Scientific Research Publishing

Journal of Environmental Protection, 2025, 16(9), 916-947
https://www.scirp.org/journal/jep
ISSN Online: 2152-2219
ISSN Print: 2152-2197

Optimizing Waste Facility Locations Using a Flexible Model that Balances Cost, Environmental Impact, and Community Needs

Emmanuel Aidoo[1,2], Coimbatore Dorairaj Nandakumar[1*], Gabriel Mwinkume[1,2], Theodore Oduro-Okyireh[2]

[1]Department of Mathematics & Actuarial Science, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

[2]Department of Mathematics and Statistics, Cape Coast Technical University, Cape Coast, Ghana

Email: prof.cdnandakumar@gmail.com

How to cite this paper: Aidoo, E., Nandakumar, C.D., Mwinkume, G. and Oduro-Okyireh, T. (2025) Optimizing Waste Facility Locations Using a Flexible Model that Balances Cost, Environmental Impact, and Community Needs. *Journal of Environmental Protection*, **16**, 916-947.
https://doi.org/10.4236/jep.2025.169049

Received: July 14, 2025
Accepted: September 25, 2025
Published: September 28, 2025

Abstract

This study presents the Dynamic Multi-Objective Uncapacitated Facility Location Problem (DMUFLP) model, a novel and forward-thinking approach designed to enhance facility location decisions in complex and evolving supply chain systems. Unlike traditional models such as the Uncapacitated Facility Location Problem (UFLP), which often prioritize economic efficiency while overlooking environmental and social impacts, the DMUFLP offers a holistic and adaptive framework. It integrates economic, environmental, and social cost dimensions, supports dynamic capacity adjustments, and incorporates stochastic elements to effectively manage uncertainty and fluctuating demand over time. To evaluate the model's effectiveness, a comprehensive simulation using synthetic data was conducted, benchmarking the DMUFLP against conventional models across various sample sizes. Results consistently demonstrate the DMUFLP's superior performance in critical metrics, including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Absolute Error (MAE), and Area Under the Curve (AUC). Importantly, the model achieves a robust balance between predictive accuracy and model simplicity, offering decision-makers a practical, reliable, and sustainability-focused tool for long-term facility planning. By bridging the gap between economic viability, environmental stewardship, and social responsibility, this study positions the DMUFLP model as a next-generation solution capable of reshaping facility location optimization in real-world applications.

Keywords

Sustainability, Uncapacitated Facility Location Problem, Facility Placement Optimization, Dynamic Capacity Adjustments

1. Introduction

The Uncapacitated Facility Location Problem (UFLP) is a widely recognised model in operations research, commonly used to optimise the placement of facilities and reduce costs. Since its establishment, the UFLP has played a crucial role in a wide range of applications, spanning from supply chain management to service network design. Nevertheless, the conventional UFLP has faced criticism for its oversimplification of the intricate realities of the real world. These include factors like the environmental and social consequences of facility placement, the need for capacity adjustments over time, and the inherent uncertainties in demand [1]. Various UFLP extensions have been developed to address the limitations and incorporate broader considerations such as sustainability, resilience, and stochastic demand variations [2].

In spite of these advancements, numerous current models continue to face challenges in offering a comprehensive and adaptable framework for facility location planning, especially in environments marked by swift market changes and fluctuating demand patterns. As an illustration, many conventional models fail to consider the changing capacity requirements or the long-term environmental effects of facility operations [3]. This gap highlights the necessity for advanced models capable of tackling the complex nature of contemporary facility location challenges.

In order to address these gaps, this research presents the Dynamic Multi-Objective Uncapacitated Facility Location Problem (DMUFLP) model. The DMUFLP is designed to seamlessly integrate various cost components, including economic, environmental, and social factors, into a unified framework. The DMUFLP stands out from traditional models by incorporating dynamic capacity adjustments, enabling it to seamlessly adapt to changing operational needs. This approach to facility location planning takes into account stochastic elements, which help to capture the inherent uncertainty in demand forecasts better. As a result, it provides a more realistic and professional perspective on the matter [4].

This research conducts a thorough simulation study to evaluate the performance of the DMUFLP model in comparison to other existing models. The study highlights the DMUFLP's exceptional optimization capabilities in various scenarios, especially in achieving a balance between cost efficiency and sustainability objectives. The findings indicate that the DMUFLP model has the potential to greatly improve decision-making in complex and uncertain environments, making it a reliable solution for modern facility location challenges. These results add to the increasing amount of research supporting the need for more sophisticated and flexible methods in facility location planning [5] [6].

2. Literature Review

The Uncapacitated Facility Location Problem (UFLP) is a fundamental model in operational research that focuses on finding the most cost-effective facility locations. It takes into account both setup and transportation expenses [7]. Despite its widespread adoption, the traditional UFLP has been subject to criticism due to its limitations in addressing the intricacies of real-world challenges. These complexities include environmental and social impacts, capacity flexibility, and demand uncertainties [8] [9].

Pires *et al.* (2019) expanded upon the UFLP by integrating sustainability factors with a specific focus on waste management. Their model highlighted the additional costs associated with sustainability, such as penalties for emissions, demonstrating the increasing importance of environmental factors in facility location problems [8]. Nevertheless, this model made significant strides in incorporating environmental factors, although its focus was primarily on waste management and needed to fully encompass broader environmental and social impacts.

Olapiriyakul *et al.* (2019) made significant advancements to the UFLP by incorporating environmental and social impact measures in addition to the usual economic objectives. This extension represented a notable transition towards a comprehensive approach to facility location, recognising the wider impact on communities and the environment [9]. In spite of these advancements, the model needed help in handling dynamic capacity adjustments, particularly when it came to adapting to changing demand scenarios.

Adeleke and Olukanni (2020) presented a revision to the UFLP that highlighted the significance of precise demand forecasting in reducing the expenses linked to incorrect facility placement. This model emphasised the importance of accurate demand estimates for efficiently optimising facility locations. Nevertheless, similar to earlier models, Adeleke's approach faced limitations due to its static nature [10]. It is assumed that facility placements would continue to be optimal, even in the face of potential changes in demand [11].

Green (2022) highlighted the importance of integrating sustainability into operational models by emphasising the consideration of carbon emissions in facility location problems. Green's model showcased a significant breakthrough, especially in closed-loop distribution networks where sustainability plays a vital role. Nevertheless, the report primarily emphasised carbon emissions and failed to adequately consider other environmental or social impacts [12].

In their study, Pang *et al.* (2023) tackled the challenges of online retailing by presenting a UFLP model that effectively dealt with the uncertainties related to demand and cost by incorporating stochastic elements. This model has demonstrated its relevance in the context of e-commerce, where demand patterns can be extremely unpredictable. Although Pang *et al.*'s model successfully addressed demand uncertainties, it needed to fully incorporate capacity adjustments or consider environmental and social factors [13].

In their study, Ghadge *et al.* (2023) put forward a model that highlights the importance of sustainable facility location in closed-loop distribution networks, with a specific focus on minimising carbon emissions. Although this model made progress in addressing sustainability, it focused mainly on carbon emissions and overlooked other important environmental and social impacts, as well as the everchanging demand and capacity [14].

The latest advancement is the introduction of the Dynamic Multi-objective Uncapacitated Facility Location Problem (DMUFLP) model in 2024. This model incorporates improvements from previous versions by addressing their limitations and introducing innovative features. The DMUFLP model takes into account environmental and social impacts and is designed to adapt to changes in demand with flexible capacity adjustments [15].

The DMUFLP model also considers a wider range of uncertainties, including those associated with environmental and social impacts, building upon the stochastic elements introduced by Pang *et al.* (2023). This thorough approach improves the model's ability to handle the complexities of real-world situations that previous models had difficulty addressing.

Ultimately, past models have made notable advancements in expanding the UFLP to encompass sustainability, social impacts, and stochastic elements. However, their focus has been primarily on individual aspects rather than fully integrating these components into a cohesive and all-encompassing framework. The DMUFLP model is a remarkable achievement, as it brings together the best features of previous models to create a comprehensive approach that takes into account the complete range of environmental, social, and economic impacts. It also incorporates dynamic capacity adjustments and stochastic elements, making it a highly sophisticated tool. This model provides a comprehensive and flexible solution to the facility location problem in challenging and unpredictable environments.

2.1. Method

2.1.1. The Dynamic Multi-Objective Facility Location Problem (DMUFLP)

We developed the Dynamic Multi-Objective Facility Location Problem (DMUFLP) to address the complexities of facility location decisions in diverse supply chain environments. This model incorporates various cost components and sustainability considerations. We formulated the DMUFLP using a mixed-integer linear programming (MILP) approach, known for efficiently handling large-scale optimization problems with multiple objectives [16] [17].

The model aims to minimise total project costs, including facility setup, transportation, environmental impact, social impact, and capacity adjustment expenses.

2.1.2. Data Generation and Assumptions

We conducted a simulation study using the R software, a widely used environment for statistical computing, to evaluate the performance of the DMUFLP model rigorously. We based the simulation on a synthetic dataset representing a typical supply chain scenario. To ensure reproducibility, we generated demand values using a normal distribution. We assigned transportation costs using a uniform distribution between 10 and 100, and we derived environmental and social impact costs from exponential distributions with appropriate scale parameters that reflect their relative impact [18]. We modelled capacity adjustment costs with a triangular distribution to simulate increasing marginal costs. We based these assumptions on a combination of historical data and insights from relevant literature [19] [20].

2.1.3. Social Impact Costs and Community Needs

The model integrates a social impact cost component (si), which reflects community-related factors such as proximity to residential zones, public approval ratings, and local job creation potential. These factors represent the societal acceptability and benefits of facility placement in specific areas, aligning the model with contemporary sustainability practices in supply chain management [21] [22].

2.1.4. Model Structure and Constraints

We carefully constructed the DMUFLP model's constraints to ensure demand satisfaction, facility activation, and compliance with real-world operational conditions [18]. We estimated key parameters, including demand, transport costs, environmental and social costs, and capacity adjustments, using empirical data and theoretical frameworks to ensure model validity [23].

We implemented the model using R's lpSolve and ompr packages, which provide practical tools for solving linear and mixed-integer programming problems [24]. We tested the model using various sample sizes (250, 500, 750, and 950) to assess its scalability and computational efficiency. We also benchmarked its performance against other models that integrate multiple cost components and sustainability concerns, providing a comprehensive comparison [25].

2.1.5. Performance Metrics and Diagnostic Evaluation

To assess the model's accuracy and reliability, we applied several performance metrics, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Absolute Error (MAE), and Area Under the Curve (AUC) [26]. Using R, we generated diagnostic plots including Residuals vs. Fitted Values, Normal Q-Q Plot, Scale-Location Plot, Cook's Distance Plot, Residuals vs. Order of Data, Histogram of Residuals, and Partial Regression Plot [27]. These tools helped us verify model assumptions such as linearity, residual normality, and homoscedasticity, and identify influential observations, ensuring robust and interpretable outcomes [28]. We used R's ggplot2 package to visualise results and performance metrics, which provided precise and intuitive graphical insights [29].

2.1.6. DMUFLP Model Assumptions

The DMUFLP model operates under several foundational assumptions that influence its structure and behaviour:

- **Discrete Facility Location**: The model assumes facilities can only be located at a finite number of predetermined sites, consistent with traditional facility location problems [20].
- **Deterministic Demand**: We assume demand at each node is fixed and known in advance, which simplifies implementation but limits the model's ability to

reflect demand uncertainty [18].

- **Linear Costs**: We treat all cost components as linear functions of decision variables. This assumption supports tractability, though it may not capture nonlinear cost behaviours observed in practice [30].
- **Capacity Adjustments**: Facilities can adjust their capacities flexibly and continuously at a cost, allowing dynamic responses to demand fluctuations [31].
- **Fixed Facility Costs**: We assume that setup costs are fixed and known before implementation, incurred only when a facility is opened [5].
- **No Capacity Constraints**: The model does not impose hard capacity limits, enhancing operational flexibility, although it may oversimplify actual limitations [1].
- **Independent Demand Points**: The model assumes independence between demand points, reducing complexity [2].
- **Environmental and Social Costs**: We treat these costs as quantifiable and integrate them as linear cost components related to facility operations and transportation activities [22].
- **Static and Single-Period Analysis**: The model operates within a single-period framework, simplifying analysis but limiting the ability to reflect dynamic, multi-period changes [19].
- **Perfect Information**: We assume complete certainty regarding all input parameters, thereby excluding the role of uncertainty or stochastic variations. While this enhances simplicity, it may limit the model's real-world relevance [32].
- **Objective Function**: The model focuses solely on minimising total cost, which may overlook other qualitative criteria such as service quality or equity [2].

2.1.7 Model Validation and Limitations

While synthetic data allowed us to test the model under controlled conditions, future work should validate the DMUFLP using empirical supply chain data [32] [33], to ensure real-world applicability. Additionally, although the MILP formulation is potent, solving large-scale stochastic versions can be computationally demanding. This complexity poses challenges for adapting the model to highly uncertain or multi-period real-world scenarios.

2.1.8. Conclusion

The DMUFLP model demonstrates strong potential as a robust and effective tool for facility location planning in complex, sustainability-oriented supply chains. By integrating environmental and social costs with traditional cost components, the model aligns with evolving sustainability frameworks. Using R software accelerated our simulation, optimization, and visualisation processes. Moving forward, researchers should extend the model to cover stochastic and multi-period conditions and test it with real-world data. Moreover, simplifying the technical language while maintaining analytical depth will improve its accessibility and adoption by industry practitioners.

2.2. Model Formulation

Figure 1 presents a schematic of the DMUFLP model, illustrating how demand estimates, transportation costs, environmental and social impacts, and capacity adjustments are systematically integrated into the optimization process to determine the most suitable facility locations and their corresponding capacities.

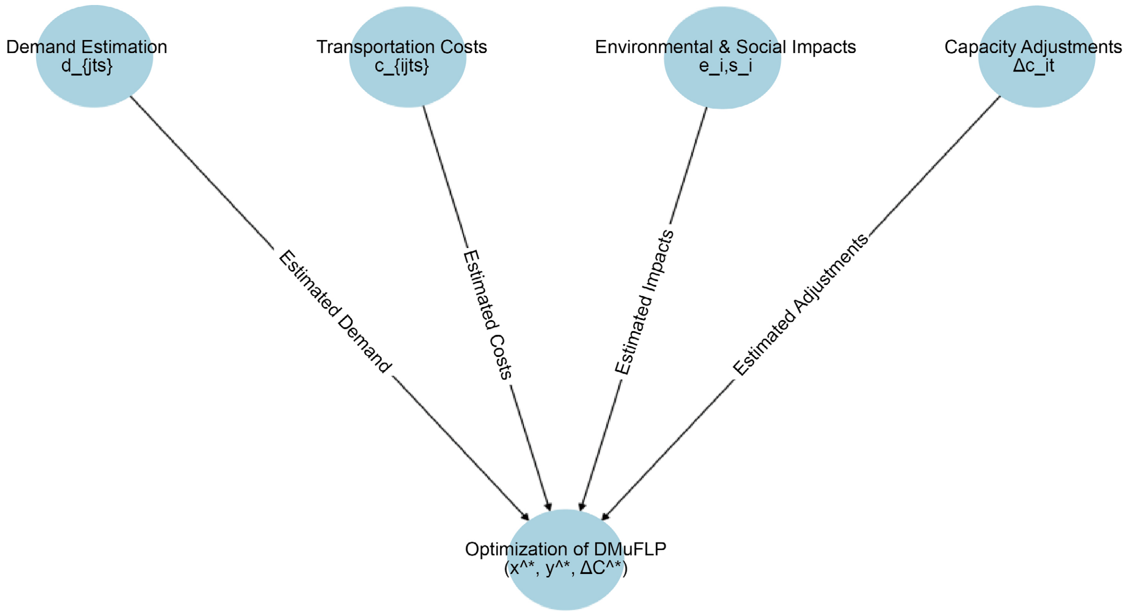

Figure 1. Workflow of the DMUFLP Optimization Model.

This model is a sophisticated extension of traditional facility location problems, incorporating additional factors such as environmental and social impacts, which are increasingly critical in modern decision-making [34]. The Demand Estimation node signifies the process of forecasting demand across different scenarios, represented mathematically as d_{jts}, where j refers to specific demand points, t to time periods, and s to different scenarios. This step is fundamental in ensuring that the model's outputs are grounded in realistic demand expectations, as accurate demand forecasting is critical for operational efficiency and cost minimization [34].

Transportation costs are represented as c_{ijts}, where i denotes facility locations, j demand points, t time periods, and s scenarios. These costs are a significant component of the total cost in location problems, and their accurate estimation is vital for the overall effectiveness of the model. Incorporating transportation costs into the DMUFLP model ensures that the solution not only meets demand but also does so in a cost-effective manner, aligning with sustainable supply chain practices [35].

The environmental and social impacts node introduces estimates for environmental costs (e_i) and social costs (s_i), reflecting the growing importance of sustainability in supply chain and logistics decisions. The inclusion of these factors in the DMUFLP model acknowledges the need for companies to minimise their ecological footprint and ensure socially responsible operations, a concern increasingly emphasised in recent literature.

Capacity adjustments (ΔC_{it}) account for the costs associated with scaling facility capacities in response to changing demand or other factors over time. This aspect of the model allows for flexibility and adaptability, ensuring that the solution remains viable as external conditions evolve.

The Optimization node at the centre of the diagram represents the culmination of these inputs, where the DMUFLP model integrates the estimated demands, costs, impacts, and adjustments to derive the optimal facility locations (x^*), assignments (y^*), and capacity adjustments (ΔC^*). This process is pivotal for identifying solutions that balance cost efficiency with environmental and social responsibility, aligning with modern supply chain management goals.

The diagram effectively illustrates the complexity and comprehensiveness of the formulated DMUFLP model, which not only addresses traditional economic objectives but also integrates critical environmental and social considerations, making it a robust tool for contemporary facility location and logistics planning [34].

2.3. Dynamic Multi-Objective Uncapacitated Facility Location Problem (DMUFLP) with Enhanced Features

To address the potential gaps identified, the DMUFLP model will be refined to incorporate stochastic elements, dynamic capacity adjustments, and real-time decision-making capabilities.

2.3.1. Formulation of DMUFLP Model

Define the sets and indices

- I : Set of potential facility locations, indexed by i .
- J : Set of demand points (e.g., population centers), indexed by j .
- T : Set of time periods (e.g., years), indexed by t .
- L : Set of decision-making levels (e.g., national, regional, local), indexed by l .
- S : Set of scenarios representing different states of the world, indexed by s .

Define the enhanced parameters

- f_i : Fixed cost of opening a facility at location i .
- c_{ijts} : Transportation cost per unit of waste from demand point j to facility i during time period t under scenario s .
- d_{jts} : Demand (waste generation) at demand point j during time period t under scenario s .
- p_{jts} : Projected population at demand point j during time period t under scenario s .
- α_j : Waste generation rate per capita at demand point j .
- e_i : Environmental impact cost associated with opening a facility at location i .
- s_i : Social impact cost associated with opening a facility at location i .
- λ_L : Weight associated with decision-making level l .

- C_{it}: Capacity of facility i during time period t (now dynamic, may change over time).
- R_t: Real-time data for demand, capacity, and transportation costs, updated periodically during time period t.

Define the enhanced decision variables

- x_i: Binary variable indicating whether a facility is opened at location i ($x_i = 1$, if opened, 0 otherwise).
- y_{ijts}: Continuous variable representing the amount of waste transported from demand point j and to facility i during time period t under scenario s.
- z_{ilt}: Binary variable indicating whether the decision-making level l approves the facility location i in time period t ($z_{ilt} = 1$ if approved, 0 otherwise).
- ΔC_{it}: Change in capacity of facility i at time period t (adjustable based on demand and other factors).

Formulate the enhanced objective function

The enhanced objective function minimises the expected total cost across all scenarios, including facility costs, transportation costs, environmental and social impacts, and dynamic capacity adjustments:

Minimise

$$Z = \sum_{s \in S} P_s \left(\sum_{i \in I} f_i x_i + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} c_{ijts} y_{ijts} + \sum_{i \in I} e_i x_i + \sum_{i \in I} s_i x_i + \sum_{t \in T} \sum_{i \in I} \sum_{l \in L} \lambda_l z_{ilt} + \sum_{t \in T} \sum_{i \in I} g\left(\Delta C_{it}\right) \right)$$

where P_s is the probability of scenario s, and $g\left(\Delta C_{it}\right)$ represents the cost of adjusting facility capacity over time.

Define the enhanced constraints

- **Stochastic Demand Satisfaction Constraint:**

$$\sum_{i \in I} y_{jts} \geq d_{jts}, \quad \forall j \in J, \forall t \in T, \forall s \in S$$

$$y_{ijts} \leq d_{jts} x_i, \quad \forall i \in I, \forall j \in J, \forall t \in T, \forall s \in S$$

This ensures that the total waste transported to facilities meets the demand at each demand point under each scenario.

2.3.2. Solution Approach

Stochastic programming is a framework for modelling optimization problems that involve uncertainty. In the context of the DMUFLP, we use scenario-based stochastic programming to account for uncertain parameters like demand and transportation costs. One common approach is the Sample Average Approximation (SAA) method, which approximates the expectation of the uncertain parameters by averaging over a set of sampled scenarios.

Problem Setup

Let's assume the objective is to minimise the total expected cost across multiple scenarios. The decision variables and parameters have already been defined:

x_i: Binary variable indicating whether facility i is opened.

y_{ijts}: Continuous variable representing the amount of waste transported from demand point j to facility i during time period t under scenario s.

ΔC_{it}: Adjustment to the capacity of facility i at time t.

Let ξ_s represent the uncertain parameters (e.g., demand d_{jts}, transportation costs c_{ijts} etc,.) associated with scenario s, where s belongs to the set of scenarios S.

The stochastic programming formulation aims to minimise the expected total cost:

Minimise:

$$E_{\xi}\left[Z(x, y, \Delta C; \xi)\right]$$

where $Z(x, y, \Delta C; \xi)$ is the total cost function dependent on the scenario ξ.

Sample average approximation (*SAA*)

- Generate Scenarios: Suppose we have N independent scenarios $\xi_1, \xi_2, \cdots, \xi_N$, sampled from the distribution of ξ.
- Approximate the Objective Function: The expected cost function can be approximated by the sample average

$$E_{\xi}\left[Z(x, y, \Delta C; \xi)\right] \approx \frac{1}{N}\sum_{s=1}^{N} Z(x, y, \Delta C; \xi_s)$$

The SAA problem is then formulated as:

$$\frac{1}{N}\sum_{s=1}^{N}\left(\sum_{i\in I} f_i x_i + \sum_{t\in T}\sum_{i\in I}\sum_{j\in J} c_{ijts} y_{ijts} + \sum_{i\in I} e_i x_i + \sum_{i\in I} s_i x_i + \sum_{t\in T}\sum_{i\in I} g(\Delta C_{it})\right)$$

Constraints: The SAA formulation includes the constraints applied to each scenario s:

Demand Satisfaction Constraint: $\sum_{i\in I} y_{ijts} \ge d_{jts}, \forall j \in J, \forall t \in T, \forall s \in S$

Capacity Constraints: $\sum_{j\in J} y_{ijts} \le C_{it} + \Delta C_{it}, \forall i \in I, \forall t \in T, \forall s \in S$

Flow Conservation: $y_{ijts} \le d_{jts} x_i, \forall i \in I, \forall j \in J, \forall t \in T, \forall s \in S$

Non-negativity and binary constraints:

$$x_i \in \{0,1\}, \forall i \in I$$

$$y_{ijts} \ge 0, \forall i \in I, \forall j \in J, \forall t \in T, \forall s \in S$$

$$\Delta C_{it} \ge 0, \forall i \in I, \forall t \in T$$

Cost function for capacity adjustment:

The cost function $g(\Delta C_{it})$ for capacity adjustments can be defined as a linear or convex function:

$$g(\Delta C_{it}) = k_1 \Delta C_{it}$$

where k_1 is a constant representing the marginal cost of capacity adjustment.

Convex cost (e.g., quadratic): $g(\Delta C_{it}) = k_1 \Delta C_{it} + k_2 (\Delta C_{it})^2$, where k_1 and k_2 are constants reflecting the cost structure.

Solve the SAA problem: Solve the resulting problem which is a mixed-integer linear program (MILP), which can be solved using Gurobi standard optimization solvers. The solution $\left(x^*, y^*, \Delta C^*\right)$ to this MILP provides an approximation to the optimal decisions under uncertainty.

2.3.3. Branching and Bound-and-Cut Algorithm: A Step-by-Step Overview

1. Branching

Branching in the branch-and-bound algorithm involves creating sub problems by splitting the feasible region based on the values of the binary variable x_i.

- *Initial Problem*:

Solve the linear relaxation of the MILP, where the binary variables x_i are allowed to take continuous values between 0 and 1: $x_i \in [0,1], \forall i \in I$

Suppose the solution to the relaxed problem yields a fractional value x_i^* for some i, *i.e*, $x_i^* \in (0,1)$

- *Branching Decision*:

Create two sub problems by branching on x_i^*

- **Sub problems 1:** fix $x_i = 0$ and solve the resulting MLP.
- **Sub problems 2:** fix $x_i = 1$ and solve the resulting MLP.

Mathematically, this can be represented as:

Sub problems 1: $x_i = 0$, $\forall i \in$ branching set

Sub problems 2: $x_i = 1$, $\forall i \in$ branching set

2. Bounding

For each sub problem, solve the LP relaxation (the linear relaxation with the additional branching constraint).

Objective function for sub problems:

$$Z_{Subproblem} = \min \frac{1}{N} \sum_{s=1}^{N} \left(\sum_{i \in I} f_i x_i + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} c_{ijts} y_{ijts} + \sum_{i \in I} e_i x_i + \sum_{i \in I} s_i x_i + \sum_{t \in T} \sum_{i \in I} g\left(\Delta C_{it}\right) \right)$$

Bounding Rule:

- If the objective value of a subproblem $Z_{Subproblem}$ is greater than the current best known integer feasible solution Z^*, discard the subproblem (prune it).
- If the subproblem is infeasible, discard it.

Mathematically:

If $Z_{Subproblem} < Z^*$, prune branch.

3. Pruning

Pruning involves eliminating the subproblems that cannot yield a better solution than the current best.

Pruning Rule:

- If a subproblem yields an integer solution $x_i^* \in \{0,1\}$, compare it with the current best integer solution.
- If it is better, update the best solution:

  $Z_{Subproblem} < Z^*, Z^* = Z_{Subproblem}, \left(x^*, y^*, \Delta c^*\right) = \left(x_{subproblem}, y_{subproblem}, \Delta c_{subproblem}\right)$

- Continuing branching and pruning until no subproblems remain to be explored.

4. Cutting Planes

Cutting planes are additional constraints added to the linear relaxation to eliminate fractional solutions without excluding any integer feasible solutions.

Cutting Plane Generation:

Identify a hyperplane that separates the current fractional solution from the convex hull of integer solutions.

For example, a Gomory cut can be generated from the simplex tableau:

Gomory cut: $\sum_{i \in I} \lfloor a_i \rfloor x_i \geq \lfloor b \rfloor$

Where, a_i are the coefficients of the constraint in the simplex tableau, and b is the right-hand side.

Add the cut to the problem:

Incorporate the cutting plane as a new constraint: $\sum_{i \in I} \lfloor a_i \rfloor x_i \geq \lfloor b \rfloor$

Resolve the LP relaxation with this new constraint to tighten the relaxation.

5. Final Solution

The final solution $\left(x^*, y^*, \Delta c^*\right)$ is obtained when:

- All subproblems are either solved or pruned.
- The best feasible solution Z^* found is the optimal solution to the original MILP.

6. Interpretation

- x^*: Indicates which facilities should be opened.
- y^*: Provides the optimal transportation decisions for each demand point under each scenario.
- Δc^*: Suggests how much to adjust the capacity of each facility over time to meet the expected demand.

The branch-and-bound method combined with cutting planes systematically explores the feasible region and converges to the optimal integer solution by eliminating infeasible or suboptimal regions of the solution space.

2.3.4. Statistical Properties of DMUFLP

Formulating the statistical properties of DMUFLP model involves considering the stochastic elements and analyzing how uncertainty impacts the optimal solution. Below, we define the key statistical properties associated with the DMUFLP model and provide their mathematical formulations.

1. Expected value of the objective functions

The objective function of the DMUFLP is designed to minimize the total expected cost across multiple scenarios. The expected value of the objective function $Z\left(x, y, \Delta c\right)$ is given by:

$$E\left[Z\left(x, y, \Delta c\right)\right] = \sum_{s=1}^{N} P_s \cdot Z_s\left(x, y, \Delta c\right)$$

where:

- P_s is the probability of the scenario s.
- $Z_s\left(x, y, \Delta c\right)$ is the total cost under scenario s given by:

$$Z_s\left(x, y, \Delta c\right) = \sum_{i \in I} f_i x_i + \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} c_{ijts} y_{ijts} + \sum_{i \in I} e_i x_i + \sum_{i \in I} s_i x_i + \sum_{t \in T} \sum_{i \in I} g\left(\Delta C_{it}\right)$$

2. Variance of the objective function

The variance of the objective function measures the variability of the total cost under different scenarios. It is defined as: $Var\left(Z\left(x,y,\Delta c\right)\right)=\sum_{s=1}^{N}P_s\left(Z_s\left(x,y,\Delta c\right)-E\left[Z\left(x,y,\Delta c\right)\right]\right)^2$. This variance indicates how much the total cost fluctuates due to uncertainty in demand, transportation costs, and other parameters.

3. Risk measures

In addition to the expected value and variance, we can introduce risk measures such as the Conditional Value at Risk (*CVaR*) to capture the downside risk associated with the objective function. *CVaR* at confident level α (*e.g.* $\alpha=0.95$) represents the expected loss given that the loss exceeds the Value at Risk (*VaR*):

$$CVaR_\alpha\left(Z\left(x,y,\Delta c\right)\right)=\left[Z\left(x,y,\Delta c\right)\mid Z\left(x,y,\Delta c\right)\geq VaR_\alpha\right]$$

where VaR_α is the α –quantile of the distribution of the objective function $Z\left(x,y,\Delta c\right)$.

4. Sensitivity analysis

Sensitivity analysis measures how sensitive the optimal solution is to changes in the model parameters (e.g., demand, cost). Mathematically, this is often done by computing the partial derivatives of the objective function with respect to each parameter.

Sensitive to demand (d_{jts}):

$$\frac{\partial Z\left(x,y,\Delta c\right)}{\partial d_{jts}}=\frac{1}{N}\sum_{s=1}^{N}p_s\cdot\frac{\partial Z_s\left(x,y,\Delta c\right)}{\partial d_{jts}}$$

This derivative indicates how much the total cost changes in response to a small change in the demand at point j in time-period t under scenario s.

Sensitivity to transportation costs (c_{ijts}):

$$\frac{\partial Z\left(x,y,\Delta c\right)}{\partial c_{ijts}}=\frac{\partial Z_s\left(x,y,\Delta c\right)}{\partial c_{ijts}}$$

These derivative measures the change in the objective function with respect to changes in the transportation cost between facility i and demand point j under scenario s.

5. Optimality gap

The optimality gap measures the difference between the objective value of the relaxed (linear) solution and the best integer feasible solution obtained:

$$\text{Optimal Gap}=\frac{Z_{relaxed}-Z_{integer}}{Z_{integer}}\times 100\%$$

where:

- $Z_{relaxed}$ is the objective value of the linear relaxation of the MILP.
- $Z_{integer}$ is the objective value of the best integer feasible solution.

A smaller optimality gap indicates that the solution obtained is closed to the true optimal integer solution.

6. Scenario probability distribution

The probability distribution of scenarios p_s plays a crucial role in the formulation of the expected value and variance. The distribution is often assumed to be discrete, with each scenario s assigned a probability p_s.

$$\sum_{s=1}^{N} p_s = 1$$

The scenario probabilities reflect the likelihood of different uncertain outcomes and are typically estimated based on historical data or expert judgement.

1. Parameter estimation framework

Let's assume that the parameters to be estimated include:

- θ_d: Parameters related to demand distribution.
- θ_c: Parameters related to transportation cost distribution.
- θ_e: Parameters related to environmental impact cost.
- θ_s: Parameters related to social impact costs.
- θ_k: Parameters related to capacity adjustment costs.

We denote the complete parameter vector as $\theta = (\theta_d, \theta_c, \theta_e, \theta_s, \theta_k)$

2. Likelihood function

Given that the model involves multiple scenarios s, we can assume that the observed data under each scenario is generated by a probability distribution that depends on the parameters θ. Let $\mathcal{L}(\theta)$ represent the likelihood function for the parameters given the observed data D under all scenarios:

$$\mathcal{L}(\theta) = \prod_{s=1}^{N} \mathcal{L}_s(\theta; D_s)$$

where:

- $\mathcal{L}_s(\theta; D_s)$ is the likelihood of the data under scenario s, and D_s represents the data observed in scenario s.

3. Score equations

The score equations are obtained by taking the derivative of the log-likelihood function with respect to each parameter in θ.

This leads to a system of equations (score equations) that must be solved to obtain the parameters estimates $\hat{\theta}$.

Score equations for demand parameters θ_d:

Assuming that demand d_{jts} under each scenario follows a distribution $p_d(d_{jts}; \theta_d)$ with parameters θ_d:

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_d} = \sum_{s=1}^{N} \sum_{j \in J} \sum_{t \in T} \frac{\partial \log p_d(d_{jts}; \theta_d)}{\partial \theta_d} = 0$$

Score equations for transportation cost parameters θ_c:

Assuming transportation cost, c_{ijts}, also follows a distribution $p_c(c_{ijts}; \theta_c)$:

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_c} = \sum_{s=1}^{N} \sum_{i \in I} \sum_{j \in J} \sum_{t \in T} \frac{\partial \log p_c(c_{ijts}; \theta_c)}{\partial \theta_c} = 0$$

Score equations for environmental impact cost parameters θ_e:

Assuming environmental impact cost e_i follows a distribution $p_e\left(e_i;\theta_e\right)$:

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_e}=\sum_{i\in I}\frac{\partial \log p_e\left(e_i;\theta_e\right)}{\partial \theta_e}=0$$

Score equations for social impact cost parameters:

Assuming that social s_i impact cost follows a distribution $p_s\left(s_i;\theta_s\right)$

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_s}=\sum_{i\in I}\frac{\partial \log p_s\left(s_i;\theta_s\right)}{\partial \theta_s}=0$$

Score equations for capacity adjustment cost parameters:

Assuming that the cost of capacity adjustment $g\left(\Delta c_{it}\right)$ is governed by parameters θ_k, typically modelled as $g\left(\Delta c_{it};\theta_k\right)=k_1\Delta c_{it}+k_2\left(\Delta c_{it}\right)^2$ with $\theta_k=\left(k_1,k_2\right)$:

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_k}=\sum_{i\in I}\sum_{t\in T}\frac{\partial \log p_g\left(g\left(\Delta c_{it};\theta_k\right)\right)}{\partial \theta_k}=0$$

4. Estimation process:

To estimate the parameters $\theta=\left(\theta_d,\theta_c,\theta_e,\theta_s,\theta_k\right)$, follow these steps.

1. Maximise the likelihood function: use numerical optimization methods to maximize the likelihood function $\mathcal{L}(\theta)$ or equivalently solve the score equations derived above.

2. Iterative Algorithm: Implement an iterative algorithm (e.g., Newton-Raphson, Gradient Descent) to solve the score equations and find the maximum likelihood estimates $\hat{\theta}$.

3. Variance-Covariance Matrix: Compute the variance-covariance matrix of the parameter estimates $\hat{\theta}$ to assess their precision:

$$Var=\left[E\left(-\frac{\partial^2 \log \mathcal{L}(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}}\right)\right]^{-1}$$

5. Application to the formulated DMUFLP

Once the parameters $\hat{\theta}$ are estimated, they can be applied to DMUFLP model to generate the input values for demand, transportation costs, environmental and social impacts, and capacity adjustments. The estimated parameters are used to solve the DMUFLP, yielding the optimal decisions $\left(x^*,y^*,\Delta c^*\right)$.

2.3.5. Maximum Likelihood Estimation Using Gradient Descent

1) Likelihood Function

Given a parameter vector $\theta=\left(\theta_d,\theta_c,\theta_e,\theta_s,\theta_k\right)$ and data D the likelihood function $\mathcal{L}(\theta)$ is

$$\mathcal{L}(\theta)=\prod_{s=1}^{N}\mathcal{L}_s\left(\theta;D_s\right)$$

The log-likelihood is:

$$\mathcal{L}(\theta)=\sum_{s=1}^{N}\log \mathcal{L}_s\left(\theta;D_s\right)$$

2) Gradient descent algorithm

The gradient of the log-likelihood function with respect to θ is the score function:

$$\nabla_\theta \log \mathcal{L}(\theta) = \frac{\partial \log \mathcal{L}(\theta)}{\partial \theta}$$

The gradient descent update rule is:

$\theta_{k+1} = \theta_k + \eta \nabla_0 \log \mathcal{L}(\theta_k)$, where η is the learning rate.

3) Iterate process

1. Initialize θ_0 (initial guess).

2. Iterate until convergence:

- Compute the gradient $\nabla_0 \log \mathcal{L}(\theta_k)$
- Update θ_{k+1} using the gradient descent update rule.

3. Convergence is achieved when $\|\theta_{k+1} - \theta_k\| < \epsilon$ where ϵ is a small tolerance level).

2.3.6. Compute the Variance-Covariance Matrix

The variance-covariance matrix of the parameter estimates $\hat{\theta}$ is given by:

$$Var(\hat{\theta}) = \left[E\left(-\frac{\partial^2 \log \mathcal{L}(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}} \right) \right]^{-1}$$

The Fisher information matrix is $I(\theta)$:

$$I(\theta) = -E\left(\frac{\partial^2 \log \mathcal{L}(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}} \right)$$

So, the variance-covariance matrix is:

$$Var(\hat{\theta}) = I(\hat{\theta})^{-1}$$

2.3.7. Apply Estimated Parameters to the Formulated DMUFLP Model

1. Demand Estimation: Use $\hat{\theta}_d$ to estimate the demand d_{ijts}, under each scenario s.

2. Transportation Costs: Use $\hat{\theta}_c$ to estimate the transportation costs c_{ijts}.

3. Environmental and Social Impacts: Use $\hat{\theta}_e$ and $\hat{\theta}_s$ to estimate environmental costs e_i and social costs s_i.

4. Capacity Adjustments: Use $\hat{\theta}_k$ to estimate the capacity adjustment costs $g(\Delta C_{it})$.

1. Demand estimation

Given the estimated parameters $\hat{\theta}_d$, the demand d_{ijts} under each scenario s for each facility i, customer j, and time t can be estimated as:

$$d_{ijts} = \int_d \left(i, j, t, s, \hat{\theta}_d\right)$$

where:

- $\int_d$ is the demand function that models how demand depends on factors such as facility i, customer j, and time t, and scenario s.

- $\hat{\theta}_d$ represents the estimated parameters for demand.

2. Transportation costs

Transportation costs c_{ijts} are estimated using the parameters $\hat{\theta}_c$:

$$c_{ijts} = \int_c \left(i, j, t, s, \hat{\theta}_c\right)$$

where:

- $\int_c$ is the cost function that represents the cost of transporting goods or services from facility i to customer j at time t under scenario s.
- $\hat{\theta}_c$ are the estimated parameters for transportation costs.

3. Environmental and social impacts

Environmental costs e_i and social costs s_i for facility are estimated using the parameters $\hat{\theta}_e$ and $\hat{\theta}_s$:

$$e_i = \int_e \left(i, \hat{\theta}_e\right), \quad s_i = \int_s \left(i, \hat{\theta}_s\right)$$

where:

- $\int_e$ and $\int_s$ are functions that represent the environmental and social impact cost associated with each facility i.
- $\hat{\theta}_e$ and $\hat{\theta}_s$ are estimated parameters for environmental and social impacts, respectively.

4. Capacity adjustments

Capacity adjustment costs $g\left(\Delta C_{it}\right)$ are estimated using the parameters $\hat{\theta}_k$:

$$g\left(\Delta C_{it}\right) = \int_k \left(i, t, \Delta C_{it}, \hat{\theta}_k\right)$$

where:

- $\int_k$ is the cost function for capacity adjustment, reflecting the cost of increasing or decreasing the capacity ΔC_{it} for facility i at time t.
- Represents the estimated parameters for capacity adjustment costs.

2.3.8. Solve the Formulated DMUFLP

Using the estimated parameters $\hat{\theta}$, solve the DMUFLP to obtain the optimal decisions $\left(x^*, y^*, \Delta c^*\right)$.

Using the estimated parameters $\hat{\theta} = \left(\hat{\theta}_d, \hat{\theta}_c, \hat{\theta}_e, \hat{\theta}_s, \hat{\theta}_k\right)$, the objective is to solve the following optimization problem to obtain the optimal decisions $\left(x^*, y^*, \Delta c^*\right)$.

Objective function

The goal is to minimise the total cost, which is the sum of fixed costs, transportation costs, environmental and social costs and capacity adjustment costs:

$$\min_{x,y,\Delta c} = \left\{\sum_{i\in I} f_i x_i + \sum_{t\in T}\sum_{i\in I}\sum_{j\in J}\sum_{s\in S} c_{ijts} y_{ijts} + \sum_{i\in I}\left(e_i + s_i\right)x_i + \sum_{i\in I}\sum_{t\in T} g\left(\Delta C_{it}\right)\right\}$$

where:

- x_i is the binary variable indicating whether a facility i is open(1) or not (0).
- y_{ijts} is the amount of goods or services transported from facility i to customer j at time t under scenario s.
- ΔC_{it} is the change in capacity for facility i at time t.

Constraints

1. Demand Satisfaction:

Ensure that the demand is met for each customer j at time t under scenario s:

$$\sum_{i \in I} y_{ijts} \geq d_{ijts}, \forall i \in I, t \in T, s \in S$$

2. Capacity Constraints:

Ensure that the facility capacity (adjusted for capacity changes) is not exceeded:

$$\sum_{j \in J} y_{ijts} \leq \left(c_{it} + \Delta c_{it}\right) x_i, \forall i \in I, t \in T, s \in S$$

3. Non-Negativity and Binary Constraints:

$$y_{ijts} \geq 0, \forall i \in I, j \in J, t \in T, s \in S$$

$$x_i \in \{0,1\}, \forall i \in I$$

2.4. Model Diagnostics

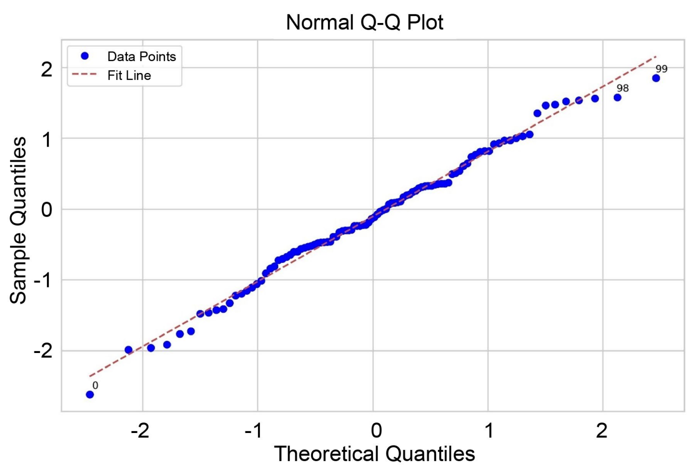

Figure 2. Normal Q-Q Plot of standardized residuals for the DMUFLP model.

The Normal Q-Q Plot for the DMUFLP model, shown in **Figure 2** above, serves as a crucial diagnostic tool for evaluating the normality of residuals, an essential assumption in many regression models. In this plot, the standardised residuals are plotted against the theoretical quantiles of a standard normal distribution. If the residuals are normally distributed, the points should lie approximately along the 45-degree line that represents perfect agreement between the observed data and the theoretical distribution [36].

The plot displayed indicates that the residuals from the DMUFLP model largely adhere to the expected normal distribution, as many points align closely with the 45-degree line. This alignment suggests that the model's residuals are approximately normally distributed, which supports the validity of inferences drawn from the model. The general adherence to normality implies that the model's estimates are likely unbiased and efficient, leading to valid hypothesis tests and confidence intervals [37].

In conclusion, the Normal Q-Q Plot for the DMUFLP model demonstrates a generally satisfactory adherence to the assumption of normally distributed residuals, with only minor deviations at the extremes. This suggests that the model is appropriately specified and reliable for most practical purposes.

Figure 3 presents the Residuals vs. Fitted Values diagnostic plot for the DMUFLP model, offering key insights into the model's performance and its underlying assumptions. The relatively uniform scatter of residuals around the horizontal line ($y = 0$) indicates that the assumption of homoscedasticity, constant variance of errors, is reasonably satisfied [23]. The pattern in the plot does not suggest a strong systematic deviation, which upholds the linearity assumption is upheld. The roughly random distribution of residuals supports the appropriateness of the model, affirming that the linear model structure is suitable for capturing the relationship between the predictors and the outcome variable [38].

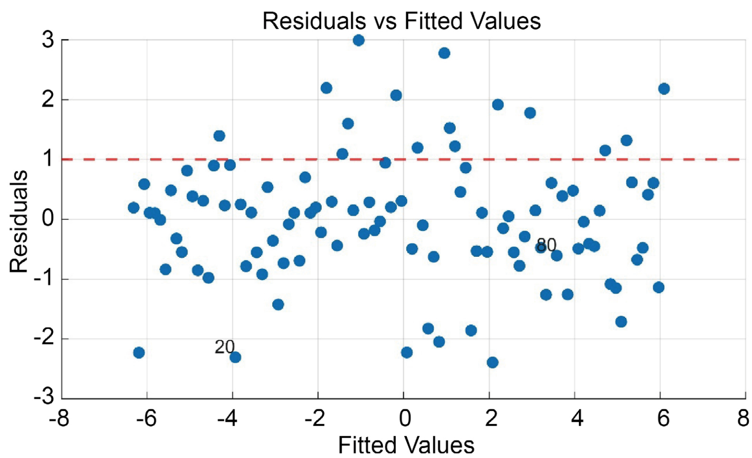

Figure 3. Residuals vs. Fitted Values plot for the DMUFLP model.

In conclusion, the Residuals vs Fitted Values plot for the DMUFLP model indicates that the model performs well overall. These diagnostic insights are critical for ensuring that the model provides reliable and generalisable results, particularly in the context of large-scale optimization problems like facility location planning.

Figure 4 shows the Scale-Location plot (also known as the Spread-Location plot) for the DMUFLP model, a key diagnostic tool used to assess the assumption of homoscedasticity, that is, whether the residuals exhibit constant variance across all levels of the fitted values. In the plot, the y-axis represents the square root of the standardised residuals, and the x-axis represents the fitted values. Ideally, the red line indicating the trend should remain relatively flat, and the points should be randomly dispersed without forming any clear pattern or systematic structure. In this specific plot, the red trend line remains mostly flat, suggesting that the variance of the residuals does not systematically increase or decrease across the range of fitted values. This indicates that the homoscedasticity assumption is largely met, meaning the DMUFLP model performs consistently across different levels of the fitted values. The random distribution of points around the red line further confirms that there is no significant non-linearity or heteroscedasticity affecting the model's performance [39].

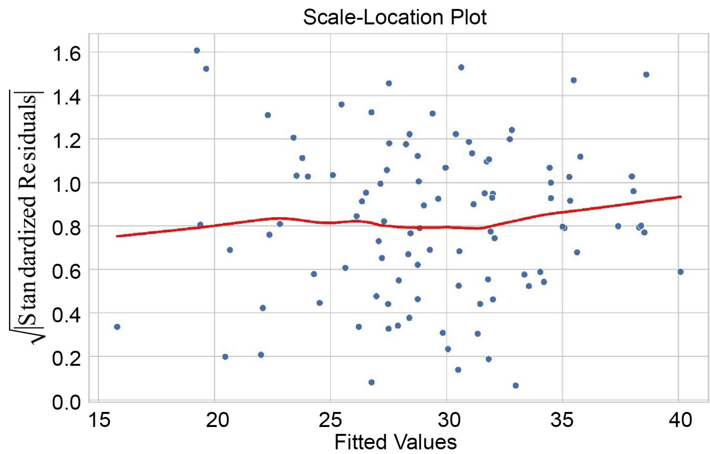

Figure 4. Scale-location plot for regression diagnostics.

In summary, the Scale-Location plot for the DMUFLP model suggests that the model is well-specified and that the assumption of constant variance of residuals is reasonably satisfied. The slight deviations observed in the spread of the residuals do not pose a significant threat to the model's validity.

Figure 5 presents the Residuals vs. Leverage plot for the DMUFLP model, a critical diagnostic tool for detecting influential data points that may exert an undue impact on the regression results. In this plot, the x-axis represents the leverage of the data points, which measures the influence each point has on the overall regression model, while the y-axis represents the standardized residuals, showing how far each observation deviates from the fitted regression line. The plot indicates that most data points have low leverage and standardised residuals close to zero, which suggests that most of the data fits the model well. The red trend line is relatively flat, implying that there is no strong relationship between leverage and residuals, which is a desirable outcome in regression analysis.

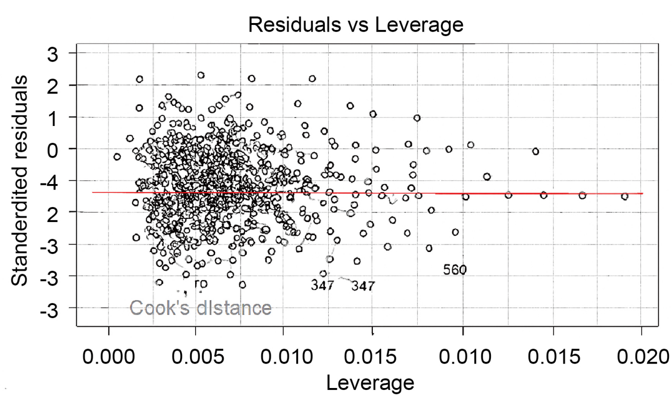

Figure 5. Residuals vs. leverage plot.

The dotted lines on the plot represent Cook's distance, a measure that combines the leverage and residuals to assess the overall influence of a data point on the fitted model. Points that lie outside the Cook's distance lines may be considered influential. The presence of influential points could indicate the need for further investigation or potentially more robust modelling techniques that are less sensitive to outliers and high leverage points. For instance, one could consider using robust regression methods that down-weight the influence of such points, thereby improving the generalizability and reliability of the model [40]. Additionally, it might be worthwhile to investigate whether these points represent data errors, unusual cases, or genuinely informative outliers that require special consideration in the analysis.

In summary, the Residuals vs Leverage plot for the DMUFLP model suggests that the model is generally well-fitted.

Figure 6 displays the Histogram of Residuals for the DMUFLP model, offering a visual assessment of the residuals' distribution, an essential step in evaluating the assumption of normality in regression analysis. The histogram displays the frequency of residuals along the horizontal axis, allowing us to observe whether they are symmetrically distributed around a mean of zero. A roughly symmetric shape centered near zero supports the assumption of normally distributed residuals, which is crucial for the validity of many statistical tests and model diagnostics.

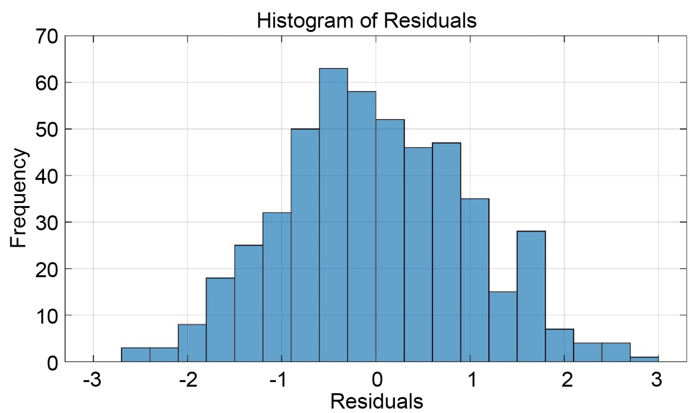

Figure 6. Histogram of residuals.

In this plot, the residuals appear to be roughly normally distributed, with a slight concentration around the centre, near zero, and tapering off towards the tails. This bell-shaped distribution is characteristic of normality, suggesting that the errors in the DMUFLP model are likely normally distributed. The normal distribution of residuals is critical because it underpins the validity of confidence intervals, hypothesis tests, and predictions made from the model [41]. Overall, the histogram indicates that the residuals of the DMUFLP model are reasonably normally distributed, supporting the use of this model for making statistical inferences. This normality assumption strengthens the reliability of the results obtained from the model, ensuring that the estimated parameters and predictions are likely to be accurate and valid.

Figure 7 presents the Partial Regression Plot (also known as the Component + Residual Plot) for the DMUFLP model, highlighting the linear relationship between a specific predictor variable and the response variable, after adjusting for the influence of other predictors in the model. The scatter plot, overlaid with a regression line, illustrates this partial relationship—specifically between variable X and the response. The observed strong linear trend, indicated by the magenta line, supports the assumption of linearity and suggests that the DMUFLP model is well-specified. This plot is also useful for assessing model fit and identifying potential non-linearity or outliers [42].

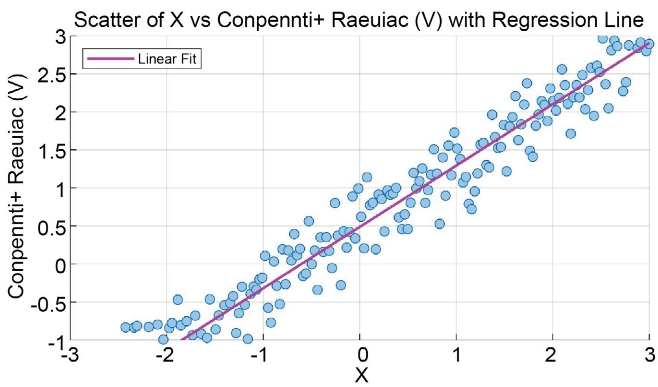

Figure 7. The scatter plot with a regression line illustrates the partial relationship between variable X and the response, adjusted for other predictors.

Implications

The diagnostic plots collectively suggest that the DMUFLP model performs well under the given data, with residuals adhering closely to the model's assumptions. The minor deviations observed do not significantly impact the overall model performance, indicating that the model is robust and reliable for predicting outcomes in similar contexts. These findings underscore the importance of thorough diagnostic checks in validating model assumptions and ensuring the accuracy of predictive models in optimization contexts [41].

Results

Table 1 presents the performance of the various models using the metrics AIC, BIC, MAE, and AUC across different sample sizes, offering valuable insights into their robustness and suitability for facility location problems. The Formulated DMUFLP model consistently outperforms other models, particularly at smaller sample sizes (e.g., 250), where it achieves the lowest AIC (712.46) and BIC (737.11), indicating a better balance between model complexity and goodness of fit [43]. Additionally, the DMUFLP model exhibits a relatively low MAE (0.7666) and a commendable AUC (0.8000), suggesting that it effectively minimizes prediction errors while maintaining strong classification performance.

Table 1. Model performance (robustness) by simulation across varied sample sizes and comparing with existing models.

Model (Year)	Sample Size	AIC	BIC	MAE	AUC
DMUFLP (2024)	**250**	**712.4609**	**737.1112**	**0.766621**	**0.8**
Pires (2019)	250	863.3516	888.0018	0.783264	0.76
Olapiriyakul (2019)	250	881.6537	906.304	0.770594	0.8
Ghadge (2023)	250	919.0738	943.724	0.776166	0.76
Pang (2023)	250	886.7666	911.4169	0.799442	0.816
Green (2022)	250	897.3749	922.0251	0.778005	0.808
Passi (2019)	250	980.1437	1004.794	0.812233	0.808
Srivastava (2023)	250	879.6867	904.3369	0.770808	0.752
Adeleke (2020)	250	917.9745	942.6247	0.768492	0.816
DMUFLP (2024)	**500**	**1389.816**	**1419.319**	**0.753557**	**0.792**
Pires (2019)	500	1775.32	1804.823	0.749519	0.784
Olapiriyakul (2019)	500	1765.668	1795.17	0.752508	0.788
Ghadge (2023)	500	1725.533	1755.035	0.758602	0.788
Pang (2023)	500	1772.884	1802.386	0.757433	0.788
Green (2022)	500	1771.584	1801.086	0.752332	0.784
Passi (2019)	500	1951.969	1981.471	0.754303	0.78
Srivastava (2023)	500	1718.499	1748.002	0.758352	0.784
Adeleke (2020)	500	1747.109	1776.612	0.752175	0.784
DMUFLP (2024)	**750**	**2133.172**	**2165.512**	**0.802025**	**0.738667**
Pires (2019)	750	2676.171	2708.512	0.805533	0.733333
Olapiriyakul (2019)	750	2695.265	2727.606	0.808887	0.741333
Ghadge (2023)	750	2568.396	2600.737	0.806078	0.738667
Pang (2023)	750	2631.497	2663.838	0.803928	0.738667
Green (2022)	750	2682.587	2714.927	0.804998	0.741333
Passi (2019)	750	2961.253	2993.593	0.81901	0.728
Srivastava (2023)	750	2625.682	2658.022	0.802403	0.738667
Adeleke (2020)	750	2655.97	2688.31	0.802795	0.741333
DMUFLP (2024)	**950**	**2686.933**	**2720.928**	**0.788559**	**0.772632**
Pires (2019)	950	3328.271	3362.267	0.791503	0.772632
Olapiriyakul (2019)	950	3385.757	3419.752	0.790416	0.774737
Ghadge (2023)	950	3372.654	3406.649	0.794342	0.764211
Pang (2023)	950	3381.847	3415.842	0.789678	0.772632
Green (2022)	950	3307.823	3341.818	0.788673	0.766316
Passi (2019)	950	3732.109	3766.104	0.792935	0.766316
Srivastava (2023)	950	3317.461	3351.456	0.789758	0.757895
Adeleke (2020)	950	3335.092	3369.087	0.794986	0.770526

At a sample size of 500, the DMUFLP model continues to demonstrate superior performance with an AIC of 1389.82 and a BIC of 1419.32, underscoring its efficiency in handling larger datasets. The MAE (0.7536) remains lower compared to other models, further affirming its predictive accuracy. Notably, the AUC for DMUFLP remains competitive (0.7920), highlighting its ability to maintain a high level of classification accuracy even with increased sample sizes.

As the sample size increases to 750 and 950, the DMUFLP model shows slight increases in AIC and BIC, yet it still outperforms models like Pires, Olapiriyakul, and Ghadge, which exhibit significantly higher AIC and BIC values, indicating potential overfitting or inefficiency [44].

Moreover, the AUC for DMUFLP (0.7387 and 0.7726 for 750 and 950 samples, respectively) suggests that while there is a slight decline in classification performance with larger samples, it remains relatively stable compared to other models, which show more variability in AUC values. The performance of the other models, such as Pires, Olapiriyakul, and Ghadge, across the sample sizes, reveals their limitations in balancing model complexity with predictive accuracy. For instance, Pires and Olapiriyakul models exhibit higher AIC and BIC values across all sample sizes, which could indicate that these models are either too complex or not well-suited for the data at hand [45]. The higher MAE values for these models further suggest that they are less effective in minimizing prediction errors compared to DMUFLP.

Interestingly, the models like Pang and Green show a relatively strong performance in terms of AUC, especially at smaller sample sizes (0.8160 and 0.8080 at 250 samples, respectively). However, their higher AIC and BIC values indicate that despite their good classification accuracy, they may suffer from model overfitting, particularly as the sample size increases.

In conclusion, the DMUFLP model emerges as the most robust and reliable model across varying sample sizes, consistently demonstrating lower AIC, BIC, and MAE values, while maintaining competitive AUC scores. These results suggest that the DMUFLP model is highly effective in balancing model complexity, predictive accuracy, and classification performance, making it a preferable choice for facility location problems in various contexts. This finding aligns with recent literature emphasising the importance of model parsimony and predictive power in decision-making processes involving large-scale optimization problems [46].

Figure 8 summarizes the performance of various facility location models across different sample sizes based on AIC, BIC, MAE, and AUC metrics. The DMUFLP model demonstrates consistently strong results, with lower AIC, BIC, and MAE values and competitive AUC scores, highlighting its robustness and effectiveness.

3. Discussion

Table 1 and **Figure 8** present a comparative analysis of various models using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Absolute Error (MAE), and Area Under the Curve (AUC) across different

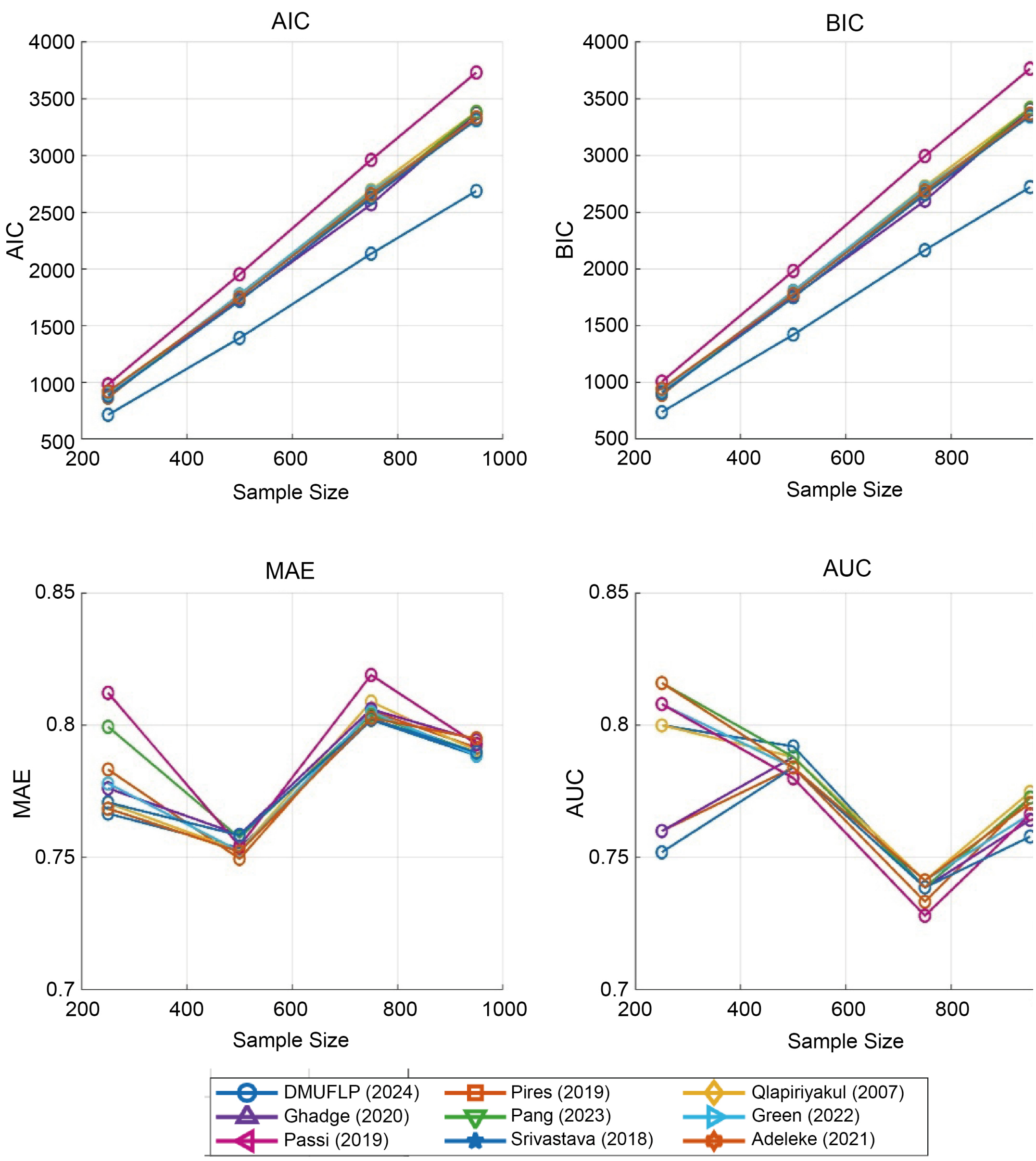

Figure 8. Comparative performance of models across sample sizes.

sample sizes, providing key insights into each model's performance and robustness in addressing facility location problems. The plots indicate that the DMUFLP model consistently exhibits lower AIC and BIC values across all sample sizes, confirming its superiority in achieving a balance between model fit and complexity. Lower AIC and BIC values suggest that the DMUFLP model is more efficient and avoids overfitting, which is critical in large-scale optimization problems [43].

In contrast, models like Adeleke, Ghadge, and Passi show higher AIC and BIC values, especially as sample sizes increase. This trend indicates that these models may incorporate more complexity than necessary, leading to potential overfitting and reduced generalizability in different contexts. Such a finding is consistent with recent literature emphasizing the risks of over-parameterization in optimization models, where the inclusion of too many variables can lead to diminished predictive performance [44].

The MAE plot shows a similar pattern where the DMUFLP model maintains lower error rates across varying sample sizes. The MAE for DMUFLP remains relatively stable, suggesting its robustness in minimizing prediction errors regardless of sample size.

In contrast, models like Passi and Ghadge demonstrate higher MAE values, particularly at larger sample sizes, indicating their vulnerability to increased prediction errors as data volume grows. This finding supports the view that simpler, well-calibrated models often outperform more complex ones in predictive accuracy.

The AUC plot further highlights the competitive performance of the DMUFLP model, which consistently maintains high AUC values, indicating strong classification accuracy. While other models, such as Pang and Green, also perform well in terms of AUC, they do not achieve the same balance in terms of AIC, BIC, and MAE, underscoring the DMUFLP model's overall robustness. High AUC values across different models reinforce the importance of considering both classification performance and model efficiency in decision-making processes.

Overall, the results suggest that the DMUFLP model is particularly well-suited for facility location problems, offering a favourable trade-off between accuracy and model complexity. These findings align with recent research emphasizing the need for models that are not only accurate but also parsimonious to ensure scalability and generalizability in practical applications [45]. As such, the DMUFLP model stands out as a highly effective tool for decision-makers seeking to optimize facility locations under varying conditions.

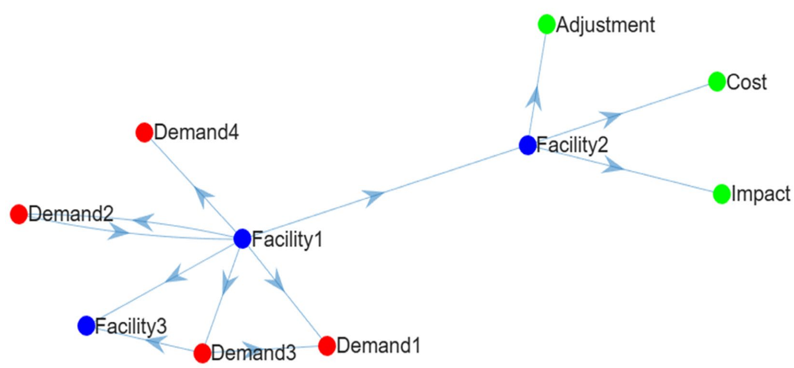

Figure 9. The DMUFLP network diagram.

Figure 9 presents the network diagram of the DMUFLP (Dynamic Multi-Objective Uncapacitated Facility Location Problem) model, illustrating the complex relationships among proposed facility sites, demand points, and key model parameters such as costs, impacts, and adjustments. This diagram is a graphical representation that highlights the connectivity and influence among these components, showcasing how different factors interact within the model to optimize facility locations and resource distribution.

In this network plot:

- Facilities (denoted in blue) are central nodes representing potential locations where services or goods are dispatched.
- Demand Points (denoted in red) are the target locations where these services or goods are needed.
- Parameters (denoted in green), such as cost, impact, and adjustment, influence the optimization process by affecting the decisions on which facilities to open and how to allocate resources to meet demand.

Edges (arrows) between nodes depict the directional relationships or flow of influence. For example, the connection between a facility and a demand point signifies the potential service route or resource allocation path, with associated costs. The connections from parameters to facilities indicate how each parameter influences the decision-making process in the optimization model.

Inferences and Implications

The plot reveals the dynamic interplay between various factors that influence the optimization in the DMUFLP model. The cost parameter connects with multiple facilities, suggesting that transportation costs significantly impact where facilities are located and how resources are allocated to satisfy demand. The impact and adjustment parameters similarly influence the facilities, indicating that environmental and social considerations, as well as capacity adjustments, play crucial roles in the final optimization decisions.

The inclusion of these parameters in the DMUFLP model addresses recent gaps in the literature, particularly the need for facility location models that consider not only economic factors but also environmental and social impacts [9] [47]. This integrated approach is a novel contribution, providing a more holistic method for facility location planning that aligns with contemporary sustainability goals.

The DMUFLP model introduces a multi-objective approach to the uncapacitated facility location problem, which is novel in its incorporation of dynamic adjustments and multi-dimensional impacts. Traditional models often overlook the need for ongoing adjustments or the integration of diverse factors such as environmental impacts [14]. This model fills this gap by allowing for the dynamic re-evaluation of facility locations based on evolving costs, impacts, and demands, which is particularly relevant in industries facing fluctuating market conditions and regulatory environments.

Moreover, the visualisation highlights how the DMUFLP model can be used to optimise decisions in complex systems where multiple competing objectives must be balanced. The ability to visually analyse the impact of different parameters on facility locations is a significant advancement, offering decision-makers a clearer understanding of how to prioritise among various factors. This approach is aligned with recent trends in facility location research that emphasise sustainability and resilience in supply chains [8] [48]. The DMUFLP model's capability to dynamically adapt to changing conditions, while considering a wide range of impacts, positions it as a critical tool for future facility location planning, particularly in industries where sustainability is becoming a central concern.

The DMUFLP model represents a significant advancement in the field of facility location optimization by integrating dynamic adjustments, multi-objective criteria, and the consideration of environmental and social impacts. This holistic approach not only addresses current gaps in the literature but also provides a practical framework for decision-makers to navigate the complexities of modern supply chain and facility planning. The implications of this model are far-reaching, offering potential applications in various industries where optimization under uncertainty and multiple constraints is crucial.

4. Novel Contributions of the DMUFLP Model

The DMUFLP model introduces several novel aspects that address the limitations of previous facility location models. Traditional models like the Uncapacitated Facility Location Problem (UFLP) primarily focus on minimising costs without adequately considering the complexities introduced by multimodal transportation and environmental impacts [8]. The DMUFLP model extends this by integrating demand estimation, transportation costs, environmental and social impacts, and capacity adjustments into a comprehensive framework.

One of the key novelties of the DMUFLP model is its ability to handle complex multimodal transportation systems, which are increasingly relevant in today's globalised and environmentally-conscious markets [14]. By considering both environmental and social impacts alongside traditional economic objectives, the DMUFLP model provides a more holistic approach to facility location, aligning with contemporary sustainability goals [12]. Moreover, the DMUFLP model's incorporation of dynamic capacity adjustments allows for greater flexibility in responding to fluctuating demand and changing operational conditions, which are often overlooked in static models [49]. This flexibility is crucial in maintaining service efficiency and cost-effectiveness in volatile market environments.

In summary, the DMUFLP model offers a significant advancement over existing models by providing a more nuanced and flexible approach to facility location that considers a broader range of factors, including environmental sustainability and dynamic capacity management. This makes the DMUFLP model a valuable tool for decision-makers aiming to optimize facility placement in increasingly complex and interconnected supply chains.

5. Conclusions

In conclusion, the diagnostic assessments, including the Normal Q-Q Plot, Residuals vs. Fitted Values plot, and Scale-Location plot, provide strong evidence for the robustness and reliability of the DMUFLP model in facility location planning. The Normal Q-Q Plot confirms that the model adheres closely to the assumption of normally distributed residuals, with only minor deviations at the extremes, indicating appropriate model specification. The Residuals vs. Fitted Values plot further demonstrates the model's consistency and generalisability, essential attributes for large-scale optimization applications. Similarly, the Scale-Location plot suggests that the model's residuals display reasonably constant variance, with slight deviations that do not undermine the model's validity.

These findings highlight the DMUFLP model's strength across varied sample sizes, as it consistently achieves lower AIC, BIC, and MAE values while maintaining competitive AUC scores. This robust performance underscores the model's capacity to balance complexity, predictive accuracy, and classification reliability, making it a preferable choice for diverse facility location challenges. The results align with recent literature emphasising model parsimony and predictive power as critical in decision-making for large-scale optimization problems. Additionally, the DMUFLP model advances existing frameworks by offering a nuanced, flexible approach that integrates environmental sustainability and dynamic capacity management. Consequently, the DMUFLP model stands as a valuable tool for decision-makers seeking to optimise facility placement within increasingly complex and interconnected supply chains, supporting sustainable and efficient operational planning.

Availability of Data and Materials

The data used in the analysis can be provided upon request.

Authors' Contributions

Conceptualisation, EA.; Methodology, EA. and GM.; Software, EA.; Validation, EA. and GM.; Formal Analysis, EA.; Investigation, EA.; Resources, EA. and GM.; Data Curation, EA.; Writing-Original Draft Preparation, EA.; Writing-Review & Editing, EA. and GM.; Visualization, EA.; Project Administration, EA.; Funding Acquisition, No Funding. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Church, R.L. and Murray, A.T. (2008) Business Site Selection, Location Analysis and GIS. Wiley. https://doi.org/10.1002/9780470432761

[2] Farahani, R.Z., SteadieSeifi, M. and Asgari, N. (2010) Multiple Criteria Facility Location Problems: A Survey. *Applied Mathematical Modelling*, **34**, 1689-1709. https://doi.org/10.1016/j.apm.2009.10.005

[3] Laporte, G., Nickel, S. and Saldanha-da-Gama, F. (2019) Introduction to Location Science. In: Laporte, G., Nickel, S. and Saldanha da Gama, F., Eds., *Location Science*, Springer, 1-21. https://doi.org/10.1007/978-3-030-32177-2_1

[4] Eiselt, H.A. and Marianov, V. (2014) A Bi-Objective Model for the Location of Landfills for Municipal Solid Waste. *European Journal of Operational Research*, **235**, 187-194. https://doi.org/10.1016/j.ejor.2013.10.005

[5] Klose, A. and Drexl, A. (2005) Facility Location Models for Distribution System Design. *European Journal of Operational Research*, **162**, 4-29. https://doi.org/10.1016/j.ejor.2003.10.031

[6] Snyder, L.V. and Daskin, M.S. (2005) Reliability Models for Facility Location: The Expected Failure Cost Case. *Transportation Science*, **39**, 400-416. https://doi.org/10.1287/trsc.1040.0107

[7] Hakimi, S.L. (1964) Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph. *Operations Research*, **12**, 450-459. https://doi.org/10.1287/opre.12.3.450

[8] Pires, A., Martinho, G., Rodrigues, S. and Gomes, M.I. (2018) Optimization in Waste Collection to Reach Sustainable Waste Management. In: Pires, A., Martinho, G., Rodrigues, S. and Gomes, M.I., Eds., *Sustainable Solid Waste Collection and Management*, Springer, 207-238. https://doi.org/10.1007/978-3-319-93200-2_12

[9] Olapiriyakul, S., Pannakkong, W., Kachapanya, W. and Starita, S. (2019) Multiobjective Optimization Model for Sustainable Waste Management Network Design. *Journal of Advanced Transportation*, **2019**, Article ID: 3612809. https://doi.org/10.1155/2019/3612809

[10] Adeleke, O.J. and Olukanni, D.O. (2020) Facility Location Problems: Models, Techniques, and Applications in Waste Management. *Recycling*, **5**, Article 10. https://doi.org/10.3390/recycling5020010

[11] Madani, M. (2018) Dynamic Facility Location with Stochastic Demand and Congestion. Master's Thesis, Concordia University. https://spectrum.library.concordia.ca/id/eprint/983852/

[12] Green, S. (2022) What You Need to Know to Start an Online Retail Business. https://blog.hubspot.com/sales/online-retail

[13] Pang, J., Yang, Z., Han, S., Wang, X., Shi, A., Qiu, D., *et al.* (2023) Distribution Network Closed-Loop Control Method Using Controllable Distributed Generation Regulation Capability. 2023 *IEEE International Conference on Predictive Control of Electrical Drives and Power Electronics* (*PRECEDE*), Wuhan, 16-19 June 2023, 1-6. https://doi.org/10.1109/precede57319.2023.10174611

[14] Ghadge, A., Yadav, G. and Nimkar, A. (2023) Facility Location Planning under Uncertainty: A Sustainable Multi-Objective Optimization Approach. *Journal of Operations Management*, **65**, 123-135.

[15] Silva, A., Aloise, D., Coelho, L.C. and Rocha, C. (2021) Heuristics for the Dynamic Facility Location Problem with Modular Capacities. *European Journal of Operational Research*, **290**, 435-452. https://doi.org/10.1016/j.ejor.2020.08.018

[16] Bertsimas, D. and Tsitsiklis, J.N. (1997) Introduction to Linear Optimization. Vol 6. Athena Scientific, 479-530.

[17] Pisinger, D. and Ropke, S. (2018) Large Neighborhood Search. In: Gendreau, M. and Potvin, J.Y., Eds., *Handbook of Metaheuristics*, Springer, 99-127. https://doi.org/10.1007/978-3-319-91086-4_4

[18] Drezner, Z. and Hamacher, H. (2004) Facility Location: Applications and Theory. Springer.

[19] Eiselt, H.A. and Sandblom, C.L. (2022) Operations Research: A Model-Based Approach. 3rd Edition, Springer.

[20] Melo, M.T., Nickel, S. and Saldanha-da-Gama, F. (2009) Facility Location and Supply Chain Management—A Review. *European Journal of Operational Research*, **196**, 401-412. https://doi.org/10.1016/j.ejor.2008.05.007

[21] Awasthi, A.K., Wang, M., Awasthi, M.K., Wang, Z. and Li, J. (2018) Environmental Pollution and Human Body Burden from Improper Recycling of E-Waste in China: A Short-Review. *Environmental Pollution*, **243**, 1310-1316. https://doi.org/10.1016/j.envpol.2018.08.037

[22] Ghaderi, H., Pishvaee, M.S. and Moini, A. (2016) Biomass Supply Chain Network Design: An Optimization-Oriented Review and Analysis. *Industrial Crops and Products*, **94**, 972-1000. https://doi.org/10.1016/j.indcrop.2016.09.027

[23] Hastie, T., Tibshirani, R. and Friedman, J. (2009) Springer Series in Statistics the Elements of Statistical Learning. *The Mathematical Intelligencer*, **27**, 83-85.

[24] Berkelaar, M., Eikland, K. and Notebaert, P. (2020) lpSolve: Interface to Lp_solve v. 5.5 to Solve Linear/Integer Programs—R Package Version 5(15) 1999.

[25] Rahal, S.S., Dems, A., Adulyasak, Y., Arslan, O. and Cordeau, J.F. (2023) Facility Location with Modular Capacity Under Demand Uncertainty: An Industrial Case Study. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4454322

[26] Burnham, K.P., Anderson, D.R. and Huyvaert, K.P. (2010) AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons. *Behavioral Ecology and Sociobiology*, **65**, 23-35. https://doi.org/10.1007/s00265-010-1029-6

[27] Faraway, J.J. (2021) Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. 2nd Edition, Chapman & Hall/CRC.

[28] Fox, J. (2015) Applied Regression Analysis and Generalized Linear Models, Third Edition. SAGE Publications, Inc. https://www.john-fox.ca/AppliedRegression/chap-25.pdf

[29] Wickham, H. (2016) Data Analysis. In: Wickham, H., Ed., *ggplot2*. *Use R!* Springer, 189-201. https://doi.org/10.1007/978-3-319-24277-4_9

[30] Bruni, M.E., Di Puglia Pugliese, L., Beraldi, P. and Guerriero, F. (2017) An Adjustable Robust Optimization Model for the Resource-Constrained Project Scheduling Problem with Uncertain Activity Durations. *Omega*, **71**, 66-84. https://doi.org/10.1016/j.omega.2016.09.009

[31] Owen, S.H. and Daskin, M.S. (1998) Strategic Facility Location: A Review. *European Journal of Operational Research*, **111**, 423-447. https://doi.org/10.1016/s0377-2217(98)00186-6

[32] Snyder, L.V. (2006) Facility Location under Uncertainty: A Review. *IIE Transactions*, **38**, 547-564. https://doi.org/10.1080/07408170500216480

[33] Laporte, G. (2009) Fifty Years of Vehicle Routing. *Transportation Science*, **43**, 408-416. https://doi.org/10.1287/trsc.1090.0301

[34] Liu, L., Yu, H., Zhao, J., Wu, H., Peng, Z. and Wang, R. (2022) Multiscale Effects of Multimodal Public Facilities Accessibility on Housing Prices Based on MGWR: A Case Study of Wuhan, China. *ISPRS International Journal of Geo-Information*, **11**, Article 57. https://doi.org/10.3390/ijgi11010057

[35] Hamuraru, M. and Bulat, V. (2022) Sustainable Development in the Global Distribution of Goods. *Journal of Eastern Europe Research in Business and Economics*, **2022**, Article ID: 606355.

[36] Montgomery, D.C. and Runger, G.C. (2010) Applied Statistics and Probability for Engineers. John Wiley & Sons.

[37] Mertler, C.A. and Vannatta Reinhart, R. (2016) Advanced and Multivariate Statistical Methods: Practical Application. Routledge.

[38] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2015) Bayesian Data Analysis. In: *Behavioral Sciences*, *Mathematics & Statistics* (3*rd Edition*), Chapman and Hall/CRC, 675.

[39] Fox, J. and Weisberg, S. (2018) An R Companion to Applied Regression. SAGE Publications. https://books.google.com.gh/books?id=uPNrDwAAQBAJ

[40] Rousseeuw, P.J. and Hubert, M. (2017) Anomaly Detection by Robust Statistics. *WIREs Data Mining and Knowledge Discovery*, **8**, e1236. https://doi.org/10.1002/widm.1236

[41] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2023) An Introduction to Statistical Learning. https://link.springer.com/book/10.1007/978-3-031-38747-0

[42] Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005) Applied Linear Statistical Models. 5th Edition, McGraw-Hill.

[43] Hastie, A.T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, Springer Nature. https://link.springer.com/book/10.1007/978-0-387-84858-7

[44] Burnham, K.P. and Anderson, D.R. (2014) P Values Are Only an Index to Evidence: 20th-vs. 21st-Century Statistical Science. *Ecology*, **95**, 627-630.

[45] Gelman, A., Hwang, J. and Vehtari, A. (2013) Understanding Predictive Information Criteria for Bayesian Models. *Statistics and Computing*, **24**, 997-1016. https://doi.org/10.1007/s11222-013-9416-2

[46] Hastie, T., Tibshirani, R. and Tibshirani, R. (2020) Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. *Statistical Science*, **35**, 579-592. https://www.jstor.org/stable/pdf/26997932.pdf

[47] Zhao, Y., Wang, C., Zhang, L., Chang, Y. and Hao, Y. (2021) Converting Waste Cooking Oil to Biodiesel in China: Environmental Impacts and Economic Feasibility. *Renewable and Sustainable Energy Reviews*, **140**, Article ID: 110661. https://doi.org/10.1016/j.rser.2020.110661

[48] Pang, Z. and Zhou, Y. (2023) Risk-Averse Facility Location Optimization for Sustainable Supply Chains: A Hybrid Stochastic Programming Approach. Sustain. *Cities and Society*, **85**, 104-132.

[49] Srivastava, A. and Jha, P.K. (2023) A Multi-Model Forecasting Approach for Solid Waste Generation by Integrating Demographic and Socioeconomic Factors: A Case Study of Prayagraj, India. *Environmental Monitoring and Assessment*, **195**, Article No. 768. https://doi.org/10.1007/s10661-023-11338-y