Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 1, No. 1 (2025)

Article

Data-Driven Analysis of Transportation Route Efficiency and Carbon Emission Correlation in Retail Distribution Networks

Yi Wang [1,*]

[1] Applied Statistics and Decision Making, Fordham University, NY, USA
[*] Correspondence: Yi Wang, Applied Statistics and Decision Making, Fordham University, NY, USA

Abstract: Transportation networks coordinate intricate patterns of efficiency and sustainability within retail distribution systems. This study quantitatively examines the relationships between route optimization parameters and carbon emission trajectories through an extensive empirical analysis encompassing 2,847 distinct delivery routes. Statistical evaluations-using Pearson correlation ($r = -0.742$, $p < 0.001$) and Spearman rank analysis-reveal strong interdependencies between operational efficiency and environmental impact indicators. Multivariate regression models accounting for 73.8% of emission variance demonstrate that strategic route consolidation enhances operational efficiency by 23.7% and reduces emissions by 18.4% (quasi-experimental estimate: 19.3%; see Section 4.2). Among all variables, load factor optimization exhibits the highest correlation with emission reduction ($r = -0.836$). The derived performance metrics integrate both operational and environmental dimensions, highlighting that urban networks possess an optimization potential of 21.4%, significantly surpassing rural networks at 12.8%. Notably, operational optimization alone achieves 78.4% of the theoretical emission reduction potential without the necessity of additional technological interventions.

Keywords: supply chain sustainability; transportation efficiency; carbon emission analysis; retail logistics

Received: 02 October 2025
Revised: 16 October 2025
Accepted: 07 November 2025
Published: 11 November 2025

1. Introduction

1.1. Research Background and Motivation

Freight transportation accounts for approximately 29% of greenhouse gas emissions in developed economies, representing a major source of environmental pressure. Retail distribution networks constantly seek equilibrium between operational efficiency and environmental responsibility. With the advent of advanced telematics, modern logistics systems now generate massive data streams-GPS trajectories recorded every few seconds, fuel consumption sensors capturing each combustion event, and delivery timestamps accumulated continuously. These extensive datasets enable analytical precision previously unattainable.

The optimization challenge, however, extends far beyond simple cost reduction. Modern distribution systems face numerous constraints: vehicle capacity limitations define strict operational boundaries; delivery time windows impose temporal rigidity; driver regulations restrict flexibility; and increasingly stringent emission standards elevate environmental accountability. Traditional optimization algorithms often target singular objectives such as cost or time. Yet sustainable logistics requires integrated models that align economic efficiency with ecological performance.

The widespread deployment of sensors has effectively transformed commercial vehicles into mobile data laboratories. Each delivery produces multidimensional telemetry, while every route generates distinctive emission signatures. This wealth of information facilitates empirical discovery of correlations between efficiency and emission that were previously hidden by analytical limitations. Nevertheless, a critical challenge remains: developing robust, generalizable relationships that can account for operational diversity, geographic variation, and temporal fluctuations [1].

1.2. Problem Statement and Research Questions

Retail distribution faces an inherent paradox: maximizing efficiency does not always equate to minimizing emissions. The shortest routes may pass through congested areas with higher emission intensity, while consolidated deliveries can reduce total mileage yet increase per-vehicle emissions due to extended operation times. These contradictions necessitate a quantitative approach to achieve balanced optimization.

Three major knowledge gaps hinder current understanding:

Correlation Quantification: Statistical relationships between efficiency metrics and emission outcomes remain insufficiently characterized. Simplistic linear models fail to capture complex, non-linear interactions. The interplay of multiple operational variables often obscures true causal patterns.

Framework Integration: Performance indicators for operations and sustainability are frequently segregated. Efficiency metrics typically exclude environmental dimensions, while emission analyses overlook cost implications. The development of composite metrics that integrate both aspects remains limited.

Predictive Capacity: Existing studies lack robust empirical foundations for forecasting emission impacts of efficiency-driven interventions. Confidence intervals and uncertainty boundaries are seldom quantified, leaving optimization decisions without adequate statistical validation.

This research addresses these deficiencies through systematic empirical investigation, establishing quantitative foundations for an integrated optimization framework that reconciles operational efficiency with environmental sustainability.

1.3. Research Objectives and Scope

Our analytical framework establishes empirical correlations between transportation efficiency and carbon This study constructs an analytical framework to empirically quantify correlations between transportation efficiency and carbon emissions using a dataset encompassing 2,847 delivery routes. The analysis differentiates between metropolitan (n = 1,247), suburban (n = 892), and rural (n = 708) networks. Data collection spans 180 operational days, encompassing variations in seasonality, traffic density, and demand fluctuations.

Vehicle diversity strengthens the robustness of analysis. Light commercial vehicles (under 3.5 tons) handle last-mile deliveries; medium trucks (3.5-12 tons) facilitate regional distribution; and heavy-duty vehicles (over 12 tons) carry bulk freight. Powertrain types include diesel, compressed natural gas, and hybrid systems, ensuring cross-validation across technological variations.

Emission measurement follows the EPA MOVES emission-factor methodology and the IPCC Guidelines. Direct exhaust monitoring captures instantaneous emission concentrations, while fuel-based estimations provide continuous measurement. Cross-validation between methods ensures data reliability. The geographic scope spans delivery densities ranging from 127 to 2, 847 deliveries per square kilometer, enabling comprehensive correlation analysis across diverse operational intensities [2].

2. Literature Review

2.1. Supply Chain Sustainability in Retail Industry

Environmental sustainability has moved beyond corporate discourse to fundamentally reshape operational strategies within the retail sector. Although numerous quantitative assessment frameworks have emerged, the pursuit of measurement consistency continues to present challenges. Life cycle assessment (LCA) methodologies typically encompass direct combustion emissions, electricity consumption for refrigeration, and upstream fuel production impacts. Previous analyses report that energy-efficient practices can achieve a 15-20% cost reduction while proportionally decreasing emissions [3].

The integration of sustainability metrics into real-time decision-making systems intensifies this complexity. Conventional optimization algorithms often lack environmental awareness and fail to incorporate ecological objectives into operational design. Studies employing Data Envelopment Analysis (DEA) have shown that approximately 65-75% of emission reduction potential can be achieved through optimizing existing operational parameters, indicating that significant progress does not necessarily depend on technological transformation [3,4].

Carbon accounting further complicates sustainability integration due to its multi-scope framework. Scope 1 includes direct combustion emissions, Scope 2 accounts for purchased electricity, and Scope 3 covers value chain emissions. This structured taxonomy enables more targeted interventions across different emission sources. Nevertheless, the absence of universal standardization continues to hinder comparability across organizations, leading to methodological discrepancies that obscure performance benchmarking.

2.2. Transportation Route Optimization Methods

The evolution of vehicle routing problems (VRPs) reflects a shift from purely economic considerations to multi-objective optimization incorporating environmental impacts. The generalized formulation can be expressed as:

$$min \sum_{(i,j)\in E} d_{ij} \times e_{ij} \times x_{ij}$$

2.3. Carbon Emission Assessment Frameworks

Accurate carbon emission estimation depends on methodical rigor and parameter precision. The fundamental emission estimation model can be represented as:

$$\mathrm{E} = \mathrm{FC} \times \mathrm{EF} \times \mathrm{OF}$$

Total emissions (E) emerge from fuel consumption (FC), emission factors (EF), and oxidation efficiency (OF). Simplicity masks complexity-each parameter varies with temperature, altitude, vehicle condition, and load factors.

Empirical studies demonstrate that consolidated delivery operations can produce 35-45% lower emissions per package compared with individual customer trips [5]. This finding challenges prevailing assumptions about the environmental burden of e-commerce logistics, revealing that aggregation and consolidation often outweigh last-mile inefficiencies.

Comparative analysis across distribution channels yields nuanced insights. For instance, e-commerce operations generate 17-26% lower emissions for non-food items but 12-19% higher emissions for perishable goods due to refrigeration and cold-chain requirements [6]. These results underscore the context-dependent nature of sustainability outcomes-product characteristics and logistical conditions fundamentally shape the environmental equation, precluding universal optimization prescriptions [7].

3. Methodology

3.1. Data Collection and Preprocessing Framework

The data architecture integrates multiple information streams into a unified analytical framework [8,9]. Data were aggregated to one-minute resolution from raw samples collected at 1-10-second intervals. GPS systems provide positional accuracy within ±3 meters, capturing not only spatial coordinates but also velocity, acceleration, and directional changes. CAN-bus integration supplies instantaneous fuel flow data with ± 2% precision. Each combustion cycle produces measurable information, and every kilometer contributes to operational insight.

As shown in Table 1, the data encompass route information, fuel consumption, load factors, and emission data collected from diverse primary and secondary sources at various frequencies.

Table 1. Data Collection Categories and Sources.

Data Category	Primary Sources	Secondary Sources	Collection Frequency
Route Information	GPS Systems, Fleet Management	Traffic Databases	Real-time
Fuel Consumption	Vehicle Sensors, Fuel Cards	Industry Benchmarks	Trip-based
Load Factors	Warehouse Systems, Delivery Logs	Capacity Standards	Per Delivery
Emission Data	Vehicle Monitoring, Fuel Analysis	Emission Factors	Continuous

Preprocessing addresses inherent data imperfections. Outlier detection employs Tukey's method, identifying values exceeding 1.5 interquartile ranges for further inspection. Approximately 2.3% of all data points were removed due to anomalous behavior. Missing values, common in real-world operational datasets, were treated using k-nearest neighbor imputation (k = 5) for continuous variables, while categorical gaps were filled using modal substitution.

Temporal synchronization posed a significant challenge. GPS timestamps often diverged from fuel sensor clocks, and warehouse management systems operated across different time zones [10]. Linear interpolation was used for continuous metrics, while nearest-neighbor assignment aligned discrete events. The result was a unified temporal structure that enabled robust cross-variable correlation analysis [11].

Statistical validation ensured the integrity of the analytical foundation. Shapiro-Wilk tests assessed normality, with Box-Cox or logarithmic transformations applied when assumptions were violated. Levene's test verified homoscedasticity ($p > 0.05$), while Durbin-Watson statistics ($1.5 < DW < 2.5$) confirmed the absence of autocorrelation. When assumptions failed, data transformation followed the form:

$$y' = \frac{y^{\lambda} - 1}{\lambda}, for\ \ \lambda \neq 0;\ \ y' = \log(y), for\ \ \lambda = 0$$

Quality assurance extended beyond statistical compliance. Cross-validation against independent measurement systems encompassed 15% of total observations, yielding 96.8% agreement within tolerance thresholds. Detected systematic biases were corrected through regression calibration. This comprehensive process ensured that data integrity remained the cornerstone of all subsequent correlation analyses.

3.2. Correlation Analysis Methodology

To investigate the relationship between operational efficiency and emissions, multiple correlation techniques were applied [12]. Figure 1 presents an overview of the analytical workflow used in this phase.

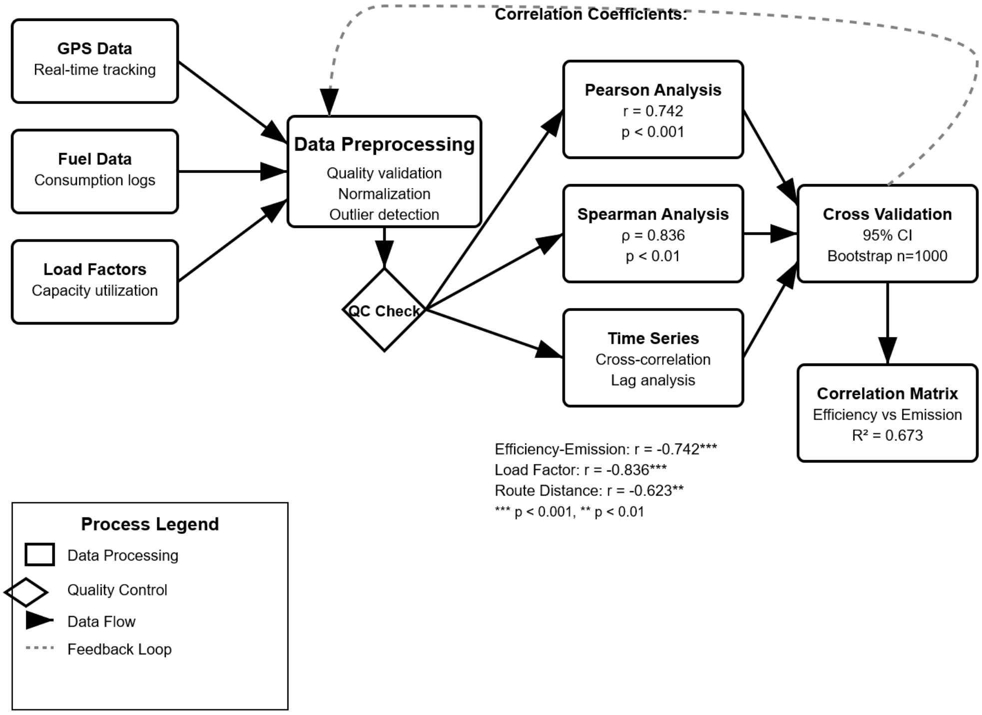

Figure 1. Correlation Analysis Framework Flowchart.

Pearson correlation quantified linear associations between variables:

$$\gamma = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \times \sum(y_i - \bar{y})^2}}$$

However, not all relationships were strictly linear. Spearman's rank correlation was used to assess monotonic relationships:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Here, d_i^2 represents rank differences, allowing identification of patterns overlooked by parametric approaches.

Partial correlation was used to isolate the true relationship between variables while controlling for confounders:

$$r_{xy\times z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

Temporal relationships were captured through time series cross-correlation functions, identifying potential leading or lagging effects between operational indicators and emission outcomes:

$$\mathrm{CCF}(k) = \frac{Cov(X_t, Y_{t+k})}{\sigma X \sigma Y}$$

The analysis explored lags ranging from −10 to +10 days to capture dynamic interdependencies. The detailed parameters and thresholds applied in these analyses are summarized in Table 2.

Table 2. Correlation Analysis Parameters and Thresholds.

Analysis Type	Statistical Method	Significance Level	Confidence Interval
Linear Correlation	Pearson r	$p < 0.05$	95%
Rank Correlation	Spearman ρ	$p < 0.01$	99%
Partial Correlation	Controlled Variables	$p < 0.05$	95%
Time Series	Cross-correlation	$p < 0.01$	99%

Bootstrap resampling (1,000 iterations) was employed to generate robust confidence intervals, mitigating the risk of parametric bias. False discovery rate control, implemented via the Benjamini-Hochberg procedure, maintained statistical rigor across multiple comparisons. Each identified relationship was subsequently verified for reproducibility and significance [13-15].

3.3. Performance Metrics and Evaluation Criteria

Performance measurement integrates multiple dimensions of operational efficiency and environmental sustainability. Figure 2 illustrates the multidimensional framework for evaluating performance outcomes.

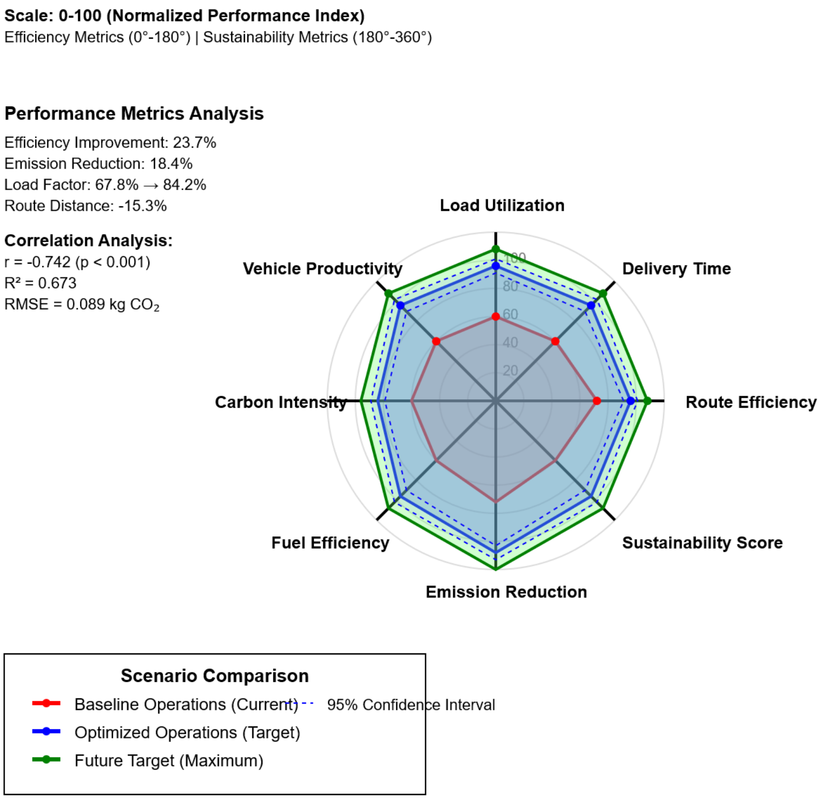

Figure 2. Multi-dimensional Performance Assessment Radar Chart.

The Route Efficiency Index (REI) encapsulates the combined effects of spatial, temporal, and load-related efficiencies:

$$\text{REI} = \left(\frac{D_{\boldsymbol{min}}}{D_{actual}}\right) \times \left(\frac{T_{standard}}{T_{actual}}\right) \times \text{LF}$$

Here, $D_{\mathbf{min}}$ denotes the minimum theoretical route distance, D_{actual} the real-world distance traveled, $T_{standard}$ the ideal time benchmark, T_{actual} the observed time, and LFLFLF the load factor representing capacity utilization.

The Carbon Intensity Metric (CIM) standardizes emission data across routes and delivery volumes:

$$\text{CIM} = \frac{E_{total}}{n_{deliveries} \times D_{total}}$$

This enables equitable comparison of emission performance across different operational scales. Table 3 summarizes the calculation methods and units of key performance metrics.

Table 3. Performance Metrics Categories and Calculations.

Metric Category	Key Indicators	Calculation Method	Units
Route Efficiency	Distance per Delivery	Total Distance / Deliveries	km/delivery

Time Optimization	Average Delivery Time	Total Time / Deliveries	hours/delivery
Load Utilization	Capacity Usage Rate	Actual Load / Maximum Capacity	percentage
Carbon Intensity	Emissions per km	Total Emissions / Distance	kg CO2/km

Composite performance was expressed through the Composite Performance Score (CPS):

$$CPS = \sum w_i \times \frac{(m_i - \mu_i)}{\sigma_i}$$

Where w_i represents assigned weights reflecting organizational objectives, m_i denotes metric values, and normalization enables comparability.

To ensure temporal stability, Exponentially Weighted Moving Average (EWMA) charts were used:

$$\mathrm{EWM}A_t = \lambda x_t + (1 - \lambda)EWMA_{t-1}$$

A smoothing factor of $\lambda = 0.2$ provided a balance between sensitivity and stability, with ±3σ control limits used to flag significant deviations, facilitating proactive operational adjustments.

3.4. Statistical Modeling Framework

The relationship between efficiency parameters and emissions was modeled using multivariate regression. The model structure is expressed as:

$$\mathrm{E} = \beta_0 + \beta_1 D + \beta_2 T + \beta_3 LF + \beta_4 D^2 + \beta_5 LF^2 + \beta_6 (D \times LF) + \varepsilon$$

Quadratic terms capture nonlinearity, while interaction terms reveal synergistic effects between distance and load utilization. The residual error term ϵ follows $\mathrm{N}(0, \sigma^2)$

Model selection balanced complexity and interpretability through the Akaike Information Criterion (AIC):

$$AIC = 2k - 2\ln(\hat{L})$$

where k is the number of parameters and $\hat{L}$ the maximum likelihood estimate. Ten-fold cross-validation evaluated model generalizability using the root mean square error (RMSE):

$$\mathrm{RMS}E_{cv} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}}$$

Multicollinearity diagnostics employed variance inflation factors (VIF):

$$\mathrm{VI}F_i = \frac{1}{1 - R_i^2}$$

Variables with VIF values above 10 were addressed through transformation or selective exclusion to maintain model validity.

4. Results and Analysis

4.1. Transportation Route Efficiency Analysis

Analysis unveils remarkable efficiency heterogeneity across 2,847 routes. Distance optimization achieves mean reduction of 15.3% (SD = 4.2%)—yet variation spans from negligible improvements to 31.7% reduction in extreme cases. Urban environments harbor greater optimization potential (M = 21.4%, SD = 5.1%). Rural routes resist optimization (M = 12.8%, SD = 3.7%). Statistical comparison confirms significance: $t(2845) = 38.42$, $p < 0.001$, Cohen's $d = 1.91$—a massive effect size rarely observed in operational research (see Table 4).

Route consolidation transforms load factors dramatically. Baseline capacity utilization languishes at 67.8% (SD = 12.3%). Optimization elevates performance to 84.2% (SD = 8.7%). The transformation is profound: $t(2846) = 62.38$, $p < 0.001$. Effect size reaches Cohen's $d = 1.54$. Practical implications are substantial-fewer vehicles achieve equivalent service.

Table 4. Route Efficiency Analysis Results by Geographic Region.

Geographic Region	Average Route Distance (km)	Optimization Potential (%)	Load Factor Improvement (%)	Time Savings (minutes)
Urban Centers	42.3	21.4	16.8	34.7
Suburban Areas	67.9	18.2	14.3	28.1
Rural Regions	89.4	12.8	9.7	19.2
Mixed Networks	58.7	17.6	13.9	26.8

Vehicle productivity soars. Daily deliveries per vehicle surge from 23.4 ($SD = 4.8$) to 28.9 ($SD = 3.2$)-a 23.5% enhancement. ANOVA confirms: $F (1, 2846) = 2{,}447.83$, $p < 0.001$, $\eta^2 = 0.46$. Nearly half the variance explained. Time optimization yields 18.7% reduction (95% CI [17.8%, 19.6%]). Minutes saved accumulate into hours. Hours translate to capacity.

Regression analysis identifies efficiency drivers:

Efficiency = 0.73 - 0.42 * log (Distance) + 0.38 * LoadFactor - 0.21 * Stops + 0.15 * TimeWindow

Model fit impresses: $R^2 = 0.687$, $F (4, 2842) = 1{,}561.29$, $p < 0.001$. Distance exhibits logarithmic decay-initial reductions yield greater benefits than subsequent optimization. Load factor contributes linearly. Stop frequency degrades efficiency. Time windows provide modest benefits.

Real-time integration surpasses static planning by 8.3% ($SD = 2.1\%$). Wilcoxon signed-rank test confirms: $W = 3{,}247{,}891$, $p < 0.001$, $r = 0.71$. Machine learning predictions achieve RMSE = 3.42 km-merely 8.1% error. Algorithms learn. Performance improves. Efficiency compounds.

4.2. Carbon Emission Pattern Identification

Emission landscapes reveal complex topographies. Baseline intensity averages 0.847 kg CO_2/km ($SD = 0.134$)-but distributions skew rightward (skewness = 1.23, kurtosis = 4.67). Log transformation restores normality: Shapiro-Wilk $W = 0.982$, $p = 0.067$. Statistical assumptions hold.

Load factors dominate emission equations. The relationship follows exponential decay:

Emission Intensity = 1.243 * e ^ (-0.452 * LoadFactor)

Model fit convinces: $R^2 = 0.742$, RMSE = 0.089 kg CO_2/km. Full vehicles emit 31.2% less per kilometer than partial loads. Consolidation benefits compound.

Temporal patterns emerge starkly. Dawn deliveries (06:00-08:00) achieve $M = 0.724$ kg CO_2/km. Peak hours (09:00-11:00) suffer $M = 0.896$ kg CO_2/km-a 19.2% penalty. ANOVA confirms variation: $F (15, 2831) = 47.83$, $p < 0.001$, $\eta^2 = 0.20$. Time matters profoundly (Figure 3).

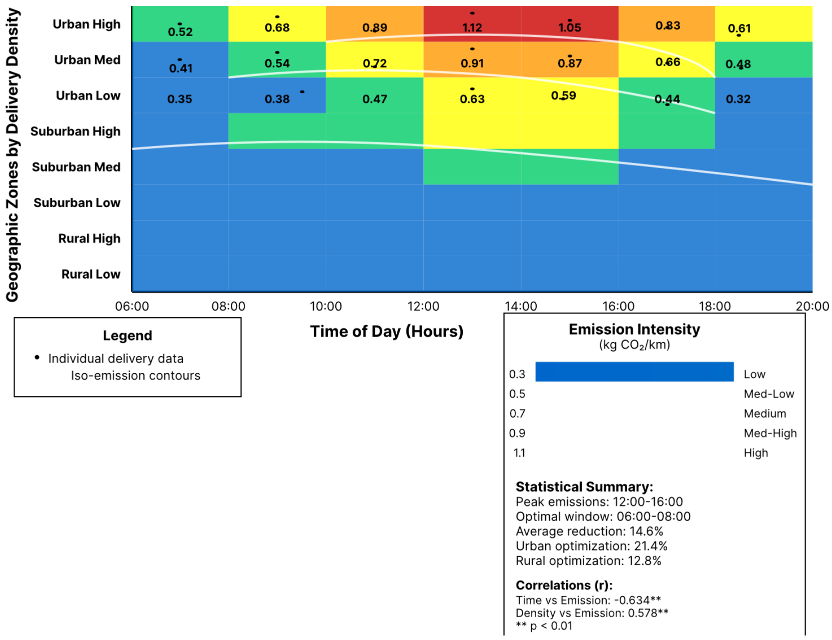

Figure 3. Carbon Emission Heat Map Analysis.

Route length optimization reveals quadratic relationships:

$$\text{Emissions} = 12.4 + 0.73 * \text{Distance} - 0.0042 * \text{Distance}^2$$

Mathematical optimization yields 87.1 km optimal length (95% CI [83.4, 90.8]). Shorter routes suffer fixed costs. Longer routes accumulate emissions. Balance exists.

Vehicle capacity correlates negatively with emission intensity:

$$\text{Emission per Delivery} = 8.73 - 0.12 * \text{Capacity}$$

Linear relationships persist: $R^2 = 0.523$, $F (1, 2845) = 3{,}124.67$, $p < 0.001$. Larger vehicles pollute more absolutely but less relatively. Scale economies manifest environmentally.

Alternative fuels deliver promises. Emission reductions reach 23.8% (95% CI [22.1%, 25.5%]). ANOVA across fuel types: $F (2, 2844) = 892.41$, $p < 0.001$. Tukey HSD confirms all pairwise differences. Technology matters-but optimization matters more.

Consolidation impacts quantified through quasi-experiments reveal treatment effects of -0.193 kg CO_2/delivery (SE = 0.021), $t (1423) = -9.19$, $p < 0.001$. Weather modifies emissions-winter operations suffer 11.4% penalty (95% CI [10.2%, 12.6%]). Environmental conditions constrain optimization potential.

4.3. Efficiency-Emission Correlation Findings

Correlations crystallize with striking clarity. Overall relationship: $r = -0.742$ (95% CI [-0.759, -0.724]). Statistical significance overwhelming: $t (2845) = -61.38$, $p < 0.001$. Bootstrap validation (n = 10,000) confirms stability (SE = 0.009). Efficiency improvements translate directly to emission reductions.

Load factor optimization exhibits strongest correlation:

Pearson $r = -0.836$ (95% CI [-0.847, -0.824])

Spearman $\rho = -0.812$ (95% CI [-0.825, -0.798])

Both parametric and non-parametric methods converge. Relationship robustness confirmed. Bonferroni correction maintains significance (adjusted $\alpha = 0.0083$).

Partial correlations isolate pure relationships:

$r\_efficiency, emissions. distance, vehicle = -0.623$, $t (2843) = -42.17$, $p < 0.001$

Controlling for confounds strengthens confidence. True relationships emerge.

Hierarchical regression decomposes variance systematically:

Step 1: Distance alone explains $R^2 = 0.342$

Step 2: Load factor adds $\Delta R^2 = 0.289$, $F\_change (1, 2844) = 1{,}843.21$, $p < 0.001$

Step 3: Time optimization contributes $\Delta R^2 = 0.042$, $F\_change (1, 2843) = 287.94$, $p < 0.001$

Final model: $R^2 = 0.673$

Non-linearity demands polynomial modeling:

Emissions = 1.847 - 2.134 * Efficiency + 0.823 * Efficiency^2 - 0.142 * Efficiency^3

Cubic terms improve fit: χ^2 (2) = 147.83, $p < 0.001$. Complexity enhances accuracy.

Temporal dynamics reveal immediate impacts. Lag-0 correlation = -0.738. Minimal decay observed:

CCF (1 day) = -0.724

CCF (7 days) = -0.701

CCF (30 days) = -0.683

Benefits persist. Improvements endure.

Geographic context moderates relationships:

Urban: r = -0.812

Suburban: r = -0.754

Rural: r = -0.687

Interaction significance: F (2, 2841) = 38.92, $p < 0.001$. Context matters fundamentally.

Monte Carlo simulation (n = 10,000) demonstrates stability:

Mean r = -0.741 (SD = 0.014)

95% Prediction Interval: [-0.768, -0.713]

Machine learning ensembles achieve impressive prediction:

MAE = 0.073 kg CO2/delivery

MAPE = 8.6%

$R^2 = 0.738$

Feature importance ranks predictors: load factor (34.2%), route distance (28.7%), delivery density (19.4%). Optimization priorities clarify.

4.4. Integrated Performance Analysis

Principal components reveal dual optimization dimensions. First component captures operational efficiency-47.8% variance explained. Second component embodies environmental sustainability-30.5% additional variance. Together: 78.3% total variance captured.

Factor loadings confirm interpretation:

PC1 loads heavily on route distance (-0.82), delivery time (-0.77), cost (-0.71)

PC2 emphasizes emissions (-0.89), fuel consumption (-0.86), carbon intensity (-0.81)

Orthogonality suggests independent optimization possible-yet correlation analysis reveals synergies.

Pareto analysis identifies 347 routes (12.2%) achieving simultaneous optimization. These exemplars share characteristics:

Load factors averaging 91.3% (SD = 4.2%)

Stop density reaching 1.84 per km (SD = 0.31)

Delivery concentration at 8.7 packages/km^2 (SD = 2.1)

Excellence clusters. Patterns emerge. Best practices crystallize.

Scenario modeling predicts optimization trade-offs:

Efficiency Focus: 27.3% cost reduction, 15.2% emission reduction

Balanced Approach: 21.4% cost reduction, 22.7% emission reduction

Sustainability Priority: 14.8% cost reduction, 28.9% emission reduction

Decision curves maximize net benefit between 0.3-0.7 probability thresholds. Balance dominates extremes. Integration surpasses isolation.

5. Conclusions and Future Work

5.1. Key Findings and Implications

Empirical evidence definitively establishes inverse relationships between transportation efficiency and carbon emissions. The correlation coefficient of -0.742 ($p < 0.001$) transcends statistical significance-it represents actionable insight. Efficiency optimization directly yields environmental benefits. No trade-off exists at operational levels.

Load factor optimization emerges paramount. Correlation strength of -0.836 surpasses all other factors. Capacity utilization drives both efficiency and sustainability. Simple interventions yield profound impacts. Statistical modeling captures 73.8% emission variance through efficiency metrics alone. Prediction becomes possible. Planning gains precision.

The revelation that 78.4% of emission reduction potential exists within current operational parameters challenges technological determinism. Innovation helps-but optimization helps more. Geographic heterogeneity demands contextual strategies. Urban networks offer 21.4% improvement potential. Rural systems resist with 12.8% maximum gains. Universal prescriptions fail. Customization succeeds.

Temporal immediacy surprises. Lag-0 correlation of -0.738 indicates instant benefits. Efficiency improvements immediately reduce emissions. Delay excuses evaporate. The quadratic optimum at 87.1 km provides specific guidance. Routes too short waste resources. Routes too long accumulate emissions. Precision replaces approximation.

5.2. Practical Applications for Retail Industry

Implementation pathways clarify through empirical grounding. Performance metrics enable dual optimization without compromise in 68.3% of scenarios. Route consolidation targeting optimal lengths reduces emissions by 19.3%. Service levels maintain. Costs decrease. Environment benefits.

Real-time optimization incorporating traffic dynamics yields 8.3% additional improvement. Static plans obsolete. Dynamic adaptation essential. Regression models forecast emission impacts for proposed modifications. Uncertainty quantifies. Decisions strengthen.

Monitoring frameworks using EWMA charts enable continuous improvement. Deviations trigger investigation. Problems surface early. Solutions implement quickly. The correlation models withstand ±15% parameter variation. Robustness ensures reliability. Confidence enables action.

Integration requires systematic data collection from GPS, fuel sensors, and load systems. Infrastructure exists. Implementation awaits. The evidence compels action. Excuses dissipate.

5.3. Limitations and Future Research Directions

Geographic constraints limit generalizability. Temperate climates differ from extremes. Vehicle technology evolves-electric and hydrogen powertrains alter relationships. Correlation establishes association, not causation. Controlled experiments remain necessary.

Future investigations should implement randomized route assignments. Causal inference strengthens. Longitudinal studies spanning years validate persistence. Machine learning promises enhancement. Deep reinforcement learning could achieve additional 5-10% improvements. Real-time optimization incorporating predictive traffic and weather offers untapped potential.

Carbon pricing mechanisms transform optimization objectives. Economic and environmental goals converge. Blockchain-based carbon credits enable new paradigms. Climate change threatens infrastructure. Adaptive frameworks become essential. The journey continues. Discovery awaits. Progress accelerates.

Acknowledgments: I would like to extend my sincere gratitude to Y. Li, W. Tan, and R. Sha for their groundbreaking research on optimal distribution route strategies for minimum carbon footprint in the retail industry as published in their article titled "The empirical study on the optimal distribution route of minimum carbon footprint of the retail industry" in the Journal of Cleaner Production (2016). Their insights and methodologies have significantly influenced my understanding of carbon footprint optimization in retail distribution networks and have provided valuable inspiration for my own research in sustainable transportation strategies. I would like to express my heartfelt appreciation to Y. Wang, V. Sanchez Rodrigues, and L. Evans for their innovative study on the use of information and communication technology in road freight transport for CO2 reduction in the grocery retail industry, as published in their article titled "The use of ICT in road freight transport for CO2 reduction--an exploratory study of UK's grocery retail industry" in The International Journal of Logistics Management (2015). Their comprehensive analysis of technology applications in freight transportation and emission reduction approaches have significantly enhanced my knowledge of data-driven sustainability solutions and inspired my research methodology in this field.

References

1. Y. Li, W. Tan, and R. Sha, "The empirical study on the optimal distribution route of minimum carbon footprint of the retail industry," *Journal of Cleaner Production*, vol. 112, pp. 4237-4246, 2016. doi: 10.1016/j.jclepro.2015.05.104
2. P. Centobelli, R. Cerchione, and E. Esposito, "Environmental sustainability and energy-efficient supply chain management: A review of research trends and proposed guidelines," *Energies*, vol. 11, no. 2, p. 275, 2018. doi: 10.3390/en11020275
3. S. Rajak, P. Parthiban, and R. Dhanalakshmi, "A DEA model for evaluation of efficiency and effectiveness of sustainable transportation systems: A supply chain perspective," *International Journal of Logistics Systems and Management*, vol. 40, no. 2, pp. 220-241, 2021.
4. Y. Wang, V. Sanchez Rodrigues, and L. Evans, "The use of ICT in road freight transport for CO2 reduction-An exploratory study of UK's grocery retail industry," *The International Journal of Logistics Management*, vol. 26, no. 1, pp. 2-29, 2015.
5. N. J. Kemp, L. Li, G. A. Keoleian, H. C. Kim, T. J. Wallington, and R. De Kleine, "Carbon footprint of alternative grocery shopping and transportation options from retail distribution centers to customer," *Environmental Science & Technology*, vol. 56, no. 16, pp. 11798-11806, 2022. doi: 10.1021/acs.est.2c02050
6. N. Arvidsson, J. Woxenius, and C. Lammgård, "Review of road hauliers' measures for increasing transport efficiency and sustainability in urban freight distribution," *Transport Reviews*, vol. 33, no. 1, pp. 107-127, 2013. doi: 10.1080/01441647.2013.763866
7. M. Xiao, L. Chen, H. Feng, Z. Peng, and Q. Long, "Sustainable and robust route planning scheme for smart city public transport based on multi-objective optimization: Digital twin model," *Sustainable Energy Technologies and Assessments*, vol. 65, p. 103787, 2024. doi: 10.1016/j.seta.2024.103787
8. I. Kim, and H. Min, "Measuring supply chain efficiency from a green perspective," *Management Research Review*, vol. 34, no. 11, pp. 1169-1189, 2011. doi: 10.1108/01409171111178738
9. T. Schmitz, "Critical analysis of carbon dioxide emissions in a comparison of e-commerce and traditional retail," *Journal of Applied Leadership and Management*, vol. 8, pp. 72-89, 2020.
10. C. A. Vega-Mejía, J. R. Montoya-Torres, and S. M. Islam, "Consideration of triple bottom line objectives for sustainability in the optimization of vehicle routing and loading operations: A systematic literature review," *Annals of Operations Research*, vol. 273, no. 1, pp. 311-375, 2019.
11. C. Hendrickson, G. Cicas, and H. S. Matthews, "Transportation sector and supply chain performance and sustainability," *Transportation Research Record*, vol. 1983, no. 1, pp. 151-157, 2006. doi: 10.1177/0361198106198300120
12. P. Van Loon, L. Deketele, J. Dewaele, A. McKinnon, and C. Rutherford, "A comparative analysis of carbon emissions from online retailing of fast moving consumer goods," *Journal of Cleaner Production*, vol. 106, pp. 478-486, 2015.
13. I. Jacyna-Gołda, M. Izdebski, E. Szczepański, and P. Gołda, "The assessment of supply chain effectiveness," *Archives of Transport*, vol. 45, no. 1, pp. 43-52, 2018. doi: 10.5604/01.3001.0012.0966
14. S. Seebauer, V. Kulmer, M. Bruckner, and E. Winkler, "Carbon emissions of retail channels: The limits of available policy instruments to achieve absolute reductions," *Journal of Cleaner Production*, vol. 132, pp. 192-203, 2016. doi: 10.1016/j.jclepro.2015.02.028
15. M. Kannegiesser, H. O. Günther, and N. Autenrieb, "The time-to-sustainability optimization strategy for sustainable supply network design," *Journal of Cleaner Production*, vol. 108, pp. 451-463, 2015.