



INFORMS Journal on Applied Analytics

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Leveraging Geospatial Analysis and Machine Learning for Optimal Green Vehicle Assignment

Josué C. Velázquez-Martínez, Laura Palacios-Argüello, Ade Barkah, Jan C. Fransoo, Karla M. Gamez-Perez

To cite this article:

Josué C. Velázquez-Martínez, Laura Palacios-Argüello, Ade Barkah, Jan C. Fransoo, Karla M. Gamez-Perez (2026) Leveraging Geospatial Analysis and Machine Learning for Optimal Green Vehicle Assignment. *INFORMS Journal on Applied Analytics*

Published online in Articles in Advance 08 Jun 2026

<https://doi.org/10.1287/inte.2021.0061>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*INFORMS Journal on Applied Analytics*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/inte.2021.0061>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2026 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Leveraging Geospatial Analysis and Machine Learning for Optimal Green Vehicle Assignment

Josué C. Velázquez-Martínez,^{a,b,*} Laura Palacios-Argüello,^c Ade Barkah,^a Jan C. Fransoo,^d Karla M. Gamez-Perez^{a,e,†}

^a Center for Transportation and Logistics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; ^b AI for Manufacturing & Supply Chain Institute, Tecnológico de Monterrey, Monterrey, Nuevo León 64700, Mexico; ^c Luxembourg Center for Transportation and Supply Chain Management, University of Luxembourg, L-1359 Luxembourg, Luxembourg; ^d Tilburg School of Economics and Management, Tilburg University, 5037 AB Tilburg, Netherlands; ^e School of Engineering & Science, Tecnológico de Monterrey, Monterrey, Nuevo León 64700, Mexico

*Corresponding author

†Deceased

Contact: josuevm@tec.mx,  <https://orcid.org/0000-0001-8863-6609> (JCVM); laura.palacios@uni.lu (LP-A); barkah@mit.edu (AB); jan.fransoo@tilburguniversity.edu,  <https://orcid.org/0000-0001-7220-0851> (JCF)

Received: June 30, 2021

Revised: February 9, 2025; August 8, 2025; January 9, 2026

Accepted: January 21, 2026

Published Online in *Articles in Advance*: June 8, 2026

<https://doi.org/10.1287/inte.2021.0061>

Copyright: © 2026 The Author(s)

Abstract. The transportation sector has been the main contributor to emissions growth in the last decade. The type of truck and its delivery characteristics largely explain the transportation carbon dioxide (CO₂) emissions and carbon intensity factors. This article introduces a novel methodology for the allocation of a fleet of vehicles to certain regions aimed at minimizing total transportation-related CO₂ emissions. Our methodology employs geospatial analysis and machine learning to assess the fuel efficiency and CO₂ emissions performance of a vehicle fleet by analyzing historical GPS data, cargo, and fuel use. Subsequently, we include these variables into a mathematical model to obtain an optimal allocation that minimizes total transportation CO₂ emissions. Our approach extends the current literature by considering detailed data for operation, such as gradient variability (road hilliness), vehicle speed, elevation/altitude, and distance between stops. We applied our methodology in Coppel, one of the largest retailers in Mexico, which operates its own fleet. Our results showed that, by exchanging 10 vehicles for one month, we observed 8% savings in fuel efficiency and transportation CO₂ emissions.

History: This paper was refereed.



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "INFORMS Journal on Applied Analytics. Copyright © 2026 The Author(s). <https://doi.org/10.1287/inte.2021.0061>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Keywords: transportation CO₂ emissions • geospatial analysis • two-stage optimization • vehicle assignment • machine learning

1. Introduction

Almost 95% of the world's transportation energy comes from petroleum-based fuels, largely gasoline and diesel (EPA 2024), which explains the large increase of global carbon dioxide (CO₂) emissions from fossil fuels in recent years. The transportation sector is one of the main contributors to greenhouse gas emissions and represents 14% of the global emissions (Lamb et al. 2021). In the European Union (EU), for instance, the transportation sector is responsible for almost 30% of the CO₂ emissions, 72% of which comes from road transport (EU 2017, Robaina and Neves 2021). Although CO₂ emissions from the different sectors have reduced by almost a quarter between 1990 and 2009, transportation CO₂ emissions have increased by almost a third over the same period (Hill et al. 2012). In the United States, the transportation sector contributes roughly 28% of the United States' total emissions and is considered the largest source of CO₂ emissions in the country, along

with the generation of electricity (EPA 2025). In addition, in developing countries, on-road vehicle emissions constitute a significant portion of CO₂ emissions from transportation-related activities (El-Fadel and Bou-Zeid 1999).

Direct emissions in transportation are mainly due to the combustion of fuel during delivery operations. This combustion is mainly driven by a variety of conditions related to the type of vehicle (e.g., engine power, torque, fuel type, and aerodynamic drag coefficient) and the characteristics of the delivery operation (e.g., type of road, slope, vehicle speed, and load). Akcelik and Besley (2003) and Demir et al. (2011) studied six emission models and showed that the main factors that affect transportation emissions are vehicle type, slope, and speed. This suggests that the same type of vehicle may have different levels of performance in terms of fuel consumption and CO₂ emissions when assigned to deliver in a variety of regions with different delivery characteristics.

The first attempt to study the effect of vehicle assignment on CO₂ emissions was presented by Velázquez-Martínez et al. (2016). The study reveals that analyzing the historical fuel performance of a fleet (i.e., transportation CO₂ intensity) unveils varying performance levels among trucks of the same type when assigned to different delivery areas. This study enables optimal vehicle assignment strategies that aim to reduce fuel consumption and transportation CO₂ emissions. However, the analysis presents certain limitations, specifically related to the effect of the delivery areas, which implies a gap in the quantitative assessment of the delivery context. We argue that using geospatial analysis enables a better characterization of the delivery regions and their influence on fuel/CO₂ emissions efficiency. Furthermore, geospatial analysis facilitates exploration into additional factors shaping transportation CO₂ emissions, including topography, congestion, and stop density.

Recent studies have demonstrated the growing relevance of advanced optimization and decision-support systems in transportation and logistics operations, focusing on both operational efficiency and sustainability outcomes. Khodabandeh et al. (2022) illustrate how heuristic-based frameworks can address complex vehicle routing problems in large-scale settings, whereas Dang et al. (2024) highlight the value of integer programming approaches for routing and fleet optimization, yielding significant cost and emissions reductions for DHL. Similarly, Abdelwahed et al. (2021) showcase how real-time optimization and discrete-event simulation can enable more sustainable operations in urban electric bus networks under uncertainty. Other research efforts in the previous literature focus primarily on routing optimization (Balamurugan et al. 2018, Schröder and Cabral 2019) and fleet replacement (Bal and Vleugel 2018, Giuliano et al. 2021) as strategies to reduce CO₂ emissions from transportation. Building on these advancements, our study complements this stream of work by integrating detailed operational and geospatial data—such as road gradient variability, vehicle speed, and stop distances—into a machine learning–driven optimization model that directly targets CO₂ emissions minimization in fleet allocation. In doing so, we address a critical gap in the literature by linking vehicle-level operational data with environmental impact, demonstrating measurable emission reductions in a real-world retail context.

In this study, we present four significant contributions. First, we extend the current literature by introducing a novel framework for performing a green vehicle assignment, utilizing geospatial analysis and machine learning to minimize transportation CO₂ emissions while studying delivery operations. In this article, the term *green vehicle assignment* refers to an assignment strategy that prioritizes the minimization of transportation-related CO₂ emissions, rather than cost or distance, using the existing vehicle fleet. Although all vehicles analyzed are

conventional diesel trucks, the “green” aspect of the model arises from the environmental objective function and the emissions-based performance criteria used to guide vehicle-to-region assignments. Second, we apply our framework in a real-case application to the last-mile fleet of one of the Mexico’s largest retailers. We analyze truck operations in 12 cities across Mexico based on 29,000 individual GPS files and more than 80 million records. Note that we use the term “cities” to collectively refer to both cities and municipalities. This rich data set improves the precision of our geospatial and machine learning analyses, enabling us to quantitatively assess and incorporate the impact of road conditions on fuel consumption and CO₂ emissions efficiency. Third, to validate our statistical findings, we conducted a large-scale field study that involved the participation of 100 students that collected 94,000 observations in the field (i.e., in the trucks). To facilitate this data collection, we designed a custom data collection tool (see Appendix A). This field study allows an accurate identification of the performance of fuel and CO₂ emissions in the vehicle fleet, facilitating optimal vehicle reassignments to reduce overall CO₂ emissions. Fourth, we provide practical insights into the relation between specific type of vehicles (i.e., brand and age) and the delivery regions and how to better allocate vehicle types to delivery regions.

The remainder of this article is organized as follows. In Section 2, we present the overview of the Mexican retailer and describe the green vehicle assignment problem. In Section 3, we present the four phases: data collection, data analysis, description of the green vehicle assignment model, and model validation. In Section 4, we describe in detail how the methodology is applied for the company under study (i.e., Mexican retailer). We include in this section all the data analysis (i.e., geospatial and machine learning analyses), mathematical formulations (i.e., a two-stage optimization model), and model validation using a field study. In Section 5, we discuss the main takeaways from the real-case application: A real-world intervention we carried out over the course of one month with the company by comparing fuel consumption and CO₂ emissions between the proposed vehicle assignment and the existing operational setup. Finally, in Section 6, we summarize the contributions of the paper and lessons learned.

2. Company Background and Problem Definition

We present a practical case of Coppel (<https://www.coppel.com>), a multinational retail company that serves both e-commerce and in-store retail markets in Mexico. Founded in 1941, the retail company sells appliances and clothing to a variety of customers and is ranked 156th largest retailer in the world. In Mexico, it operates 200 distribution centers (DCs) and

serves approximately 1,500 retail stores and performs an average of 25,000 home deliveries daily; that is, considering an operation of five days per week and 52 weeks per year, the company conducts more than 6 million logistics movements annually. The company manages a private fleet of more than 1,200 vehicles for last-mile deliveries (i.e., transportation activity from the store or distribution center to the end customer). The company's delivery trucks operate in diverse geographies and road conditions throughout Mexico. Each truck is assigned to an origin regional DC and to a set of retail stores and consumers (via home delivery). The company operates a heterogeneous fleet in terms of brand and age (capacity remains the same), and delivers under different conditions of traffic, altitude, road gradient, utilization, and so on.

We utilize geospatial data to evaluate how different combinations of road conditions influence fuel consumption and CO₂ emissions. Our objective is to optimize vehicle-to-region assignments to improve fleet performance across the entire distribution network. In this context, a region refers to a delivery area assigned to a specific truck and anchored to a nearby DC. The study focuses on a subset of 89 trucks operating in various cities and municipalities across Mexico, including Los Mochis, Tecámac, Cuautitlán Izcalli, Azcapotzalco, Iztapalapa, Toluca, Querétaro, León, Veracruz, Puebla, Culiacán, and Monterrey (Figure 1).

To analyze the transportation activity and the fleet performance of the company's vehicles, we organized all the information into the following six data sets (see Appendix B for all the details regarding the data sets).

- *Data Set 1: Fleet vehicle information*—Baseline fleet characteristics provided by the company, including vehicle brand, model, age, and each truck's assigned delivery region.

- *Data Set 2: GPS track data*—GPS coordinates for all pickup and delivery operations of each vehicle.

- *Data Set 3: Field study operational data*—Additional operational and contextual information collected during the field study, supplementing Data Set 1 with variables not available in the company's records, including road conditions and gradient, local traffic conditions, and weather observations.

- *Data Set 4: Fuel data*—Fuel consumption records for the vehicle fleet, including the fuel card with date and mileage information.

- *Data Set 5: Logistics movements*—Information related to daily movements of items for home and store deliveries.

- *Data Set 6: Stock-Keeping Unit (SKU) data*—SKU catalog used to estimate truck utilization.

3. Methodology

We propose a four-phase methodology: (i) data collection, (ii) data analysis, (iii) green vehicle assignment model, and (iv) model validation. Figure 2 describes our proposed methodology.

3.1. Data Collection

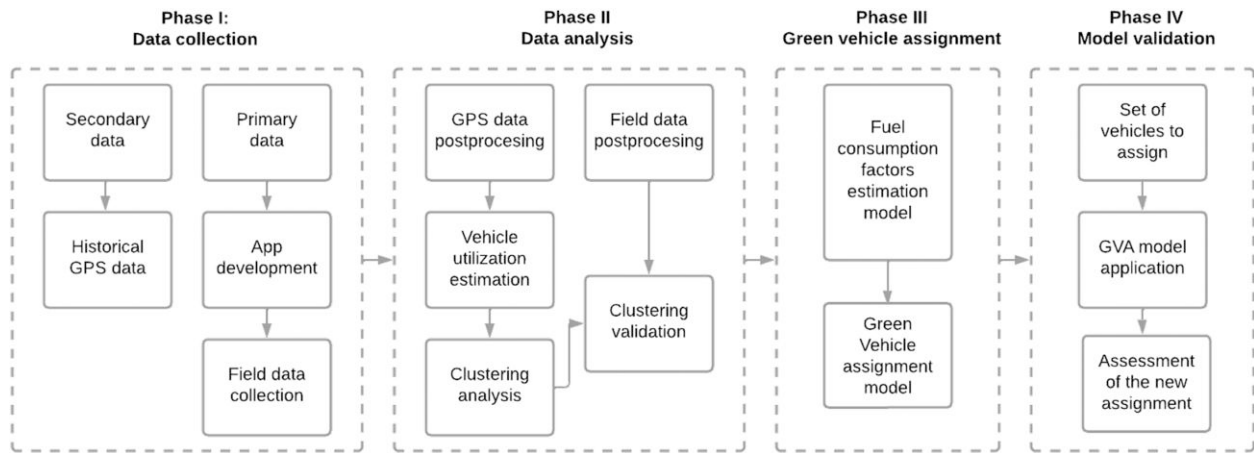
This phase combines primary and secondary data sources to analyze multiple factors related to road conditions and their impact on delivery efficiency, fuel consumption, and CO₂ emissions.

For the primary data collection, we implemented a dual data collection strategy.

Figure 1. Overview of the Study Area in the Real-Case Application, Highlighting the Regions Across Mexico Where Coopel's Distribution Centers Are Located



Figure 2. Proposed Four-Phase Methodology



- *Quantitative data collection:* We extracted GPS traces and topographical information to capture measurable features of the road network, such as elevation profiles, road curvature, and travel speeds. In addition, we characterized the delivery fleet by vehicle brand and years of use.

- *Field study and app-based observations:* We conducted a three-week validation field study in collaboration with the Instituto Tecnológico y de Estudios Superiores de Monterrey, Mexico. More than 100 undergraduate and graduate students participated in the study, accompanying Coppel delivery drivers across 146 last-mile delivery routes. These routes originated from nine distribution centers, selected to represent geographic and economic diversity across the country.

Using a custom mobile data collection platform (see Appendix A), students captured more than 94,000 geo-tagged observations through four dedicated mobile applications:

- The *Starting Point* app recorded vehicle conditions at departure.
- The *Event-Based* app logged real-time road and traffic events, including steep inclines, congestion, and intersections.
- The *Pickup and Delivery* app documented all delivery-related activities and stops.
- The *Fuel Consumption* app tracked fuel usage, which we later converted into CO₂ emissions.

Students collected these data during full-day ride-alongs with Coppel drivers, providing detailed route-level information on last-mile operational conditions. In addition to observations based on apps in real time, each student conducted a qualitative assessment of the delivery area at the end of each route. They recorded contextual features such as traffic intensity, infrastructure quality, delivery density, and the presence of informal settlements. Based on this assessment, they manually assigned each area to a predefined cluster type—such as *Urban High Density* or *Remote Low*

Access—and uploaded this classification via the app interface. We developed these cluster definitions in an earlier pilot phase and refined them in consultation with Coppel’s logistics team.

We later used these qualitative classifications to validate the results of our clustering analysis. In particular, we compared the clusters generated by the unsupervised machine learning algorithm with the clusters assigned to students from the field study. We present and discuss this comparison in detail in Section 4.

3.2. Data Analysis

This phase includes data analysis from the GPS and field data. For analyzing the GPS data, three steps are performed to clean and analyze the GPS data, and then the vehicle utilization is calculated to finally perform a clustering analysis.

3.2.1. GPS Data Analysis. Given the large volume and variety of data, we conducted a quality process to minimize noise in the calculation, specifically to avoid outliers that are usually caused by errors in the GPS tracking process. Because GPS data are subject to various accuracy limitations, we use the “Snap to Roads” API from Google Maps to correct position data. Because of satellite position geometry limitations, GPS altitude data are not fully reliable. Therefore, in this substep, we obtained corrected elevation data for each position using the Google Elevation API.

Then we segmented each route, based on the number of stops the truck made during the day, considering the timestamp between two consecutive GPS position points. Based on these segments, the speed and distance traveled by the vehicle in each segment are calculated. Finally, at various points along the route, the road gradient is calculated at each GPS position, and then we divided each segment into 100-meter subsegments and estimated the average slope

along the entire subsegment. We used this information to generate a topographic profile of the segment. Figure 3 shows the postprocessing process for GPS data analysis.

Once we analyze the GPS data, we compute the load transported by each vehicle, measured in kilograms, and the utilization ratio, defined as the ratio of load to truck capacity. We then group delivery routes into bins with similar utilization levels to account for the effect of vehicle load on fuel consumption and CO₂ emissions. This binning process allows us to isolate the impact of load variations when estimating fuel efficiency and emissions performance across different delivery routes.

Although the regions in our study are anchored to known distribution centers and their associated vehicle assignments, the operational characteristics within these regions, such as traffic intensity, road gradients, delivery stop frequency, and average travel speed, vary considerably. To uncover these hidden patterns, we apply a clustering analysis using geospatial and vehicle utilization data. This unsupervised learning approach groups regions based on operational similarity rather than geographic proximity or administrative boundaries. By identifying clusters with distinct delivery conditions, we highlight performance variability across regions for each truck type. The analysis reveals that certain vehicle types operate more efficiently in specific cluster profiles. These insights would be difficult to obtain through manual classification alone and provide operational value by informing vehicle assignment strategies that improve fuel efficiency and reduce CO₂ emissions.

We employ an unsupervised machine learning algorithm, k-means clustering (Ball and Hall 1965, MacQueen 1967), to identify operational factors that may influence fleet performance, particularly in terms of fuel consumption and CO₂ emissions. The k-means

algorithm partitions the data set into k clusters, where k is a user-defined parameter. It does so by minimizing the within-cluster variance, using Euclidean distance as the similarity measure between each data point and the corresponding cluster centroid (Singh et al. 2011). Among the factors identified are the following:

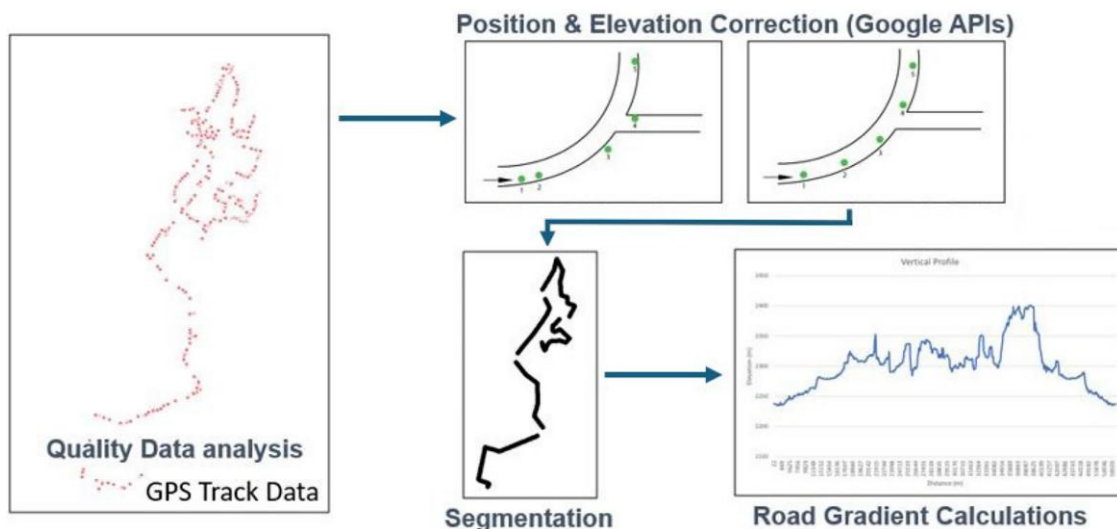
- Effect of slope: We selected the parameters of the gradient variance, defined as the variance of segment-level average road gradients within a given delivery route, percentage of the flat route (road gradient is less than $\pm 1\%$), and percentage of the steep route (road gradient is 4% or greater).
- Effect of speed: We selected the average velocity and the average segment length. The latter provides an estimate of the emissions efficiency changes associated with the number of stops in a route, which also affects the overall speed.
- Effect of the city elevation: We selected the mean elevation of the city as a parameter that describes the characteristics of the delivery areas.

3.2.2. Field Data Analysis. To analyze similarities and/or discrepancies between the GPS and vehicle utilization data findings and the field data collected, we propose a clustering validation. To do this, we estimated a fit ratio defined as the total number of routes that belong to a cluster according to the k-means, divided by the number of these routes that are also classified, via the field study, in the same type of cluster.

3.3. Green Vehicle Assignment Model

This phase includes the development of two mathematical models: a linear optimization model that estimates the emission factors per cluster and an integer program that minimizes the overall CO₂ emissions by assigning the vehicle to the delivery areas. Note that

Figure 3. GPS Data Analysis Postprocessing Process



our proposed green vehicle assignment model does not assume the use of alternative fuel or low-emission vehicle technologies. Instead, it focuses on reducing overall CO₂ emissions by optimally matching fuel-based vehicles to delivery regions based on empirically estimated emissions performance under different operational conditions.

3.3.1. Fuel Consumption Factors Estimation Model. To estimate CO₂ emissions and evaluate alternative vehicle-to-region assignments, we developed a fuel consumption model based on transportation activity factors. The data underlying this model were obtained from the company’s ECO system, which integrates automated fuel card transactions—recording time, location, volume, and vehicle ID—with manually reported odometer readings and distances traveled. We performed a systematic data validation process using descriptive statistics, histograms, and box plots to detect anomalies such as negative mileage values, implausible fuel volumes, and unit inconsistencies. To ensure data quality, we applied two rule-based filters: a fuel volume threshold of 2–140 liters (aligned with the vehicles’ tank capacity) and a distance traveled range of 2–1,200 kilometers (representing the feasible operating range under unloaded conditions). We used the cleaned data set to estimate region-specific fuel consumption factors, normalized by distance traveled, vehicle load, and utilization rates. These region-specific factors formed the basis for our emissions estimation framework. To ensure methodological consistency, we followed the guidelines established by the Network for Transport Measures (NTM 2018), which provide standardized procedures for calculating transport-related CO₂ emissions under real-world conditions.

In the fuel consumption factors estimation model (see Appendix C), we denote by I the set of vehicles and J the set of delivery regions.

Our mathematical formulation allows us to obtain the values for $FC_{i,j}^e$ and $FC_{i,j}^f$, such that the absolute difference with respect to the real fuel consumption factor $FC_{i,j}^r$ is minimized, through Objective Function (C.2). Constraints (C.3) describe the fuel consumption factor error. Constraints (C.4) ensure that the fuel consumption factor for an empty vehicle is lower than for a fully loaded vehicle, which is consistent with the NTM formula. Constraints (C.5) are the nonnegativity constraints. Note that, in order to avoid solutions with equal values for the fuel emission factors empty and full, we initialized the decision variables by using the parameters provided by NTM (2018) for a small/medium truck with a capacity lower than 7.5 tons.

3.3.2. Green Vehicle Assignment Model. The second stage focuses on defining the mathematical formulation

for the optimal CO₂ fleet assignment based on the results obtained in the estimation of the emission factor. This model assigns a vehicle i from the set of vehicles I to serve a delivery region j from a set of delivery regions J , by minimizing total CO₂ emissions. We formulate the green vehicle assignment model (see Appendix D).

Objective Function (D.1) is the sum of the total CO₂ emissions corresponding to all vehicles $i \in I$ assigned to all delivery regions $j \in J$. Constraints (D.2) and (D.3) stipulate that each vehicle is assigned to a single delivery region and vice versa (i.e., each delivery region is assigned to a single vehicle). Constraints (D.4) assure that vehicles do not exceed their weight capacity. Finally, Constraints (D.5) are integrality constraints.

3.4. Green Vehicle Assignment Model Validation

To validate the model, we used information from the Mexican retailer to obtain the number of vehicles to be exchanged from one region to another to assess the performance of the fleet and to compare the fuel consumption and CO₂ emissions of the current assignment versus the new assignment provided by the model.

4. Results

Based on GPS data collection, we analyzed six months of historical data for 89 trucks and 16,500 GPS files, each representing an entire route of a specific truck for a full day. Additionally, we classified the fleet into seven vehicle types based on the brand model (from one to seven) and vehicle age using three ranges: zero to three, four to seven, and more than eight years (Table 1). Note that the company operates other vehicle types. In this study, we focus solely on the vehicle types involved in the trip analysis of the 89 trucks under study.

Data collection from the field involves three-week data collection in nine different cities of Mexico (i.e., Azcapotzalco, Iztapalapa, Toluca, Queretaro, Leon, Veracruz, Puebla, Culiacan, and Monterrey); each had one DC, with significant geographical and economic diversity to provide a good representation of last-mile delivery operations in Mexico.

Table 1. Classification of the Vehicles Analyzed in the Study Based on Brand Model and Usage Age

Classification	Brand	Years of usage
Type 1	1	4–7
Type 2	2	0–3
Type 3	2	4–7
Type 4	3	4–7
Type 5	4	4–7
Type 6	4	+8
Type 7	5	+8

Table 2. Variance Results for Six Operational Parameters Across the Four Clusters Identified Using the k-Means Algorithm

Variable name	Units of measure	F-value	SSE	df	p-value
Gradient Variance	Squared units	145.1	10.94	3	0.0
Average Velocity	Meters/second	48.03	1.99	3	0.0
Mean Elevation	Meters	181.8	92.2	3	0.0
Average Segment Length	Meters	33.35	4.782	3	0.0
Percentage of road gradient ± 1	Percentage	123.2	17.27	3	0.0
Percentage of road gradient ≥ 4	Percentage	165.6	22.94	3	0.0

4.1. Data Analysis

After cleaning all the data and performing position normalization and elevation correction, we segmented the routes. To perform this segmentation, we used a threshold based on operational expectations: The company under study estimates that a truck typically spends 16 minutes at a customer location. Therefore, when the time gap between two consecutive GPS points exceeded 16 minutes, we introduced a segment break. Using these segments, we calculated the vehicle’s speed and distance traveled within each segment. Subsequently, we conducted a load analysis by examining the number of available records and their impact on the aggregate fuel efficiency factor. From this, we calculated each vehicle’s load (in kilograms) and utilization rate (defined as load divided by truck capacity). For our analysis, we focused on a utilization range of 34%–66%, which encompasses the majority of observed delivery routes.

4.1.1. Clustering Analysis. We computed the k-means algorithm until we obtained a sufficient confidence level in the analysis of variance ($\geq 95\%$). To determine the adequate number of clusters, we used the elbow method that focuses on minimizing the sum of squared errors (SSEs) across all clusters for k values between 1 and 10. By balancing the minimization of SSEs with a representative number of elements within the clusters, we selected a value of $k = 4$. Table 2 shows the analysis of variance of the six parameters against the four clusters.

As a result of the clustering analysis, we obtained four groups of delivery areas with statistically equivalent road characteristics. Table 3 shows the four clusters (i.e., A, B, C, D) with their respective k-means centers (i.e., the arithmetic mean position of all the points in each cluster).

Cluster A includes areas with high elevation (more than 2,000 meters), hilly (gradient variance of 32%), low average speed (5.73 m/s), and short segment lengths (average of 265 meters between stops). These characteristics suggest that the delivery areas included in this cluster are highly urbanized and congested. Cluster B includes areas with low elevation, flat, and with medium average speed and segment lengths, whereas Cluster C has the highest average speed and longest segment length. This suggests that Cluster C covers delivery areas that are accessible via highways, with low density of stops and congestion. Last, Cluster D is another version of Cluster A (highly urbanized and congested), but less hilly (e.g., 68% of road gradient is between -1% and 1%). Figure 4 shows some map snapshots of examples of real delivery routes for the four clusters to better understand the applicability of the current approach.

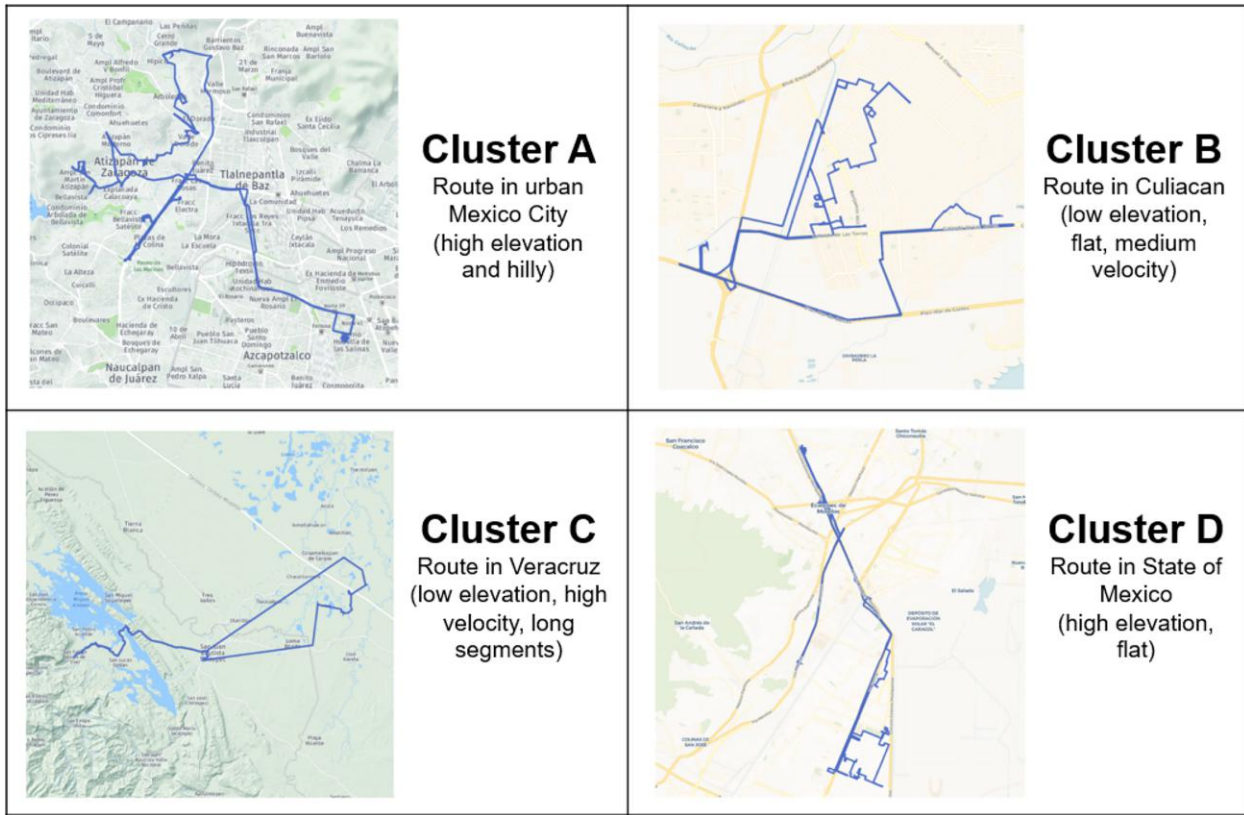
Then, considering the vehicle type classification, we identify the performance of each type of vehicle in each cluster (Figure 5). Cluster A (high elevation, hilly, short segments, low velocity) is the cluster with the highest CO₂ emissions per ton-kilometer (8.5% higher than the average of all clusters), whereas Cluster C (low elevation, flat, long segments, high velocity) shows the lowest (5% lower than the average of all clusters). We calculated the average emission factors per type of truck, cluster, and model year (see Appendix E).

This trend is consistently evident across all clusters (as seen in Types 6 and 7), regardless of the vehicle brand. However, on closer examination, we find that when comparing Type 1 vehicles (with four to seven years of usage) to Type 6 (with eight or more years of usage), Type 1 demonstrates superior performance in Clusters B and D, whereas Type 6 shows marginally better performance in Cluster A. Thus, the number of

Table 3. Cluster Centers Representing the Average Road Characteristics Identified Through k-Means

Cluster	Gradient variance squared units	Average velocity meters/second	Elevation meters	Segment length meters	Percentage road gradient ± 1 %	Percent road gradient ≥ 4 %
A	32.76	5.73	2,391	265	31.8	16.9
B	2.72	8.39	150	570	83.3	0.54
C	1.67	10.68	60.1	828	85.1	0.72
D	23.41	5.64	2,255	220	68.4	7.8

Figure 4. Clusters Description: Map Snapshots Illustrating Example Delivery Routes for the Four Identified Clusters



Notes. Cluster A represents high-elevation, hilly, highly urbanized, and congested areas with low speeds and short stop distances. Cluster B covers low-elevation, flat regions with moderate speeds and segment lengths. Cluster C includes routes with the highest speeds and longest stop distances, typical of highway deliveries with low congestion. Cluster D resembles Cluster A but with less hilly terrain.

years in operation does not appear to be a definitive factor in determining vehicle performance within a specific cluster.

In addition, we notice that some old vehicles, for example Vehicle Type 6, show the best carbon emissions performance in Clusters A and D, which refer to delivery areas that have high elevation, low velocity, short segment lengths, and hilly and flat roads, whereas Type 7 has the best performance in Cluster C, that is, in an area with high elevation, high velocity, long segment lengths, and flat roads. These are relevant insights that can help companies efficiently allocate their fleet of old vehicles in regions with these characteristics to improve the performance of carbon emissions.

4.1.2. Clustering Validation. To validate the clustering analysis based on GPS data and vehicle utilization, we compared the results of the k-means algorithm with the qualitative classifications provided by students during the field study—that is, the cluster types manually assigned and uploaded through the mobile apps. For this comparison, we defined a fit ratio, calculated as the number of routes assigned to a specific

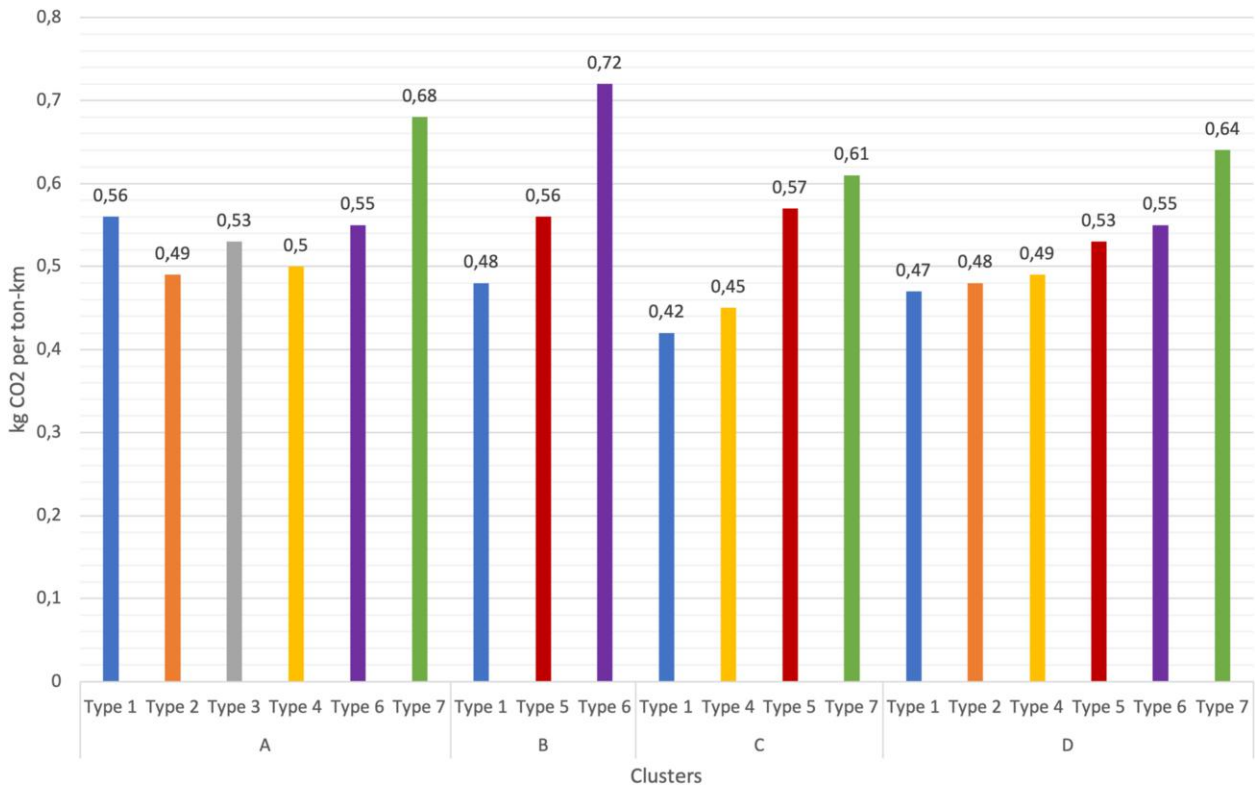
cluster by the k-means algorithm, divided by the number of those routes that were also assigned the same cluster type through qualitative field validation. Note that the fit ratio is estimated using only the routes collected during the field study. For example, a region with a fit ratio of 50% indicates that half of the routes in that region were classified into the same cluster by both the k-means algorithm and the field study. Table 4 presents the fit ratio for each of the nine regions included in the analysis.

We notice that the fit ratio in all regions where we conducted the field study is higher than 84%; in fact, the total average is 90.2%. By estimating the 95% confidence interval of the mean, we obtained a range of 87.2–93.1. Therefore, we concluded that there is statistical evidence that the cluster classification obtained by our methodology is consistent with the unbiased classification conducted by the qualitative analysis.

4.2. Green Vehicle Assignment Model

This consists of a two-stage optimization model, detailed in Appendices C and D, which aims to minimize fuel consumption and CO₂ emissions by matching the 89 vehicles to 89 delivery areas.

Figure 5. Performance of Different Vehicle Types Across Clusters, Highlighting CO₂ Emissions per Ton-Kilometer



4.2.1. Fuel Consumption Factors Estimation Model. The first-stage model for the estimation of emissions factors considers the estimation of accurate emissions factors (i.e., carbon intensity) of vehicles for the context of Mexico. We ran our model per each set of delivery areas belonging to each cluster, and as a result, we defined four different average fuel consumption factors per each cluster (A, B, C, D; Table 5).

We notice that Cluster A (hilly and highly congested) has the fuel consumption factors with the worst performance, whereas Cluster C (flat and low elevation and high average speed) shows the best performance (12% better). This estimation of carbon emission factors is key for the process and another relevant contribution of our

Table 4. Fit Ratio Between k-Means Clustering Results and Qualitative Field Classifications, Measuring the Alignment of Data-Driven and Manually Validated Cluster Assignments Across Nine Regions

Region	Fit ratio (%)
Azcapotzalco	90
Iztapalapa	95
Culiacan	84
Leon	88
Monterrey	91
Puebla	87
Queretaro	90
Toluca	96
Veracruz	90

study because, to the best of our knowledge, no other academic publication or available source related to transportation emissions provides carbon intensity factors for the context of Mexico.

4.2.2. Green Vehicle Assignment Model. Using the previous emission factors, we formulate the optimal vehicle assignment per delivery area (i.e., the mathematical formulation that matches vehicles to delivery regions), assigning 89 vehicles into 89 delivery regions. We used General Algebraic Modeling System to solve the problem on a Dell laptop with two Intel Core i7 CPUs and 16 GB memory.

After conducting our analysis, the results indicate that one in four vehicles should be reassigned, leading to an overall improvement of 4.9% in CO₂ emissions savings over the course of a year. This optimal reassignment strategy reduces CO₂ emissions for 25% of the vehicles, whereas another 25% experience an

Table 5. Average Fuel Consumption Factors Estimated for Each Delivery Area Cluster (A, B, C, D)

Fuel consumption factors (empty & full)	Cluster A	Cluster B	Cluster C	Cluster D
FCe	0.1856	0.1691	0.1648	0.1741
FCf	0.2090	0.1959	0.1874	0.2026

Source. Own elaboration.

increase in emissions. The remaining 50% of the vehicles show no change in emissions with the new assignment. Figure 6 shows the tradeoff between the number of vehicles relocated between distribution centers and the maximal total reduction in CO₂ emissions. It helps to clearly demonstrate the relationship and potential tradeoffs, offering a more comprehensive understanding of the impact of vehicle relocation on emissions reduction.

This is a consequence of the mathematical model described, where, for example, a specific vehicle can be exchanged to a region where it has a worse performance, if that implies that a better-performing vehicle is assigned to take its place instead. By focusing on minimizing overall CO₂ emissions, our model also analyses solutions where the best-performing vehicles are assigned to regions with the most active operations, that is, in delivery areas where vehicles are required to travel more distance.

4.3. Model Validation

In this section, we describe the implementation of our proposed methodology with the company under study. This real implementation has two main objectives: first, to illustrate the potential savings in fuel consumption and CO₂ emissions with empirical data observed during the transportation activity, and second, to validate the output (i.e., savings) estimated by

our mathematical model and compare it with the results observed in the operations.

The company requested us to carry out a small-scale implementation of the methodology, which involved a set of vehicles and exchanges, evaluating their performance over a month. To ensure proper vehicle pair exchanges, we incorporated a symmetry constraint into our mathematical formulation (see Appendix F). The symmetry requirement supports coordinated vehicle exchanges between regions. After running the model with this constraint, we sorted out the solution exchanges from the highest to the lowest, with respect to carbon emissions savings. The company then selected five exchanges for implementation, using their own qualitative operational feasibility. These qualitative considerations included factors not explicitly modeled, such as temporary vehicle unavailability due to maintenance, driver familiarity with specific regions, and internal operational policies. Figure 7 shows the information for the selected vehicles and delivery areas to perform the exchange of operations.

After one month of operations, we collected data for the 10 vehicles exchanged in the intervention, specifically their fuel consumption, loads transported, and GPS data per truck. Using this information, we calculated the performance of each truck in each delivery area (i.e., the average kilograms of CO₂ emissions per ton-kilometer) and compared the vehicles

Figure 6. Tradeoff Between the Number of Vehicles Relocated Between Distribution Centers and the Maximum Total CO₂ Emissions Reduction, Illustrating the Balance Between Operational Changes and Environmental Benefits

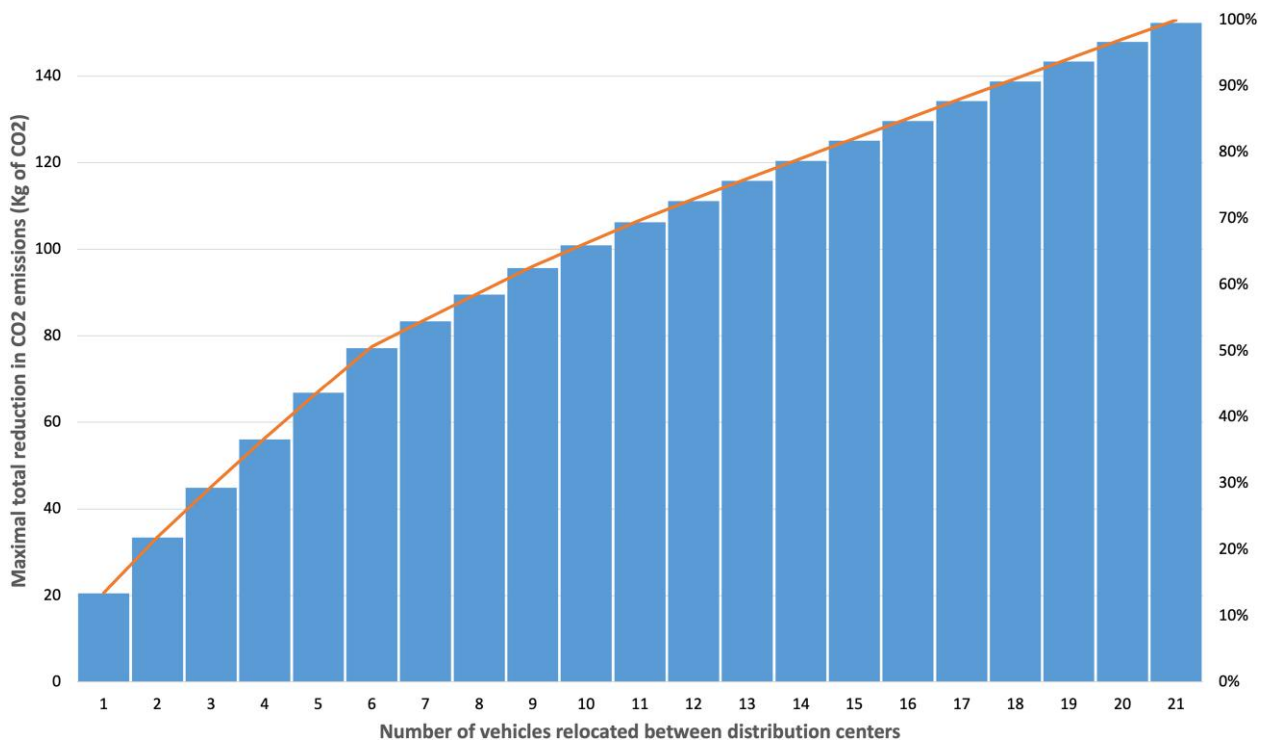












Figure 7. Details of the Five Selected Vehicle Exchanges Implemented During a One-Month Field Study, Including the Assigned Vehicles and Delivery Regions

	Assigned Vehicle to Delivery Area		Assigned Vehicle to Delivery Area	
Exchange 1	Vehicle 1: Type 1 (Brand 1; Model 2011) CLUSTER B		Vehicle 2: Type 1 (Brand 1; Model 2013) CLUSTER D	
Exchange 2	Vehicle 3: Type 3 (Brand 2; Model 2012) CLUSTER A		Vehicle 4: Type 2 (Brand 1; Model 2016) CLUSTER D	
Exchange 3	Vehicle 5: Type 5 (Brand 4; Model 2014) CLUSTER D		Vehicle 6: Type 4 (Brand 2; Model 2014) CLUSTER D	
Exchange 4	Vehicle 8: Type 1 (Brand 1; Model 2011) CLUSTER D		Vehicle 7: Type 4 (Brand 3; Model 2014) CLUSTER D	
Exchange 5	Vehicle 10: Type 1 (Brand 1; Model 2014) CLUSTER C		Vehicle 9: Type 5 (Brand 4; Model 2011) CLUSTER B	

and the performance of their initial assignment. Our results show that the exchanges provided by the green vehicle assignment model obtained an overall reduction of fuel and CO₂ emissions of 2,200 liters of diesel and 5,200 kg of CO₂, respectively. In total, we observed savings of 8% for both absolute emissions (i.e., kilograms of CO₂) and emission factors (i.e., kilograms of CO₂ per ton-kilometer). Table 6 shows a detailed comparison of each vehicle exchange, their initial and observed carbon emissions factors, respectively, and the savings (%) expected by the model and observed in the experiment. These results show evidence that the improvements estimated by our model were, in fact,

similar to the results observed in the intervention, that is, the mean error of 1%.

5. Discussion and Practical Implications

The performance of each exchange provides relevant insights.

Insight 1. Older vehicles perform better when operating in delivery areas with few stops and no congestion.

Exchange 1 reveals an unexpected outcome: An older vehicle (Vehicle 1, Model 2011) achieved 15% higher CO₂ emissions efficiency (0.50 kg CO₂ per ton-km) than a newer counterpart (Vehicle 2, Model 2013)

Table 6. Comparison of Vehicle Assignments by Distance, Carbon Emission Factors, and Percentage Savings for Each Exchange

Exchange	Vehicle ID	Distance (km)		Cluster assignment		kg CO ₂ /ton-km		Savings (%)	
		Before	After	Before	After	Before	After	Model	Observed
1	Vehicle 1	93	95	D	B	0.54	0.50	6	5.6
	Vehicle 2	97	131	B	D	0.59	0.44	25	23.4
2	Vehicle 3	94	108	D	A	0.80	0.71	10	10
	Vehicle 4	93	110	A	D	0.78	0.63	18	18
3	Vehicle 5	70	87	D	D	0.72	0.58	20	20
	Vehicle 6	112	97	D	D	0.70	0.84	-16	-13
4	Vehicle 7	77	88	D	D	0.56	0.49	12	10
	Vehicle 8	88	88	D	D	0.63	0.63	0	1
5	Vehicle 9	87	82	C	B	0.60	0.66	-11	-10
	Vehicle 10	86	86	B	C	0.63	0.52	18	20

when assigned to a delivery region with relatively low congestion and longer distances between stops, resembling highway driving patterns. Although newer vehicles typically outperform older ones in fuel and emissions efficiency, this result suggests that operational context can strongly influence performance. Possible contributing factors (e.g., gear ratios optimized for constant speed driving, differences in engine tuning, or specific maintenance histories) could explain the superior performance of the older vehicle in this context. Although these factors require further investigation beyond the scope of this study, the result may indicate that companies could benefit from assigning older vehicles to operate in regions with similar characteristics, where they might match or even outperform newer models.

Insight 2. *Newer vehicles perform better when operating in delivery areas with congestion and many stops.*

Although newer vehicles have noticeable better fuel and CO₂ emissions efficiencies, our intervention (Exchange 1) shows evidence that a younger vehicle performs significantly better (19%) when assigned to a region that is more congested and with shorter distance between stops (i.e., Cluster D). Therefore, these results suggest that companies might benefit from assigning a newer fleet to deliver within more complex delivery areas.

Insight 3. *Increasing the mileage of best performing vehicles reduces the overall CO₂ emissions.*

An important result from both the model and the experiment involved reassigning the best-performing vehicles to regions in the same cluster category but with the requirement to travel more distance. This result implies overall improvements in the efficiency of fuel and carbon emissions, leading to a reduction in CO₂ emissions. To illustrate this insight, note that Exchanges 3 and 4 were conducted in regions characterized by Cluster D but with a difference in the distance traveled. As anticipated, we assigned vehicles with superior performance to regions that required these vehicles to travel longer distances, resulting in an overall improvement in fuel and carbon emissions efficiencies. Our analysis reveals that older vehicles perform better in regions with limited congestion and high speeds, such as highways. However, Exchange 5 assigns the old vehicle to Cluster B, thus allocating a better vehicle to Cluster C, where it can significantly reduce fuel and carbon emissions.

Beyond these managerial insights, the novelty of our study focuses on quantifying these effects and translating them into systematic operational decisions. By integrating geospatial clustering of delivery regions, real fuel consumption data, and an optimization-based

allocation framework, we provide a replicable method for matching vehicles to operational contexts in a way that minimizes overall CO₂ emissions. This approach moves beyond anecdotal or experience-based decision making by quantifying the performance impact of each vehicle-to-region assignment and explicitly modeling the tradeoffs involved. Our proposed approach allows managers to assess whether allocating a higher-performing vehicle to a region with longer routes yields greater emission reductions than deploying it in a more congested area, and to do so using a consistent data-driven methodology. The combination of empirical field validation and optimization enables companies to embed these insights into routine operational planning, ensuring that vehicle assignment decisions systematically capture the efficiency gains revealed by our analysis.

6. Conclusions

Companies often prioritize updating their fleet when they have the opportunity, aiming to keep it as modern as possible. However, challenges arise when it comes to deciding what to do with older vehicles or how to optimize operations with the existing fleet. These issues remain to be effectively resolved. To do this, in this article, we studied the vehicle assignment problem using geospatial analysis and machine learning and presented a new approach that allows a fleet of vehicles to be assigned to delivery regions with the objective of reducing the overall CO₂ emissions of transportation.

We applied our methodology with Coppel, one of the largest retailers in Mexico, where we focused on studying 89 last-mile trucks that operate in 12 of the largest cities in the country. We presented a model that combines statistical analysis and mathematical programming that allows for a better characterization of the delivery areas and their effect on fuel efficiency and CO₂ emissions. Our model considers the delivery context such as gradient variability (the hilliness of roads), vehicle speed, the elevation (altitude) in which vehicles operate, and delivery segment length. By using the k-Means algorithm, we analyzed all these factors and identified four types of delivery areas.

Our analysis included a three-week field study in nine regions in Mexico and used more than 100,000 records collected through a mobile app that served to validate the results of the statistical analysis. Using the historical information of 160 vehicles, our approach shows estimates of annual savings of fuel consumption of approximately 31,250 liters of diesel (73,362 kg of CO₂ emissions) or USD35,000.

In addition, we also conducted a real intervention in which we tested and validated our approach by swapping 10 vehicles for one month. We documented

absolute savings of 8% in carbon emissions and fuel efficiency and present relevant managerial insights into the best fleet vehicle allocation. Although older vehicles tend to have worse performance than new vehicles, we observed evidence, both numerical and empirical, in which old vehicles that were reassigned to regions with few stops and no congestion significantly outperformed newer vehicles (i.e., 15% better).

However, this study did not consider other factors such as the effect of the vehicle brand and various technical specifications (e.g., horsepower). These elements could significantly influence the outcomes and therefore are interesting avenues for future research.

Finally, by proposing the assignment of specific vehicles to particular regions, companies can optimize their fleet selection based on the unique geospatial characteristics of each delivery area. This research offers companies a thorough analysis of the broader applicability of green vehicle assignments through geospatial analysis, enabling more efficient and sustainable transportation solutions. The objective of this article was to describe vehicle performance and emission factors within a specific region and not to develop a predictive model for vehicle emissions in the delivery regions. Instead, we focused on identifying the fuel emission factors that best align with company operations by

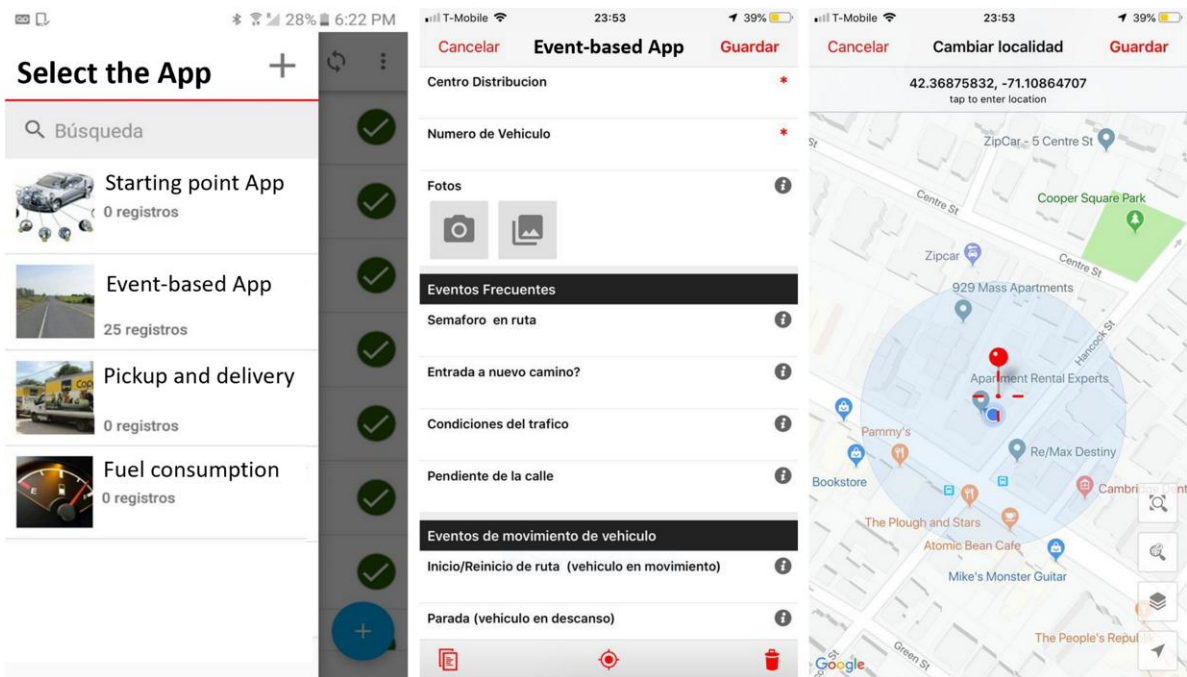
minimizing error. Through optimization, our research ensures that the values obtained are as accurate as possible for the intended application within the existing fleet infrastructure of the company. Although strategies such as acquiring a new fleet or considering fleet composition could further enhance performance, we leave these aspects as fruitful research avenues. Likewise, future work could extend the proposed green vehicle assignment model into a multiobjective formulation—considering, for example, operating costs, delivery times, or other performance metrics alongside emissions—allowing decision makers to explore and visualize tradeoffs and assess fleet assignment strategies in a more comprehensive manner.

Acknowledgments

The authors thank the sponsoring company Coppel and the undergraduate students from Tecnológico de Monterrey in Mexico, who supported the data collection during the field validation of our analysis. This article is dedicated to Karla Gámez Pérez, a thoughtful and invaluable colleague, who was the lead author of this research project. Karla Gámez Pérez was a beloved mother, wife, daughter, and dear friend who left a lasting impact on our lives. This research was supported by Coppel through a sponsored research project, which contributed to the development of this paper.

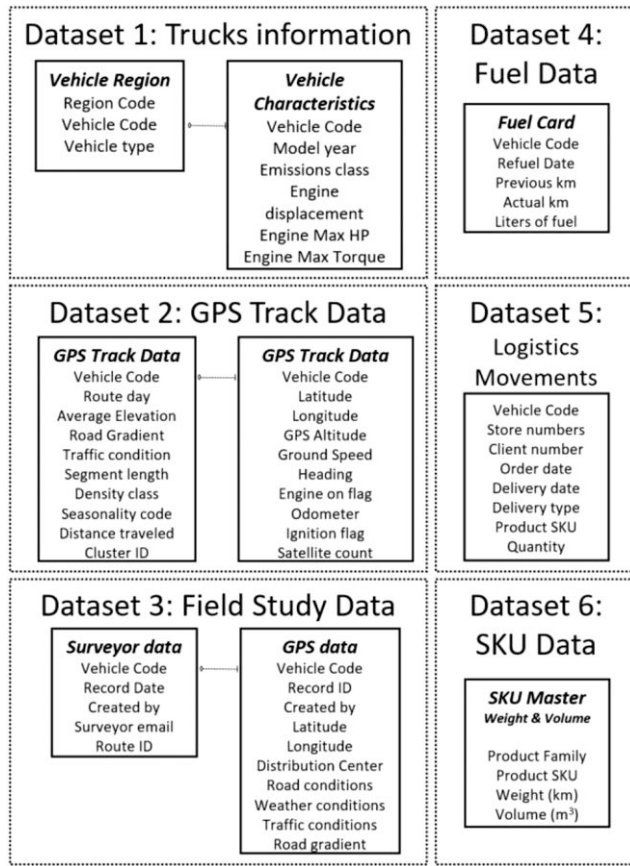
Appendix A. App Design: Field Study

Figure A.1. Interface for Field Data Collection App



Appendix B. Description of Data Sets

Figure B.1. Data Sets for Transportation Activity & Fuel Efficiency



Appendix C. Fuel Consumption Factors Estimation Model

The objective of the model is to characterize these detailed emission factors by using the real consumption and minimizing the overall error between them. We describe the model as follows.

We denote by i a vehicle from the set of vehicles I and j a delivery region from a set of delivery regions J . We consider the following parameters and decision variables:

Parameters

q_{ij} : Average weight per vehicle i at each delivery region j (tons)

C : Vehicle capacity (tons)

$FC_{i,j}^r$: Real fuel consumption factor provided by the company (liters per kilometer)

Decision variables

FC_{ij}^e : Fuel consumption factor of the empty vehicle (liters per kilometer)

FC_{ij}^f : Fuel consumption factor of the fully loaded vehicle (liters per kilometer)

$e_{i,j}$: Fuel consumption factor error (liters per kilometer)

$$\min \sum_{i \in I} \sum_{j \in J} e_{i,j}^2 \quad (C.1)$$

s.t.

$$e_{i,j} = FC_{i,j}^r - \left(FC_{ij}^e + [FC_{ij}^f - FC_{ij}^e] \frac{q_{ij}}{C} \right) \quad \forall i \in I, j \in J \quad (C.2)$$

$$FC_{ij}^e \leq FC_{ij}^f \quad \forall i \in I, j \in J \quad (C.3)$$

$$FC_{ij}^e, FC_{ij}^f \geq 0 \quad \forall i \in I, j \in J \quad (C.4)$$

Appendix D. Green Vehicle Assignment Model

We formulate the green vehicle assignment model as follows:

Parameters

k : CO₂ emission factor (2.61 kg of CO₂ per liter of diesel)

d_{ij} : Distance travelled per vehicle i at each delivery region j

Variables

X_{ij} : Binary variable for truck-route assignment

Objective function

$$\min k \sum_{i \in I} \sum_{j \in J} d_{ij} \left[FC_{ij}^e + [FC_{ij}^f - FC_{ij}^e] \frac{q_{ij}}{C} \right] X_{ij} \quad (D.1)$$

$$\text{s.t.} \sum_{j \in J} X_{ij} = 1 \quad \forall i \in I \quad (D.2)$$

$$\sum_{i \in I} X_{ij} = 1 \quad \forall j \in J \quad (D.3)$$

$$\sum_{j \in J} X_{ij} q_{ij} \leq C_i \quad \forall i \in I \quad (D.4)$$

$$X_{ji} \in \{0, 1\} \quad \forall i \in I, j \in J \quad (D.5)$$

Appendix E. Average Emission Factors per Type of Truck, Cluster, and Model Year

Table E.1. Average Emission Factors per Type of Truck, Cluster, and Model Year

Year	Cluster A						Cluster B			Cluster C				Cluster D						
	Type 1	Type 2	Type 3	Type 4	Type 6	Type 7	Type 1	Type 5	Type 6	Type 1	Type 4	Type 5	Type 7	Type 1	Type 2	Type 4	Type 5	Type 6	Type 7	
2008						0.8006														
2009									0.7233				0.2529							
2010						0.5278												0.5588	0.6391	
2011	0.5028		0.5720		0.5470		0.4817				0.4482	0.5786		0.5077						
2012	0.5801		0.4889				0.6065										0.5130			
2014				0.5138			0.4030	0.5668		0.4352							0.4675	0.5351		
2015				0.4932			0.3811			0.3977				0.3446		0.4462				
2016	0.3086	0.4957													0.4761					
2018																				

Source. Own elaboration.

Appendix F. Constraint Integrated to Guarantee Pair Exchanges Between Vehicles Year

To guarantee vehicle pair exchanges, as requested by the company, we incorporated the following constraint into our mathematical formulation. This constraint enforces a one-to-one correspondence: each vehicle is assigned to exactly one delivery region, and each delivery region is assigned to exactly one vehicle.

Variables

X_{ij} : Binary variable for truck-route assignment

X_{ji} : Binary variable for route-truck assignment

$$X_{ij} = X_{ji} \quad \forall i \in I, j \in J \quad (\text{F.1})$$

References

- Abdelwahed A, van den Berg PL, Brandt T, Ketter W, Mulder J (2021) A boost for urban sustainability: Optimizing electric transit bus networks in Rotterdam. *INFORMS J. Appl. Anal.* 51(5): 391–407.
- Akcelik R, Besley M (2003) Operating cost, fuel consumption, and emission models in aaSIDRA and aaMOTION. *Proc. 25th Conf. Australian Institutes Transport Res.* (University of South Australia, Adelaide, Australia), 1–15.
- Bal F, Vleugel JM (2018) Heavy-duty trucks and new engine technology: Impact on fuel consumption, emissions and trip cost. *Internat. J. Energy Production Management* 3(3):167–178.
- Balamurugan T, Karunamoorthy L, Arunkumar N, Santhosh D (2018) Optimization of inventory routing problem to minimize carbon dioxide emission. *Internat. J. Simulation Model* 17(1):42–54.
- Ball G, Hall D (1965) A novel method of data analysis and pattern classification. Technical Report NO. NTIS AD 699616, Stanford Research Institute, Menlo Park, CA.
- Dang Y, Allen TT, Singh M, Gillespie J, Cox J, Monkmeyer J (2024) Innovative Integer programming software and methods for large-scale routing at DHL supply chain. *INFORMS J. Appl. Anal.* 54(1):20–36.
- Demir E, Bektaş T, Laporte G (2011) A comparative analysis of several vehicle emission models for road freight transportation. *Transportation Res. Part D Transport Environment* 16(5): 347–357.
- El-Fadel M, Bou-Zeid E (1999) Transportation GHG emissions in developing countries: The case of Lebanon. *Transportation Res. Part D Transport Environment* 4(4):251–264.
- EPA (2024) Inventory of U.S. greenhouse gas emissions and sinks: 1990–2022. Retrieved April 15, <https://www.epa.gov/ghge/missions/inventory-us-greenhouse-gas-emissions-and-sinks>.
- EPA (2025) Greenhouse gases equivalencies calculator: Calculations and references. Retrieved December 12, <https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator-calculations-and-references>.
- EU (2017) Assessment of the modalities for LDV CO₂ regulations beyond 2020. Retrieved April 1, https://climate.ec.europa.eu/system/files/2017-11/ldv_co2_modalities_for_regulations_beyond_2020_en.pdf.
- Giuliano G, Dessouky M, Dexter S, Fang J, Hu S, Miller M (2021) Heavy-duty trucks: The challenge of getting to zero. *Transportation Res. Part D: Transport Environment* 93:102742.
- Hill N, Brannigan C, Smokers R, Schrotten A, van Essen H, Skinner I (2012) Routes to 2050: Developing a better understanding of the secondary impacts and key sensitivities for the decarbonisation of the EU's transport sector by 2050. <http://www.eutransportghg2050.eu>.
- Khodabandeh E, Snyder LV, Dennis J, Hammond J, Wanless C (2022) CH Robinson uses heuristics to solve rich vehicle routing problems. *INFORMS J. Appl. Anal.* 52(2):173–188.
- Lamb WF, Wiedmann T, Pongratz J, Andrew R, Crippa M, Olivier JG, Wiedenhofer D, et al. (2021) A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. *Environment. Res. Lett.* 16(7):073005.
- MacQueen J (1967) Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Sympos. Math Statist. Probability*, 281–297.
- NTM (2018) Emission factors for greenhouse gas inventories: US EPA. Retrieved January 7, https://www.epa.gov/sites/production/files/2018-03/documents/emission-factors_mar_2018_0.pdf.
- Robaina M, Neves A (2021) Complete decomposition analysis of CO₂ emissions intensity in the transport sector in Europe. *Res. Transportation Econom.* 90:101074.
- Schröder M, Cabral P (2019) Eco-friendly 3D-routing: A GIS based 3D-routing-model to estimate and reduce CO₂-emissions of distribution transports. *Comput. Environment. Urban Systems* 73:40–55.
- Singh K, Dimple M, Naveen S (2011) Evolving limitations in k-means algorithm in data mining and their removal. *Internat. J. Comput. Engrg. Management* 12:105–109.
- Velázquez-Martínez JC, Fransoo JC, Blanco EE, Valenzuela-Ocaña KB (2016) A new statistical method of assigning vehicles to delivery areas for CO₂ emissions reduction. *Transportation Res. Part D Transport Environment* 43:33–144.

Verification Letter

Mr. Joaquin Alberto Ortiz Millan, Logistics Planning, Coppel S.A. de C.V., Calle Republica #2855 PTE, Culiacan, Sinaloa 80105, Mexico, writes:

“Coppel is a multinational Mexican company that serves both e-commerce and in-store retail markets in Mexico and Argentina. Coppel is based in Culiacán, Sinaloa, and was founded in 1941 (www.coppel.com). Coppel sells appliances and clothing products for a variety of customers and is ranked 156th largest retailer in the world [Deloitte (2017). ‘Global Power of Retailing’ (PDF): 20. Retrieved August 8, 2017]. In Mexico, Coppel operates with 200 distribution centers and serves approximately 1,500 retail stores and performs more than 20,000 home deliveries (with an average of 25,000), that is, more than 6 million logistics movements annually. Coppel manages a private fleet of more than 1,200 vehicles only for last-mile delivery operation.

“Coppel has a strong sustainability commitment; due to this reason, since 2017, we started a research collaboration with the MIT Sustainable Logistics Initiative (SLI) working on improving logistics operations via better logistics sustainability. Under the umbrella of this collaboration, between 2018 and 2019, MIT SLI conducted a research project with 160 vehicles for last-mile delivery and proposed a methodology that minimizes fuel consumption (i.e., CO₂ emissions) by improving the assignment of vehicles in all delivery regions. The result of the vehicle assignment shows estimates of annual savings of fuel consumption of approximately 3.5% (~31,250 liters of diesel or 73,362 kg of CO₂ emissions or ~USD\$35,000).

“To validate the proposed model, the MIT SLI helped us conduct a pilot that included the exchange of 10 vehicles in 10 locations at four different distribution centers at Coppel. We ran the experiment for one month in October of 2018 (for more details, see Appendix A). The experiment generated savings of 8% on fuel efficiency (kilometers per liter), achieving in some cases savings up to 20%, equivalent to ~2,200 liters of diesel (5,163 kg of CO₂). We expect to

conduct a full implementation of the proposed model in the entire fleet at Coppel, and we estimate to obtain savings of at least ~USD\$238,000 annually.

“If you have any additional questions regarding this verification letter, please do not hesitate to contact me.”

Josué C. Velázquez Martínez is a research scientist at the MIT Center for Transportation & Logistics and incoming professor at Tecnológico de Monterrey, Mexico. He leads research initiatives on sustainable supply chains and data-driven logistics, working at the intersection of operations research, analytics, and artificial intelligence.

Laura Palacios-Argüello is a research scientist at the Luxembourg Centre for Logistics and Supply Chain Management at the University of Luxembourg. She leads the LCL City Logistics Freight Lab. Her research focuses on freight transportation, urban logistics, distribution network design, food supply chains, and sustainability.

Ade Barkah is an engineering manager at PayPay Corporation in Tokyo, Japan. He works on large-scale systems engineering and analytics-driven operational problems in industry. He holds an MS degree in supply chain management from the Massachusetts Institute of Technology.

Jan C. Fransoo is a professor of operations and logistics management at the School of Economics and Management, Tilburg University, Netherlands. His research studies global supply chains, retail operations, and logistics using quantitative, model-based, and qualitative methods.

Karla M. Gamez-Perez was a postdoctoral associate at the Massachusetts Institute of Technology, where this research began. She later became a professor at Tecnológico de Monterrey and Manager of R&D in Logistics at Sigma Alimentos in Mexico. Karla was a recognized expert in sustainable logistics and mathematical modeling, known for her ability to connect rigorous research with practical implementation. Her contributions to this work and to the broader field continue to have lasting impact.