convergence stability become apparent. NSGA-II + ALNS (RL)

achieves the lowest distance trajectory and stabilizes earlier, whereas the NSGA-II baseline and the random ALNS variant exhibit slower stabilization and larger fluctuations. The RF-guided and XGBoost-guided variants provide moderate improvements but still show noticeable oscillations in later generations.

The spikes observed in all hybrid approaches are expected, as ALNS is applied periodically and uses a simulated annealing acceptance rule. During ALNS iterations, destroy–repair perturbations can temporarily worsen solution quality, and such moves may be accepted probabilistically to promote exploration. Because Figure 3 reports the mean population performance, these temporary deteriorations appear as spikes. Notably, the RL-guided strategy yields smaller oscillations and faster recovery, resulting in more stable convergence and a lower final mean distance than that of NSGA-II by the last generation.

Table 6 Multi-objective performance metrics (HV and IGD)

Method	HV $\pm$ SD	IGD $\pm$ SD	Avg. Pareto Points.
NSGA-II + ALNS (RL)	192,021.87 $\pm$ 17,037.33	0.95 $\pm$ 0.22	1.20
NSGA-II + ALNS (Random)	175,245.26 $\pm$ 21,961.69	1.0 $\pm$ 0.00	1.25
NSGA-II + ALNS (RF)	163,208.89 $\pm$ 38,677.83	1.0 $\pm$ 0.00	1.45
NSGA-II + ALNS (XGB)	161,489.70 $\pm$ 31,939.27	1.0 $\pm$ 0.00	1.40
NSGA-II (only)	141,999.62 $\pm$ 34,275.53	1.0 $\pm$ 0.00	1.45

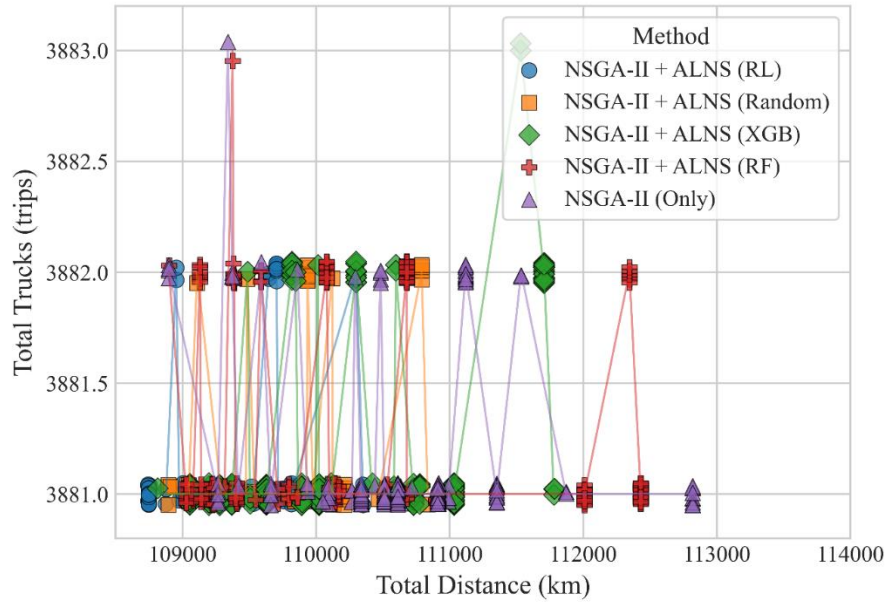


Figure 4 Pareto fronts across optimization methods

4.4 Pareto Front and Multi-objective Performance

The multi-objective performance of the five optimization methods was evaluated by analyzing their Pareto fronts and two standard indicators: HV and IGD. Figure 4 illustrates the obtained Pareto fronts, representing the trade-off between total transportation distance and the number of truck trips.

As shown in Figure 4, all methods generate a limited number of non-dominated solutions, reflecting the highly correlated nature of the two objectives in large-scale sugarcane transportation planning. Nevertheless, clear differences in Pareto quality can be observed. NSGA-II + ALNS (RL) consistently produces solutions with shorter distances while maintaining comparable truck usage, indicating more effective exploration of the trade-off surface. In contrast, the Pareto fronts obtained by NSGA-II (only) and the random ALNS variant are more scattered toward higher-distance regions, suggesting weaker convergence behavior. The RF-guided and XGBoost-guided variants show intermediate performance but do not match the consistency of the RL-guided approach.

Table 6 summarizes the multi-objective performance over 20 independent runs using HV and IGD. The NSGA-II + ALNS (RL) approach achieves the highest mean HV (192,021.87) with relatively low variability, indicating superior Pareto-front coverage and robustness compared with those of the other methods.

The IGD values are identical or close to 1.0 for several algorithms, which is expected due to the sparsity of Pareto-optimal solutions in the large-scale instance, in which each run typically yields only one or two non-dominated points. Under such conditions, IGD becomes less discriminative because multiple methods produce overlapping or similarly distant Pareto sets relative to the reference front.

Despite this limitation, the RL-guided method attains the lowest mean IGD, suggesting slightly closer convergence toward the reference front. Overall, these results confirm that reinforcement learning-guided operator selection enhances Pareto-front quality and stability, with hypervolume serving as the more informative indicator in this large-scale, tightly constrained setting.

Table 7 Scenario-based resilience analysis (RL-only)

Scenario	Total Distance (km)	Total Trips	Runtime (s)
Baseline	109,712.45	3,881	180.82
Outage F1	103,265.21	3,234	155.39
Outage F2	137,252.96	3,881	181.09
Outage F3	78,214.18	2,326	116.39
Outage F4	145,125.02	3,636	170.78

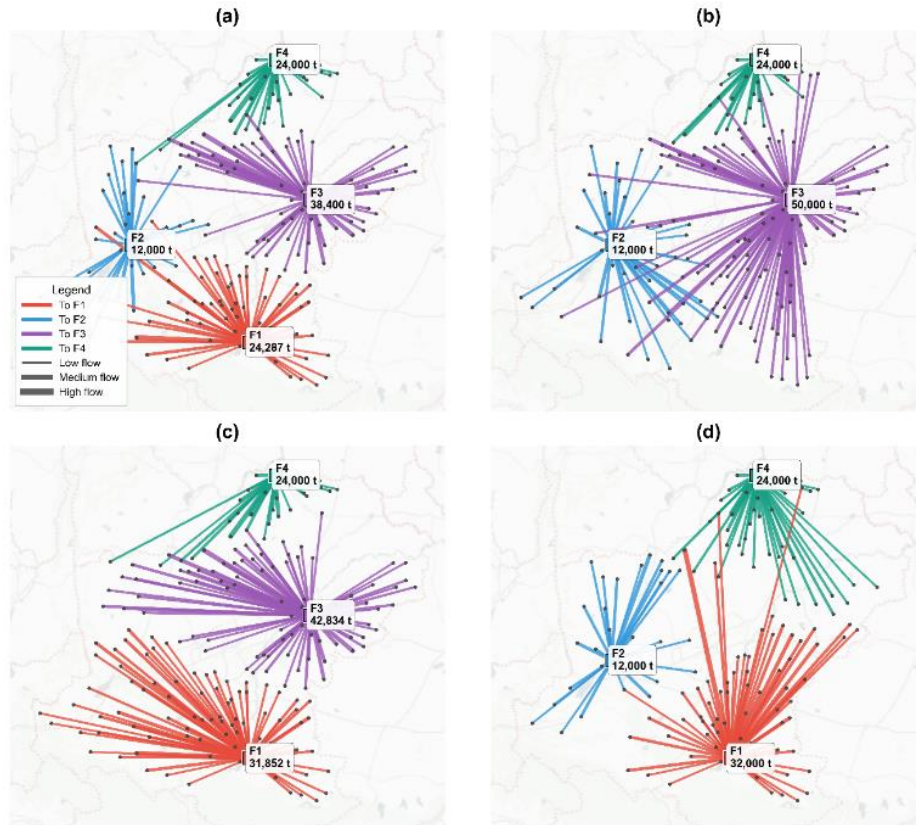


Figure 5 Flow maps under baseline and representative factory-outage scenarios: (a) baseline, (b) outage F1, (c) outage F2, and (d) outage F3

4.5 Scenario-based Resilience (RL-only)

To assess the adaptability of the reinforcement learning-guided hybrid algorithm under operational disruptions, four factory-outage scenarios were examined. In each case, one processing mill (F1 – F4) was assumed to be temporarily unavailable, and the NSGA-II + ALNS (RL) framework re-optimized sugarcane allocations to the remaining factories while satisfying capacity constraints. The objective was to evaluate whether the method could maintain feasible operations and limit adverse impacts on transportation performance.

Table 7 summarizes the results of the RL-guided method under the four outage scenarios relative to

the baseline. This analysis focuses exclusively on the RL-based framework, as it demonstrated the most stable and competitive performance in the primary experiments. When factory F1 was unavailable, reassignment to nearby mills reduced total transportation distance by approximately 5.9%. The outage of factory F3 led to the largest improvement (28.7%), reflecting effective redistribution toward geographically closer facilities. In contrast, outages at factories F2 and F4 increased transportation distance due to longer hauling requirements. Nevertheless, all solutions remained feasible, and factory capacity limits were consistently respected.

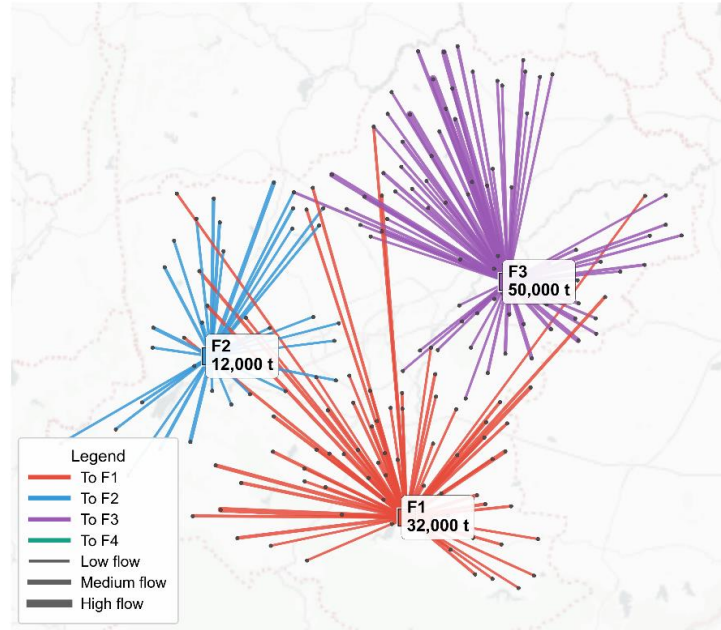


Figure 6 Flow map under outage at Factory F4

Table 8 Top five subdistricts with largest distance reductions in the baseline scenario (RL-only)

Subdistrict	Reassigned From → To	Before (km)	After (km)	Reduction (km, %)
T30	F1 → F2	83.31	10.30	73.01 (−87.6%)
T38	F1 → F2	68.38	10.19	58.19 (−85.1%)
T39	F1 → F2	85.68	14.91	70.77 (−82.6%)
T40	F1 → F2	90.02	17.36	72.66 (−80.7%)
T41	F1 → F2	75.60	15.26	60.34 (−79.8%)

Spatial Resilience Visualization

The spatial flow patterns corresponding to the resilience analysis are shown in Figures 5 and 6. Figure 5(a–d) presents the baseline configuration together with representative outage scenarios (F1–F3), illustrating how transportation flows are redistributed in response to moderate disruptions. In contrast, Figure 6 highlights the outage of factory F4 as the most severe case, where longer redirection paths dominate the network. Despite this extreme disruption, the RL-guided framework preserves balanced and feasible routing structures without over-assignment at any factory.

Overall, the numerical results in Table 7 and the spatial flow maps confirm that the RL-guided NSGA-II + ALNS framework exhibits strong resilience, maintaining operational feasibility and adaptive routing behavior under both moderate and severe factory-outage conditions.

Example of Local Adaptation

Table 8 lists five representative subdistricts that experienced the most significant reductions in travel

distance during the baseline optimization (no outage). All were initially assigned to factory F1 but were reallocated to a nearer mill (F2), resulting in distance reductions of 70–87%. This illustrates the RL agent’s ability to identify low-cost allocation opportunities and refine assignments based on learned patterns in the search space.

Overall, the hybrid RL approach demonstrated strong resilience under factory-shutdown scenarios, consistently maintaining feasible operations and achieving substantial distance reductions when rerouting opportunities existed. Even in the most disruptive cases, the algorithm sustained balanced allocations across the remaining mills without exceeding capacity constraints. These results highlight the potential of RL-guided optimization as a decision-support tool for robust and sustainable sugarcane-transportation planning under real-world uncertainty.

4.6 Statistical Analysis

A one-way analysis of variance (ANOVA) was performed on the final total transportation distance obtained from 20 independent runs to examine whether statistically significant differences exist among the five optimization methods: NSGA-II (only), NSGA-II + ALNS (random), NSGA-II + ALNS (RF), NSGA-II + ALNS (XGB), and NSGA-II + ALNS (RL).

The ANOVA results indicate a significant overall effect of algorithm type on the final distance, $F(4, 95) = 7.62, p < 0.001$, confirming that at least one method differs significantly from the others.

Post-hoc pairwise comparisons using Tukey's HSD test ($\alpha = 0.05$) reveal that NSGA-II + ALNS (RL) achieves a significantly lower mean distance than NSGA-II (only) ($p < 0.001$) and NSGA-II + ALNS (RF) ($p = 0.024$). In addition, NSGA-II + ALNS (RL) significantly outperforms NSGA-II + ALNS (XGB) ($p = 0.015$).

No statistically significant differences are observed between NSGA-II + ALNS (RL) and NSGA-II + ALNS (Random) ($p > 0.05$), indicating comparable mean performance between these two hybrid variants under the chosen confidence level.

Furthermore, the comparison between NSGA-II (Only) and NSGA-II + ALNS (Random) shows a statistically significant improvement in favor of the Random ALNS hybrid ($p = 0.006$), highlighting the benefit of incorporating adaptive large-neighborhood search even without learning-based operator control.

Overall, the statistical analysis confirms that the reinforcement learning-guided hybrid method (NSGA-II + ALNS (RL)) provides statistically significant improvements over the classical NSGA-II baseline and selected learning-based hybrids, while maintaining competitive performance relative to other ALNS-enhanced variants. These results support the effectiveness of RL-guided operator selection for large-scale capacitated transportation optimization.

4.7 Environmental and Economic Impact

To assess the environmental and economic implications of the proposed framework, four indicators were evaluated before and after optimization: total transportation distance, fuel consumption, CO₂ emissions, and fuel cost. The *Before* case represents the baseline solution generated from a fixed seeded random assignment. Table 9 summarizes the comparative results across all methods.

Table 9 Comparison of total distance, fuel, CO₂, and fuel-cost reductions among optimization methods

method	km_total	Δ distance (%)	fuel (L)	Δ fuel (%)	CO ₂ _kg	Δ CO ₂ (%)	fuel_cost_baht	Δ fuel_cost_baht
Before	123,313.52	–	28,131.83	–	75,955.94	–	898,530.69	–
NSGA-II + ALNS (RL)	109,245.22	–11.41	25,108.96	–10.75	67,794.20	–10.75	801,980.29	–96,550.40
NSGA-II + ALNS (Random)	109,656.32	–11.08	25,205.02	–10.40	68,053.55	–10.40	805,048.24	–93,482.45
NSGA-II + ALNS (XGB)	109,991.07	–10.80	25,268.68	–10.18	68,225.43	–10.18	807,081.58	–91,449.11
NSGA-II + ALNS (RF)	109,950.67	–10.84	25,282.22	–10.13	68,262.00	–10.13	807,514.17	–91,016.52
NSGA-II (Only)	110,469.27	–10.42	25,414.14	–9.66	68,618.18	–9.66	811,727.66	–86,803.03

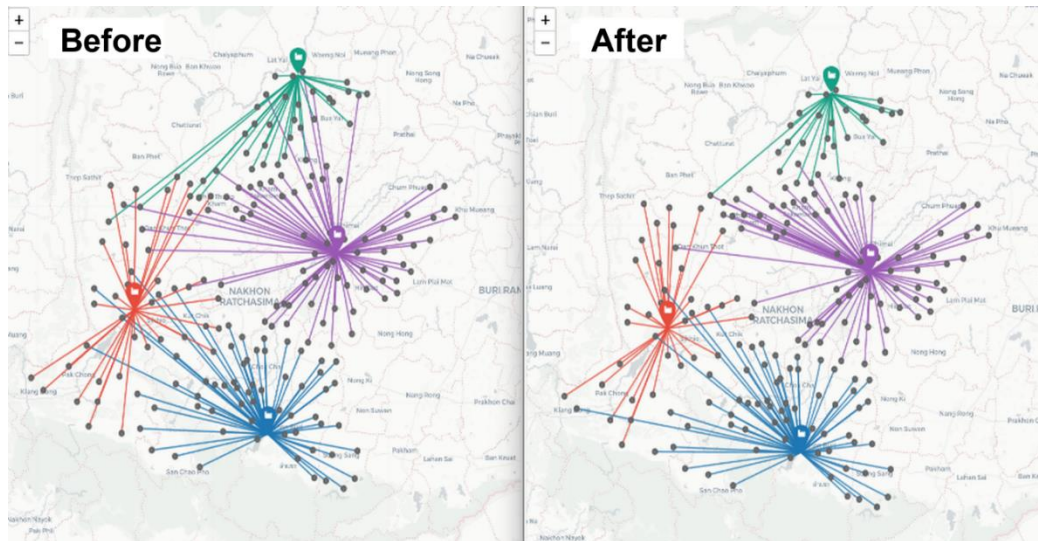


Figure 7 Sugarcane transportation flow maps before (left) and after (right) optimization using the hybrid RL approach

Table 10 Overall performance comparison across optimization methods

Method	Distance Rank ↓	Stability Rank ↓	Runtime Rank ↓	Overall Assessment
NSGA-II + ALNS (RL)	1st	1st	5th	Best multi-objective quality
NSGA-II + ALNS (random)	2nd	2nd	2nd	Balanced
NSGA-II + ALNS (RF)	3rd	5th	3rd	Moderate
NSGA-II + ALNS (XGB)	4th	4th	4th	Moderate
NSGA-II (Only)	5th	3rd	1st	Fast but inferior quality

The NSGA-II + ALNS (RL) method achieved the largest overall improvements, reducing total transportation distance by 11.41%, fuel consumption and CO₂ emissions by 10.75%, and yielding an absolute fuel-cost savings of approximately 96,550 Baht per planning cycle. These results demonstrate the effectiveness of reinforcement learning in enhancing adaptive neighborhood search within the hybrid framework.

The random, XGBoost, and RF variants also delivered substantial benefits, with distance reductions of 10.8 – 11.1%, fuel and emission reductions of 10.1 – 10.4%, and fuel-cost savings ranging from 91,000 to 93,000 Baht. In contrast, the NSGA-II (only) approach produced the smallest improvements, achieving a 10.42% reduction in distance and fuel cost savings of approximately 86,700 Baht.

Overall, the results confirm that integrating ALNS-based neighborhood search significantly enhances both environmental efficiency and economic performance, with the reinforcement learning-guided variant providing the most consistent and impactful gains

Figure 7 visualizes the spatial distribution of sugarcane transportation flows before and after optimization. The optimized configuration exhibits significantly shorter and more localized routing patterns, with long-haul paths substantially reduced. The resulting transport network becomes more compact and geographically clustered, indicating the improved allocation of subdistricts to nearer processing locations. Overall, these outcomes confirm that incorporating reinforcement learning into the NSGA-II + ALNS framework provides measurable environmental and economic benefits, supporting more sustainable and efficient sugarcane logistics planning.

5. Discussion

The experimental results clearly demonstrate the effectiveness of incorporating reinforcement learning into the NSGA-II + ALNS hybrid framework for large-scale agricultural transportation planning. Among all tested variants, the RL-guided hybrid consistently achieves the lowest total transportation distance, fuel consumption, and CO₂ emissions while also exhibiting the smallest variability across

independent runs. These findings indicate that adaptive operator selection significantly enhances search stability and prevents premature convergence in complex, capacity-constrained routing problems.

To further interpret the trade-offs among solution quality, robustness, and computational effort, Table 10 provides a consolidated comparison across distance performance, stability, and runtime. While the baseline NSGA-II remains the fastest method, it produces inferior solution quality. In contrast, the RL-guided hybrid delivers the best overall balance by achieving the shortest distance and the highest stability at the expense of a moderate increase in computational time. The random, RF-guided, and XGBoost-guided variants exhibit intermediate performance, suggesting that static or pre-trained operator-selection mechanisms are less effective in adapting to evolving search conditions.

From an operational perspective, the RL-enhanced framework offers tangible benefits for real-world logistics planning. The observed reductions of approximately 11% in total transportation distance and approximately 10–11% in fuel consumption and CO₂ emissions directly support energy-efficiency and emission-reduction objectives in Thailand's agricultural sector. Moreover, under factory-outage scenarios, the RL-guided model successfully reallocates transportation flows while maintaining feasibility with respect to capacity constraints, demonstrating robustness to operational disruptions. Overall, these findings confirm that reinforcement learning provides a scalable and effective mechanism for guiding neighborhood search, thereby strengthening the practical applicability of hybrid optimization approaches in large-scale agricultural logistics systems.

6. Conclusion

This study developed a hybrid optimization framework that integrates a multi-objective evolutionary algorithm with reinforcement learning-guided adaptive large neighborhood search to address large-scale sugarcane transportation planning. The problem was formulated as a bi-objective capacitated transportation problem, aiming to minimize total transportation distance and the number of truck trips while satisfying factory capacity constraints across multiple subdistrict origins and processing facilities.

Computational experiments were conducted using a real-world sugarcane transportation network in northeastern Thailand consisting of 199 subdistricts and four processing plants. Compared with the baseline logistics configuration, the proposed hybrid

RL framework reduced total transportation distance from 123,313.52 km to 109,245.22 km, representing an improvement of 11.41%. This reduction led to a decrease in fuel consumption from 28,131.83 L to 25,108.96 L (10.75%), corresponding to a reduction in CO₂ emissions from 75,955.94 kg to 67,794.20 kg, and an absolute fuel-cost savings of approximately 96,500 Thai Baht per planning cycle. Statistical analysis using ANOVA and Tukey's HSD test confirmed that these improvements were significant at the 95% confidence level.

Beyond numerical performance gains, the proposed framework demonstrated strong adaptability under disruption scenarios in which one processing factory was temporarily unavailable. The reinforcement learning-guided model successfully reallocated sugarcane flows to the remaining facilities without violating capacity constraints, while maintaining competitive transportation efficiency. This highlights the robustness of adaptive operator selection in handling capacity disruptions and operational uncertainty.

Overall, the results indicate that integrating reinforcement learning with evolutionary and neighborhood search mechanisms provides a practical and scalable solution for complex agricultural transportation systems. The proposed framework supports more sustainable logistics management by consistently reducing transportation distance, fuel consumption, and emissions, and it can be extended to other large-scale capacitated transportation and distribution problems. Future research may incorporate real-time traffic conditions, seasonal yield variability, and dynamic fleet information to further enhance responsiveness and support real-world deployment in smart agricultural logistics systems.

7. 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) (2023), Thailand, for providing the real-world data used in this research.

8. Abbreviations

Abbreviation	Full Term
ALNS	Adaptive Large Neighborhood Search
ANOVA	Analysis of Variance
CO ₂	Carbon Dioxide

Abbreviation	Full Term
CTP	Capacitated Transportation Problem
MILP	Mixed-Integer Linear Programming
NSGA-II	Non-dominated Sorting Genetic Algorithm II
RF	Random Forest
RL	Reinforcement Learning
THB	Thai Baht
XGB	eXtreme Gradient Boosting

9. CRediT Statement

Putis Wittayasin: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft.
Nuchsa Kriengkarakot: Validation, Resources, Writing – Review & Editing, Supervision.
Preecha Kriengkarakot: Conceptualization, Methodology, Resources, Writing – Review & Editing, Supervision, Project Administration.

10. References

- Aliano Filho, A., Melo, T., & Pato, M. V. (2020). *Tactical planning of sugarcane harvest and transport operations*. Retrieved from <https://www.econstor.eu/handle/10419/224073>
- Arabani, A. B., & Farahani, R. Z. (2012). Facility location dynamics: An overview of classifications and applications. *Computers & Industrial Engineering*, 62(1), 408-420. <https://doi.org/10.1016/j.cie.2011.09.018>
- Bengio, Y., Lodi, A., & Prouvost, A. (2021). Machine learning for combinatorial optimization: A methodological tour d'horizon. *European Journal of Operational Research*, 290(2), 405-421. <https://doi.org/10.1016/j.ejor.2020.07.063>
- Chetthamrongchai, P., Auansakul, A., & Supawan, D. (2001). Assessing the transportation problems of the sugar cane industry in Thailand. *Transport and Communications Bulletin for Asia and the Pacific*, 70(2001), 31-39.
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182-197. <https://doi.org/10.1109/4235.996017>
- Energy Policy and Planning Office. (2025). *Retail diesel price in Thailand*. Ministry of Energy. Retrieved from <https://www.eppo.go.th>
- Higgins, A. (2006). Scheduling of road vehicles in sugarcane transport: A case study at an Australian sugar mill. *European Journal of Operational Research*, 170(3), 987-1000. <https://doi.org/10.1016/j.ejor.2004.07.055>
- Hoff, A., Andersson, H., Christiansen, M., Hasle, G., & Løkketangen, A. (2010). Industrial aspects and literature survey: Fleet composition and routing. *Computers & Operations Research*, 37(12), 2041-2061. <https://doi.org/10.1016/j.cor.2010.03.015>
- Jena, S. D., & Poggi, M. (2013). Harvest planning in the Brazilian sugar cane industry via mixed integer programming. *European Journal of Operational Research*, 230(2), 374-384. <https://doi.org/10.1016/j.ejor.2013.04.011>
- Johnn, S. N., Darvariu, V. A., Handl, J., & Kalcsics, J. (2024). A graph reinforcement learning framework for neural adaptive large neighbourhood search. *Computers & Operations Research*, 172, Article 106791. <https://doi.org/10.1016/j.cor.2024.106791>
- Karimi-Mamaghan, M., Mohammadi, M., Meyer, P., Karimi-Mamaghan, A. M., & Talbi, E. G. (2022). Machine learning at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art. *European Journal of Operational Research*, 296(2), 393-422. <https://doi.org/10.1016/j.ejor.2021.04.032>
- Krungsri Research. (2023). *Business/industry outlook 2024–2026: Sugar industry*. Krungsri Bank. Retrieved from https://www.krungsri.com/getmedia/cb3b7846-61c0-4e0b-bffc-b1a44380f9d6/IO_Sugar_240725_EN_EX.pdf.aspx
- Naparswad, T. (2013). *A study of fuel consumption of trucks* [Master's thesis]. Suranaree University of Technology. Retrieved from https://eng.sut.ac.th/ce/ce_course/download/project/7-1-55/20THOSSAPOL%20NAPARSWAD.pdf
- Office of the Cane and Sugar Board (OCSB). (2023). *Sugarcane cultivation report 2023/24*. Ministry of Industry, Thailand. Retrieved from <https://www.ocsb.go.th/2024/reports-articles/area-yield/27524/>

- Pisinger, D., & Ropke, S. (2018). Large neighborhood search. *Handbook of metaheuristics*. Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-91086-4_4
- Reijnen, R., Zhang, Y., Lau, H. C., & Bukhsh, Z. (2024). *Online control of adaptive large neighborhood search using deep reinforcement learning* [Conference presentation]. Proceedings of the international conference on automated planning and scheduling, Washington, DC, US. <https://doi.org/10.1609/icaps.v34i1.31507>
- Rein, P. (2016). *Cane sugar engineering*. Retrieved from <https://www.cabidigitallibrary.org/doi/full/10.5555/5/20173154754>
- Ropke, S., & Pisinger, D. (2006). An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science*, 40(4), 455-472. <https://doi.org/10.1287/trsc.1050.0135>
- Solomon, S., Banerji, R., Shrivastava, A. K., Singh, P., Singh, I., Verma, M., ... & Sawnani, A. (2006). Post-harvest deterioration of sugarcane and chemical methods to minimize sucrose losses. *Sugar Tech*, 8(1), 74-78. <https://doi.org/10.1007/BF02943746>
- Sun, Y., & Lang, M. (2015). Bi-objective optimization for multi-modal transportation routing planning problem based on Pareto optimality. *Journal of Industrial Engineering and Management*, 8(4), 1195-1217. <https://doi.org/10.3926/jiem.1562>
- Teixeira, E. D. S., Rangel, S., Florentino, H. D. O., & de Araujo, S. A. (2023). A review of mathematical optimization models applied to the sugarcane supply chain. *International Transactions in Operational Research*, 30(4), 1755-1788. <https://doi.org/10.1111/itor.13056>
- U.S. Environmental Protection Agency. (2025). *GHG emission factors hub*. Office of Climate Change Programs. Retrieved from <https://www.epa.gov/system/files/documents/2025-01/ghg-emission-factors-hub-2025.pdf>
- Zitzler, E., & Thiele, L. (1998). *Multiobjective optimization using evolutionary algorithms a comparative case study* [Conference presentation]. International Conference on Parallel Problem Solving from Nature. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0056872>
- Zitzler, E., Deb, K., & Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2), 173-195. <https://doi.org/10.1162/106365600568202>