International Journal of Combinatorial Optimization Problems and Informatics, 16(3), May-Aug 2025, 216-228. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i3.608

www.editada.org

Integrated Allocation of Gasoline Stations and Bio-gasoline Facilities: A Logistical Approach for Sustainable Gasoline Distribution in Mexico

Santiago Omar Caballero Morales[1], Gladys Bonilla Enriquez[2], Felipe de Jesus Trujillo Romero[3]
[1] Universidad Popular Autonoma del Estado de Puebla A.C. - UPAEP University, Facultad de Ingeniería Industrial y Logística, Puebla, Puebla, 72410, Mexico
[2] Tecnologico Nacional de Mexico TECNM / Instituto Tecnológico de Puebla ITP, Puebla, Puebla, 72220, Mexico
[3] Universidad de Guanajuato - Department of Engineering in Electronics, Salamanca, Guanajuato, 36787, Mexico
santiagoomar.caballero@upaep.mx, gladys.bonilla@puebla.tecnm.mx, fdj.trujillo@ugto.mx

Abstract. Optimizing the allocation of gasoline stations to refineries and depots improves fuel distribution efficiency and service coverage, particularly under contingency scenarios. This study presents a hybrid metaheuristic approach that integrates *k*-Means clustering with Simulated Annealing (SA) and Tabu Search (TS) to simultaneously optimize station allocation and plan the expansion of biofuel facilities. The model considers geographic distribution, capacity constraints, balanced service loads, and the costs associated with infrastructure expansion. Using actual data from Mexico, the algorithm was applied to evaluate the integration of new bio-gasoline plants into the existing fuel network. Results show that the addition of 17-18 new biofuel facilities can significantly reduce the total service distance and balance the workload across the network, thereby easing the demand for existing infrastructure. This contributes to both logistical efficiency and environmental sustainability. The proposed approach offers a practical decision-support tool for policymakers and energy planners working to enhance national fuel distribution systems and advance renewable energy strategies.

Keywords: Gasoline Distribution, Bio-fuel Plants, *k*-Means Algorithm, Tabu Search, Simulated Annealing

Article Info
Received January 20, 2025
Accepted April 01, 2025

1 Introduction

The efficient distribution of gasoline is essential for ensuring energy security and driving economic activity in any nation. A well-functioning distribution network depends not only on the logistical delivery of fuel but also on the production and storage infrastructure that supports it. However, these networks face significant challenges, including rising demand, geographic dispersion, and capacity constraints in production facilities and storage systems. Additionally, environmental concerns linked to traditional fossil fuels further complicate the distribution landscape, necessitating innovative approaches to modernize and optimize these critical systems (Hilpert et al., 2015; Ihsan et al., 2020). These issues are particularly pronounced in Mexico, a country that relies heavily on imported gasoline and has a highly dispersed energy distribution system (Vivoda et al., 2023). Addressing these challenges requires innovative approaches that enhance the efficiency of existing infrastructure and integrate sustainable energy solutions.

Bio-gasoline, derived from renewable biomass such as jatropha, sunflower, and recycled oils, is a sustainable alternative to traditional fossil fuels (Paredes-Cervantes et al., 2020; Vega et al., 2024). Its ability to reduce greenhouse gas emissions - used alone or blended with diesel or gasoline - and its high biodegradability offer significant environmental advantages (Luque et al., 2010; Naimah et al., 2020; Bradu et al., 2023). Blending bio-gasoline with conventional fuels further improves combustion efficiency while enhancing sustainability, making it a vital component in advancing cleaner energy solutions for the transportation sector (Zandie et al., 2022). These qualities make it a promising option for reducing the ecological footprint of fuel use, particularly in applications where sustainability is a priority (Ng & Maravelias, 2016; Perez-Lechuga et al., 2019; Aba et al., 2022).

From both infrastructure and social development perspectives, the existing gasoline production and distribution systems are highly compatible with bio-gasoline, enabling seamless integration with minimal modifications. This compatibility minimizes the need for significant new infrastructure investments, providing a cost-effective pathway toward sustainable fuel solutions (Kim et al., 2019; Naimah et al., 2020). Furthermore, bio-gasoline production can drive rural economic growth by creating employment opportunities, supporting local agriculture, and reducing dependence on imported gasoline, contributing to enhanced energy security and social development (Silalahi et al., 2020).

Efficient bio-gasoline distribution and production are critical to developing a sustainable and economically viable alternative fuel supply chain. Bio-gasoline logistics involve a multi-stage process that includes feedstock cultivation, biomass transportation, biofuel production, storage, and distribution to end-users. Each stage presents unique challenges that must be addressed to ensure the reliability and cost-effectiveness of the supply chain (Poh-Ying et al., 2020; Leon-Olivares et al., 2020).

One of the most critical aspects of bio-gasoline logistics is the strategic location of production and distribution facilities. Facility locations directly influence transportation distances, production costs, and network efficiency. Placing facilities near feedstock sources reduces raw material transportation costs while positioning them closer to demand centers minimizes the distance bio-gasoline must travel to reach consumers. This dual consideration of feedstock and demand is essential for optimizing logistics networks and enhancing sustainability (Duarte et al., 2012; Hong & Mwakalonge, 2020).

Moreover, efficient bio-gasoline logistics require robust supply chain strategies, including optimizing transportation routes, inventory management, and balancing production capacity with demand fluctuations. Advanced decision-making tools and models, such as mixed-integer programming and geographic information systems (GIS), are increasingly being used to effectively design and manage bio-gasoline distribution networks (Lim & Ouyang, 2016). These tools enable stakeholders to assess various scenarios, minimize costs, and enhance resilience to disruptions.

By integrating production, storage, and distribution strategies with a focus on logistical efficiency, bio-gasoline supply chains can overcome challenges related to geographic dispersion, seasonal variability in feedstock supply, and fluctuations in market demand (Aba et al., 2022). This integration not only enhances the economic viability of bio-gasoline but also contributes to the broader goal of transitioning to cleaner and more sustainable energy solutions (Laners et al., 2015; Kim et al., 2018).

In this context, the present work advances the logistical framework for gasoline distribution by establishing a comprehensive approach to redesign the network, aiming to improve service coverage, minimize travel distances, and integrate bio-gasoline production facilities. Leveraging real-world data from Mexico, this study simultaneously addresses the allocation of gasoline stations to existing refineries and depots while incorporating future bio-gasoline infrastructure. The hybrid *k*-Means algorithm with Simulated Annealing (SA) and Tabu-Search (TS) is designed to incorporate essential factors such as geographic distribution, balanced load allocation, and strategic facility placement. This approach can reduce the service load on existing infrastructure, promoting sustainability and environmental benefits in fuel supply chains.

2. Strategies for Bio-Gasoline Distribution Optimization

Strategic facility placement is crucial for efficient bio-gasoline logistics, as it directly impacts costs and supply chain (SC) performance. Positioning facilities near feedstock sources lowers raw material transport costs, while proximity to demand centers minimizes delivery distances. Allocation algorithms enhance this process by identifying optimal facility locations and efficiently distributing resources to the nearest production and distribution centers, ensuring cost-effective and streamlined operations.

Recent studies have investigated various optimization techniques for allocating bioenergy facilities and distributed generation units, with a growing focus on clustering-based approaches, such as k-Means, for facility location. *k*-Means clustering is a widely used unsupervised learning method that partitions spatial data into clusters, minimizing intra-cluster variance and improving logistical efficiency. In facility location problems, *k*-Means clustering is widely applied in logistics to group demand points and resource sources, facilitating strategic facility placement that reduces transportation costs, optimizes resource allocation, and effectively manages capacity constraints, ultimately improving supply chain efficiency and performance (Caballero-Morales et al., 2018; Razi, 2019).

Castillo-Villar (2014) provided a comprehensive review of metaheuristic optimization methods, including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Simulated Annealing (SA), and Tabu Search (TS), focusing on their application in supply chain network design (SCND) and facility location. While these methods offer robust optimization capabilities, clustering-based approaches such as *k*-Means provide computationally efficient solutions for large-scale spatial data, especially in early-stage planning.

Zhang et al. (2017) developed an integrated GIS-based approach combining *k*-Means clustering with optimization modeling to determine cost-effective bioethanol facility locations. Their study demonstrated that *k*-Means effectively grouped supply and demand nodes, reducing bioethanol SC costs under demand and supply uncertainties. The results showed that clustering facilities based on geographic proximity mitigated cost increases resulting from fluctuations in biomass availability.

Rodrigues-Costa et al. (2020) introduced a two-phase spatial analysis method that integrated *k*-Means clustering with the Maximize Capacitated Coverage technique to optimize the location of bioenergy facilities. They used AHP-weighted criteria and a Natural Breaks classification to identify potential facility locations, demonstrating that clustering-based selection reduced biomass transport distances by 31% compared to traditional optimization methods.

Sharma et al. (2024) proposed a multi-objective framework that incorporated *k*-Means clustering into the allocation of bio-refineries, depositories, and harvesting centers for a sustainable micro-algal biodiesel supply chain. Their study showed that *k*-Means-based facility location lowered overall economic costs and environmental impact, with facility installation accounting for 59% of total costs and 67% of environmental burden. Sensitivity analysis confirmed that clustering-based facility placement improves adaptability to changing supply chain conditions, making it a feasible approach for industry stakeholders.

Although these studies underscore *k*-Means as a valuable tool for facility location optimization, the algorithm has inherent limitations. Random centroid initialization can lead to inconsistent convergence, affecting solution stability. Moreover, *k*-Means requires a predefined number of clusters, which can result in suboptimal cluster shapes and increased sensitivity to outliers. For balanced assignment and capacity constraints, *k*-Means requires modifications to its clustering process, including constraint-aware centroid updates, penalty-based balancing, and hybrid optimization techniques.

This work adapts *k*-Means for balanced station allocation, optimizing service load distribution and minimum distance. Unlike Caballero-Morales et al. (2018), this approach integrates Simulated Annealing (SA) and Tabu Search (TS) to enhance performance using the balanced score metric introduced by Caballero-Morales et al. (2024). Additionally, the model incorporates the nationwide network of gasoline stations and depots to develop an optimized logistical framework.

3. Mathematical Formulation

The *k*-Means algorithm clusters N data points into k groups, minimizing within-cluster variance to ensure similarity among points (Hung et al., 2005). It partitions data into k distinct, non-overlapping clusters, with centroids iteratively updated based on point assignments. As shown in Figure 1, each point is assigned to the nearest centroid, and cluster centers are recalculated until convergence (Caballero-Morales et al., 2018).

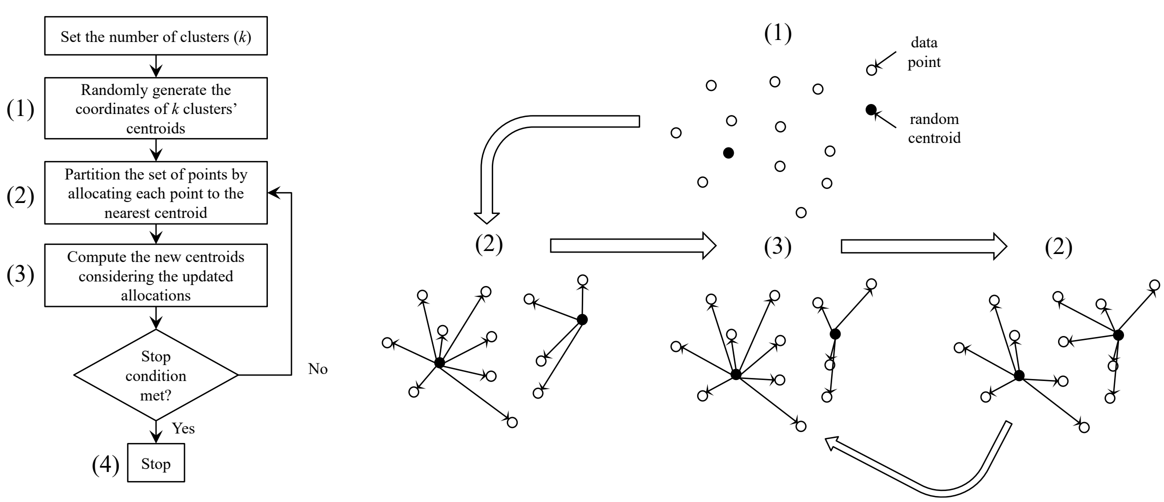


Fig. 1. Structure of the standard *k*-Means algorithm (example with k=2 clusters and N=12 data points).

3.1 Balanced Score Metric

The allocation generated by the k-Means algorithm leads to a minimum distance clustering. However, as discussed in (Caballero-Morales et al., 2024), the k-Means algorithm can create clusters with high allocation load while allocating few data points to other clusters. This unbalanced allocation may lead to a high coefficient of variability:

$$CV = \frac{\sigma}{\mu}, \tag{1}$$

where σ and μ are the standard deviation and average number of data points allocated to each cluster. Within gasoline distribution, a high CV implies a large variability in allocations, meaning that some facilities are overloaded with gasoline stations while others have few stations to serve. In the presence of a disruptive event (e.g., pipeline closure), this variability can result in significant economic losses for the stations allocated to the affected and overloaded facilities. In such cases, a balanced allocation can reduce these risks.

An upper bound H is set for the maximum number of allocated gasoline stations per facility to control allocation variability. Note that adding this capacity restriction can compromise the minimum distance aspect. This is because some stations must be allocated to facilities that are available far away to comply with the H restriction.

As presented in (Caballero-Morales et al., 2024), if D is the minimum distance of the unrestricted allocation problem, and D_H is the minimum distance of the H restricted allocation problem, then a minimum distance error gap can be computed as:

$$e_H = \frac{D_H - D}{D}. \tag{2}$$

If the H-restricted allocation results in a minimum distance equal to that of the unrestricted case (i.e., no increase in service distance), then $e_H = 0$. In contrast, if the restricted allocation increases service distance, then $e_H > 0$. Because each H-restriction leads to allocations that can be assessed with e_H and $CV \rightarrow CV_H$, integrating both metrics can balance the allocation of stations to facilities appropriately while reducing its impact on the minimum distance aspect. As analyzed in Caballero-Morales et al. (2024), an appropriate balance can be achieved when $CV_H \approx e_H$.

3.2 Objective Function and Restrictions

Once the balanced score metric is defined, the mixed-integer linear programming (MILP) model for the facility location and allocation problem is formulated. For this purpose, the parameters and variables presented in Table 1 are defined.

Table 1. Set of variables and parameters of the MILP model.

Parameters and Variables	Description
I	Set of gasoline stations
J_{fixed}	Set of pre-existing facilities (fixed locations)
J_{new}	Set of candidate locations for new facilities
$J = J_{fixed} \cup J_{new}$	Set of facilities
H	Capacity limit of allocated stations per facility.
D	Total distance of the unconstrained case (without capacity restriction)
D_H	Total distance of the H-constrained case (with capacity restriction)
β	Penalty score for $e_H > 0$
γ	Penalty score for $CV > 0$
λ	Cost of new facility.
$x_{ij} \in \{0,1\}$	$x_{ij} = 1$ if station i is assigned to facility j; $x_{ij} = 0$ otherwise.
$y_j \in \{0,1\}$	$y_j = 1$ if facility j (existing or new depot) is active; otherwise, $y_j = 0$.
$C_i = (X_i, Y_i)$	X-Y coordinates of station i
$C_j = (X_j, Y_j)$	X-Y coordinates of facility j
N_j	Number of stations assigned to facility j
d_{ij}	Distance between station i and facility j
$v \in [0, 1, 2, \ldots, UB\}$	Number of new facilities with a fixed upper bound UB.

The objective function is set to minimize the total distance while achieving minimum e_H, CV, and controlling the number of new facilities:

$$\min \sum_{i\in I}\sum_{j\in J} x_{ij}d_{ij} + \beta e_H + \gamma CV + \lambda v. \tag{3}$$

Note that $\sum_{i\in I}\sum_{j\in J} x_{ij}d_{ij} = D_H$, and D is computed by the unconstrained k-Means algorithm. The β and γ penalties are scale factors to prevent D_H from dominating the objective function, making the contributions of e_H and CV negligible. Hence, these parameters are computed as fractions within the range of D_H. Regarding λ, it represents the activation or opening cost of each new facility (v). It is important to consider the necessary investment in network design to maintain control over the new facilities (Paredes-Cervantes et al., 2020; Sosa-Rodriguez & Vazquez-Arenas, 2021).

The restrictions are listed as follows:

$$\sum_{j\in J} x_{ij} = 1, \quad \forall i \in I. \tag{4}$$

$$\sum_{i\in I} x_{ij} \leq H, \quad \forall j \in J. \tag{5}$$

$$x_{ij} \leq y_j, \quad \forall i \in I, \forall j \in J. \tag{6}$$

$$y_j = 1, \quad \forall j \in J_{fixed}. \tag{7}$$

$$\sum_{j\in J_{new}} y_j = v, \quad 0 \leq v \leq UB. \tag{8}$$

$$\frac{1}{|J|}\sum_{i\in I} x_{ij} - \varepsilon \leq \sum_{i\in I} x_{ij} \leq \frac{1}{|J|}\sum_{i\in I} x_{ij} + \varepsilon, \quad \forall j \in J. \tag{9}$$

(4) establishes that each station is assigned to exactly one facility; (5) limits the number of allocated stations per facility to a maximum of H; (6) enforces facility activation: a facility must be active if it has allocated stations; (7) mandates that pre-existing (fixed) facilities are always activated; (8) establishes that the number of new facilities (v) must not exceed the upper bound UB; and (9) is a balance constraint to ensure similar workloads across facilities with ε acting as a deviation tolerance factor: a small ε enforces nearly equal station allocations, while a larger ε allows more flexibility in distribution.

Finally, the new facility locations (X_j, Y_j) are computed as centroids of the allocated stations with coordinates (X_i, Y_i):

$$X_j = \frac{\sum_{i\in I} x_{ij}X_i}{N_j}, \quad \forall j \in J_{new} \tag{10}$$

$$Y_j = \frac{\sum_{i\in I} x_{ij}Y_i}{N_j}, \quad \forall j \in J_{new} \tag{11}$$

4. Hybrid Metaheuristic

The k-Means algorithm is initially executed to initialize the search process and establish the reference value for D. Subsequently, Tabu Search (TS) and Simulated Annealing (SA) are applied to optimize v while ensuring that CV and e_H remain as small as possible.

4.1 *k*-Means + TS

Figure 2 presents the structure of the hybrid metaheuristic k-Means + TS. TS is a metaheuristic optimization technique that enhances local search by using a memory structure to avoid cycling back to previously visited solutions. As presented in Figure 2, the TS algorithm begins with an initial unconstrained solution. It iteratively explores a set of neighboring constrained solutions generated through predefined moves, such as reassigning stations to different facilities or swapping station allocations.

At each iteration, the best non-tabu move is selected based on the objective function (3), which minimizes transportation costs while ensuring capacity constraints and workload balance are met. To prevent immediate reversals of recent moves, TS maintains a tabu list, which records previously applied moves and prohibits them for a fixed number of iterations τ known as the tabu tenure. This restriction ensures that the search process does not stagnate in locally optimal regions, instead encouraging exploration of diverse areas of the solution space. The algorithm terminates after reaching a maximum number of T iterations.

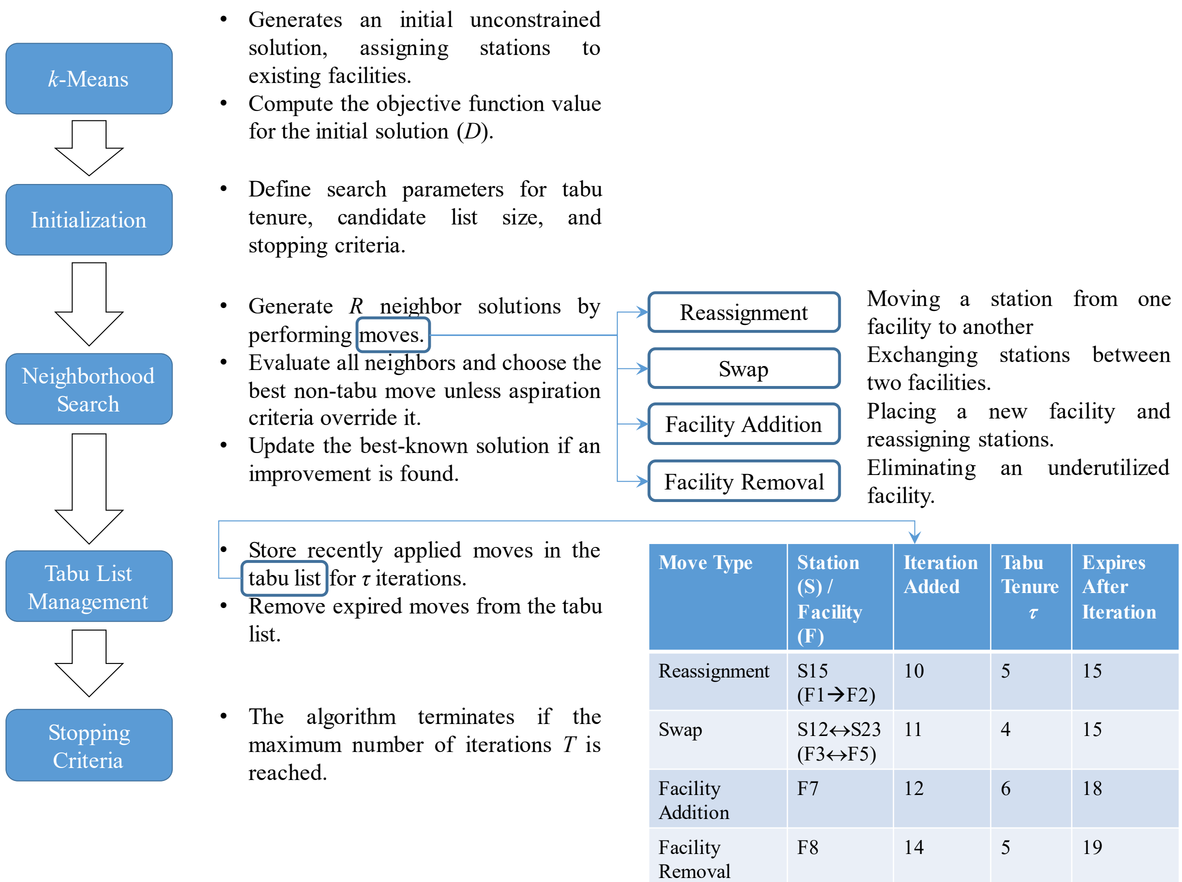


Fig. 2. Structure of the hybrid *k*-Means + TS algorithm.

4.2 *k*-Means + SA

Figure 3 presents the structure of the hybrid metaheuristic *k*-Means + SA. SA is a probabilistic metaheuristic optimization algorithm inspired by the annealing process in metallurgy, where a material is heated and gradually cooled to achieve a stable crystalline structure. In the context of combinatorial optimization, SA mimics this process by iteratively refining solutions as it explores the solution space while avoiding premature convergence to local optima.

SA begins with an initial solution and an initial temperature parameter K_0, which determines the probability of accepting worse solutions. At each iteration, a neighboring solution is generated by applying a perturbation move, such as reassigning a station to a different facility, swapping station allocations, or adding/removing facility locations. The objective function evaluates the new solution, and it is accepted if it improves the current solution. However, if the new solution is worse, it is accepted with a probability given by the Metropolis criterion:

$$P = e^{(f(Z)-f(Z'))/K}, \tag{12}$$

where $f(Z)$ and $f(Z')$ represent the objective function values of the current and new solutions, respectively, and K is the current temperature. This probabilistic acceptance mechanism enables SA to escape local optima by temporarily accepting suboptimal solutions, allowing exploration of a broader search space.

The effectiveness of SA relies on a carefully designed cooling schedule that gradually reduces the temperature K to refine the search around promising solutions. The temperature is updated at each iteration according to a predefined cooling function, typically expressed as $K=\alpha K$, where α is a cooling rate parameter (e.g., 0.95). A slower cooling rate encourages extensive exploration in the early stages, while a faster reduction focuses on exploitation. The stopping criteria for SA are typically defined based on reaching a minimum temperature threshold K_{min}, a maximum number of iterations, or stagnation in solution improvement.

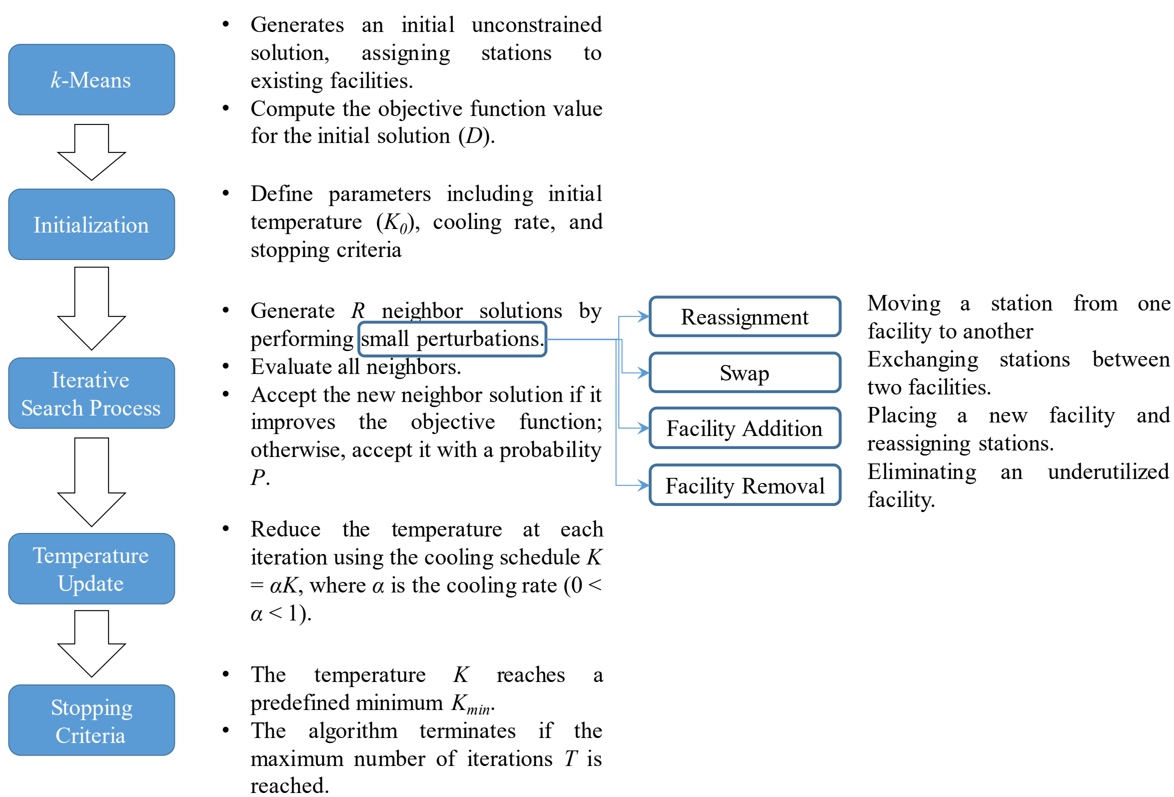


Fig. 3. Structure of the hybrid *k*-Means + SA algorithm.

5. Results

The hybrid metaheuristics were implemented using Octave and MATLAB on an HP Workstation with an Intel Zeon CPU at 3.40 GHz and 8 GB RAM. Table 2 presents the TS/SA parameters for the hybrid metaheuristics.

Table 2. Parameters of the *k*-Means + TS/SA Metaheuristics

k-Means + TS	
T	500
τ	10
R	10
k-Means + SA	
K_0	1000
Stop Condition: K_{min}	10
α	0.95
Objective Function	
β, γ, λ	100, 100, 50

As a test instance, we considered the network of approximately 12,200 gasoline stations supplied by 87 gasoline depots and six refineries across Mexican territory (Cuellar, 2024). For this approach, all gasoline stations were considered suitable for receiving standard gasoline and bio-gasoline. This is because one key advantage of bio-gasoline is its compatibility with the existing infrastructure and vehicle fleets designed for conventional gasoline (Lee-Stafford et al., 2019; Dahlgren, 2022). Figure 4 presents the locations of the stations and current depot facilities.

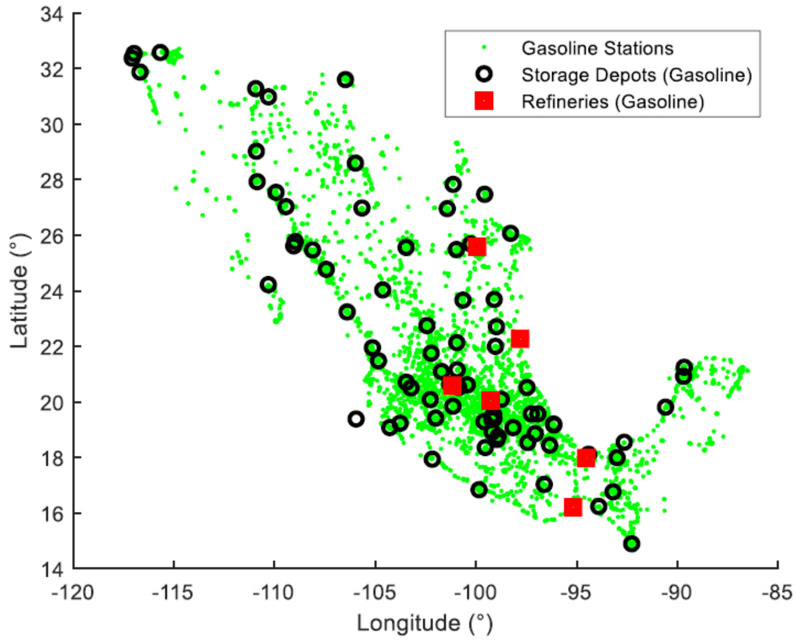


Fig. 4. Current location of gasoline stations, refineries and storage depots in Mexico.

5.1 Determining *H* and *D*

As mentioned in Section 3.1, H plays a crucial role in balancing station allocation. To determine an appropriate value for H, a capacitated k-Means algorithm was applied while accounting for the existing depot facilities. Figure 5 presents the effect of H on the balancing metrics CV_H and e_H with the existing facilities. As observed, although CV_H reaches a minimum value at approximately H=125, e_H remains higher than 1.0. The minimum value for e_H is achieved at $H \geq 260$ with a slight increase in CV_H. $H = 200$ is selected for the metaheuristics because at this value $CV_H \approx e_H$. Regarding the baseline value for D, it was estimated with the unrestricted k-Means algorithm. This led to $D = 4084.07$.

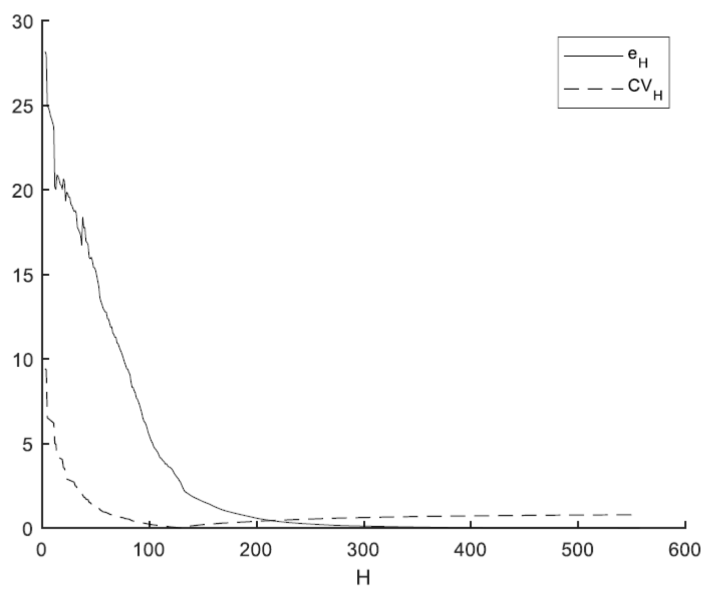
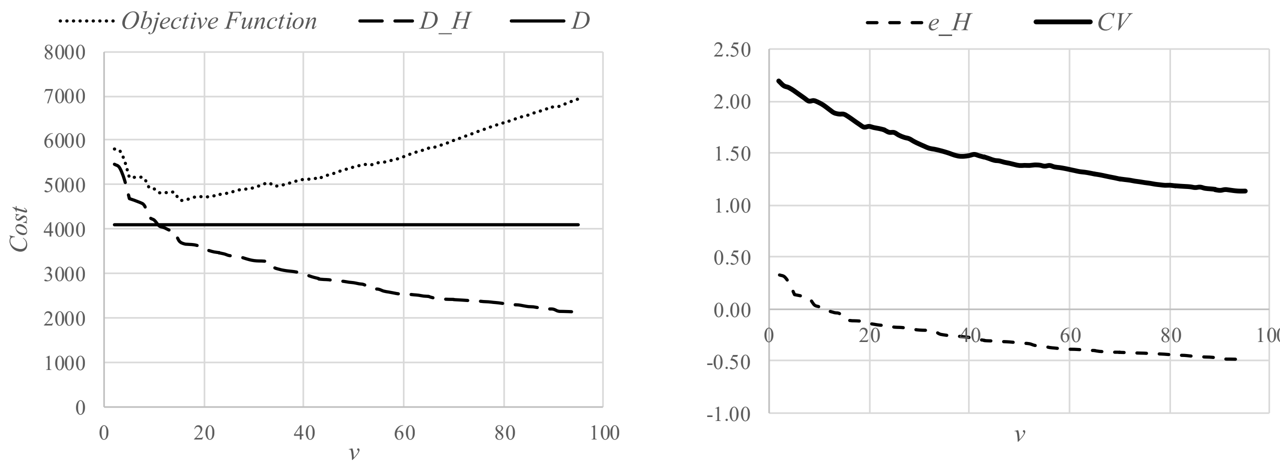


Fig. 5. Effect of H on the balance metrics CV_H and e_H (pre-existing facilities).

5.2 Performance of the *k*-Means + TS Algorithm

Figure 6 illustrates the performance of the *k*-Means + TS algorithm. Figure 6(a) illustrates the convergence of the Objective Function (3) and the total service distance D_H as the number of additional facilities, v, increases. The reference value D=4084.07, computed using the unrestricted *k*-Means algorithm without additional facilities, remains constant across all values of v. As expected, operational costs increase with the additional infrastructure. However, this increase is more pronounced when the number of additional facilities is either small (v<10) or large (v>20). In the case of small v, the limited number of facilities results in high total service distances and unbalanced station allocations, which drive operational costs above the savings from reduced infrastructure. Conversely, for large values of v, although service distances are reduced, the cost of installing and operating a greater number of facilities outweighs these savings, leading to higher overall costs. For $v \in [10,20]$, the model achieves a balance where costs are minimized while also reducing total service distance, resulting in $D_H < D$. This suggests that within this range, the trade-off between infrastructure expansion and service efficiency is optimized.



(a) (b)

Fig. 6. Performance of the *k*-Means+TS algorithm.

Figure 6(b) presents the changes in e_H and CV as v increases. CV continually decreases while e_H transitions from positive to negative values. This transition is due to D_H becoming larger and smaller when compared to D. Note that the transition of e_H to negative values is at $v \approx 10$. Figure 7 presents the location and allocation results with v = 18 and D_H = 3769. Table 3 presents the recommended locations for the v = 18 new bio-gasoline depot/refinery facilities.

Table 3. Locations of v=18 bio-gasoline facilities as estimated by the *k*-Means+TS algorithm

Longitude	Latitude	Longitude	Latitude	Longitude	Latitude
-97.8182	16.5996	-113.0840	31.0140	-111.6567	25.2450
-87.3915	20.4989	-103.1997	20.8204	-105.2277	20.6771
-100.3293	25.7799	-99.1534	17.4024	-109.6964	29.8342
-107.1425	28.6341	-113.8635	27.8379	-103.7562	20.3181
-107.7347	30.3728	-109.8473	23.0728	-97.6284	25.7418
-104.4435	29.1922	-105.3954	28.0657	-108.2434	27.9535

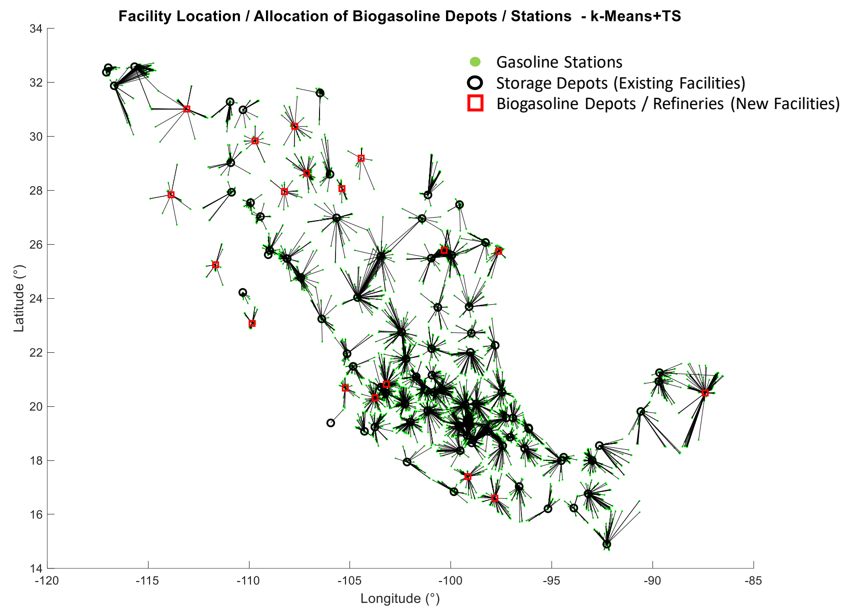


Fig. 7. Location of v=18 new bio-gasoline facilities and allocation of gasoline stations as estimated by the k-Means+TS algorithm.

5.3 Performance of the *k*-Means + SA Algorithm

Figure 8 illustrates the performance of the k-Means + SA algorithm. Figure 8(a) illustrates the convergence of the Objective Function (3) and the total service distance D_H as the number of additional facilities, v, increases. As with the k-Means + TS algorithm, the increase in operational costs is more pronounced when the number of additional facilities is either small (v<10) or large (v>30). For $v \in [10,30]$, the model achieves a balance where costs are minimized while also reducing total service distance, resulting in $D_H < D$. Figure 8(b) presents the changes in e_H and CV as v increases. The pattern closely resembles the performance of the k-Means + TS algorithm. Figure 9 illustrates the allocation results with v=17 and D_H=3861. Table 4 presents the recommended locations for the v = 17 new bio-gasoline depot/refinery facilities.

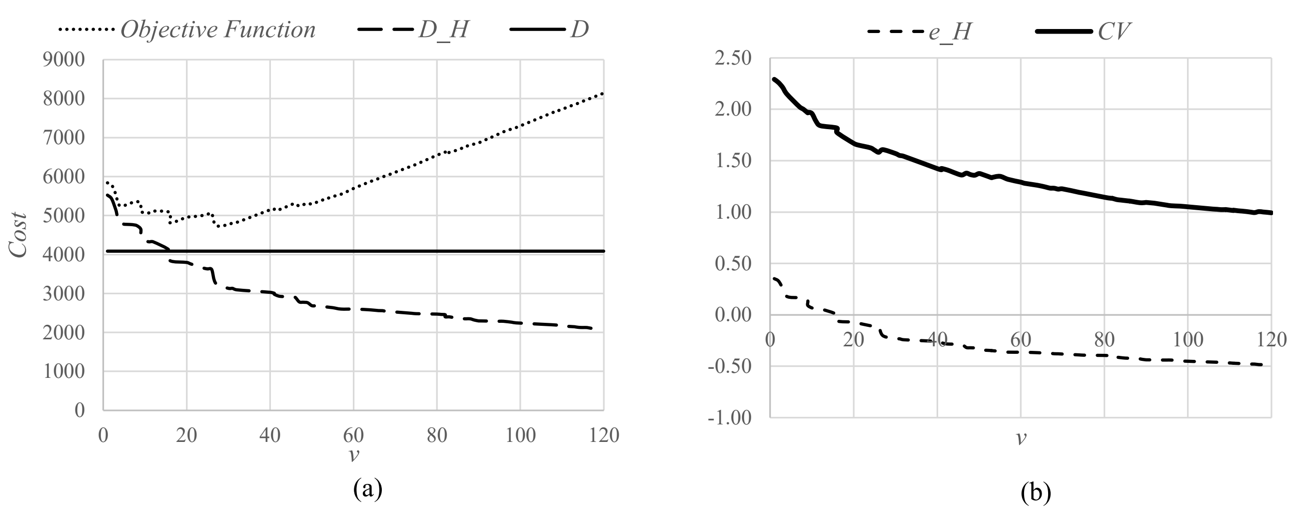


Fig. 8. Performance of the k-Means+SA algorithm.

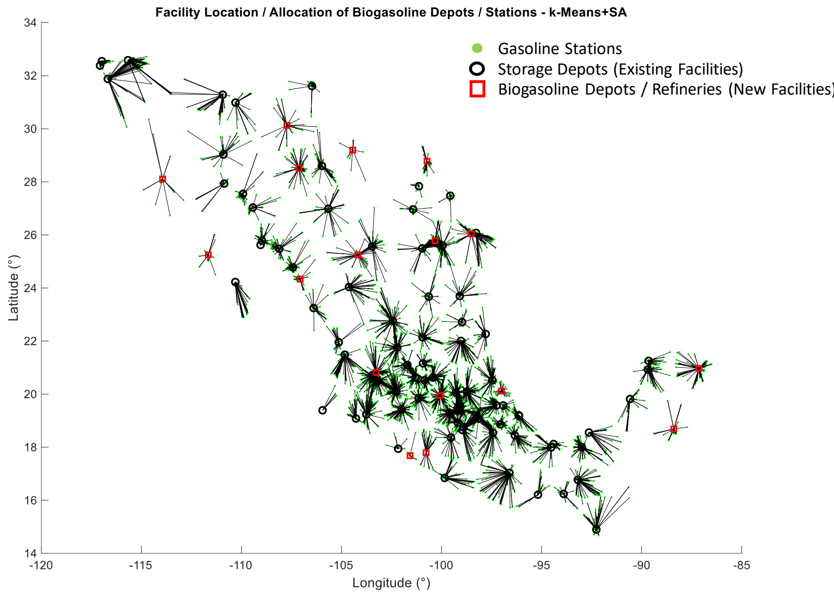


Fig. 9. Location of v=17 new bio-gasoline facilities and allocation of gasoline stations as estimated by the k-Means+SA algorithm.

Table 4. Locations of v=17 bio-gasoline facilities as estimated by the k-Means+SA algorithm

Longitude	Latitude	Longitude	Latitude	Longitude	Latitude
-100.3293	25.7799	-113.9416	28.1007	-103.2726	20.8157
-107.7313	30.1235	-107.1312	28.5085	-104.4435	29.1922
-100.0845	19.9520	-100.7312	28.7714	-88.3955	18.6862
-107.0689	24.3324	-98.5220	26.0340	-101.5791	17.6777
-97.0129	20.1107	-87.1509	20.9656	-100.7777	17.7931
-104.1886	25.2448	-111.6567	25.2450		

6. Conclusions

This work presented hybrid metaheuristic approaches to address the complex problem of integrating the bio-gasoline production infrastructure into the existing network of refineries and depots in Mexico. Specifically, k-Means clustering was adapted and combined with Tabu Search (TS) and Simulated Annealing (SA) to optimize both the allocation of gasoline stations to current facilities and the placement of new bio-gasoline plants, under constraints of service distance, capacity, and balanced load distribution. The method aimed to minimize operational costs while supporting the strategic transition toward more sustainable fuel logistics.

Using real geographical and infrastructure data, the approach demonstrated that the addition of 17-18 new bio-gasoline plants can achieve a favorable balance between infrastructure expansion and distribution efficiency. This configuration not only can reduce the total service distance D_H, but also can contribute to a more uniform allocation of demand across facilities, ensuring scalability and operational feasibility.

Nevertheless, the performance of the proposed hybrid metaheuristics could be improved through adaptive parameter tuning, as the current use of static cooling schedules (SA) and tabu tenures (TS) may limit early exploration or overly constrain the search in later stages. Integrating self-adaptive mechanisms - such as reinforcement learning or dynamic tabu tenure - would enable the algorithm to adjust based on performance feedback. Additionally, incorporating multi-objective optimization to balance trade-offs among distance, installation costs, and workload distribution, along with introducing stochastic elements like biomass supply variability or demand fluctuations, would enhance model realism and robustness for long-term strategic planning.

Regarding the applicability of these algorithms, the results demonstrate their potential as a valuable complementary decision-support tool for policymakers and industry stakeholders involved in the deployment of sustainable fuels. By providing data-driven insights for infrastructure planning and optimization, the proposed approach can support strategic initiatives such as:

- Establish a supply chain for the emerging market for all by-products generated during the bio-gasoline transformation process. For example, waste materials from jatropha biomass can be utilized to produce methane, which can be used to generate electricity for the self-consumption of the bio-refinery (Paredes-Cervantes et al., 2020).

- Develop national policies, such as the elimination of the value-added tax (VAT) and the special tax on products and services (Impuesto Especial de Productos y Servicios, IEPS), to promote the success of the renewable energy and biofuel markets (Ayala-Bautista, 2013; Sosa-Rodriguez & Vazquez-Arenas, 2021).

- Plan the strategic biomass production infrastructure to supply the new bio-gasoline plants. This can generate jobs and economic development in the regions close to the bio-gasoline plants.

References

Aba, M. M., Le-Roux, G. A. C., & Menezes, B. C. (2022). Integrated ethanol and gasoline supply chain planning under environmental constraints: A case study of Brazil. *Computers & Chemical Engineering, 164*, Article 108287.

Bradu, P., Biswas, A., Nair, C., Sreevalsakumar, S., Patil, M., Kannampuzha, S., Mukherjee, A. G., Wanjari, U. R., Renu, K., Vellingiri, B., & Gopalakrishnan, A. V. (2023). Recent advances in green technology and Industrial Revolution 4.0 for a sustainable future. *Environmental Science and Pollution Research, 30*, 124488–124519.

Caballero-Morales, S. O., Barojas-Payan, E., Sánchez-Partida, D., & Martínez-Flores, J. L. (2018). Extended GRASP-Capacitated K-Means clustering algorithm to establish humanitarian support centers in large regions at risk in Mexico. *Journal of Optimization*, 1–14.

Caballero-Morales, S. O., Bonilla-Enríquez, G., & Martínez-Flores, J. L. (2024). Approaches to improve the distribution aspects of gasoline in Mexico. In Garrido et al. (Eds.), *Proceedings of Computational Logistics* (pp. 117–129). Springer-Verlag.

Castillo-Villar, K. K. (2014). Metaheuristic algorithms applied to bioenergy supply chain problems: Theory, review, challenges, and future. *Energies, 7*, 7640–7672.

Coronado-Ortega, M. A., Montero-Alpírez, G., Eliezer, A., García-González, C., Cerezo-Román, J., Pérez-Pelay o, L. J., & Ayala-Bautista, J. R. (2013). Materials technological challenges for the biodiesel industry development in Mexico. In Mendez-Vilas, A. (Ed.), *Materials and Processes for Energy: Communicating Current Research and Technological Developments* (pp. 279–288). Formatex Research Center.

Cuellar, J. (2024). *Ubicación de gasolineras y precios comerciales de gasolina y diésel por estación*. Gobierno de México: Comisión Reguladora de Energía (CRE). https://datos.gob.mx/busca/dataset/ubicacion-de-gasolineras-y-precios-comerciales-de-gasolina-y-diesel-por-estacion

Dahlgren, S. (2022). Biogas-based fuels as renewable energy in the transport sector: An overview of the potential of using CBG, LBG and other vehicle fuels produced from biogas. *Biofuels, 13*(5), 587–599.

Duarte, A., Sarache, W., & Cardona, C. (2012). Cost analysis of the location of Colombian biofuels plants. *DYNA, 79*(176), 71–80.

Hilpert, M., Mora, B. A., Ni, J., Rule, A. M., & Nachman, K. E. (2015). Hydrocarbon release during fuel storage and transfer at gas stations: Environmental and health effects. *Current Environmental Health Reports, 2*, 412–422.

Hong, J. D., & Mwakalonge, J. L. (2020). Biofuel logistics network scheme design with combined data envelopment analysis approach. *Energy, 209*, Article 118441.

Hung, M. C., Wu, J., Chang, J. H., & Yang, D. L. (2005). An efficient k-means clustering algorithm using simple partitioning. *Journal of Information Science and Engineering, 21*, 1157–1177.

Ihsan, A. J., Mustafa, A. E., & Akeel, A. A. (2020). The impact of the distribution of fuel stations on the urban environment: Case study Kut-Iraq. *IOP Conference Series: Materials Science and Engineering, 737*, 012024.

Kim, S., Dale, B. E., Jin, M., Thelen, K. D., Zhang, X., Meier, P., Reddy, A. D., Jones, C. D., Izaurralde, R. C., Balan, V., Runge, T., & Sharara, M. (2019). Integration in a depot-based decentralized biorefinery system: Corn stover-based cellulosic biofuel. *Global Change Biology Bioenergy, 11*(7), 871–882.

Kim, S., Zhang, X., Dale, B. E., Reddy, A. D., Jones, C. D., & Izaurralde, R. C. (2018). EISA-compliant ethanol fuel from corn stover in a depot-based decentralized system. *Biofuels, Bioproducts and Biorefining, 12*(5), 873–881.

Laners, P., Tan, E. C. D., Searcy, E. M., Scarlata, C. J., Cafferty, K. G., & Jacobson, J. J. (2015). Strategic supply system design: A holistic evaluation of operational and production cost for a biorefinery supply chain. *Biofuels, Bioproducts and Biorefining, 9*(6), 648–660.

Lee-Stafford, W. H., Lotter, G. A., Maltitz, G. P., & Brent, A. C. (2019). Biofuels technology development in Southern Africa. *Development Southern Africa, 36*(2), 155–174.

Leon-Olivares, E., Minor-Popocatl, H., Aguilar-Mejía, O., & Sánchez-Partida, D. (2020). Optimization of the supply chain in the production of ethanol from agricultural biomass using mixed-integer linear programming (MILP): A case study. *Mathematical Problems in Engineering*, 1–25.

Lim, M. K., & Ouyang, Y. (2016). Biofuel supply chain network design and operations. In A. Atasu (Ed.), *Environmentally Responsible Supply Chains* (Vol. 3, pp. 143–162). Springer-Verlag.

Luque, R., Lovett, J. C., Datta, B., Clancy, J., Campelo, J. M., & Romero, A. A. (2010). Biodiesel as feasible petrol fuel replacement: A multidisciplinary overview. *Energy & Environmental Science, 3*(11), 1706–1721.

Naimah, K., Mardanie, B. R., Dwi, L. M., Adi, S. N., & Sulistiyanto. (2020). A review of technology assessment of green gasoline processing. *Energy Reports, 6*, 1641–1649.

Ng, R. T. L., & Maravelias, C. T. (2016). Design of cellulosic ethanol supply chains with regional depots. *Industrial & Engineering Chemistry Research, 55*(12), 3420–3432.

Paredes-Cervantes, S. A., Barahona-Pérez, L. F., Barroso-Tanoira, F. G., & Ponce-Marbán, D. V. (2020). Biofuels and their potential in the Mexican energy market. *Revista de Economía – Universidad Autónoma de Yucatán, 37*, 35–56.

Pérez-Lechuga, G., Aguilar-Velázquez, S. L., Cisneros-López, M. A., & Venegas-Martínez, F. (2019). A model for the location and scheduling of the operation of second-generation ethanol biorefineries. *Journal of Mathematics in Industry, 9*, 1–14.

Poh-Ying, H., Phun-Chien, C. B., & Yee-Van, F. (2020). Operational management implemented in biofuel upstream supply chain and downstream international trading: current issues in Southeast Asia. *Energies, 13*(7), 1–26.

Razi, F. F. (2019). A hybrid DEA-based k-means and invasive weed optimization for facility location problem. *Journal of Industrial Engineering International, 15*(3), 499–511.

Rodrigues Costa, F., Alvares Soares Ribeiro, C. A., Marcatti, G. E., Lorenzon, A. S., Teixeira, T. R., Fonseca Domingues, G., Martins de Castro, N. L., Rosa dos Santos, A., Paulo Soares, V., & Martins da Costa de Menezes, S. J. (2020). GIS applied to location of bioenergy plants in tropical agricultural areas. *Renewable Energy, 153*, 911–918.

Sharma, N., Chauhan, A., Pratap Singh, A., & Arora, R. (2024). Optimizing sustainable bio-fuel supply chain: A multi-objective model utilizing micro-algae resources. *Discover Sustainability, 5*(1), 1–28.

Silalahi, F. T. R., Simatupang, T. M., & Siallagan, M. P. (2020). Biodiesel produced from palm oil in Indonesia: Current status and opportunities. *AIMS Energy, 8*(1), 81–101.

Sosa-Rodríguez, F. S., & Vázquez-Arenas, J. (2021). The biodiesel market in Mexico: Challenges and perspectives to overcome in Latin-American countries. *Energy Conversion and Management, 12*, 1–7.

Vega, L. P., Bautista, K. T., Campos, H., Daza, S., & Vargas, G. (2024). Biofuel production in Latin America: A review for Argentina, Brazil, Mexico, Chile, Costa Rica and Colombia. *Energy Reports, 11*, 28–38.

Vivoda, V., Krame, G., & Spraggon, M. (2023). Oil theft, energy security and energy transition in Mexico. *Resources, 12*, 1–15.

Zandie, M., Ng, H. K., Gan, S., Muhamad-Said, M. F., & Cheng, X. (2022). The viability of using gasoline-integrated biodiesel–diesel mixtures in engines as a solution to greenhouse gas emissions: A review. *Clean Energy, 6*, 848–868.

Zhang, F., Wang, J., Liu, S., Zhang, S., & Sutherland, J. W. (2017). Integrating GIS with optimization method for a biofuel feedstock supply chain. *Biomass and Bioenergy, 98*, 194–205.