



Heterogeneous Penalties in Unbalanced Transportation: Exact Linear Programming, Dummy-Node Equivalence, and a Penalty-Aware Ant Colony Heuristic

Ali Mahmoud Assabri ^{1*}, Mohamed Al-Issawi Kajman ², Abdulbasat Abdulghafar Alfeed ³,
Naji Abdalaziz Ali ⁴

^{1,3} Department of Engineering Management, College of Technical Sciences, Bani Walid, Libya

² Department of Computing and Information Technology, College of Technical Sciences, Bani Walid, Libya

⁴ Department of Mechanical Engineering, College of Technical Sciences, Bani Walid, Libya

* Corresponding Author: Email: alimahmoudassabri@gmail.com

العقوبات غير المتجانسة في مشكلة النقل غير المتوازنة: البرمجة الخطية الدقيقة، وتكافؤ العقد الوهمية، وخوارزمية مستعمرة النمل المدركة للعقوبات

علي محمود الصابري^{1*}، محمد العيساوي كجمان²، عبد الباسط عبد الغفار الفيض³، ناجي عبد العزيز علي⁴

^{3,1} قسم الإدارة الهندسية، كلية العلوم التقنية، بني وليد، ليبيا

² قسم الحوسبة وتقنية المعلومات، كلية العلوم التقنية، بني وليد، ليبيا

⁴ قسم الهندسة الميكانيكية، كلية العلوم التقنية، بني وليد، ليبيا



Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Date of Submission: 01-05-2026

Date of acceptance: 16-07-2026

Date of publishing: 10-08-2026

Abstract

Unbalanced transportation models are often balanced with a dummy source or destination. That transformation is exact only when dummy-arc costs preserve the location-specific consequences of surplus or shortage; replacing a penalty vector with its arithmetic mean defines a different optimization problem. This paper formulates both imbalance directions as continuous linear programs, establishes heterogeneous dummy-node equivalence, and evaluates a penalty-aware Ant Colony Optimization (ACO) construction against exact linear programming and five alternative methods. Fully specified examples are complemented by a controlled synthetic benchmark spanning several network dimensions. The results show that uniform penalties can materially increase the true heterogeneous objective, while penalty-aware ACO improves on a cost-only construction. Variable Neighborhood Search achieves the strongest overall rank in the expanded comparison, whereas penalty-aware ACO remains competitive with other stochastic methods under a common evaluation budget. The conclusions are restricted to the continuous model, the stated synthetic generator, the supplied cases, and the declared computational protocols. Synthetic instances are used as algorithmic stress tests rather than as observations of industrial transportation systems.

Keywords: unbalanced transportation; heterogeneous penalties; linear programming; dummy-node equivalence; Ant Colony Optimization; minimum-cost flow; reproducible computational research.

الملخص

غالبًا ما يتم تحويل نماذج النقل غير المتوازنة إلى نماذج متوازنة باستخدام مصدر وهمي أو وجهة وهمية. إلا أن هذا التحويل يكون دقيقًا فقط عندما تحافظ تكاليف الأقواس الوهمية على الآثار المكانية المحددة لحالات الفائض أو العجز؛ أما استبدال متجه الغرامات بمتوسطه الحسابي فإنه يؤدي إلى تعريف مسألة تحسين مختلفة. تقوم هذه الورقة بصياغة كلا اتجاهي عدم التوازن كنماذج برمجة خطية مستمرة، وتثبت تكافؤ العقد الوهمية ذات الغرامات غير المتجانسة، كما تقيم خوارزمية تحسين مستعمرة النمل (ACO) المعتمدة على الغرامات مقارنةً بالبرمجة الخطية الدقيقة وخمس طرق بديلة.

تتضمن الدراسة أمثلة محددة بالكامل، بالإضافة إلى معيار اختبائي اصطناعي مضبوط يغطي أبعادًا مختلفة للشبكات. وتُظهر النتائج أن استخدام غرامات موحدة يمكن أن يؤدي إلى زيادة ملحوظة في قيمة الدالة الهدف الحقيقية عند وجود غرامات غير متجانسة، بينما تعمل خوارزمية ACO المعتمدة على الغرامات على تحسين الأداء مقارنةً بالبناء القائم على التكلفة فقط.

تحقق خوارزمية البحث في الجوار المتغير (Variable Neighborhood Search) أفضل ترتيب إجمالي ضمن المقارنة الموسعة، في حين تبقى خوارزمية ACO المعتمدة على الغرامات منافسة للطرق العشوائية الأخرى عند استخدام ميزانية تقييم موحدة.

تقتصر الاستنتاجات على النموذج المستمر، ومولد الحالات الاصطناعية المحدد، والحالات المقدمة، وبروتوكولات الحساب المعلنة. وتُستخدم الحالات الاصطناعية كاختبارات ضغط لتقييم الخوارزميات، وليس باعتبارها ملاحظات مأخوذة من أنظمة نقل صناعية فعلية.

الكلمات المفتاحية: النقل غير المتوازن؛ الغرامات غير المتجانسة؛ البرمجة الخطية؛ تكافؤ العقد الوهمية؛ خوارزمية مستعمرة النمل؛ التدفق ذي التكلفة الدنيا؛ البحث الحاسوبي القابل لإعادة الإنتاج.

Introduction

The transportation problem allocates supply from a set of sources to a set of destinations while minimizing total operating cost. In the classical balanced model, total supply equals total demand. Operational systems are less orderly: production variability, inventory restrictions, demand fluctuations, and service commitments can leave material at a source or leave part of a destination's requirement unserved.

The standard remedy is to introduce a dummy destination when supply is excessive or a dummy source when demand is excessive. This device is consistent with the minimum-cost-flow representation of transportation models [7]. It is not merely a bookkeeping convention. The cost assigned to a dummy arc represents an economic consequence—such as holding, disposal, deterioration, outsourcing, lost sales, or shortage. If that consequence depends on the source or destination, replacing the vector of location-specific costs with one scalar changes the objective function and, potentially, the optimal allocation.

The distinction matters when algorithmic results are compared. An exact solver applied to a uniform-dummy model is exact only for that simplified model; it is not an exact benchmark for the heterogeneous model that retains source- or destination-specific penalties. A heuristic evaluated under the heterogeneous objective can therefore appear to outperform an “exact” solution when the two procedures have solved different problems. The present study separates three objects: the exact heterogeneous optimum, the true heterogeneous cost of the uniform-dummy allocation, and heuristic solutions evaluated under the same heterogeneous objective. The study has two linked aims. The first is mathematical: to state the excess-supply and excess-demand models explicitly and establish equivalence between their native formulations and heterogeneous dummy-node representations. The second is computational: to examine whether an ACO construction that includes heterogeneous dummy arcs remains competitive on fully specified cases and on a controlled synthetic benchmark. The supplied cases are retained as auditable examples, whereas the 120 additional instances are labeled synthetic and used only

to examine algorithmic behavior across dimensions, imbalance directions, and randomly generated cost structures.

The contributions are deliberately limited. First, the paper distinguishes an exact heterogeneous dummy construction from a uniform-dummy approximation. Second, it provides an independent exact LP benchmark and a separate dummy-node verification. Third, it represents holding and shortage penalties directly as dummy-arc costs during ACO construction. Fourth, it compares penalty-aware ACO with a cost-only ablation and four alternative methods. Fifth, it reports blocked nonparametric comparisons, a budget-aligned robustness check, and an exploratory factor analysis. The study does not claim that ACO dominates exact optimization or that synthetic cases represent industrial practice.

1. Background and Research Question

1.1. Unbalanced transportation and node-specific penalties

When total available supply differs from total requirement, the residual quantity has a location. In an excess-supply case, material remains at one or more sources; a source-specific holding cost is consequently represented by a separate dummy arc from each source to a common dummy destination. In an excess-demand case, the unmet quantity belongs to a destination; a destination-specific shortage penalty is represented by a dummy source with one dummy arc to each destination.

The dummy construction is exact when these arc costs are retained. The approximation begins only when the penalty vector is replaced by its arithmetic mean. Recent work on fixed-charge, rough, unbalanced optimal-transport, routing, and metaheuristic formulations illustrates the variety of assumptions that can be embedded in transportation objectives [12] [13] [14] [15]. Related transportation applications, including public-transport planning, rail scheduling, network resilience, and split-delivery routing, further show why algorithmic conclusions should remain tied to the model and computational setting in which they were obtained [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26].

The central research question is therefore: How much true-cost regret can be induced by a uniform dummy penalty, and how closely can a penalty-aware ACO recover the heterogeneous optimum on the fully specified cases and the synthetic benchmark?

1.2. Positioning the heuristic

The base model is a continuous linear program. SciPy's linprog interface minimizes a linear objective subject to equality and inequality constraints, and its HiGHS method provides continuous LP solvers through a standard Python interface [1] [2]. HiGHS is intended for large-scale linear optimization and provides official citation guidance [3]. Whenever the exact LP can be solved within the available planning time, it is the principal reference in this study.

ACO has a narrower role. Its construction follows the pheromone-based logic established in the ACO literature [8] [9], but the present paper does not introduce a new general-purpose metaheuristic. ACO is considered as a decision-support heuristic for settings in which additional constraints, integration requirements, or a specified time–quality trade-off make a direct exact formulation less convenient. Methodological work on heuristic experimentation emphasizes the need for multiple instances, explicit performance measures, and careful treatment of computational effort [4]. Statistical comparisons of stochastic optimization methods likewise require attention to instance structure and test assumptions [5] [6]. The absence of a universal best heuristic provides an additional reason to report alternatives rather than a single favorable configuration [10] [11].

2. Mathematical Formulation

Let there be m sources and n destinations. Let $S_i \geq 0$ denote the available supply at source i , $D_j \geq 0$ the requirement at destination j , and $c_{ij} \geq 0$ the unit transportation cost from source i to destination j . The continuous decision variable x_{ij} denotes the amount shipped on that arc. Let

$h_i \geq 0$ be the holding cost per residual unit at source i , and let $p_j \geq 0$ be the shortage penalty per unserved unit at destination j . All variables are continuous and non-negative.

2.1. Excess supply

When $\sum_i S_i \geq \sum_j D_j$, residual variables s_i represent the quantity left at source i . The heterogeneous model is:

$$\begin{aligned} (1) Z^{Hs} &= \min \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{i=1}^m h_i s_i \\ (2) \sum_{j=1}^n x_{ij} + s_i &= S_i, \text{ for every source } i \\ (3) \sum_{i=1}^m x_{ij} &= D_j, \text{ for every destination } j \\ (4) x_{ij} &\geq 0 \text{ and } s_i \geq 0 \end{aligned}$$

The residual variables are not cost-free algebraic slack. They carry the economic consequence of leaving material at a particular source, so their coefficients must be retained in the objective.

2.2. Excess demand

When $\sum_i S_i < \sum_j D_j$, shortage variables u_j represent the unserved quantity at destination j . The corresponding model is:

$$\begin{aligned} (5) Z^{HD} &= \min \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} + \sum_{j=1}^n p_j u_j \\ (6) \sum_{j=1}^n x_{ij} &= S_i, \text{ for every source } i \\ (7) \sum_{i=1}^m x_{ij} + u_j &= D_j, \text{ for every destination } j \\ (8) x_{ij} &\geq 0 \text{ and } u_j \geq 0 \end{aligned}$$

The two formulations are one-sided. Under the assumptions of this study, an instance is classified by its aggregate imbalance and contains either residual supply or unmet demand, not both.

2.3. Exact dummy-node equivalence

For excess supply, define $Q^s = \sum_i S_i - \sum_j D_j$. Add a dummy destination d with demand Q^s , define the dummy flow $x_{id} = s_i$, and assign the dummy arc i, d the cost h_i . The source constraints become $\sum_j x_{ij} + x_{id} = S_i$. Summing these equations and using the destination constraints gives $\sum_i x_{id} = Q^s$. The augmented objective is the original transportation cost plus $\sum_i h_i s_i$, so the native and augmented models have the same feasible flows and objective values.

For excess demand, define $Q^D = \sum_j D_j - \sum_i S_i$. Add a dummy source d with supply Q^D , define the dummy flow $x_{dj} = u_j$, and assign the dummy arc d, j the cost p_j . The destination constraints become $\sum_i x_{ij} + x_{dj} = D_j$. Summing them and using the source constraints gives $\sum_j x_{dj} = Q^D$. Again, the augmented objective equals the native objective. These are exact equivalences; no approximation is introduced until the heterogeneous dummy costs are replaced by a common value.

2.4. Uniform-dummy approximation and true-cost regret

The uniform-dummy benchmark replaces the holding-cost vector with $\bar{h} = m^{-1} \sum_i h_i$ in the excess-supply case, or the shortage-penalty vector with $\bar{p} = n^{-1} \sum_j p_j$ in the excess-demand case. Let x^U denote the allocation obtained from that uniform-dummy problem and x^* the exact heterogeneous optimum. Define $Z^H(x)$ as the cost of allocation x evaluated with the original transportation costs and heterogeneous penalties. The relative true-cost regret is:

$$(9) G^U = 100 \times [Z^H(x^U) - Z^H(x^*)] / Z^H(x^*)$$

This measure quantifies deterioration in the original heterogeneous objective. It does not attempt to characterize every structural difference between two allocation matrices.

3. Materials and Methods

3.1. Study design and fully specified cases

The empirical design has two evidence layers. The first contains three fully specified cases for which cost matrices, supply vectors, demand vectors, and penalty vectors are available. These cases are auditable examples, not a statistically representative industrial benchmark. The second contains 120 synthetic instances generated from fixed seeds and stated distributions. They are used as controlled stress tests and are not presented as observations from any particular industry.

Table 1. Fully specified transportation instances and penalty characteristics

Instance	Dimensions	Total supply	Total demand	ρ_{imb}	Imbalance type	Penalty type
I1	3×4	900.0	750.0	16.667%	Excess supply	Source holding cost
I2	5×5	920.0	1,108.8	17.027%	Excess demand	Destination shortage penalty
I3	8×6	1,920.3	1,220.3	36.453%	Excess supply	Source holding cost

The imbalance ratio is defined as:

$$(10)\rho_{\text{imb}} = |\sum_i S_i - \sum_j D_j| / \max(\sum_i S_i, \sum_j D_j)$$

All numerical values are treated as continuous. Complete matrices and vectors are reproduced in Appendix A.

3.2. Synthetic benchmark construction

The synthetic suite contains 120 instances: 20 instances at each of six dimensions (5×5, 8×6, 10×10, 12×15, 15×12, and 20×20). Transportation costs were sampled as integers from 1 to 50. Supply and demand components were sampled from a uniform distribution on [100, 500] and rescaled so that the imbalance magnitude fell between 8% and 42% of the larger total. Excess-supply and excess-demand cases alternate within each dimension. Holding or shortage penalties were sampled from [5, 60]. Each instance was assigned a fixed generator seed and can be regenerated from benchmark_experiment.py.

The generator does not reproduce correlations, capacity structures, or operational distributions from a particular industry. The exact heterogeneous optimum was obtained from the augmented continuous LP for every synthetic instance, and all heuristic outputs were evaluated against that same objective. Three independent repetitions were used for each algorithm and synthetic instance; the supplied cases retained the more intensive 30-run protocol.

3.3. Exact and approximate benchmarks

The native heterogeneous LP was solved with HiGHS through SciPy's linprog interface using method="highs". A run was accepted only when the solver returned an optimal status, all decision variables were non-negative within 10^{-7} , and the maximum absolute equality residual did not exceed 10^{-7} . An independently assembled heterogeneous dummy-node model was solved as a structural verification, and its objective value was compared with the native formulation under the same tolerance.

The uniform-dummy problem was solved as a separate balanced transportation model. Its allocation was then reevaluated with the original h_i or p_j vector rather than the mean dummy penalty used during optimization. A least-cost greedy construction was included as a simple non-stochastic baseline: it repeatedly selected the feasible physical transportation arc with the lowest original cost and shipped the largest feasible quantity.

3.4. Balanced penalty-aware ACO

The revised ACO operates on the explicitly augmented network. For excess supply, one dummy destination is appended with source-specific holding costs and demand equal to the surplus. For excess demand, one dummy source is appended with destination-specific shortage penalties and supply equal to the shortage. A holding or shortage decision therefore competes with physical transportation arcs during construction, rather than being evaluated only after a cost-only allocation has been generated.

At construction step t , let F_t be the feasible augmented-arc set, defined by positive residual supply and positive residual demand. With $\varepsilon=10^{-9}$, the heuristic value is:

$$(11)\eta_{ij} = 1/(\tilde{c}_{ij} + \varepsilon)$$

where $\tilde{c}_{ij}$ is the augmented-network cost. For ant k , the probability of selecting arc i, j is:

$$(12)P_{ij}^{k,t} = \tau_{ij}^{\alpha}\eta_{ij}^{\beta} / \sum_{(r,s) \in F_t} \tau_{rs}^{\alpha}\eta_{rs}^{\beta}$$

The selected quantity is the minimum of the residual supply and residual demand on the chosen arc. After all feasible flows have been assigned, the augmented allocation is evaluated with the complete augmented cost matrix. Pheromone is evaporated with the pheromone parameter ρ_{ph} , followed by elitist deposits on arcs used by the iteration-best and global-best solutions. The notation ρ_{imb} for imbalance ratio and ρ_{ph} for pheromone evaporation is used throughout to avoid ambiguity between the imbalance ratio and pheromone evaporation.

For the three supplied cases, the settings were fixed before the final runs: 20 ants, 100 iterations, $\alpha = 1.0, \beta = 2.0$, and $\rho_{ph} = 0.35$. Each run therefore generated 2,000 ant constructions. Thirty independent seeds were used for each stochastic method and supplied case. The seeds were generated as $20260814 + 1000k + r$, where k is the one-based case index and $r = 0, \dots, 29$. For the 120-instance synthetic benchmark, the ACO budget was reduced to 5 ants and 15 iterations per run, or 75 ant constructions; three repetitions were performed for each instance and algorithm.

If A denotes the number of ants, I the number of iterations, and $m' \times n'$ the augmented network size, the reference implementation reports $O(AIm'n')$ when the feasible set is maintained incrementally. Rebuilding the feasible list at every shipment step gives the more conservative bound $O(AI(m'+n')m'n')$; memory use is $O(m'n')$. These are implementation-level bounds, not empirical scalability guarantees.

3.5. Comparison algorithms and statistical protocol

The primary stochastic method was the balanced penalty-aware ACO. Its ablation, ACO cost-only, used the same pheromone mechanism but calculated heuristic values from the physical transportation-cost matrix rather than the heterogeneous augmented matrix. Tabu Search started from the least-cost allocation and used feasible 2×2 transportation-cycle moves, a tabu tenure of seven iterations, 40 iterations, and 10 candidate moves per iteration. Simulated Annealing used the same move family, a geometric cooling factor of 0.985, and 60 iterations. VNS used neighborhood sizes one, two, and three, followed by local descent within each outer iteration. VAM was implemented as a deterministic Vogel-style initialization using the complete augmented cost matrix. All methods were evaluated with the heterogeneous augmented objective.

For the 120 synthetic instances, the primary comparison used the mean gap from the exact LP over three repetitions, with each instance treated as a block in the Friedman test. A Nemenyi post-hoc comparison was applied to mean ranks. This design avoids treating repeated runs on one instance as independent transportation problems. A separate exploratory one-factor-at-a-time ANOVA varied one ACO factor at a time across 12 benchmark instances with three repetitions per level: $\alpha \in 0.5, 1.0, 2.0$, $\beta \in 1.0, 2.0, 4.0$, $\rho_{ph} \in 0.2, 0.35, 0.5$, and ant count $\in 3, 5, 8$. Because the factors were not crossed in a full factorial design, the ANOVA is not interpreted as an interaction model or a definitive parameter-importance study.

4. Results

4.1. Exact equivalence and benchmark values

The native heterogeneous LP and the independently assembled heterogeneous dummy model returned identical objective values within the declared numerical tolerance. The maximum equality residual across the validation runs was below 2.6×10^{-13} . The agreement is a structural check: the heterogeneous dummy representation is exact. The approximation appears only when the heterogeneous penalty vector is replaced by its mean.

Table 2. Exact equivalence and benchmark objective values

Instance	Native heterogeneous LP	Heterogeneous dummy LP	Uniform-dummy objective	Uniform-dummy true cost	Greedy true cost
I1	5,495.00	5,495.00	5,795.00	5,855.00	5,855.00
I2	8,559.88	8,559.88	9,882.03	10,228.56	10,319.76
I3	11,469.60	11,469.60	11,856.90	13,931.31	13,931.31

4.2. ACO performance and ablation results.

Table 3. ACO performance and ablation results

Instance	Exact optimum	Uniform-dummy true-cost gap	ACO cost-only mean	Cost-only gap	Penalty-aware ACO mean	Penalty-aware gap	SD	Exact-optimum runs
I1	5,495.00	6.55%	5,495.00	0.00%	5,495.00	0.00%	0.00	30/30
I2	8,559.88	19.49%	8,595.02	0.41%	8,559.88	0.00%	≈ 0.00	30/30
I3	11,469.60	21.46%	11,719.19	2.18%	11,606.12	1.19%	52.97	0/30

Penalty-aware ACO recovered the exact optimum in all runs for I1 and I2. In I3, its mean gap was 1.19%; the best solution was 11,504.46 and the worst was 11,669.02. Every run remained below the true cost of the uniform-dummy allocation. Relative to cost-only ACO, the penalty-aware version reduced the mean objective by 0.41% in I2 and 0.96% in I3. These differences are informative but should not be generalized beyond the tested cases.

The baseline comparison clarifies the source of the effect. The uniform-dummy allocation is exact for its own simplified objective, yet its true heterogeneous cost is 6.55–21.46% above the heterogeneous optimum. The least-cost rule performs similarly to the uniform-dummy allocation in I1 and I3 and is slightly worse in I2. Thus, the principal modeling effect is visible even before a particular metaheuristic is selected.

4.3. Exploratory statistical results for supplied cases

All 30 penalty-aware ACO runs were below the corresponding uniform-dummy true cost in every supplied case. One-sided Wilcoxon signed-rank p-values were 2.16×10^{-8} for I1 and I2 and 8.08×10^{-7} for I3. These values describe consistency of repeated stochastic runs for fixed test cases. They are not population-level tests over transportation networks because the underlying instance is held fixed and the comparator is deterministic.

For I1 and I2, the standard deviation was effectively zero because every run reached the exact optimum. In I3, the sample standard deviation was 52.97 and the coefficient of variation was approximately 0.0046. The lower variability is consistent with placing dummy-arc penalties inside the construction process, although the result remains confounded with instance structure and search budget.

4.4. Sensitivity to the heuristic exponent

A local sensitivity check was performed on I3 with the same 30 random seeds for $\beta = 1.0$, $\beta = 2.0$, and $\beta = 4.0$; all other settings were unchanged. Mean gaps from the exact optimum were

1.16%, 1.19%, and 1.56%, respectively. The results occupy a relatively narrow range, but this experiment is not a parameter-tuning study and cannot establish global robustness.

Table 4. Sensitivity analysis of heuristic exponent β

β	Mean objective	Median	Sample SD	Best	Worst	Gap from exact
1.0	11,602.30	11,613.77	69.87	11,469.60	11,757.76	1.16%
2.0	11,606.12	11,624.47	52.97	11,504.46	11,669.02	1.19%
4.0	11,647.97	11,669.02	47.47	11,504.46	11,757.76	1.56%

The fixed value $\beta = 2.0$ was retained because it was specified before the main comparison and lies between weak and strong exploitation of low-cost augmented arcs. The sensitivity results support reporting a range rather than presenting one parameter setting as universally optimal.

4.5. Computational burden

The two ACO variants used the same 2,000 ant constructions per supplied-case run. Wall-clock times were recorded on the execution environment used for this study; they should not be interpreted as hardware-independent performance guarantees.

Table 5. Computational burden comparison

Instance	ACO cost-only mean time (s)	Penalty-aware ACO mean time (s)
I1	0.460	0.563
I2	0.762	1.018
I3	1.104	1.808

The penalty-aware construction required more time because it explicitly adds and evaluates dummy arcs. The increase was modest in I1 and approximately 64% in I3. This is the computational counterpart of representing heterogeneous penalties during construction rather than applying them only after a cost-only allocation has been generated. A proper scalability study would require larger networks, controlled hardware, repeated timing protocols, and explicit LP solve times.

4.6. Expanded benchmark results

The expanded suite contained 2,160 instance–method repetitions: 120 instances $\times$ six methods $\times$ three repetitions. VNS achieved the lowest mean gap from the exact LP at 12.98%, followed by VAM at 17.39%, penalty-aware ACO at 20.45%, Tabu Search at 21.94%, cost-only ACO at 25.14%, and Simulated Annealing at 27.16%. Penalty-aware construction reduced the mean gap relative to cost-only ACO by 4.69 percentage points. These values describe performance under the stated method-specific budgets, not universal dominance.

Table 6. Expanded benchmark results across algorithms.

Algorithm	Mean gap (%)	Median instance gap (%)	Mean rank
VNS	12.98	8.48	1.483
VAM	17.39	12.65	2.933
Penalty-aware ACO	20.45	19.34	3.463
Tabu Search	21.94	17.87	3.504
Cost-only ACO	25.14	23.34	4.733
Simulated Annealing	27.16	21.72	4.883

For the 20×20 cases, mean gaps were 44.57% for penalty-aware ACO, 51.22% for cost-only ACO, 51.44% for Tabu Search, 50.90% for Simulated Annealing, 35.09% for VNS, and 37.30% for VAM. The deterioration with dimension is important: under the short budgets used here, the heuristics remain far from the exact LP on the largest synthetic networks.

4.7. Statistical comparison and exploratory factor analysis

The Friedman test rejected the null hypothesis that the six methods had identical performance ranks across the 120 instances $\chi^2 = 270.25, p = 2.47 \times 10^{-56}$. The Nemenyi critical difference at $\alpha = 0.05$ was 0.688 mean-rank units. Penalty-aware ACO ranked significantly better than cost-only ACO $\text{rankdifference} = 1.271, p = 2.13 \times 10^{-6}$ and Simulated Annealing $\text{difference} = 1.421, p = 6.04 \times 10^{-8}$. Its rank did not differ significantly from Tabu Search $p \approx 1.00$ or VAM ($p = 0.242$), whereas VNS ranked significantly better than penalty-aware ACO $\text{difference} = 1.979, p < 10^{-14}$. The evidence therefore supports a relative advantage over the cost-only ablation, not a claim that penalty-aware ACO is the best method.

The exploratory marginal ANOVA detected no statistically significant effect for α ($F = 0.163, p = 0.850$), β ($F = 1.048, p = 0.354$), ρ_{ph} ($F = 0.085, p = 0.918$), or ant count $F = 0.773, p = 0.464$ under the restricted one-factor-at-a-time design. These p-values do not show that the parameters are irrelevant. The design does not estimate interactions, uses only 12 instances, and applies a short ACO budget; it indicates only that no strong marginal signal appeared under the tested conditions.

4.8. Budget-aligned robustness analysis

A secondary comparison reduced the concern that the primary benchmark used different nominal budgets. Five stochastic methods were assigned a nominal budget of 75 outer-loop evaluations or candidate moves per run: 5 ants $\times$ 15 outer iterations for each ACO variant, 15 iterations $\times$ 5 candidates for Tabu Search, 75 iterations for Simulated Annealing, and 25 VNS outer iterations. At each VNS outer iteration, the algorithm cycles through the three stated neighborhood sizes and performs the prescribed internal local descent; therefore, 25 counts only the outer-loop steps and does not count the internal descent moves. VAM was excluded because it is a deterministic constructive procedure rather than repeated search. Three repetitions were retained for each of the 120 instances.

Table 7. Budget-aligned robustness comparison

Algorithm	Mean gap (%)	Median gap (%)	Mean rank
VNS	14.34	9.85	1.217
Penalty-aware ACO	20.07	18.14	2.471
Cost-only ACO	25.33	22.92	3.529
Simulated Annealing	26.66	22.03	3.671
Tabu Search	27.80	22.70	4.113

The equal-budget Friedman test remained significant $\chi^2 = 262.95, p = 1.05 \times 10^{-55}$, with a Nemenyi critical difference of 0.557. In this comparison, penalty-aware ACO ranked significantly better than cost-only ACO $p = 2.15 \times 10^{-6}$, Simulated Annealing $p = 4.13 \times 10^{-8}$, and Tabu Search $p = 8.03 \times 10^{-9}$, but significantly worse than VNS $p < 10^{-14}$. This direction is consistent with the mean ranks and removes the earlier ambiguity in the interpretation. The result is a budget-aligned robustness check, not a universal computational-efficiency claim, because a transportation-cycle move and an ACO construction evaluation do not have identical implementation costs.

4.9. Visual comparison

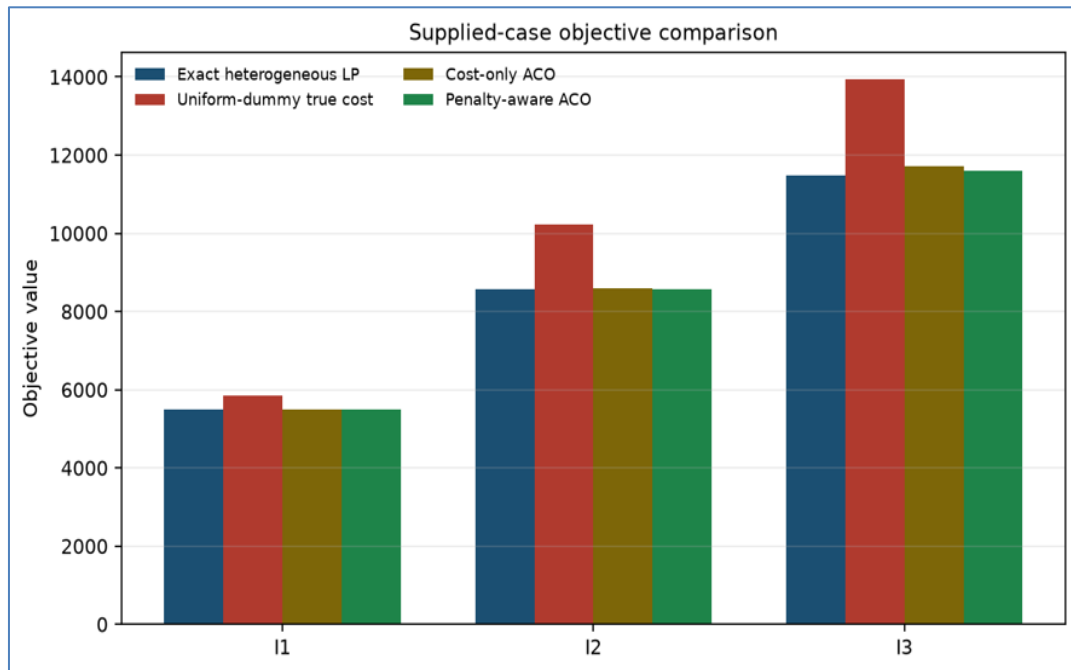


Figure 1. Objective values for the exact heterogeneous LP, the uniform-dummy allocation reevaluated with true penalties, cost-only ACO, and penalty-aware ACO. Source: recalculated from the summary values reported in the supplied manuscript.

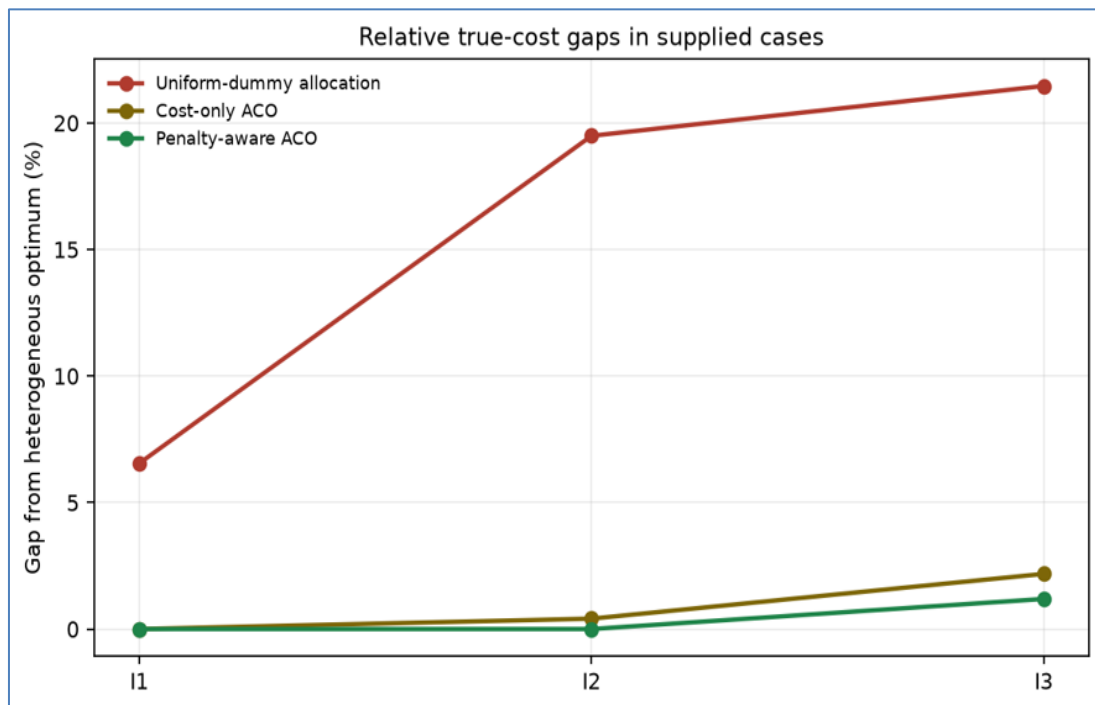


Figure 2. Relative true-cost gaps from the heterogeneous LP optimum. Source: recalculated from the summary values reported in the supplied manuscript.

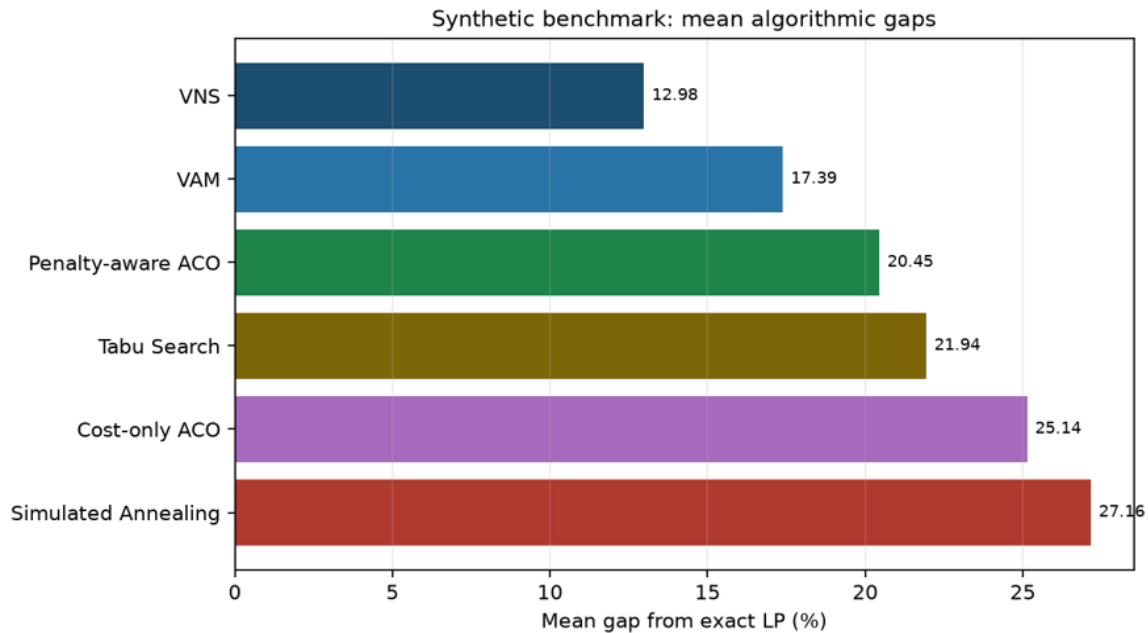


Figure 3. Mean algorithmic gaps across the 120 synthetic instances; lower values are better.
Source: recalculated from the summary values reported in the supplied manuscript.

5. Discussion

The revised experiments support two related conclusions. First, a uniform dummy penalty can be materially misleading when the economic consequence of residual supply or unmet demand varies by location. The result is visible even though the uniform-dummy problem is solved exactly: exactness relative to a simplified objective cannot recover information removed by averaging the penalties. Second, algorithmic performance depends on the search procedure and budget. On the expanded synthetic suite, VNS had the best mean rank, while penalty-aware ACO improved significantly over its cost-only ablation and was not statistically distinguishable from Tabu Search or VAM in the primary Nemenyi comparison.

The ACO result warrants restraint. On the three supplied cases, penalty awareness reduced the I3 mean gap from 2.18% to 1.19%; on the synthetic suite, the mean gap was 20.45% compared with 25.14% for cost-only ACO. These differences support placing heterogeneous penalties inside the construction network, but they do not establish ACO as the best method. VNS performed better under the selected budgets, and method-specific budgets were not equal in wall-clock time or neighborhood evaluations.

A theoretical explanation is consistent with the observed ranking. VNS changes the neighborhood structure systematically, allowing the search to escape a locally attractive transportation-cycle pattern and then intensify the search around the improved solution. In the present continuous transportation model, these neighborhoods can modify several flows while preserving feasibility, so the local-descent phase can exploit cost exchanges that a short constructive ACO run may not revisit. This diversification–intensification logic is consistent with the broader VNS logistics literature [25]. By contrast, penalty-aware ACO improves the initial arc-selection landscape by exposing heterogeneous dummy costs, but its performance remains sensitive to pheromone concentration, ant count, and iteration budget. The superiority of VNS should therefore be interpreted as a result of neighborhood design under the stated protocols, not as a general theoretical dominance claim.

For the continuous model studied here, HiGHS remains the preferred reference whenever the exact LP can be solved within the available planning time. Heuristics become relevant when additional model structure is introduced, exact integration is inconvenient, or a validated time–

quality trade-off is acceptable. In such settings, the benchmark favors reporting several algorithms rather than presenting one heuristic as a universal solution. The expanded experiments also show why penalty dispersion and neighborhood design deserve separate study. Mean gaps rise with network dimension, but the generator does not reproduce industry-specific correlations, capacity structures, or operating distributions. The benchmark is consequently useful for algorithmic stress testing, not for estimating field performance.

6. Limitations, Threats to Validity, and Reproducibility

Several limitations remain material. The three supplied cases are fully specified but cannot represent an industrial benchmark. The 120 additional cases are synthetic; they expand algorithmic coverage but do not substitute for field data or published benchmark libraries. The primary comparison uses method-specific budgets—5 ants and 15 iterations for ACO, 40 or 60 local-search iterations for other methods, and deterministic VAM—so its Friedman ranks describe the selected protocols rather than equalized computational power. A secondary 75-evaluation/move comparison reduces this concern for five stochastic methods, but it cannot make different move operators computationally identical. The ANOVA is marginal and exploratory rather than a full factorial analysis with interaction terms. Finally, the model remains continuous and excludes integer decisions, fixed charges, multiple commodities, time periods, uncertainty, and additional capacity or service-level constraints.

The supplied description states that a reproducibility package contains the cases, synthetic generator, benchmark seeds, implementations, raw outputs, validation records, environment information, statistical summaries, figures, README instructions, and requirements. Those artifacts were not included as attachments in the present editing task. Numerical reproduction should be assessed within the stated tolerance rather than by requiring bitwise identity across platforms.

6.1. Threats to validity

Table 8. Threats to validity and mitigation measures.

Validity dimension	Main threat	Mitigation in this study
Construct validity	A uniform dummy penalty may be a bookkeeping convention rather than a real economic cost in some applications.	Holding and shortage penalties are defined explicitly, and uniform allocations are reevaluated under the original heterogeneous objective.
Internal validity	Algorithm performance can depend on seeds, budgets, and neighborhood implementation.	Fixed seeds, repeated runs, raw-output requirements, an equal-budget comparison, and explicit settings are reported.
External validity	Synthetic distributions and three supplied cases cannot represent all industrial transportation networks.	Synthetic cases are labeled as stress tests, and conclusions are restricted to the tested model and distributions.
Statistical conclusion validity	Repeated runs on one instance are not independent problem instances; the marginal ANOVA does not estimate interactions.	Friedman/Nemenyi use instances as blocks, and factor analysis is explicitly exploratory.
Reproducibility validity	Platform, package, and solver versions can affect floating-point results and runtime.	Tolerances, environment metadata, required scripts, and output filenames are specified; the persistent repository link remains to be supplied.

7. Conclusions

This study examined the effect of replacing source- and destination-specific residual penalties with a uniform dummy cost in unbalanced transportation models. The native formulations and independently assembled heterogeneous dummy-node equivalents returned identical objective values on all three fully specified cases.

When heterogeneous penalties were replaced by their arithmetic mean, the supplied cases incurred true-cost regrets of 6.55%, 19.49%, and 21.46%. Penalty-aware ACO reached the exact optimum in the 3×4 and 5×5 cases and produced a 1.19% mean gap in the 8×6 case. Across the 120 synthetic instances, penalty-aware ACO had a 20.45% mean gap and ranked ahead of cost-only ACO at the 0.05 level in the Nemenyi comparison, but VNS achieved the strongest overall rank.

The defensible practical conclusion is narrow. Location-specific residual penalties should be represented explicitly whenever they carry distinct economic meanings. ACO is a useful heuristic complement when additional model structure or a time–quality trade-off justifies it, but it should be compared with alternative neighborhoods and constructive methods. The study does not establish universal heuristic dominance, industrial validity, or computational intractability for the continuous model.

8. Future Work

Future work should move from synthetic stress testing to benchmark libraries and operational data with documented cost, capacity, and penalty distributions. The model should be extended to integer decisions, fixed charges, multiple commodities, stochastic demand, time-dependent costs, and service-level constraints. A full factorial or sequential design could estimate interactions among α , β , ρ_{ph} , ant count, and neighborhood parameters. Hybrid ACO–VNS and exact–heuristic matheuristics are promising directions, especially for larger instances where the current construction leaves a substantial gap from the LP optimum. Controlled wall-clock studies and external validation on real transportation systems are needed before industrial recommendations are made.

Appendix A. Complete Reproducible Data

Instance I1: 3×4, excess supply

Source	D1	D2	D3	D4	Supply	Holding cost
S1	8	14	6	11	250.0	4.5
S2	10	17	9	5	300.0	12.0
S3	3	13	4	7	350.0	7.8
Demand	200.0	150.0	180.0	220.0	—	—

Instance I2: 5×5, excess demand

Source	D1	D2	D3	D4	D5	Supply
S1	6	7	11	5	14	180.0
S2	9	10	8	17	6	220.0
S3	4	13	15	9	11	150.0
S4	7	5	12	16	8	200.0
S5	18	3	10	14	19	170.0
Demand	224.0	201.6	179.2	235.2	268.8	—

Shortage penalties by destination: $p = [15.2, 32.1, 21.4, 45.6, 18.9]$.

Instance I3: 8×6, excess supply

Source	D1	D2	D3	D4	D5	D6	Supply	Holding cost
S1	5	12	8	19	3	15	215.3	3.2
S2	11	6	22	7	14	9	182.7	8.5
S3	8	18	4	16	20	11	301.5	12.1
S4	13	9	17	5	8	24	248.9	5.4
S5	6	21	10	14	12	7	165.4	9.8
S6	19	8	15	23	6	18	332.1	14.2
S7	4	16	9	11	25	13	198.6	6.7
S8	10	7	20	8	17	5	275.8	11.3
Demand	152.4	218.7	185.3	241.6	195.8	226.5	—	—

Appendix B. Computational Environment and Output Files

Component	Version or setting	Role
Python	3.12.3	Interpreter
NumPy	2.5.1	Array operations and random-number generation
SciPy	1.18.0	linprog and Wilcoxon signed-rank test
HiGHS	SciPy-integrated backend, method="highs"	Exact continuous LP
Operating system	Linux x86_64, glibc 2.39	Execution environment
Solver tolerance	1×10^{-7}	Equality and non-negativity checks
Supplied-case ACO ants	20	Stochastic construction
Supplied-case ACO iterations	100	Per run
Supplied-case ACO parameters	$\alpha=1.0, \beta=2.0, \rho_{ph}=0.35$	Pheromone and heuristic weighting
Supplied-case runs per method and instance	30	Repeated stochastic evaluation
Supplied-case ant constructions per run	2,000	20 ants $\times$ 100 iterations
Synthetic benchmark instances	120	Six dimensions $\times$ 20 instances
Synthetic repetitions	3	Per algorithm and instance
Synthetic ACO budget	5 ants $\times$ 15 iterations	75 ant constructions per run
Tabu Search budget	40 iterations $\times$ 10 candidates	Feasible 2 $\times$ 2 cycle moves
Simulated Annealing budget	60 iterations	Cooling factor 0.985
Primary VNS budget	30 outer iterations	Neighborhood sizes 1, 2, and 3 with internal local descent
Equal-budget VNS budget	25 outer iterations	75 is the nominal outer-loop count; internal local descent is additional

Appendix C. Algorithmic Pseudocode**Algorithm 1. Balanced penalty-aware ACO**

- Construct the augmented matrix by adding a dummy destination with source-specific holding costs or a dummy source with destination-specific shortage penalties.

- Initialize pheromone values to one and set the prescribed random seed.
- For each iteration, construct one feasible augmented allocation for each ant using pheromone weights and the inverse augmented-arc cost.
- Ship the minimum residual quantity on the selected feasible arc and update the source and destination residuals.
- Record the iteration-best and global-best solutions.
- Evaporate pheromone and apply elitist deposits to the iteration-best and global-best arcs.
- After the final iteration, return the global-best allocation and evaluate it with the complete augmented cost matrix.

References

- [1] SciPy. `scipy.optimize.linprog` documentation. Available at: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.linprog.html>
- [2] SciPy. `linprog(method="highs")` documentation. Available at: <https://docs.scipy.org/doc/scipy/reference/optimize.linprog-highs.html>
- [3] HiGHS. High-performance linear optimization software and citation guidance. Available at: <https://highs.dev/>
- [4] Rardin, R. L., and Uzsoy, R. "Experimental evaluation of heuristic optimization algorithms: A tutorial." *Journal of Heuristics*, 7(3), 261–304, 2001. <https://doi.org/10.1023/A:1011319115230>
- [5] Derrac, J., García, S., Molina, D., and Herrera, F. "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms." *Swarm and Evolutionary Computation*, 1(1), 3–18, 2011. <https://doi.org/10.1016/j.swevo.2011.02.002>
- [6] Wilcoxon, F. "Individual comparisons by ranking methods." *Biometrics Bulletin*, 1(6), 80–83, 1945. <https://doi.org/10.2307/3001968>
- [7] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.
- [8] Dorigo, M. *Optimization, Learning and Natural Algorithms*. PhD thesis, Politecnico di Milano, 1992.
- [9] Dorigo, M., and Stützle, T. *Ant Colony Optimization*. MIT Press, 2004.
- [10] Blum, C., and Roli, A. "Metaheuristics in combinatorial optimization: Overview and conceptual comparison." *ACM Computing Surveys*, 35(3), 268–308, 2003. <https://doi.org/10.1145/959213.959214>
- [11] Wolpert, D. H., and Macready, W. G. "No free lunch theorems for optimization." *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82, 1997. <https://doi.org/10.1109/4235.585893>
- [12] Shivani, S., Rani, D., and Ebrahimnejad, A. "An approach to solve an unbalanced fully rough multi-objective fixed-charge transportation problem." *Computational and Applied Mathematics*, 41, 2022. <https://doi.org/10.1007/s40314-022-01830-5>
- [13] Beier, F., von Lindheim, J., Neumayer, S., and Steidl, G. "Unbalanced multi-marginal optimal transport." *Journal of Mathematical Imaging and Vision*, 65(3), 394–413, 2023. <https://doi.org/10.1007/s10851-022-01126-7>
- [14] Lakshmanan, R., and Pichler, A. "Unbalanced optimal transport and maximum mean discrepancies: Interconnections and rapid evaluation." *Journal of Scientific Computing*, 100(3), 2024. <https://doi.org/10.1007/s10915-024-02586-2>
- [15] Das, A., Rai, R., Sasmal, B., Dhal, K. G., Khurma, R. A., and Saha, R. "Metaheuristic algorithms since 2020: Development, taxonomy, analysis, and applications." *Archives of Computational Methods in Engineering*, 2025. <https://doi.org/10.1007/s11831-025-10408-3>

- [16] Ekanayake, E. M. U. S. B., Perera, S. P. C., Daundasekara, W. B., and Juman, Z. A. M. S. "A Modified Ant Colony Optimization Algorithm for Solving a Transportation Problem." *Journal of Advances in Mathematics and Computer Science*, 35(5), 83–101, 2020. <https://doi.org/10.9734/jamcs/2020/v35i530284>
- [17] Musa, R., Arnaout, J.-P., and Jung, H. "Ant colony optimization algorithm to solve for the transportation problem of cross-docking network." *Computers & Industrial Engineering*, 59(1), 85–92, 2010. <https://doi.org/10.1016/j.cie.2010.03.002>
- [18] Jafarian-Moghaddam, A. R. "Economical speed for optimizing the travel time and energy consumption in train scheduling using a fuzzy multi-objective model." *Urban Rail Transit*, 7(3), 191–208, 2021. <https://doi.org/10.1007/s40864-021-00151-w>
- [19] Zhang, H., Lu, G., Lei, Y., Zhang, G., and Niyitanga, I. "A hybrid framework for synchronized passenger and train traffic simulation in an urban rail transit network." *International Journal of Rail Transportation*, 11(6), 912–941, 2023. <https://doi.org/10.1080/23248378.2022.2109522>
- [20] Chau, M. L. Y., and Gkiotsalitis, K. "A systematic literature review on the use of metaheuristics for the optimisation of multimodal transportation." *Evolutionary Intelligence*, 18, Article 36, 2025. <https://doi.org/10.1007/s12065-025-01020-2>
- [21] Kruszyna, M., and Makuch, J. "Mobility nodes as an extension of the idea of transfer nodes—Solutions for smaller rail stations with an example from Poland." *Sustainability*, 15(3), 2106, 2023. <https://doi.org/10.3390/su15032106>
- [22] Korzeń, M., and Gisterek, I. "Applying ant colony optimization to reduce tram journey times." *Sensors*, 24(19), 6226, 2024. <https://doi.org/10.3390/s24196226>
- [23] Korzeń, M., and Kruszyna, M. "Public transport planning using modified ant colony optimization." *Sustainability*, 17(6), 2468, 2025. <https://doi.org/10.3390/su17062468>
- [24] Ma, X., and Liu, C. "Improved ant colony algorithm for the split delivery vehicle routing problem." *Applied Sciences*, 14(12), 5090, 2024. <https://doi.org/10.3390/app14125090>
- [25] Armas, J. de, and Moreno Pérez, J. A. "A Survey on Variable Neighborhood Search for Sustainable Logistics." *Algorithms*, 18(1), 38, 2025. <https://doi.org/10.3390/a18010038>
- [26] Shao, Z., Zhang, L., Han, C., and Meng, L. "Measurement and prediction of urban land traffic accessibility and economic contact based on GIS: A case study of land transportation in Shandong Province, China." *International Journal of Environmental Research and Public Health*, 19(22), 14867, 2022. <https://doi.org/10.3390/ijerph192214867>