Transportation Research Part D 152 (2026) 105173

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

A three-stage evaluation of airline low-carbon competitiveness

Mengyuan Lu [a,b,**], Jinglei Yu [a,*], Fei Li [a], Yue Zhang [a], Quanxing Jia [a]

[a] China Academy of Civil Aviation Science and Technology, Beijing, China
[b] Centre for Air Transport Management, Cranfield University, Bedford, UK

ARTICLE INFO

Keywords:
Airline sustainability
Low-carbon competitiveness
SBM-DDF model
Random Forest
Carbon efficiency

ABSTRACT

Achieving deep decarbonisation in aviation requires a systematic understanding of how airlines differ in their carbon efficiency and what operational or technological factors drive these differences. This study proposes an integrated three-stage analytical framework that combines a non-oriented SBM-DDF model, the Global Malmquist–Luenberger (GML) productivity index, Random Forest regression and a cloud-model heterogeneity assessment to evaluate the low-carbon competitiveness of 20 major global airlines from 2018 to 2022. The SBM-DDF model benchmarks multi-input–multi-output environmental efficiency, while the GML index captures dynamic productivity changes and decomposes them into efficiency change and technical change. Random Forest analysis identifies the key operational determinants of carbon efficiency, and the cloud model characterises heterogeneity in performance level, volatility and uncertainty.

Results reveal pronounced cross-airline and intertemporal heterogeneity. Mean efficiency declined markedly during 2020–2021 and rebounded in 2021–2022, although absolute CO_2-slack reductions lagged behind relative technical efficiency improvements. GML analysis shows that productivity changes were modest and mainly driven by managerial and operational efficiency rather than technological progress, indicating the limited short-run impact of fleet renewal on low-carbon competitiveness. Airline-level patterns demonstrate consistently strong and stable performance among Singapore Airlines, EasyJet and Cathay Pacific. Random Forest results identify revenue tonne-kilometres (RTK), employment scale and fleet size as dominant drivers of emission efficiency, while the cloud-model heterogeneity typology reveals four distinct groups ranging from efficient-stable to inefficient-volatile. Policy implications emphasise the need for performance-sensitive benchmarks in ETS and SAF-crediting schemes, and for managerial efficiency improvements to complement long-run technological transitions.

1. Introduction

The aviation industry plays a pivotal role in enabling global mobility and economic integration, serving over 4.5 billion passengers annually and supporting more than 65 million jobs worldwide (ICAO, 2023). However, it is also a significant and growing contributor to climate change, accounting for approximately 2.5 % of global carbon dioxide (CO_2) emissions and potentially rising to 20–25 % by 2050 if mitigation efforts are not accelerated (IEA, 2018; Owen et al., 2010). The carbon intensity of air travel is magnified by the high-altitude emissions of aircraft, which exert a warming effect up to five times greater than equivalent ground-level sources (Lee et al.,

* Corresponding author.
** Corresponding author at: China Academy of Civil Aviation Science and Technology, Beijing, China.
E-mail addresses: mengyuan.lu@cranfield.ac.uk (M. Lu), yujingleink@126.com (J. Yu).

https://doi.org/10.1016/j.trd.2025.105173
Received 27 July 2025; Received in revised form 7 December 2025; Accepted 14 December 2025
Available online 2 January 2026
1361-9209/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2009).Fig. 7.2.

In response to this challenge, international and regional regulatory bodies have developed a suite of market-based mechanisms and technological standards aimed at controlling aviation emissions. The European Union launched its Emissions Trading System (EU ETS) for aviation in 2012, followed by the UK ETS after Brexit, while the International Civil Aviation Organization (ICAO) introduced the Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) to stabilize net emissions from international flights at 2019 levels (ICAO, 2016). These programs emphasize the need for cleaner fleets, improved operational efficiency, and uptake of Sustainable Aviation Fuel (SAF).

At the same time, airlines are navigating complex trade-offs between environmental goals, economic pressures, and operational resilience. Fuel typically accounts for 20–30 % of an airline's operating costs, making decarbonization both a financial and environmental imperative (Brueckner & Zhang, 2010). The COVID-19 pandemic further exposed vulnerabilities in airline business models, prompting strategic re-evaluations of fleet composition, route planning, and technology investment (Sharma et al., 2021; Liu et al., 2020). As civil aviation faces heightened scrutiny, the ability of airlines to decouple profitability from emissions has become a critical dimension of competitiveness.

Existing studies highlight substantial heterogeneity in airline environmental performance, driven by differences in fleet composition, network structure, business models, and regulatory exposure (Liu & Jiang, 2023; Huang et al., 2020). However, much of this literature relies on single-dimension metrics (e.g., fuel burn per RTK) or static frontier methods, which inadequately capture dynamic productivity, multi-dimensional heterogeneity, and the operational drivers underlying emission efficiency. Furthermore, the dramatic shock of the COVID-19 pandemic fundamentally altered utilisation patterns, capacity decisions, and efficiency trajectories, raising questions about how airlines' low-carbon competitiveness evolved through this unprecedented period.

To address these gaps, this study develops an integrated three-stage analytical framework (Fig. 1.1) combining: (1) a cloud-model assessment to characterise heterogeneity not only in performance level (Ex) but also in volatility (En) and uncertainty (He); (2) Random Forest analysis to identify key operational drivers of carbon efficiency; (3) a non-oriented SBM-DDF model to benchmark environmental efficiency with desirable and undesirable outputs; and (4) the Global Malmquist–Luenberger (GML) index to capture dynamic productivity changes and decompose them into efficiency change (GEC) and technical change (GTC);

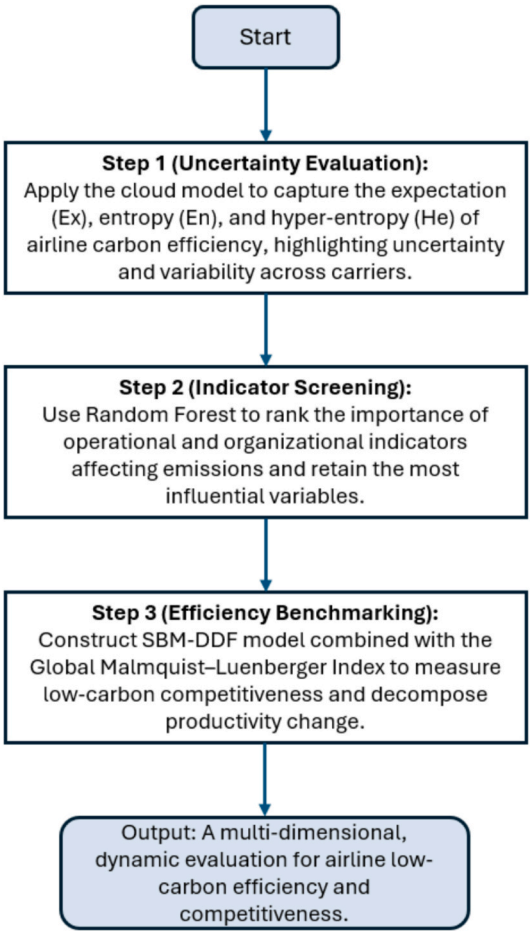


Fig. 1.1. Logic flowchart of methodology.

Using data from 20 major global airlines between 2018 and 2022, the framework provides a unified evaluation of static efficiency, dynamic productivity, underlying drivers, and heterogeneity typologies. This multi-layered approach advances existing research by integrating machine-learning-based driver identification with frontier and cloud-based heterogeneity analyses, while also enabling a clearer understanding of how COVID-19 disruptions interacted with structural efficiency patterns. The results offer new insights for both academic research on aviation decarbonisation and policy design in ETS and SAF-credit markets.

2. Literature Review

Aviation is a significant and fast-growing source of carbon emissions, prompting intense research into carbon efficiency metrics for airlines and how these relate to decarbonization strategies. In broad terms, carbon efficiency refers to the amount of CO_2 emitted per unit of service delivered. Recent literature has explored a range of indicators and methods to evaluate airlines' carbon efficiency. This includes straightforward intensity measures as well as advanced modeling approaches (Data Envelopment Analysis, Malmquist productivity indices, Life Cycle Assessment, etc.), often with an aim to inform policy or management strategies for decarbonization.

2.1. Airline carbon efficiency indicators and metrics

A fundamental way to gauge an airline's carbon efficiency is via intensity indicators – for example, CO_2 emissions per passenger-kilometer or per ton-kilometer. These metrics capture how effectively an airline transports people or cargo relative to its carbon footprint. Reducing this intensity (through fuel-efficient aircraft, higher load factors, or operational improvements) is a core industry goal. However, recent findings suggest that operational intensity improvements have not kept pace with overall growth in air traffic, implying that absolute emissions continued to rise pre-pandemic (Huang et al., 2023).

More sophisticated frameworks now include multi-factor productivity and frontier benchmarking. DEA-based models (Liu & Jiang, 2023) and Global Malmquist Productivity Indices (Huang et al., 2020) have become standard tools to assess emission efficiency. Importantly, Zhang & Wang (2024) use difference-in-differences analysis to show that EU ETS enforcement moderately reduced intra-European aviation emissions, driven by changes in operational behavior rather than demand suppression. Nguyen et al. (2022) employed a non-convex GML index to examine eco-productivity of EU airlines, finding slow but positive progress post-ETS inclusion—mostly due to gradual efficiency gains and technological improvement.

Recent research also integrates machine learning. Zhao (2025) combined Random Forest with DEA to weight operational variables in port logistics, and a similar logic has been applied in aviation by this study to address subjective indicator selection and improve interpretability.

From a system-wide carbon perspective, Life Cycle Assessment (LCA) continues to provide value. Su-ungkavatin et al. (2023) and Yang & Yao (2025) emphasize the importance of evaluating full upstream emissions and SAF scenarios using LCA frameworks, suggesting that carbon efficiency should reflect not only in-flight performance but also fuel source characteristics. These frameworks are now central to SAF evaluation under CORSIA and emerging ETS extensions.

2.2. Empirical findings on airline carbon efficiency

Empirical literature documents considerable heterogeneity across airline types and regions. Liu et al. (2020) decompose emission changes into TLC and ROP components, showing production- and operation-stage inefficiencies as key differentiators. Hamdan et al. (2022) integrate SAF and carbon pricing in an optimization model, demonstrating that carbon cost mechanisms are essential to making low-emission alternatives like SAF economically viable. Zhang et al. (2023) highlight how digital transformation—particularly in Asia-Pacific LCCs—acts as a green efficiency driver.

Meanwhile, Malmquist-based research (Huang et al., 2020; Chen et al., 2021) finds that productivity improvements vary by region and often derive more from service improvements than carbon gains. Gualini et al. (2023) and Amankwah-Amoah (2020) document how COVID-19 shock responses differed between FSCs and LCCs, leading to divergence in efficiency recovery paths.

Chen et al. (2024) also use a game-theoretic model to show how SAF mandates shift competition between LCCs and FSCs. Their findings demonstrate that more efficient carriers can gain under SAF policy regimes, providing further evidence that efficiency differentials carry strategic implications.

2.3. Airline decarbonization strategies

In terms of strategy, recent literature illustrates a shift toward integrating carbon efficiency metrics into broader aviation policy instruments. Zhang & Wang (2024) and Nguyen et al. (2022) both provide evidence that emission trading systems (ETS) produce modest but meaningful changes in behavior.

Studies like Sharma et al. (2021) simulate CORSIA outcomes for U.S. carriers and highlight that only combining SAF with strong compliance metrics yields emission stabilization. Pang & Chen (2023) support the inclusion of decoupling and DEA scores in regulatory benchmark design. Strauss et al. (2021) and Lu et al. (2025) examine modal shift to HSR and carbon tax synergies, respectively, showing non-airline levers can also support aviation decarbonization.

Fuel switching remains a critical area. Yang and Yao (2025) find that SAF pathways can achieve 60–85 % GHG reductions, particularly with carbon capture, but highlight cost barriers. Integrated metrics combining service quality, emissions, and cost (e.g. GML with bad outputs) are increasingly seen as superior to single-dimension assessments.

2.4. Research gaps and contributions

Despite advancements, existing studies exhibit several shortcomings. First, while DEA- and GML-based approaches are widely used, most do not explicitly address uncertainty in efficiency measurement, leading to limited interpretation of heterogeneity. Second, variable selection in benchmarking studies often relies on subjective judgment, resulting in indicator redundancy and reduced robustness. Third, dynamic decomposition of carbon productivity—particularly under disruptive shocks such as COVID-19—remains underexplored in airline research. Finally, heterogeneity is typically described in broad terms (e.g., FSC vs. LCC), without systematic quantification of its form, drivers, and evolution.

This study contributes to the literature in four ways. First, it introduces a cloud model to capture expectation, entropy, and hyper-entropy parameters, thereby quantifying both central tendency and variability in airline low-carbon efficiency. Second, it employs a Random Forest model to rank the relative importance of operational indicators and to perform sensitivity analysis, reducing subjectivity in variable selection. Third, it applies a SBM-DDF model integrated with the Global Malmquist–Luenberger (GML) index, with bootstrapping, to provide dynamic benchmarking and statistical confidence intervals. Finally, it offers a multi-dimensional characterization of heterogeneity, linking uncertainty (cloud model), operational drivers (Random Forest), and dynamic decomposition (GML), thereby delivering new insights for policy and managerial practice.

3. Methodology

To ensure both diagnostic precision and policy relevance, this study adopts a three-stage analytical framework that integrates a cloud model, random forest (RF), and the Slack-Based Measure with Directional Distance Function and Global Malmquist–Luenberger (SBM-DDF-GML) model. Rather than functioning as a linear sequence of statistical tools, this structure forms a synergistic methodology tailored to progressively refine insights across exploratory, explanatory, and evaluative dimensions. The cloud model, rooted in uncertainty theory, enables nuanced screening of airline carbon performance under conditions of variability and ambiguity—particularly valuable for preliminary diagnostics in policy contexts with data heterogeneity. The second stage deploys random forest, a robust ensemble machine learning algorithm, to quantify and rank the importance of numerous operational and environmental indicators. This enhances model interpretability and transparency, aligning with policymaker needs for evidence-based prioritization. Finally, the SBM-DDF-GML model facilitates intertemporal benchmarking of carbon efficiency by integrating desirable and undesirable outputs under a dynamic production frontier, capturing performance trajectories that are essential for long-term policy design. Together, this three-stage framework offers a coherent mechanism to identify, explain, and evaluate the drivers of low-carbon competitiveness in the airline sector.

3.1. Cloud Model-Based preliminary assessment

The cloud model was adopted to evaluate the statistical properties of airline carbon efficiency beyond simple descriptive statistics. Unlike histograms or distribution plots, which only capture distributional shape, the cloud model characterizes efficiency distributions using three interpretable parameters: expectation (Ex), entropy (En), and hyper-entropy (He). Together, these parameters reflect the average level, variability, and stability of variability, providing a multidimensional representation of performance.

3.1.1. Parameter Definitions

For an efficiency series {x1,x2,…,xn}, the parameters are defined as follows:

$$E_x = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\chi_i - E_x)^2}$$

$$H_l = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(E_n^b - \overline{E_n}\right)^2}$$

where n is the number of observations, x_i is the i-th observed efficiency value, E_n^b is the entropy estimated from the b-th bootstrap resample, $\overline{E_n}$ is the mean of bootstrap entropies, and B is the number of bootstrap iterations (set to 1,000 in this study).

- Ex (Expectation) represents the mean efficiency level; lower values indicate stronger carbon efficiency performance.
- En (Entropy) reflects variability or volatility in efficiency; lower values imply more stable operations.
- He (Hyper-entropy) quantifies the uncertainty of variability itself, measuring whether fluctuations are consistent or erratic. Lower He values suggest more reliable performance stability.

3.1.2. Estimation procedure

To ensure comparability across airlines, all efficiency series were first normalized to a consistent scale. Parameter estimation

proceeded in two steps:

1) **Direct calculation of Ex and En** from observed efficiency values.
2) **Estimation of He** using block bootstrap resampling of efficiency series to preserve temporal dependence. A total of 1,000 bootstrap samples were generated, producing a distribution of entropy values from which the standard deviation was calculated as He.

3.1.3. Advantages of the cloud model

This approach extends conventional distributional analysis in two ways. First, the parameter set (Ex, En, He) allows formal comparison of both average efficiency and its stability across airlines, rather than relying on graphical inspection alone. Second, the bootstrap-based estimation of He introduces a dimension of robustness by quantifying the reliability of observed volatility. These advantages position the cloud model as a valuable complement to efficiency frontiers, enabling a richer interpretation of operational and environmental performance.

3.2. Random Forest-Based feature importance selection

To improve the robustness of the traditional DEA-based efficiency evaluation and reduce the dimensionality of input–output variables, this study employs a Random Forest (RF) regression model to identify the most significant determinants of CO_2 emissions among a wide set of airline operational indicators.

Random Forest, introduced by (Breiman, 2001), is a supervised ensemble learning algorithm composed of multiple decision trees. Its strengths lie in handling high-dimensional data, resisting overfitting, and offering built-in feature importance scores (Louppe et al., 2013). The RF model was trained using airline-level panel data from 2018 to 2022, with CO_2 emissions as the dependent variable, and operational variables such as: Available Seat Kilometers (ASK), Revenue Passenger Kilometers (RPK), Operating Costs, Fuel Consumption, Number of Employees, Number of Flights, Fleet Size, Passenger Numbers, Operating Revenue, as candidate independent variables.

To determine the relative importance of each variable, the Mean Decrease in Impurity (MDI) was calculated for all trees in the ensemble (Louppe et al., 2013). Variables with higher importance scores were retained for the subsequent SBM-DDF-GML analysis.

3.3. SBM-DDF-GML model

The SBM-DDF model could evaluate airlines' environmental and operational efficiency by constructing a non-radial, non-oriented directional distance function, accounting for both desirable and undesirable outputs simultaneously (Fukuyama & Weber, 2010). Inefficiencies are decomposed into specific inefficiency indicators related to CO_2 emissions, thus constructing the Emission Efficiency Index (EEI).

SBM: The SBM model is a non-radial DEA method that accounts for slacks in inputs and outputs, meaning it can handle situations where some inputs or outputs are not proportionally reduced or increased (Tone, 2001). This is critical for the study of airline efficiency since slacks in factors such as labor, capital, and undesirable outputs like CO_2 emissions often exist. The model does not assume proportional reduction across all inputs, allowing for more accurate efficiency measurements in complex operational environments (Li & Shi, 2014).

DDF: The Directional Distance Function (DDF) is used to model the desirable outputs (e.g., passenger kilometers) and undesirable outputs (e.g., carbon emissions) (Chung et al., 1997). The DDF model allows for simultaneous expansion of desirable outputs and contraction of undesirable outputs, capturing the trade-offs airlines face when improving both operational performance and environmental impact (Zhang & Choi, 2014).

GML Index: The Global Malmquist-Luenberger (GML) index extends the static evaluation provided by DEA and DDF by measuring productivity changes over time (Chung et al., 1997; Oh, 2010), separating efficiency shifts into two components: Green Efficiency Change (GEC) and Global Technological Change (GTC) (Li & Li, 2021). This approach allows us to determine whether airlines improved efficiency through operational optimization (GEC $>$ 1) or technology adoption (GTC $>$ 1).

3.3.1. Model Formulation

To evaluate airline efficiency, we employ a non-radial, non-oriented Slack-Based Measure with Directional Distance Function (SBM-DDF) model. The model is specified under the assumption of Constant Returns to Scale (CRS), which posits that a proportional increase in all inputs leads to a proportional increase in all outputs.

The model's key specifications are as follows:

- Direction Vector (g): The model uses a common, global direction vector based on the range of the observed data for each variable: $g=(\max(x)-\min(x),\max(y)-\min(y),\max(b)-\min(b))$
- Normalization: The objective function maximizes the sum of normalized slacks, where slack variables are scaled by their respective direction vectors and the number of variables in their category.
- Solver Settings: The linear programming problems were solved using the dual-simplex algorithm, configured with a convergence tolerance of 1e-6 and a maximum of 1000 iterations.

Let airline i in period t have the following input and output structure:

Inputs: $.x \in R^N$
Desirable outputs:$.y \in R^M$
Undesirable outputs:$.b \in R^K$

Let s^-, s^+, and s^b denote the slack variables for inputs, desirable outputs, and undesirable outputs, respectively. Let $\lambda \in R^n$ represent the intensity vector.

Objective function:

$$\rho = \frac{1 - \frac{1}{N}\sum_{n=1}^{N}\frac{s_n^-}{x_n}}{1 + \frac{1}{M+K}\left(\sum_{m=1}^{M}\frac{s_m^+}{y_m} + \sum_{k=1}^{K}\frac{s_k^b}{b_k}\right)}$$

- Numerator: Measures average input slacks (inefficiencies) relative to each input x_n.
- Denominator: Combines slack (shortfall or surplus) in desirable outputs (y_m) and undesirable outputs (b_k) like emissions.
- The model maximizes efficiency by minimizing input excess and desirable output shortfall, while also reducing undesirable outputs.

Subject to:

$x_i = X\lambda + s^-, .y_i = Y\lambda - s^-, b_i = B\lambda + s^b, \lambda \geq 0$

Where:

- x_i, y_i, b_i are the input, desirable output, and undesirable output vectors of DMU i.
- X, Y, B are the matrices formed by all DMUs.
- s^-, s^+, s^b are slack variables.
- λ a vector of weights corresponding to peer DMUs.

These constraints ensure that the evaluated DMU i is compared to a convex combination of other DMUs. The λ-values indicate which efficient DMUs contribute to the efficiency projection and to what extent. A non-zero λ value signifies that the corresponding airline acts as a benchmark for the evaluated airline. By analyzing the distribution of λ-values across inefficient DMUs, it is possible to identify industry leaders, observe regional or structural benchmarking patterns, and support strategic learning and peer emulation.

The inefficiency index ρ can be decomposed into three components:

- Input inefficiency:

$$\rho_x = \frac{1}{N}\sum_{n=1}^{N}\frac{s_n^-}{x_n}$$

- Desirable output inefficiency:

$$\rho_y = \frac{1}{M}\sum_{m=1}^{M}\frac{s_m^+}{y_m}$$

- Undesirable output inefficiency:

$$\rho_b = \frac{1}{K}\sum_{k=1}^{K}\frac{s_k^b}{b_k}$$

To assess environmental inefficiency, the Emission Efficiency Index (EEI) is defined as:

$$EEI_i = 1 - \theta_c$$

Where:

- θ_c is the CO_2 emission slack

A higher EEI value (closer to 1) indicates better performance in emissions reduction. EEI values range from 0 to 1.

3.3.1.1. Super-SBM analysis. While the SBM-DDF model provides baseline efficiency scores by jointly considering desirable and undesirable outputs, it assigns an efficiency value of unity to all frontier decision-making units (DMUs), making it impossible to distinguish among the best performers. To enhance discrimination and identify benchmark leaders, the Super-SBM model is subsequently applied, allowing efficiency scores greater than one and enabling ranking of frontier airlines (Li et al., 2013). All decision-

making units (DMUs) achieved feasible solutions across 100 observations (20 DMUs × 5 years), with super-efficiency values ranging from 0.80 to 2.50.

The statistical inference and robustness assessment were performed through two complementary resampling procedures. A jackknife (leave-one-out) analysis with 20 iterations evaluated the sensitivity of frontier positions to the exclusion of individual DMUs, confirming full convergence and frontier stability. A bootstrap procedure with 10 replications was further applied to generate resampled distributions of super-efficiency and GML indices.

The overall mean super-efficiency score was 1.1145 (standard deviation 0.3866), with 46.0 % of DMUs exceeding unity, indicating a balanced discrimination capacity of the model. After bias correction, the adjusted mean decreased slightly to 1.0588 (standard deviation 0.3673), showing that the bias-corrected results remain close to the original estimates. Confidence intervals (90 %, 95 %, and 99 %) were computed based on the bootstrap samples, confirming that efficiency estimates are statistically robust and not unduly influenced by near-frontier observations or small-sample effects.

3.3.2. Global Malmquist-Luenberger index

To evaluate the dynamic performance of airlines over time, the Global Malmquist-Luenberger (GML) Index is used (Chung et al., 1997; Oh, 2010). The GML index measures the change in total factor productivity between period t and $t+1$, accounting for both efficiency improvements and technological progress (Li & Li, 2021).

$$GML_{t\to t+1} = \frac{D_{t+1}^{SBM}(x_{t+1}, y_{t+1}, b_{t+1})}{D_t^{SBM}(x_t, y_t, b_t)} \times \left[\frac{D_{t+1}^{SBM}(x_t, y_t, b_t)}{D_t^{SBM}(x_t, y_t, b_t)} \times \frac{D_{t+1}^{SBM}(x_{t+1}, y_{t+1}, b_{t+1})}{D_t^{SBM}(x_{t+1}, y_{t+1}, b_{t+1})}\right]^{1/2}$$

Where:

- D_t^{SBM} is the SBM-based directional distance function in period t
- The GML index can be decomposed into two components:
- Green Efficiency Change (GEC):

$$GEC = \frac{D_{t+1}^{SBM}(x_{t+1}, y_{t+1}, b_{t+1})}{D_t^{SBM}(x_t, y_t, b_t)}$$

- Global Technological Change (GTC):

$$GTC = \left[\frac{D_{t+1}^{SBM}(x_t, y_t, b_t)}{D_t^{SBM}(x_t, y_t, b_t)} \times \frac{D_{t+1}^{SBM}(x_{t+1}, y_{t+1}, b_{t+1})}{D_t^{SBM}(x_{t+1}, y_{t+1}, b_{t+1})}\right]^{1/2}$$

Thus, the full decomposition is:
$GML = GEC \times GTC$

- If $GML > 1$: Productivity has improved.
- If $GML < 1$: Productivity has declined.
- If $GML = 1$: No change in performance.

3.4. Indicator redundancy and dimensionality reduction

The selection of input–output indicators is a critical step in ensuring both the validity and discriminatory power of data envelopment analysis (DEA). In efficiency studies with a relatively small number of decision-making units (DMUs), the inclusion of multiple highly correlated indicators can introduce redundancy, inflate dimensionality, and reduce the ability of the DEA model to distinguish performance differences. To address this issue, a systematic procedure was implemented to identify, evaluate, and mitigate indicator collinearity, as illustrated in chapter 5.2 and 5.3.

3.5. Model justification and interpretive scope

The integrated modelling framework—combining the non-oriented slack-based measure directional distance function (SBM-DDF) with the Global Malmquist–Luenberger (GML) productivity index, Random Forest, and the Cloud model—was selected to capture both the static and dynamic dimensions of airlines' low-carbon efficiency. The SBM-DDF is particularly appropriate for this context because it simultaneously accounts for desirable (RTK) and undesirable (CO_2) outputs, allowing the evaluation of environmental and operational performance within a single production frontier. Unlike radial or oriented DEA models, the non-oriented DDF avoids proportional input–output assumptions, which is crucial in airline operations where emission reductions and capacity changes may not occur uniformly.

The GML extension enables inter-temporal comparison and decomposition into technical change (GTC) and efficiency change (GEC), providing insights into whether observed productivity shifts arise from managerial practices or technological advancement. The Random Forest and Cloud models further enhance interpretive robustness: the former identifies key operational drivers of

efficiency heterogeneity, while the latter characterises the stability and reliability of airline performance across dimensions (Ex, En, He).

This integrated design does not seek to infer strict causality but rather to provide relative benchmarking evidence that can inform managerial and policy decisions. By triangulating results from multiple complementary models, the framework enhances the credibility, interpretive depth, and reproducibility of efficiency assessment in complex, multi-output airline systems.

4. Data Collection and variables

This study assesses the low-carbon competitiveness of 20 major global airlines over the period 2018–2022 using a three-stage analytical framework. A wide range of operational, financial, and environmental indicators were collected to support the analysis, beginning with preliminary visualization via cloud models, followed by feature importance ranking using Random Forest, and concluding with efficiency evaluation through the SBM-DDF-GML model.

4.1. Full indicator framework across analytical stages

To ensure a comprehensive and multidimensional evaluation of airline low-carbon efficiency, this study selected 12 indicators (shown in Table 4.1). These included traditional efficiency metrics as well as flight distance and load factor, which provide additional dimensions of operational and environmental performance. Available data on revenue passenger kilometres (RPK) and revenue tonne kilometres (RTK) were consolidated into a single RTK measure covering both passenger and cargo traffic, in order to avoid overlapping representation of capacity. The selection criteria were based on theoretical relevance, data availability, industry practices, and precedent in environmental efficiency studies in aviation and transport sectors (Zhou et al., 2021). Each indicator was mapped to one or more stages of the analytical framework—cloud model, Random Forest feature ranking, and SBM-DDF-GML modeling—depending on its explanatory strength and alignment with modeling assumptions.

4.2. Data Collection and Processing Protocol

For all indicators, the primary sources were each airline's official Annual Report. If data for a specific indicator or year was missing from these official disclosures, the data was supplemented using statistics from the CAPA (Centre for Aviation) Airline Profile database. When using official reports, all figures were converted to standardised units for cross-airline comparison (e.g., fuel consumption reported in million gallons was converted to tonnes). For financial indicators, data were first extracted in their local currency as reported in the annual report, then converted to U.S. Dollars (USD) using the annual average exchange rate for the corresponding year, as provided by CAPA; no adjustments were made for inflation. All data for a given year were collected according to each company's specific financial year as stated in their annual report. CO_2 emissions data were collected from annual and sustainability reports, based on Scope 1 CO_2 emissions from flight operations where available. If this was not specified, total Scope 1 CO_2 emissions were used.

All input and output variables were harmonized to consistent units across airlines (see Appendix A). Specifically, CO_2 emissions were expressed in tonnes, RTK in tonne-kilometres, employees as headcount, aircraft as fleet size, load factor in %, and flight distance as average stage length in kilometres. To ensure comparability and to prevent scale effects, all variables were normalized prior to DEA and RF analysis using Min–Max scaling to the [0,1] range:

Table 4.1
Full Indicator Framework Across Analytical Stages.

Category	Indicator	Source	Used in Cloud / RF / Correlation	Used in SBM-DDF-GML
Environmental	CO_2 Emissions	Annual Report/CSR	Yes	Yes (undesirable output)
Operational	Fuel Consumption	Annual Report/CSR	Yes	No
	Flight distance (stage length)	Annual Report/ CAPA	Yes	Yes (input)
	Load factor	Annual Report/ CAPA	Yes	Yes (input)
	RTK (Revenue Ton Kilometers – both passenger and cargo)	Annual Report/ CAPA	Yes	Yes (desirable output)
	Number of Passengers	Annual Report/CSR	Yes	No
Capacity	ASK (Available Seat Kilometers)	Annual Report/ CAPA	Yes	No
	Number of Flights	Annual Report/ CAPA	Yes	No
	Number of Employees	Annual Report/CSR	Yes	Yes (input)
	Number of Aircraft	Annual Report/ CAPA	Yes	Yes (input)
	Operating Revenue	Annual Report	Yes	No
	Operating Cost	Annual Report	Yes	No

$$y^{*}_{train} = \frac{y_{train} - min(y)}{max(y) - min(y)}$$

Predictions from RF were subsequently converted back to the original scale using:

$$\hat{y}_{test} = \hat{y}^{*}_{test} \times (max(y) - min(y)) + min(y)$$

This preprocessing step ensures consistent comparability across variables while retaining interpretability in the original units.

5. Empirical analysis

5.1. Cloud Model-Based visualization of carbon efficiency

The cloud model was applied to visualize the distribution of airline carbon intensity from two perspectives: carbon emissions per revenue tonne-kilometre (RTK) and carbon emissions per unit of operating revenue. In both cases, the parameters Ex, En, and He provide a statistical characterization of performance: Ex reflects the average level, En represents variability, and He captures the stability of variability. This enables not only visualization but also quantification of differences across carriers.

5.1.1. Carbon emissions per RTK

Fig. 5.1 presents the cloud distributions of CO_2/RTK across the 20 airlines. The industry average is marked with a dashed blue line, and each airline's expectation (Ex) is shown as a vertical magenta line.

Efficient and stable performers — Airlines such as Cathay Pacific, Korean Air, EasyJet, Ryanair, China Southern, and Singapore Airlines exhibit low Ex values (close to or below 1.0) with narrow entropy (En) and relatively low hyper-entropy (He). These features indicate consistently low emissions per RTK with minimal volatility, reflecting modern fleets, simplified route networks, and standardized aircraft types.

Volatile or inconsistent performers — Emirates, ANA, and Air New Zealand display wider cloud spreads and higher En values, with relatively large He, suggesting unstable performance. This may be associated with reliance on long-haul networks and mixed fleet strategies, where efficiency fluctuates depending on route structure and load factors.

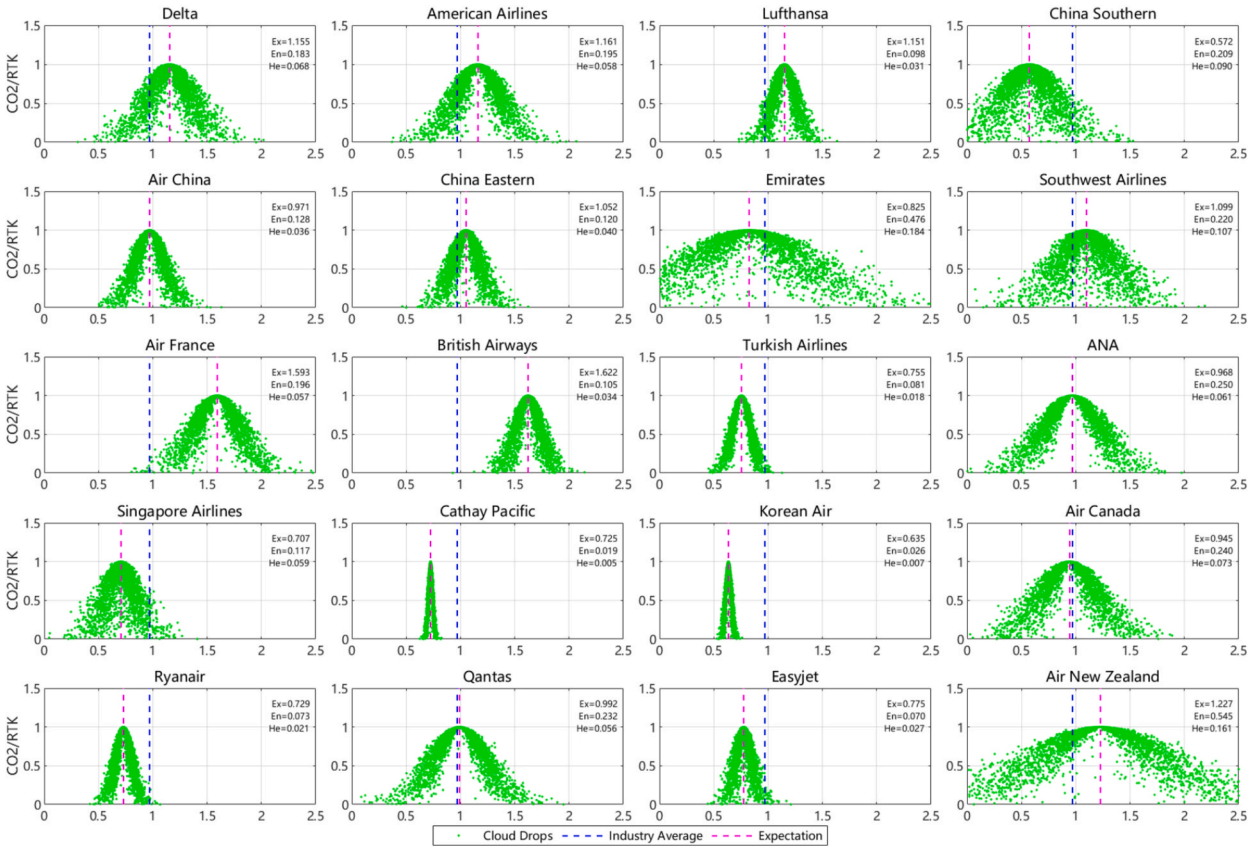


Fig. 5.1. Carbon Intensity Index of CO_2/RTK for 20 global airlines.

Less efficient performers — Legacy European carriers such as Lufthansa, Air France, and British Airways have Ex values well above the industry average and relatively broad distributions, indicating persistent inefficiencies in aircraft utilization and energy intensity.

The Ex–En scatterplot (Fig. 5.2) further highlights these patterns. Airlines positioned in the lower-left quadrant (low Ex and low En) can be considered efficiency benchmarks. In contrast, airlines with high Ex and high En (e.g., Air New Zealand, Emirates) face both structural inefficiency and volatility. The ellipse sizes, proportional to He, illustrate that even airlines with moderate averages may differ significantly in the reliability of their performance.

5.1.2. *Carbon emissions per unit of revenue*

Fig. 5.3 illustrates carbon intensity relative to revenue. This measure combines operational efficiency with economic outcomes, reflecting an airline's ability to generate revenue under a carbon-constrained environment.

High carbon–profitability efficiency — Carriers such as EasyJet, Qantas, and ANA achieve low Ex values with narrow En, suggesting both efficient emissions management and strong revenue generation. Their favorable positioning demonstrates competitive resilience in a context of rising carbon costs.

Volatile revenue efficiency — British Airways, Lufthansa, Air France show larger En and He, indicating greater instability in carbon intensity per revenue unit. This volatility may result from fluctuating yields, heterogeneous network structures, and market-specific pricing challenges.

High Ex values, implying greater emissions per unit revenue. This outcome reflects both operational inefficiencies and the challenges of competing in markets with high fixed costs and older fleet compositions.

The Ex–En scatterplot for CO_2/Revenue (Fig. 5.4) provides a comparative overview. Airlines in the lower-left area (low Ex, low En) represent best practice in balancing carbon performance with revenue generation. By contrast, those in the upper-right, such as British

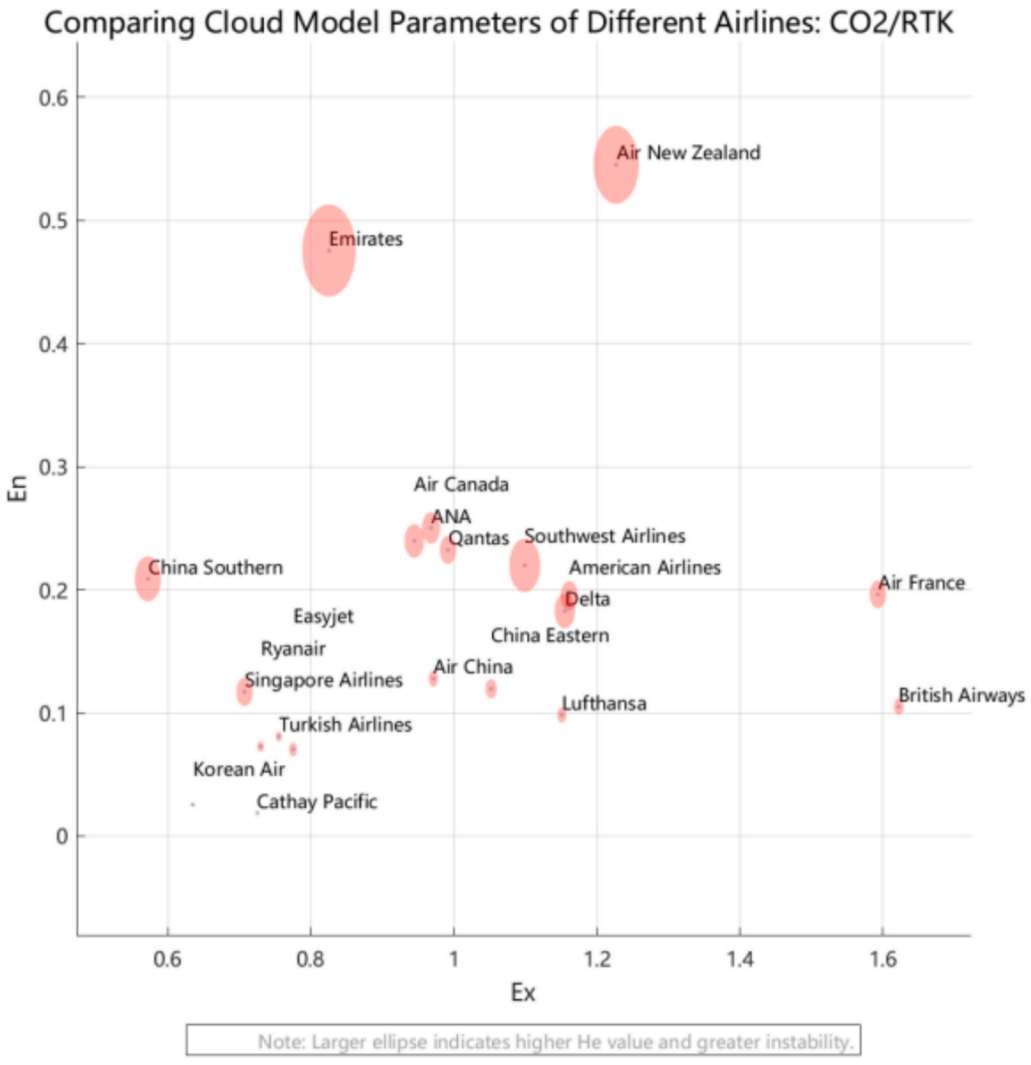


Fig. 5.2. Comparing cloud model parameters of different airlines: CO2/RTK.

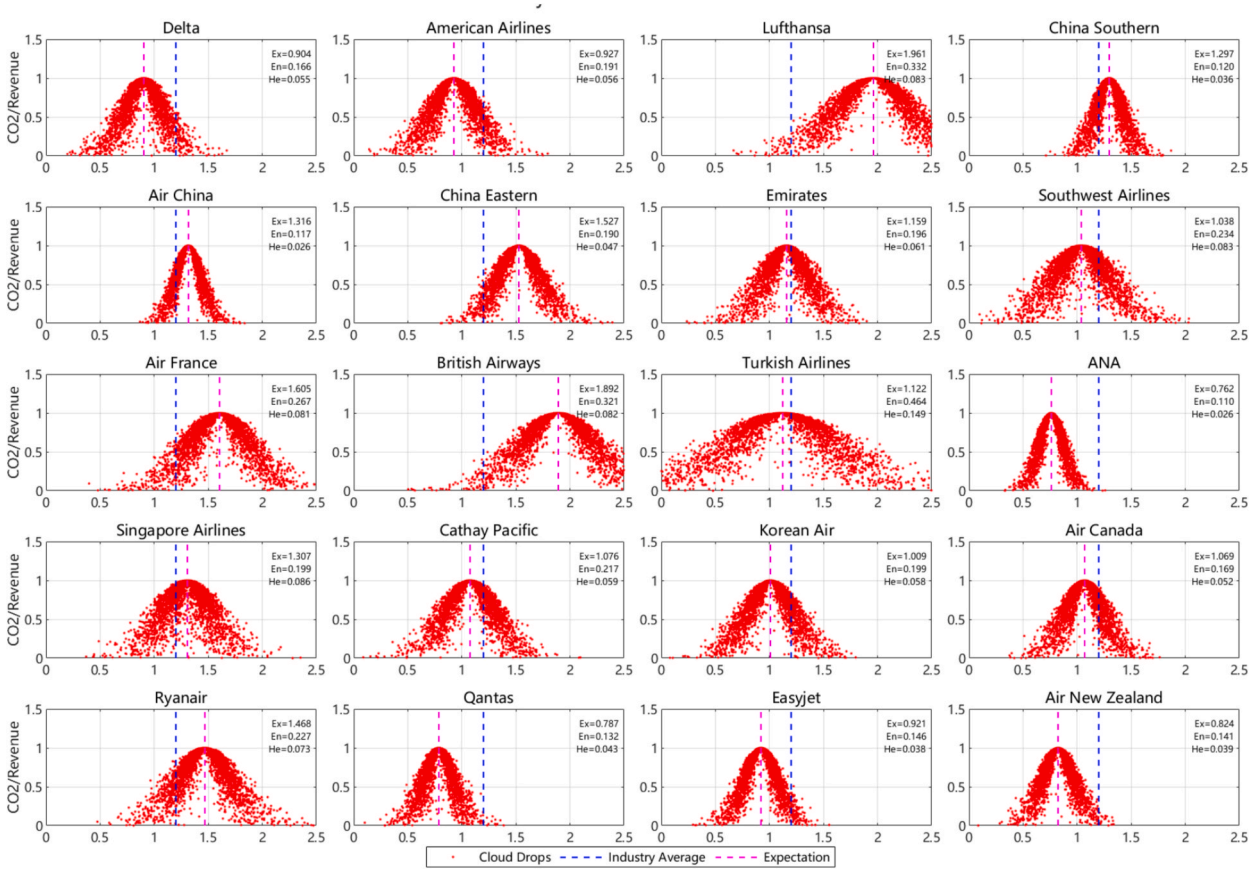


Fig. 5.3. Carbon Intensity Index of CO2 per unit of operating revenue for 20 global airlines.

Airways and Lufthansa, combine higher average carbon intensity with greater instability, underscoring the need for both efficiency improvements and revenue model adjustments.

5.2. *Correlation analysis*

The correlation analysis revealed widespread collinearity among the 12 candidate indicators. The Pearson correlation matrix (Fig. 5.5, top) demonstrates particularly strong associations between several pairs, such as ASK and operating revenue ($r = 0.91$, $p < 0.001$), fuel consumption and CO_2 emissions ($r = 1.00$, $p < 0.001$), and number of flights and number of aircraft ($r = 0.81$, $p < 0.001$). The statistical significance map (Fig. 5.5, bottom left) confirms that most of these correlations are highly significant, with –log10 (p-values) exceeding 2 in many cases. The distribution of correlation coefficients (Fig. 5.5, bottom right) shows an average correlation of 0.51, with the majority of values skewed towards strong positive relationships. Together, these results provide clear evidence of multicollinearity across the candidate indicator set, underscoring the need for dimensionality reduction prior to DEA modelling.

5.3. *Dimensionality reduction procedure*

To systematically reduce redundancy while retaining the most informative variables, a three-stage procedure was adopted:

1) Clustering analysis – Hierarchical clustering was conducted using Ward linkage, and the elbow method was applied to determine an appropriate threshold for cluster separation (shown in Fig. 5.6). This step grouped variables into clusters based on similarity in correlation structure.
2) Feature importance evaluation – Within each cluster, Random Forest regression was employed to evaluate the relative importance of each variable (shown in Fig. 5.7). This approach identifies indicators that contribute most to explaining variance in the dataset and thus ensures that retained variables are informative.
3) Collinearity testing – Variance inflation factor (VIF) values were calculated for each variable. Variables with VIF values above the established threshold were excluded from the final set. Fig. 5.8 illustrates the VIF analysis, with retained variables lying to the left of the threshold line (yellow line).

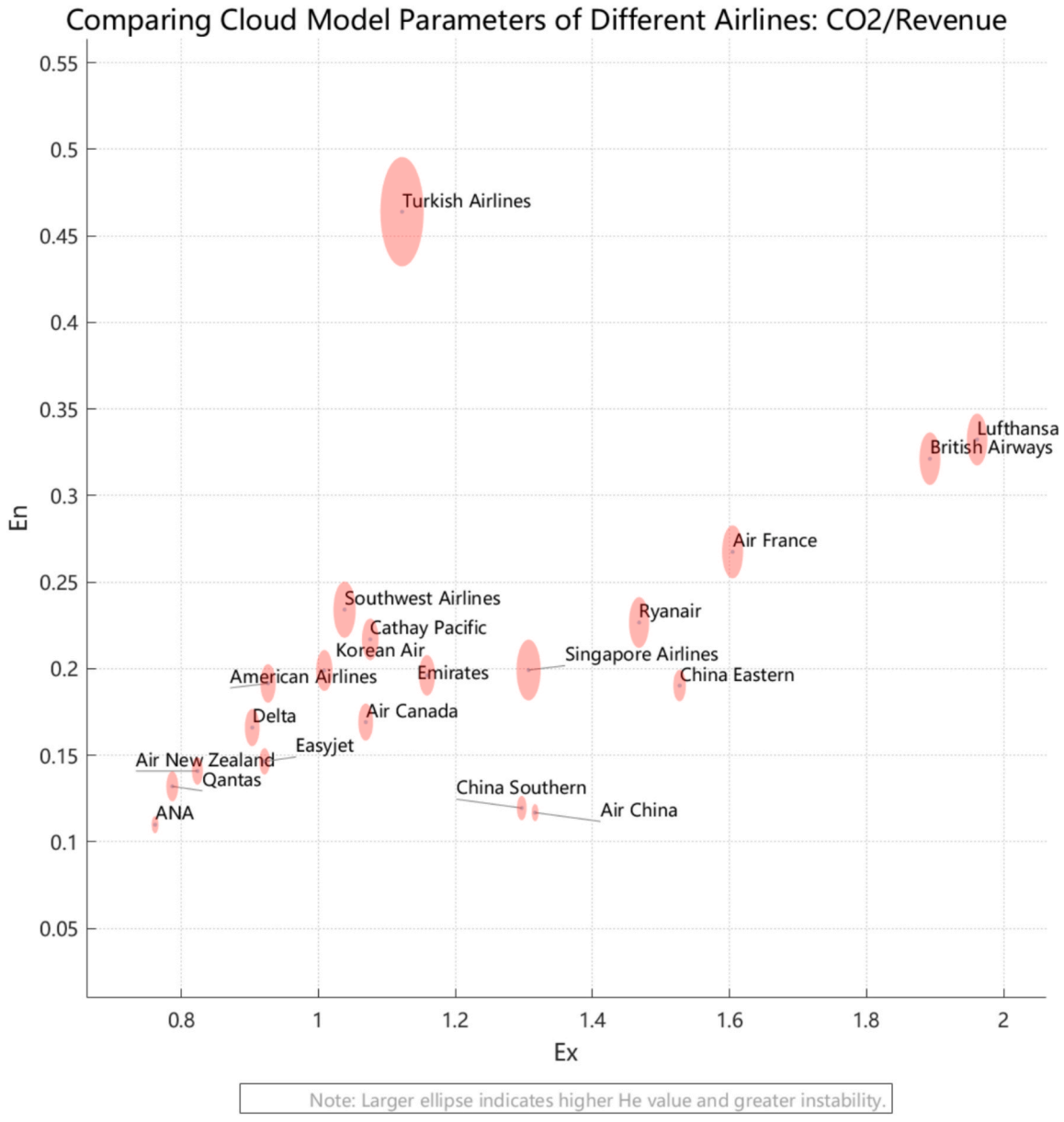


Fig. 5.4. Comparing cloud model parameters of different airlines: CO2/Revenue.

By combining clustering, feature importance evaluation, and VIF analysis, the procedure enabled systematic removal of redundant variables while prioritizing those with both high informational value and low multicollinearity.

5.4. Predictive performance of the Random Forest model

Random Forest regression was used to identify the relative contribution of explanatory variables to airline carbon efficiency. The model was configured with the following hyperparameters:

- Number of trees: 500
- mtry (features considered per split): square root of the number of features (default setting)
- Tree depth: unrestricted, with minimum leaf size optimized via OOB error
- Split criterion: mean squared error (MSE)
- Random seed: 42 for reproducibility

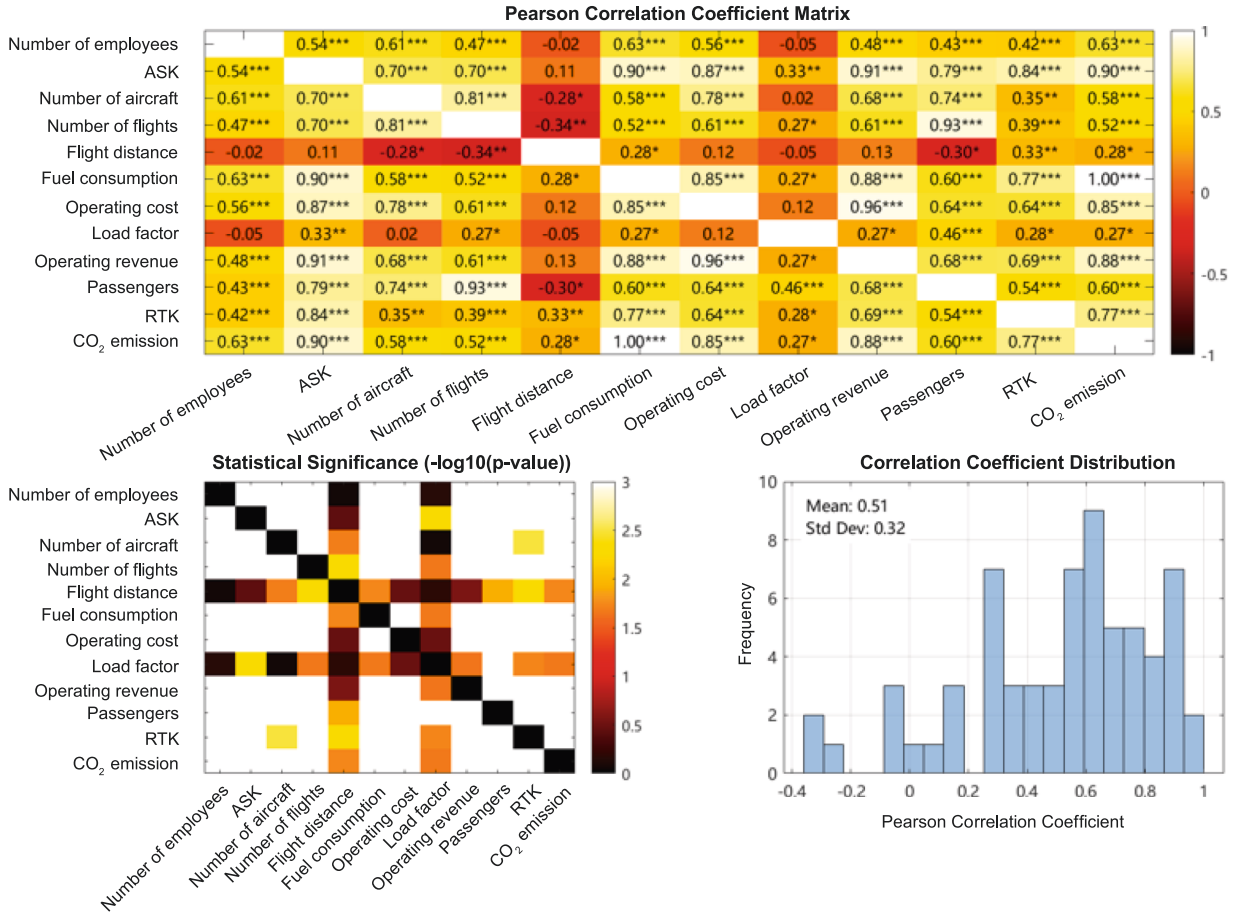


Fig. 5.5. Pearson correlation matrix of candidate indicators.

The minimum leaf size was optimized by evaluating OOB error across candidate values {5, 10, 20, 50, 100, 200, 500}. The OOB error curve (Fig. 5.9) shows that error decreased as leaf size was reduced, with the optimal balance achieved at 5.

Model accuracy was assessed using both OOB error and 5-fold cross-validation. Cross-validation results indicated strong predictive performance: average Test R^2 of 0.84, Test RMSE of 4.01, and Test MAE of 3.17 (Fig. 5.10). These results confirm that the RF model achieved stable generalization.

Feature importance was quantified using both Mean Decrease Impurity (MDI) and permutation importance. The resulting rankings (Figs. 5.11 and 5.12) consistently identified RTK as the most influential predictors of CO_2 emissions. Lower-ranked variables included number of employees and number of aircraft, though their importance was not negligible. The similarity between MDI and permutation-based rankings confirms the robustness of the importance assessment, while minor differences are observed, reflecting the known bias of MDI toward correlated variables.

5.5. *Sensitivity analysis of key emission drivers*

To examine the robustness of the Random Forest findings and the elasticity of emission outcomes to changes in operational drivers, a sensitivity analysis was performed with MATLAB. Each input variable was perturbed sequentially within a ± 100 % range of its observed values, while holding other variables constant. This adjustment ensures that the perturbation span remains within a plausible operational domain, avoiding the extrapolation risks noted in earlier drafts.

The results are presented in Fig. 5.13. Across all variables, positive perturbations (increases) led to higher predicted CO_2 emissions, while negative perturbations (decreases) produced emission reductions. The magnitude of change varied by variable:

- Revenue tonne kilometres (RTK): The strongest driver, showing nearly proportional elasticity. A 100 % increase in RTK corresponded to a ~ 30 % increase in CO_2 emissions, whereas halving RTK reduced emissions by nearly the same proportion.
- Number of aircraft and number of employees: Both exhibited moderate effects, with approximately 15–20 % changes in emissions at full perturbation. These reflect resource scaling effects on emissions.

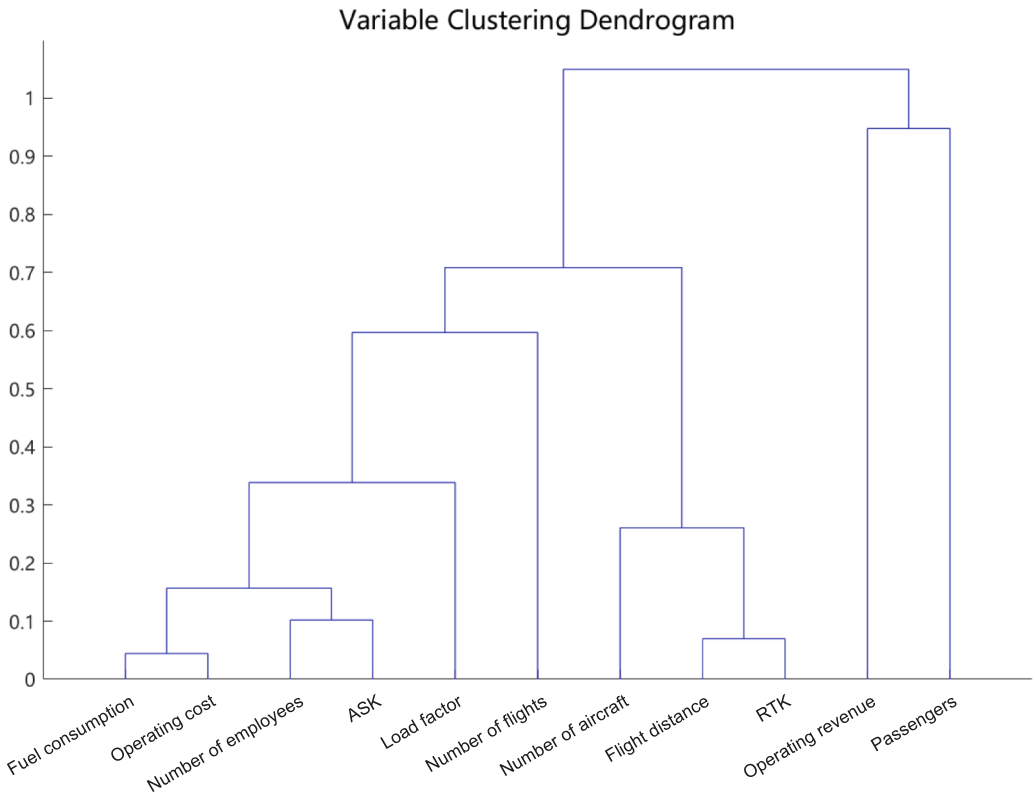


Fig. 5.6. Variable Clustering Dendrogram.

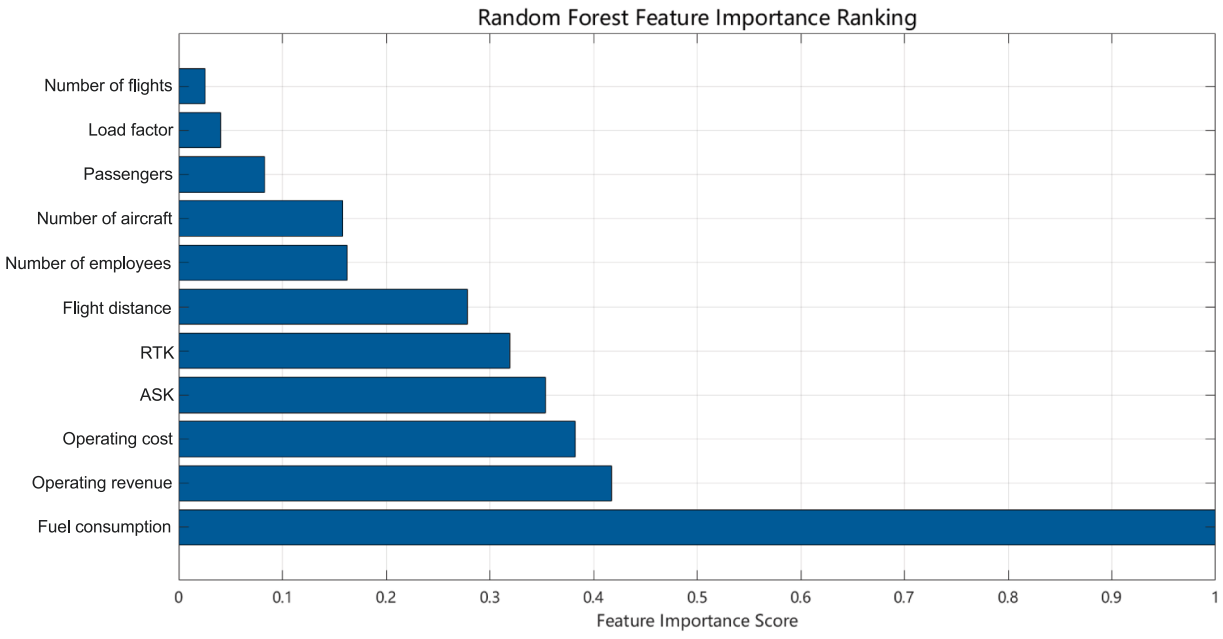


Fig. 5.7. Random Forest importance ranking.

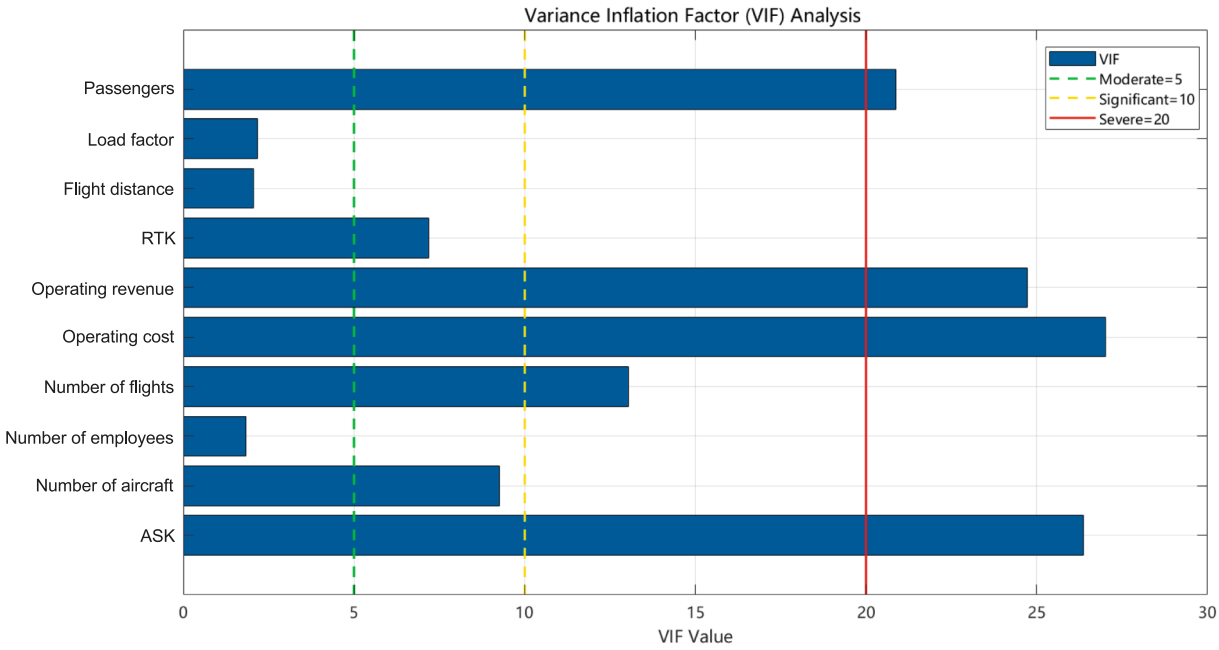


Fig. 5.8. Variance Inflation Factor (VIF) analysis for dimensionality reduction.

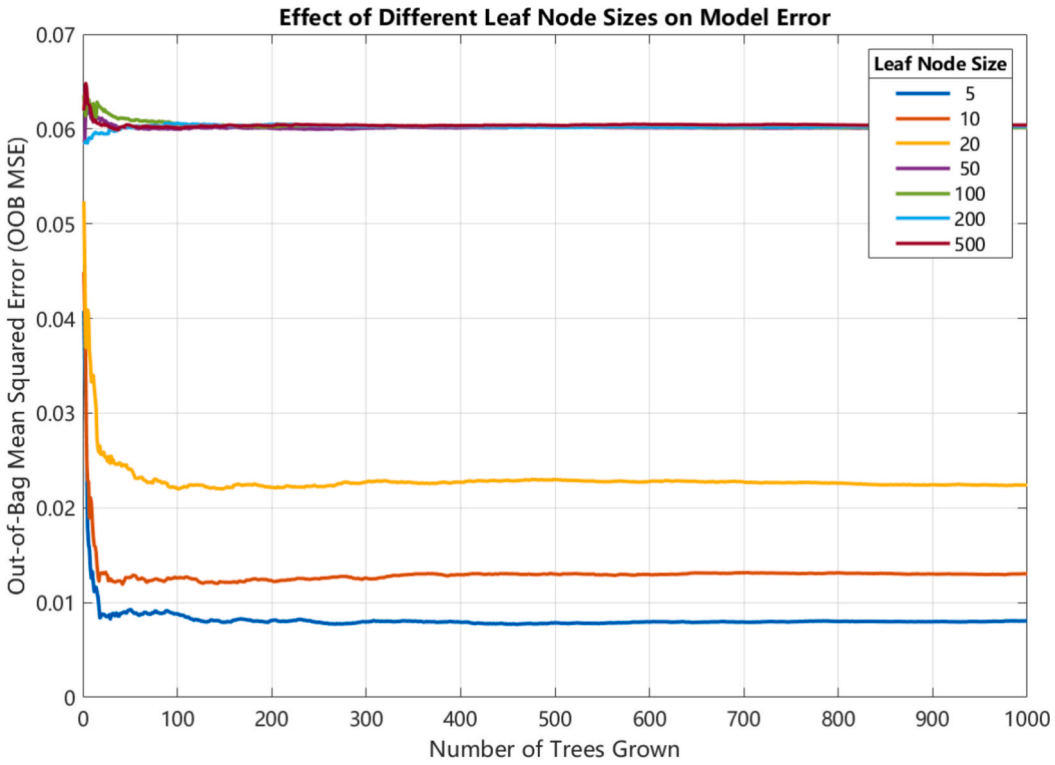


Fig. 5.9. Out-of-bag error curves for different minimum leaf sizes.

- Flight distance (average stage length): Longer stage lengths were associated with modestly higher emissions (10–15 %), consistent with the balance between longer flight sectors and reduced frequency.
- Load factor: The weakest driver in the sensitivity test, with ± 100 % perturbations leading to comparatively smaller changes (5–10 %). This is consistent with RF feature importance rankings, which placed load factor below structural and scale variables.

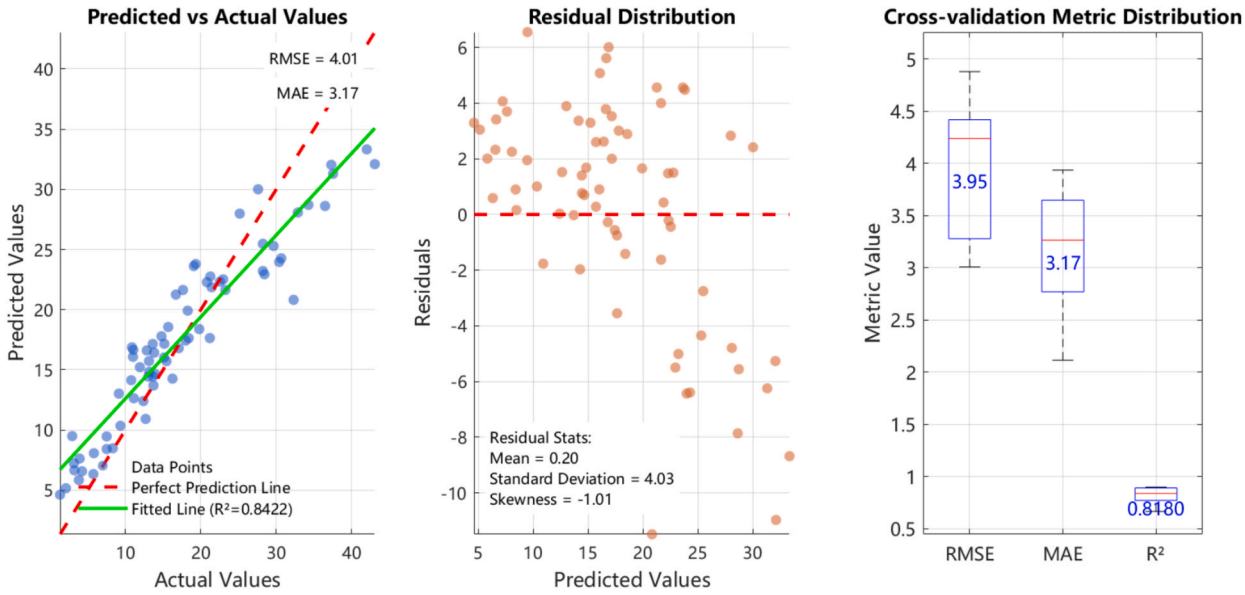


Fig. 5.10. Cross-validation results for the Random Forest model.

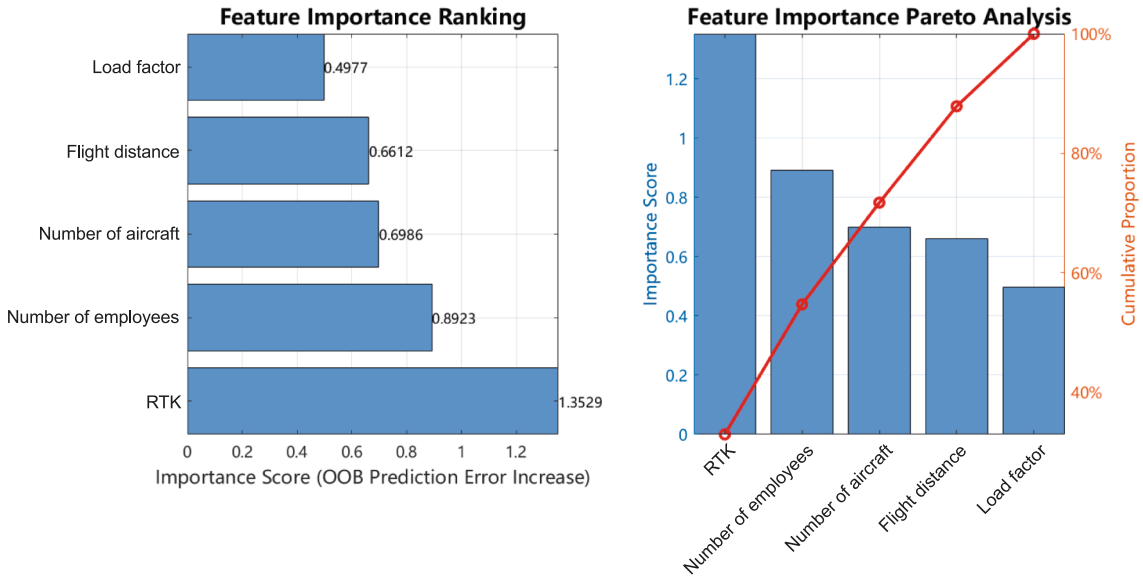


Fig. 5.11. Mean Decrease Impurity (MDI) variable importance rankings.

Overall, the analysis confirms that RTK is the dominant emission driver, followed by fleet and labour scale indicators, while load factor plays a secondary role.

Informed by these findings, the airline low-carbon efficiency evaluation indicator framework for the SBM-DDF model was finalized. As illustrated in Table 5.1, the selected system incorporates five input indicators: labor input (number of employees), capital input (number of aircraft), and operational input (load factor and flight distance). RTK is designated as the desirable output and CO_2 emissions is designated as the single undesirable output.

This indicator system ensures alignment between methodological design and empirical relevance. It captures the multidimensional nature of airline operations while maintaining model parsimony and interpretability. The variables selected provide a robust basis for evaluating low-carbon efficiency across heterogeneous carriers and will serve as the foundation for the subsequent SBM-DDF-GML modeling and benchmarking in the next chapter.

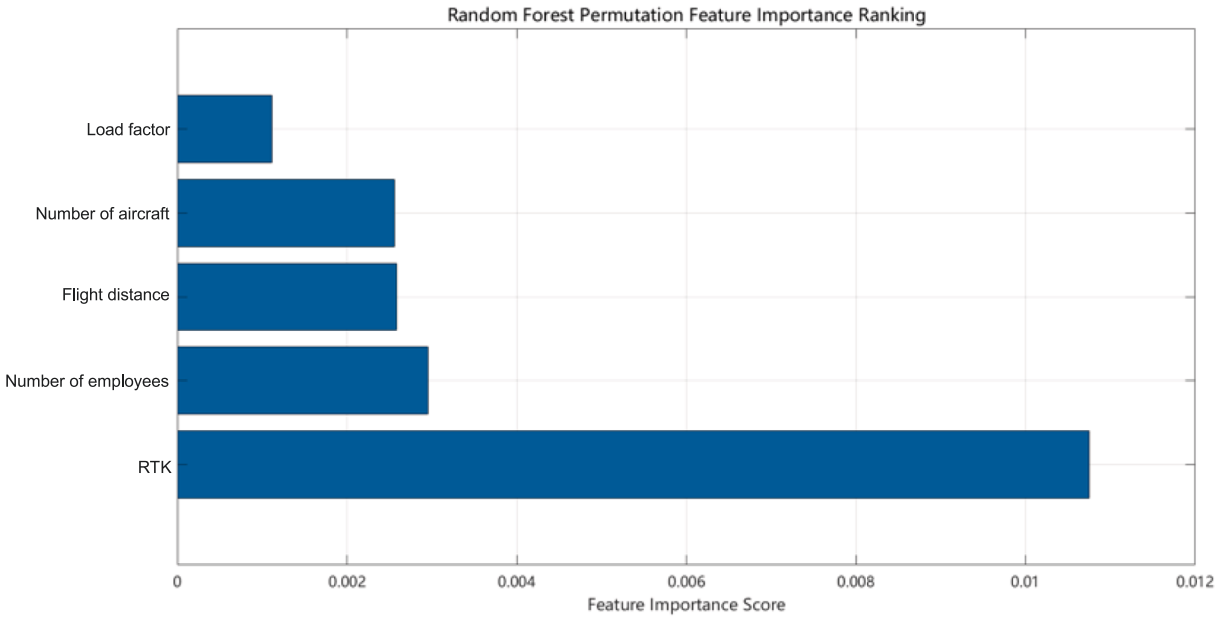


Fig. 5.12. Permutation importance rankings.

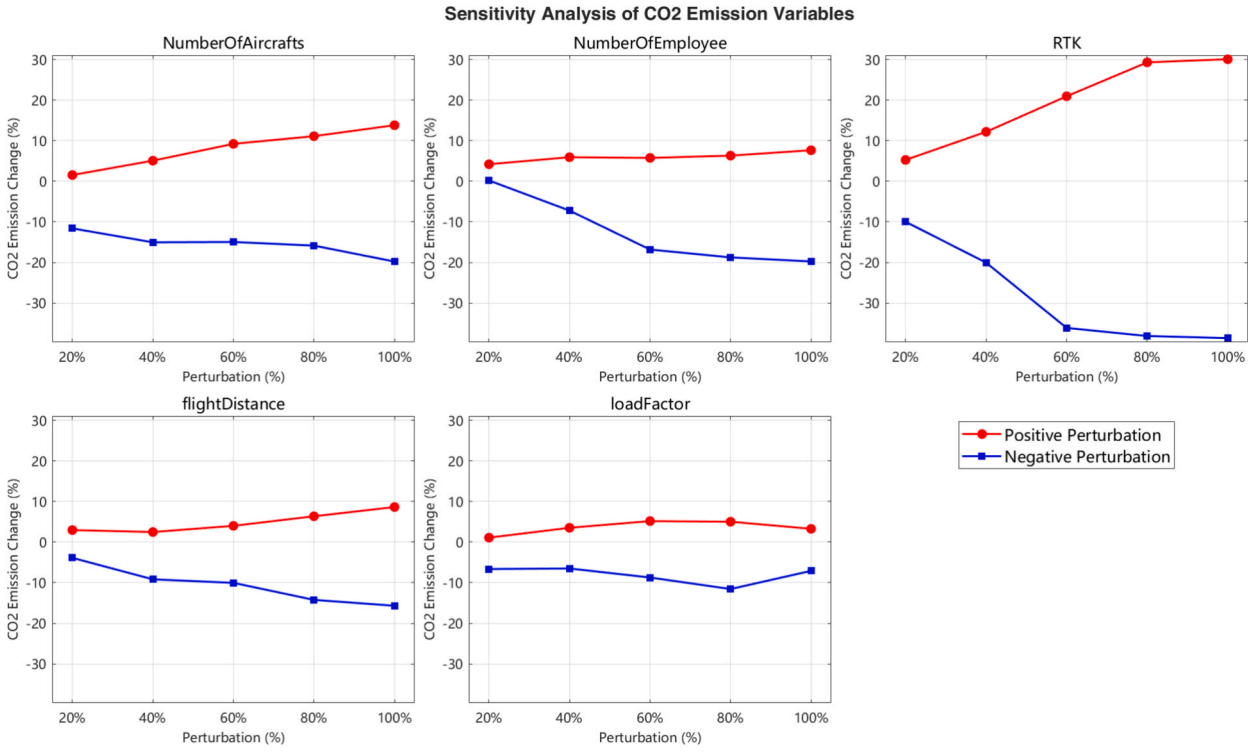


Fig. 5.13. Sensitivity analysis of CO_2 emission variables.

6. Results and Discussion

6.1. *Efficiency analysis results*

The static SBM-DDF and Super-SBM models were applied to evaluate the low-carbon operational efficiency of twenty global airlines

Table 5.1
Low-carbon efficiency evaluation indicator framework.

Input/Output	Variable	Interpretation
Labor input	Number of employees	Workforce resources used
Capital input	Number of aircraft	Fleet capacity resource
Operational input	Load factor	Reflects seat/tonne utilization efficiency
	Flight distance (stage length)	Captures network configuration and route mix
Desirable output	RTK	Transport work performed
Environmental output	CO_2 emission	

from 2018 to 2022. Both models provide complementary perspectives on airlines' relative performance: the SBM-DDF captures static efficiency based on input–output–emission slack reduction, whereas the Super-SBM extends discrimination beyond the efficiency frontier to identify benchmark-leading DMUs. Fig. 6.1 presents the results of the SBM-DDF model, and Fig. 6.2 compares the corresponding Super-SBM super-efficiency outcomes.

The SBM-DDF results reveal that most airlines operated below the efficient frontier, with efficiency scores generally ranging between 0.30 and 0.97. The pre-pandemic years (2018–2019) exhibit comparatively high and stable efficiencies, where carriers such as China Eastern (0.97), Air China (0.93–0.96), and American Airlines (0.90) consistently approached the frontier. In contrast, European network carriers—such as Lufthansa (0.37) and Air France (0.39)—displayed greater inefficiencies, likely reflecting structural differences in network complexity and fleet composition. During the 2020–2021 period, the pandemic shock caused a pronounced compression of efficiency scores and greater dispersion across the sample. Efficiency reductions were particularly evident for China Eastern (0.41 in 2020), American Airlines (0.36 in 2021), and China Southern (0.33 in 2021), while some network carriers—such as Lufthansa (0.85–0.91) and Air France (0.76–0.82)—maintained comparatively stable performance. A moderate recovery emerged in 2022, with airlines such as American (0.92), Southwest (0.92), and Emirates (0.95) returning toward pre-COVID efficiency levels.

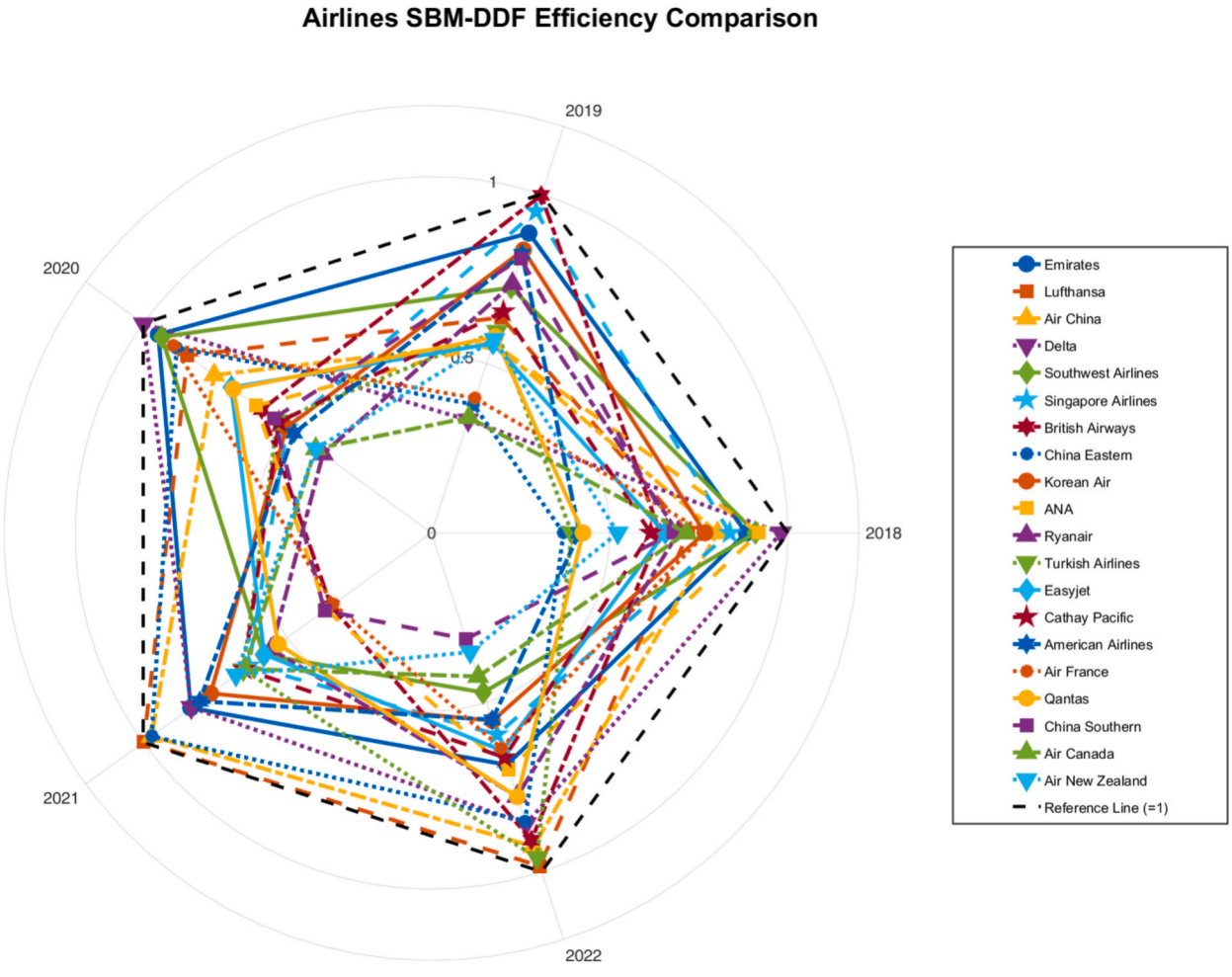


Fig. 6.1. Efficiency scores of the 20 airlines from the SBM-DDF model (2018–2022).

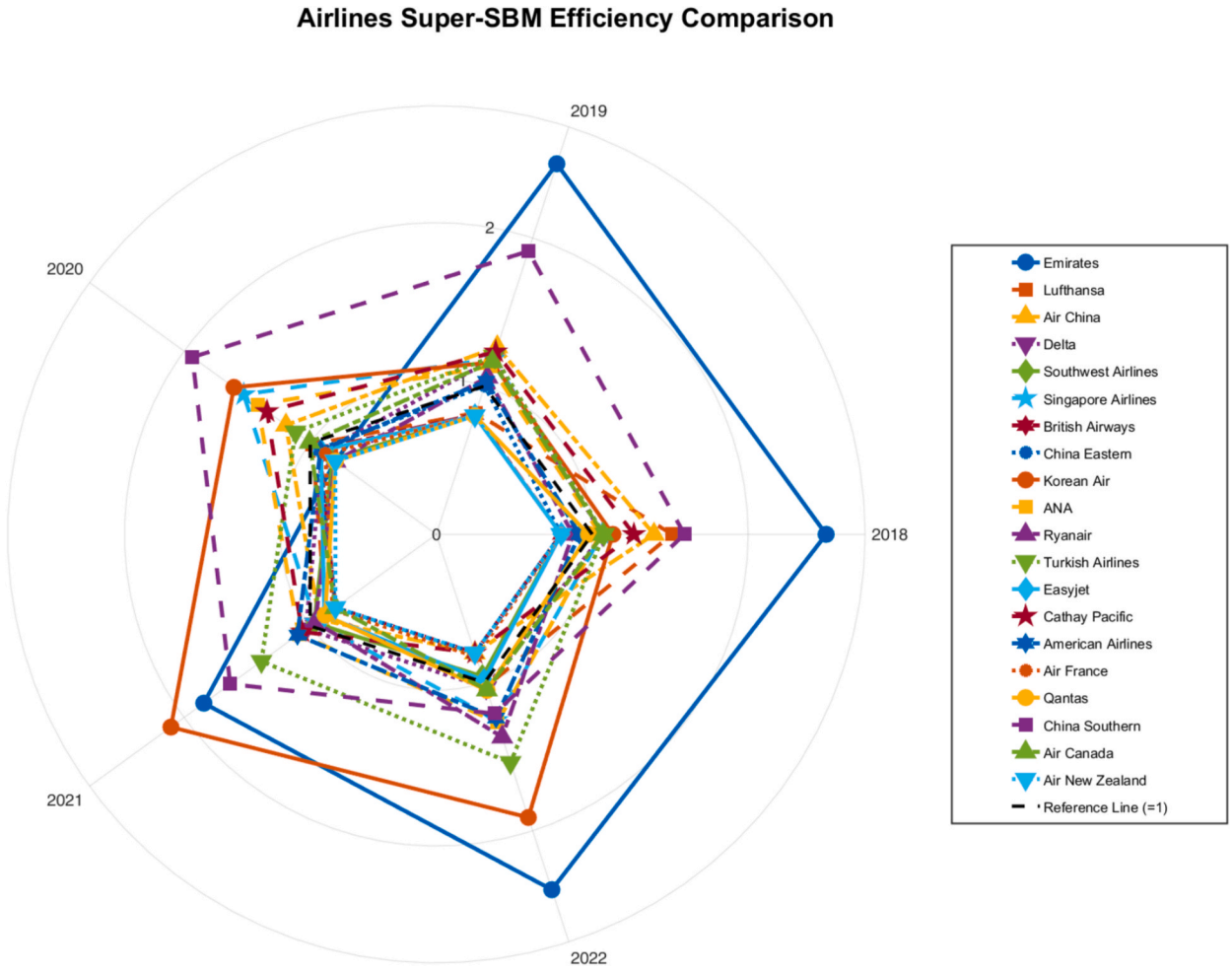


Fig. 6.2. Efficiency scores of the 20 airlines from the Super-SBM model (2018–2022).

However, others—including Korean Air (0.31) and Air New Zealand (0.44)—remained below their historical benchmarks. It should be noted that several extreme observations in the pre- or post-pandemic years may partly reflect fiscal year discrepancies across airlines, meaning that reported values may implicitly capture pandemic effects depending on each carrier's financial reporting cycle.

The Super-SBM results extend the frontier to allow efficiency scores greater than unity, thereby enabling clearer discrimination among high-performing airlines. The observed super-efficiency values range from the lower bound of 0.80 to a maximum of 2.50, indicating substantial heterogeneity in frontier positions. Emirates consistently appears as a benchmark leader, with super-efficiency values of 2.50 in 2018–2019, a pandemic-related decline to 0.80 in 2020, and a strong rebound to 1.79 in 2021 and 2.37 in 2022. Similarly, China Southern demonstrates stable frontier dominance with values between 1.67 and 1.99 across the study period. Korean Air shows a sustained upward trajectory from 1.21 in 2018 to 2.17 in 2022, while Singapore Airlines and All Nippon Airways (ANA) achieved super-efficiency levels above 1.0 even during the most challenging pandemic years. Turkish Airlines also improved significantly, increasing from 1.28 in 2021 to 1.65 in 2022.

A number of major full-service carriers, including American, Air Canada, British Airways, Air France, and Qantas, maintained stable yet moderate super-efficiency values around unity, reflecting relatively balanced but non-frontier performance. The clustering of values at the imposed lower limit of 0.80—such as for Southwest, Air France, EasyJet, and Air New Zealand in the early years—indicates consistent underperformance relative to the best-practice frontier.

Comparing the two models, several key contrasts emerge. The SBM-DDF model yields smoother, more conservative trends across years, emphasizing proportional slack reduction in inputs and emissions. The Super-SBM model, in turn, amplifies the differentiation among efficient carriers, particularly during periods of market turbulence. For instance, Emirates, China Southern, and Korean Air exhibit high super-efficiency values (above 2.0) while their SBM-DDF scores remain below unity, implying that these carriers set the best-practice frontier in specific operational or emission dimensions, even though residual inefficiencies persist when evaluated holistically. Conversely, American Airlines and Air Canada achieve SBM-DDF scores close to 0.90–0.95 but super-efficiency values around 1.0, suggesting consistent yet non-dominant operational efficiency.

During the 2020–2021 shock, SBM-DDF results reveal an industry-wide efficiency contraction, while the Super-SBM model shows a

reordering of the frontier: long-haul carriers with strong cargo operations, such as Korean Air, China Southern, and Turkish Airlines, achieved relative super-efficiency gains, whereas network carriers heavily dependent on passenger traffic (e.g., Emirates and American Airlines) experienced temporary declines.

Overall, the comparative evidence indicates that the SBM-DDF model provides a robust and conservative benchmark for multi-dimensional efficiency assessment, whereas the Super-SBM offers enhanced discriminatory power to identify and characterize frontier leaders. The divergence between the two models highlights that even benchmark airlines retain measurable improvement potential, especially regarding input utilization, fleet energy efficiency, and CO_2 intensity management. This finding aligns with the subsequent GML decomposition results, which show that variations in managerial efficiency (GEC) rather than technological progress (GTC) were the primary source of productivity fluctuations during the pandemic period.

6.2. Environmental efficiency (EEI) analysis

To further isolate the environmental dimension of efficiency, the CO_2 emission slack from the SBM-DDF model was converted into a single-dimensional Environmental Efficiency Index (EEI). This measure captures the extent to which airlines minimize CO_2 emissions relative to their operational frontier. Across the sample, EEI values range from 0.86 to 1.00, with a mean of approximately 0.97, indicating generally high environmental efficiency but notable inter-airline and temporal variation (shown in Fig. 6.3).

In 2018–2019, several carriers—including Lufthansa, Emirates, ANA, Singapore Airlines, Ryanair, Qantas, and EasyJet—achieved perfect environmental efficiency (EEI = 1.00), reflecting effective emission control prior to the pandemic. North American carriers such as Delta (0.86–0.87) and American Airlines (0.89) consistently lagged slightly behind, likely due to larger network scope and older fleet structures. The pandemic period (2020–2021) produced a marked convergence toward full environmental efficiency: nearly all airlines reached EEI $\geq$ 0.96, and seven attained EEI = 1.00. This improvement largely reflects the sharp contraction in flight activity during COVID-19, which lowered absolute CO_2 output, rather than pure technological or managerial advancement. In 2022, as global traffic rebounded, EEI values remained high (mean = 0.97), with Asian full-service carriers such as ANA, Cathay Pacific, Korean Air and Singapore Airlines maintaining perfect scores, while U.S. and European network carriers experienced modest declines (e.g., Delta 0.89; American Airlines 0.88; Air France 0.92).

The EEI trajectory indicates that CO_2-specific performance was less volatile than overall SBM-DDF efficiency, confirming that operational recovery and emission efficiency followed different paths after the pandemic. While most carriers approached the environmental frontier, small differences persist, likely attributable to variations in fleet renewal, energy mix (SAF adoption), and fiscal-year reporting differences that influence observed CO_2 totals. The results demonstrate that environmental performance improvements during 2020–2021 should be interpreted cautiously, as they may capture temporary reductions in flight operations rather than sustained efficiency gains.

6.3. Global Malmquist–Luenberger analysis

The Global Malmquist–Luenberger (GML) productivity index was applied to capture dynamic changes in airlines' low-carbon performance between 2018 and 2022. The decomposition into efficiency change (GEC) and technical change (GTC) allows for distinguishing between managerial and technological sources of productivity variation.

6.3.1. Overall GML productivity trends

As shown in Fig. 12.1 and 12.4, the GML index across the sample exhibits moderate fluctuations, with a mean of 1.0039, standard deviation of 0.0289, and a range from 0.9501 to 1.0499. Approximately 55 % of airline–year observations show GML $>$ 1, indicating productivity improvements, while 45 % show GML $<$ 1, reflecting temporary regressions, particularly during the 2020–2021 pandemic period. The aggregate trend reveals that productivity declined slightly in 2019–2021 and rebounded in 2021–2022, consistent with global air transport recovery patterns.

6.3.2. Decomposition: Technical and efficiency change

The technical change (GTC) component remained close to unity (mean = 0.9992, σ = 0.0113), implying limited technological progress at the industry level. Only 35 % of the airlines exhibited GTC $>$ 1 across periods, suggesting slow diffusion of cleaner or more fuel-efficient aircraft technologies.

By contrast, the efficiency change (GEC) component shows slightly higher variability (mean = 0.9989, σ = 0.0176), with 40 % of cases improving managerial efficiency and 60 % experiencing declines. The alternation between GEC gains and losses across years implies that operational adjustments (e.g., fleet utilization, route restructuring) played a larger role in short-term fluctuations than structural technological advances.

6.3.3. Airline-Level productivity patterns

As detailed in Table 6.1-6.3, Singapore Airlines, Qantas, and Air New Zealand achieved consistently higher GML indices ($>$ 1.02 on average), driven by steady GEC improvements and resilience during pandemic years. In contrast, carriers such as Delta and Air China show notable drops (GML $\approx$ 0.95 in 2020–2021), indicating pandemic-related operational inefficiencies. The decomposition reveals that, while GTC values for most airlines remained around 1.00, GEC fluctuated more strongly, confirming that short-term managerial performance differences dominate overall productivity dynamics in this period.

The model's robustness was verified through bootstrap analysis (Fig. 6.4) provides 95 % confidence intervals for efficiency and

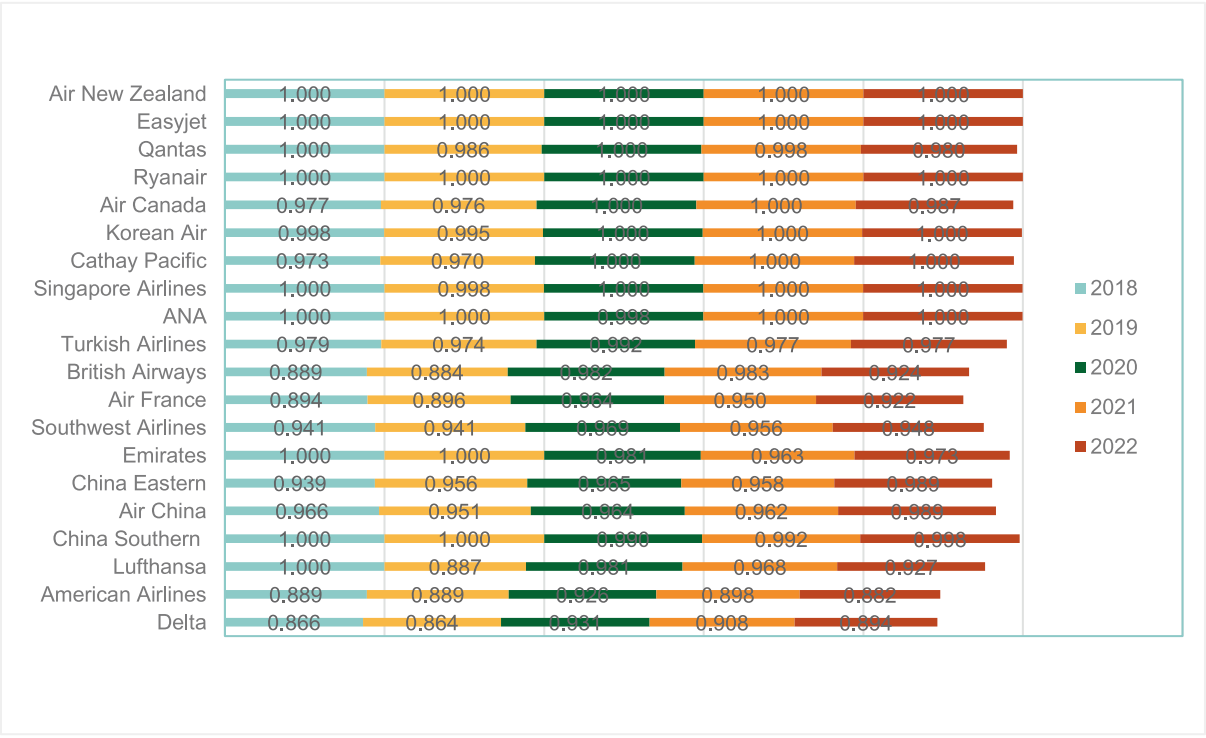


Fig. 6.3. EEI scores for the 20 global airlines during 2018–2022.

Table 6.1
GML index values (2018–2022).

GML	2018_2019	2019_2020	2020_2021	2021_2022	
Delta	0.9519	1.0278	0.9513	1.0349	
American Airlines	0.9678	0.9993	0.9678	1.0108	
Lufthansa	1.0168	0.9762	0.9719	1.0227	
China Southern	1.0192	1.0273	0.9910	0.9608	
Air China	1.0137	0.9953	0.9577	1.0255	
China Eastern	0.9819	0.9710	0.9573	0.9727	
Emirates	1.0202	0.9513	1.0398	1.0310	
Southwest Airlines	1.0206	1.0331	0.9969	1.0203	
Air France	1.0379	0.9552	1.0035	1.0272	
British Airways	0.9830	0.9789	1.0476	1.0159	
Turkish Airlines	0.9888	0.9776	1.0385	0.9878	
ANA	1.0116	1.0203	1.0280	0.9756	
Singapore Airlines	0.9996	1.0302	1.0216	1.0387	
Cathay Pacific	0.9924	0.9972	0.9950	0.9892	
Korean Air	0.9796	1.0358	1.0101	1.0393	
Air Canada	1.0101	0.9501	1.0376	1.0078	
Ryanair	1.0173	0.9689	0.9791	1.0208	
Qantas	1.0292	1.0499	0.9859	1.0222	1.05
Easyjet	0.9778	1.0238	1.0452	1.0141	1.00
Air New Zealand	1.0261	1.0407	0.9918	1.0237	0.95

GML estimates, with narrow bands (± 0.03) indicating model stability. The mean bootstrap efficiency (0.6755) and mean GML (1.0039) closely match the original estimates, confirming that random sampling variation does not materially affect the conclusions.

To enhance the interpretability of the DEA benchmarking results, a λ-intensity heatmap and benchmark frequency matrix were constructed (Table 6.4; Fig. 6.5).

The λ values quantify the relative contribution of efficient DMUs to the efficiency projection of inefficient airlines. Non-zero λ values therefore identify benchmark carriers serving as performance frontiers within the sample.

As shown in Table 6.4, Korean Air, Emirates, China Southern and Air New Zealand consistently appear as dominant benchmark airlines, exhibiting the highest λ-intensities (0.13–0.36). These carriers demonstrate strong operational efficiency and carbon-management capability, frequently serving as reference peers for other airlines. Singapore Airlines, ANA, and Ryanair also show

Table 6.2
GTC index values (2018–2022).

GTC	2018_2019	2019_2020	2020_2021	2021_2022	
ANA	1.0155	1.0178	1.0103	1.0091	
Delta	1.0091	1.0044	1.0086	1.0131	
Singapore Airlines	1.0115	1.0076	1.0041	1.0032	
Emirates	1.0145	0.9964	1.0143	0.9927	
Qantas	0.9939	0.9850	1.0195	1.0170	
Turkish Airlines	1.0038	0.9974	1.0045	1.0094	
Air New Zealand	0.9840	1.0050	1.0085	1.0101	
Southwest Airlines	0.9817	0.9980	1.0160	1.0039	
China Eastern	1.0131	0.9973	1.0073	0.9800	
China Southern	1.0157	0.9938	0.9879	0.9975	
Air Canada	1.0104	0.9884	1.0148	0.9800	
Cathay Pacific	1.0137	0.9899	0.9846	0.9999	
Air France	0.9975	0.9844	1.0067	0.9975	
Ryanair	0.9873	1.0079	0.9955	0.9928	
Easyjet	0.9965	0.9874	1.0046	0.9943	
Air China	1.0078	1.0001	0.9905	0.9815	
Korean Air	1.0042	0.9877	1.0014	0.9831	
Lufthansa	1.0135	0.9809	0.9808	0.9979	1.02
American Airlines	0.9808	0.9915	1.0000	0.9883	1.00
British Airways	0.9813	0.9897	0.9831	1.0006	0.98

Table 6.3
GEC index values (2018–2022).

GEC	2018_2019	2019_2020	2020_2021	2021_2022	
Air New Zealand	1.0239	1.0247	1.0137	0.9744	
Turkish Airlines	1.0250	1.0126	0.9877	1.0008	
Emirates	1.0129	1.0193	1.0027	0.9905	
Easyjet	1.0063	0.9952	1.0244	0.9984	
Qantas	1.0056	1.0286	1.0053	0.9789	
American Airlines	1.0152	0.9765	0.9985	1.0272	
China Eastern	0.9909	1.0230	1.0067	0.9925	
Singapore Airlines	0.9940	1.0014	0.9925	1.0205	
Cathay Pacific	0.9935	1.0096	0.9926	1.0002	
Southwest Airlines	1.0296	0.9920	0.9748	0.9988	
Air China	0.9918	0.9995	1.0011	0.9989	
Korean Air	1.0230	0.9886	0.9770	0.9978	
China Southern	0.9850	1.0277	0.9783	0.9942	
Air France	0.9791	1.0292	0.9865	0.9897	
Lufthansa	1.0002	0.9879	1.0015	0.9942	
Air Canada	0.9921	0.9810	1.0072	1.0025	
British Airways	0.9960	0.9770	1.0019	1.0023	
Delta	1.0038	0.9828	1.0115	0.9751	1.03
Ryanair	0.9824	0.9800	0.9862	1.0163	1.00
ANA	0.9883	0.9777	0.9808	0.9777	0.97

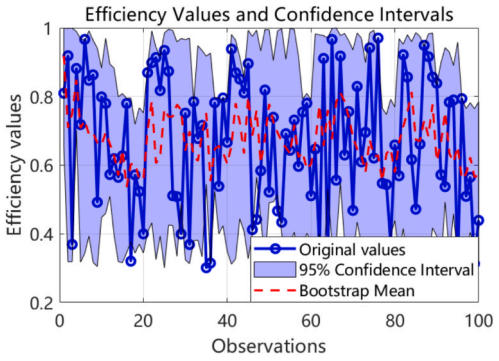


Fig. 6.4. GML Bootstrap efficiency values and confidence intervals.

Table 6.4
λ-index of 20 global airlines learning from benchmarking.

λ	Lufthansa	China Southern	Emirates	ANA	Singapore Airlines	Korean Air	Ryanair	Air New Zealand
Delta	0.00	0.02	0.04	0.00	0.00	0.13	0.00	0.00
American Airlines	0.00	0.02	0.04	0.00	0.00	0.14	0.00	0.00
Lufthansa	0.04	0.00	0.01	0.00	0.00	0.15	0.00	0.00
China Southern	0.00	0.12	0.01	0.00	0.00	0.07	0.00	0.00
Air China	0.00	0.36	0.02	0.00	0.00	0.15	0.02	0.00
China Eastern	0.00	0.00	0.18	0.00	0.00	0.18	0.00	0.00
Emirates	0.00	0.00	0.11	0.00	0.00	0.09	0.00	0.00
Southwest Airlines	0.00	0.01	0.01	0.00	0.00	0.18	0.00	0.00
Air France	0.00	0.00	0.01	0.00	0.00	0.19	0.00	0.00
British Airways	0.00	0.00	0.01	0.00	0.00	0.19	0.00	0.00
Turkish Airlines	0.00	0.00	0.03	0.00	0.00	0.17	0.00	0.00
ANA	0.01	0.00	0.00	0.11	0.00	0.05	0.00	0.03
Singapore Airlines	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.01
Cathay Pacific	0.00	0.00	0.02	0.00	0.00	0.12	0.00	0.02
Korean Air	0.00	0.00	0.01	0.00	0.03	0.13	0.00	0.02
Air Canada	0.00	0.00	0.01	0.00	0.00	0.16	0.00	0.03
Ryanair	0.01	0.00	0.00	0.01	0.02	0.03	0.11	0.03
Qantas	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.12
EasyJet	0.01	0.00	0.00	0.01	0.04	0.07	0.00	0.08
Air New Zealand	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20

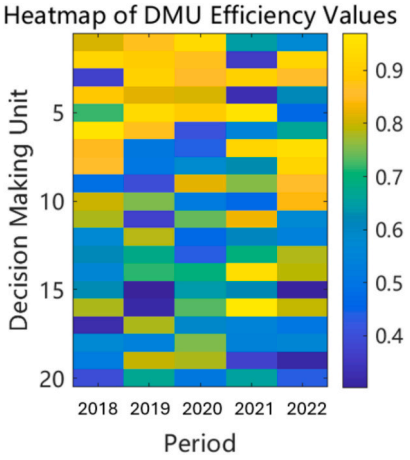


Fig. 6.5. Heatmap of DMU Efficiency Values (2018–2022).

moderate benchmark roles (λ ≈ 0.10–0.19), indicating region-specific leadership—particularly within Asia and the Europe. In contrast, several European full-service airlines (e.g., British Airways, Air France) appear mainly as followers rather than benchmarks.

The heatmap (Fig. 6.5) illustrates the temporal evolution of airline efficiency between 2018 and 2022. Periodic shifts in dominant benchmark carriers coincide with external shocks (notably 2020–2021), highlighting changes in network adaptability and capacity management. Overall, λ-intensity dispersion remains moderate (average 0.35, concentration 0.93), confirming a relatively stable benchmark structure across periods despite pandemic disruptions.

6.4. Comparative Discussion with existing literature

The empirical patterns uncovered in this paper both confirm and extend existing findings in airline carbon-efficiency research. For instance, Liu et al. (2020) demonstrate that production- and operational inefficiencies—rather than external factors—constitute the primary sources of emission change in global aviation, aligning with our finding that managerial (GEC) factors dominate over technological progress (GTC) in the GML decomposition. See et al. (2025) adopt a dynamic network DEA approach for 24 global airlines and uncover regionally diverse efficiency trajectories, mirroring our observation of heterogeneous rebound paths in 2021–2022 across Asian, European, and North American carriers. In the U.S. context, Kaffash et al. (2024) use directional distance functions to link environmental efficiency with financial and operational performance, and they too find that size and scale correlate with emission outcomes—a parallel to our Random Forest result that RTK and fleet size are key drivers of heterogeneity.

The Cloud-model typology of volatility (En) and uncertainty (He) extends the frontier of comparative literature by introducing a stability dimension seldom considered in conventional DEA or GML studies. This is a methodological contribution beyond static

interpretations in earlier works. Additionally, the persistence of CO_2 slack—even for benchmark carriers—echoes the inefficiency remnants identified by Chen et al. (2017) in Chinese airlines, where large firms remain imperfectly efficient in emission control. Together, these comparisons underscore that while the basic insight—that operational scale and management strategy matter—is a shared theme, the integration of volatility, uncertainty, and machine-learning driver diagnostics provides a more comprehensive and actionable lens on airline decarbonisation.

7. Robustness analysis

7.1. Impact of excluding pandemic years

To address the potential distortion caused by the COVID-19 pandemic (2020–2021), which led to unprecedented traffic collapses and grounded fleets, a robustness test was conducted by re-estimating airline efficiency using only pre-pandemic data (2018–2019). For each sample, we calculated annual super-efficiency scores and bias-corrected super-efficiency scores, and then derived an average score for each airline across the corresponding period. Rankings of the 20 airlines were based on these average scores.

The results show a high degree of consistency between the two samples. The Spearman's rank correlation coefficient between the rankings derived from the two-year and five-year datasets is 0.9162, and the correlation of the bias-corrected super-efficiency scores is 0.9208. These strong correlations confirm that, although absolute efficiency levels were lower during the pandemic, the relative benchmarking across airlines remains robust.

Fig. 7.1 illustrates the comparison of super-efficiency and bias-corrected super-efficiency scores for both the pre-COVID (2018–2019) and full-sample (2018–2022) models. The chart shows that airlines with high performance under the five-year specification (e.g., Emirates, China Southern) also retained leading positions in the pre-COVID sample, while those with lower relative efficiency (e.g., Air France, British Airways) consistently lagged behind in both periods.

This analysis confirms that the inclusion of 2020–2021 does not fundamentally alter the relative efficiency ordering across airlines, and thus the main results are not driven by pandemic distortions. At the same time, keeping the five-year dataset allows the model to capture both pre- and post-shock dynamics, offering a more comprehensive perspective on airline low-carbon competitiveness.

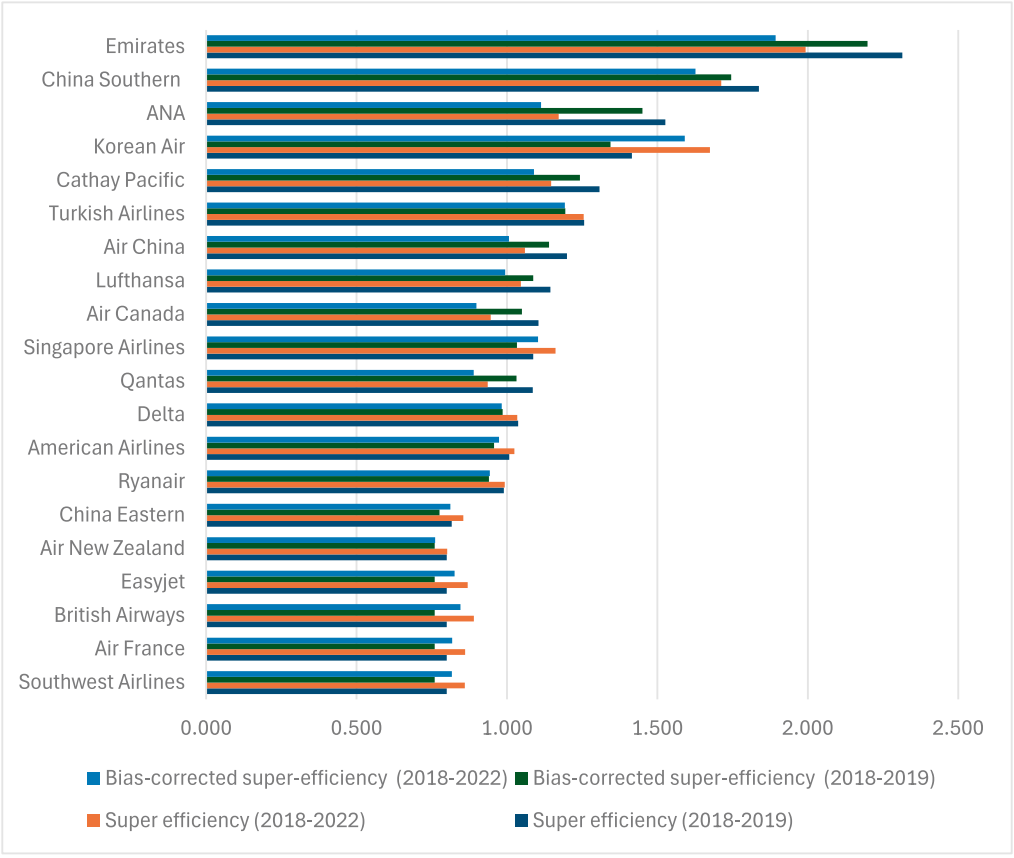


Fig. 7.1. Comparison of average super-efficiency and bias-corrected super-efficiency scores under pre-COVID (2018–2019) and full-sample (2018–2022) datasets.

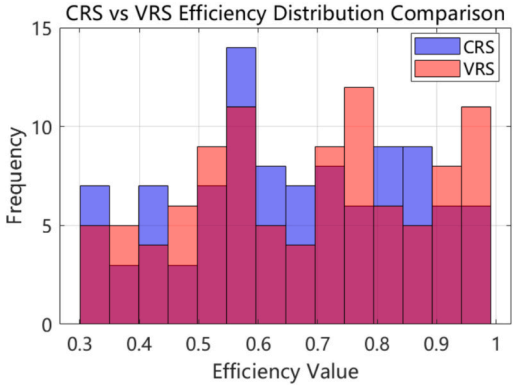


Fig. 7.2. Comparison of CRS and VRS efficiency results under the SBM-DDF model.

7.2. Alternative input specifications

To examine the stability of the efficiency results against alternative input specifications, an additional test was conducted by substituting the input variable number of aircraft with number of flights. This adjustment addresses potential redundancy among operational scale indicators while retaining the other input and output variables unchanged. For both specifications, super-efficiency and bias-corrected super-efficiency scores were calculated for the five-year period (2018–2022), and average scores were used to derive airline rankings. The results show a high degree of consistency, with the Spearman's rank correlation coefficient between the two sets of bias-corrected super-efficiency rankings equal to 0.94. This confirms that the relative efficiency ordering is robust to the choice between aircraft count and flight frequency as a scale measure, reinforcing the reliability of the benchmarking framework.

7.3. Returns-to-scale assumptions

To examine the sensitivity of efficiency and productivity estimates to returns-to-scale (RTS) assumptions, an additional test was conducted by re-estimating the SBM-DDF-GML model under variable returns to scale (VRS), while the baseline analysis assumed constant returns to scale (CRS). The CRS specification reflects the relatively homogeneous scale and cost structures of major international airlines, consistent with prior studies in airline efficiency benchmarking.

Under CRS, the mean efficiency was 0.6603 (SD = 0.1880) and the mean GML index was 1.0014 (SD = 0.0286). Under VRS, the mean efficiency was 0.6791 (SD = 0.1960) and the mean GML index was 0.9969 (SD = 0.0306). The correlation between CRS and VRS efficiency values (−0.0467) and GML indices (−0.0188) indicates sensitivity of absolute scores to the RTS assumption. However, paired t-tests for both efficiency ($p = 0.4996$) and GML ($p = 0.3389$) confirm that these differences are not statistically significant. The average efficiency difference between CRS and VRS models was 0.2216 (33.56 %), and the average GML difference was 0.0348 (3.48 %), suggesting that while absolute efficiency magnitudes vary, productivity trends remain consistent.

Fig. 7.2 compares the distributions of efficiency scores under CRS and VRS assumptions, showing largely overlapping patterns across most airlines. Based on these results, the CRS specification is retained in the main analysis for consistency and comparability, while the VRS-based results are reported as a robustness check. Detailed numerical outputs are provided in Appendix C.

8. Policy implications

The integrated results from the SBM-DDF, Super-SBM, GML decomposition, Cloud model, Random Forest regression, and the derived Environmental Efficiency Index (EEI) provide a comprehensive understanding of airlines' low-carbon competitiveness between 2018 and 2022. Together, these analyses reveal that performance heterogeneity across airlines is multi-dimensional, evolving, and strongly influenced by operational scale and resource allocation rather than by pure technological advancement. The findings yield several key implications for policy design and industry practice.

8.1. Recognising multi-dimensional heterogeneity in airline performance

The Cloud model analysis demonstrates that heterogeneity extends beyond mean efficiency levels (Ex) to encompass volatility (En) and the uncertainty of volatility (He), defining a three-dimensional structure of efficiency reliability. Airlines such as Ryanair, EasyJet, and Singapore Airlines exhibit low Ex–low En profiles, signifying efficient and stable operations that can serve as benchmark references for the industry. By contrast, Emirates and Air Canada (low Ex–high En) display strong performance but unstable volatility, indicating a need for operational stabilisation policies, while Air New Zealand and Turkish Airlines (high Ex–high He) show both inefficiency and instability, suggesting deeper systemic reforms. This multi-dimensional heterogeneity framework provides regulators with a diagnostic tool that distinguishes between structural inefficiency, management volatility, and data uncertainty—dimensions often overlooked in traditional frontier analyses.

8.2. Identifying operational drivers of heterogeneity

The Random Forest analysis complements the Cloud model by revealing which variables drive observed differences in efficiency. Revenue tonne-kilometres (RTK) emerged as the most influential determinant, followed by employment scale, flight distance and number of aircrafts, whereas load factor played a comparatively minor role. This pattern indicates that resource scale and network capacity management are key levers of carbon efficiency heterogeneity. Sensitivity analysis confirmed that emissions are highly elastic to RTK and fleet composition, implying that regulatory and managerial interventions targeting network rationalisation, aircraft utilisation, and resource allocation can yield measurable carbon gains. These findings reinforce that heterogeneity is not random but systematically linked to operational scale and management strategies.

8.3. Strengthening dynamic benchmarking and performance monitoring

The SBM-DDF and GML results reveal temporal instability, especially during 2020–2021, underscoring the need for dynamic performance tracking rather than static compliance indicators. Frontier-based benchmarking systems could be institutionalised within schemes such as the EU ETS, UK ETS, and ICAO CORSIA, enabling the differentiation between temporary operational shocks and persistent inefficiency. Integrating volatility (En) and uncertainty (He) metrics from the Cloud model would further allow regulators to monitor stability and reliability of carbon performance, providing early warning signals of systemic inefficiency.

8.4. Incentivising managerial innovation and fleet optimisation

Decomposition results indicate that managerial efficiency improvements, rather than technological change, drive most productivity variations. Policies should therefore promote managerial innovation—for instance, through dynamic slot allocation, route optimisation, and incentive schemes tied to verified operational improvements. Linking such measures with the RF-identified key drivers (RTK and fleet size) would allow data-driven performance contracts that reward carbon reductions achieved through improved resource management.

8.5. Avoiding over-interpretation of short-term environmental gains

The EEI results show that most airlines reached near-perfect environmental efficiency during 2020–2021, but largely because of pandemic-induced traffic contraction rather than genuine emission-intensity improvements. Policymakers should interpret these temporary gains with caution and prioritise structural indicators of sustainable efficiency, such as bias-corrected super-efficiency or GML-based productivity change, which better capture enduring improvements in technology and management.

9. Conclusion

This study introduced a unified framework integrating SBM-DDF efficiency analysis, the GML productivity index, Random Forest driver identification, and cloud-model heterogeneity assessment to evaluate the low-carbon competitiveness of 20 global airlines from 2018 to 2022. The combined results reveal substantial variation in environmental performance across airlines and over time, shaped by both structural characteristics and pandemic-related disruptions.

Efficiency levels declined sharply in 2020–2021 and rebounded in 2021–2022, but CO_2-slack patterns indicate that absolute emission reductions remain uneven across carriers. GML decomposition shows that productivity changes were modest and primarily driven by efficiency change, underscoring the continuing importance of operational and managerial practices relative to technological progress. This suggests that fleet renewal and SAF adoption—while crucial long-term strategies—were not yet widespread or impactful enough during the study period to dominate short-run productivity dynamics.

The cloud-model results extend the interpretation of heterogeneity by identifying four distinct typologies of performance: efficient-stable, efficient-volatile, inefficient-stable, and inefficient-volatile airlines. Meanwhile, Random Forest analysis highlights RTK, employment scale, and fleet size as the principal determinants of carbon efficiency, revealing significant elasticity of emission performance to utilisation and scale factors.

Together, these findings contribute to the literature by demonstrating that low-carbon competitiveness is multi-dimensional, dynamic, and shaped by both operational scale and volatility patterns. The integrated methodological framework offers a replicable tool for monitoring airline performance under ETS and CORSIA, and the heterogeneity typologies provide actionable insights for regulators designing performance-based incentives. While the study identifies strong associations between airline strategies and efficiency outcomes, the frontier-based methods do not establish causality; therefore, the policy implications should be interpreted as indicative rather than causal. Future research combining quasi-experimental identification or firm-level SAF adoption data would further strengthen causal inference on decarbonisation strategies.

CRediT authorship contribution statement

Mengyuan Lu: Writing – original draft, Formal analysis, Data curation, Conceptualization. **Jinglei Yu:** Visualization, Software, Methodology, Investigation. **Fei Li:** Supervision, Investigation. **Yue Zhang:** Supervision, Investigation. **Quanxing Jia:** Supervision, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the National Key Research and Development Project (NO.2022YFB2602000), the National Nature Science Foundation of China (U2333217), and the Fundamental Research Funds of CASTC (XXX252060302025029).

Appendix A:. Variable sources and specifications

A1. Variable Definitions and units

Variable	Definition	Unit	Notes
Fuel consumption	Total annual jet fuel consumption	tonnes	Converted from gallons/litres where required using ICAO standards
CO_2 emissions	Total annual CO_2 emitted from operations	tonnes	Calculated directly from fuel or taken from sustainability reports
ASK (Available Seat Kilometres)	Seat capacity × distance flown	million km	Revenue-related output
RPK (Revenue Passenger Kilometres)	Passenger demand × distance	million km	Desirable output in SBM-DDF
RTK (Revenue Tonne Kilometres)	Cargo + passenger tonne-km	million tonne-km	Desirable output in SBM-DDF
Number of flights	Total annual departures	units	Aggregated from airline reports/CAPA
Flight distance	Average stage length	km	Sometimes listed as miles, then converted
Passengers	Total number of passengers	million	Aggregated from airline reports/CAPA
Operating revenue	Total annual operating revenue	USD billion	All converted to USD for comparison
Operating cost	Total annual operating cost	USD billion	All converted to USD for comparison
Number of employees	Total workforce	thousand	Aggregated from airline reports
Number of aircraft	Total fleet size	units	Aggregated from airline reports/CAPA
Load factor	RPK/ASK	%	Aggregated from airline reports/CAPA

A2. Data sources by airline

Airline	Primary Data Sources	Notes on Transformations
Delta	Annual Report; CAPA	Fuel converted from gallons to tonnes; distance converted from miles to km
American Airlines	Annual Report; CAPA	Fuel converted from gallons to tonnes; distance converted from miles to km
Lufthansa	Annual Report; CAPA	EUR converted to USD
China Southern	Annual Report; CSR Report; CAPA	CNY converted to USD
Air China	Annual Report; CSR Report; CAPA	CNY converted to USD
China Eastern	Annual Report; CSR Report; CAPA	CNY converted to USD
Emirates	Annual Report; CAPA	All operational KPIs directly reported
Southwest Airlines	Annual Report; ESG Report; CAPA	Fuel converted from gallons to tonnes; distance converted from miles to km
Air France–KLM	Annual Report; CAPA	EUR converted to USD
British Airways (IAG)	Annual Report; CAPA	GBP converted to USD
Turkish Airlines	Annual Report; CAPA	TRY converted to USD
ANA	Annual Report; CAPA	All operational KPIs directly reported
Singapore Airlines	Annual Report; CAPA	All operational KPIs directly reported
Cathay Pacific	Annual Report; CAPA	All operational KPIs directly reported
Korean Air	Annual Report; CAPA	Cargo contributes significantly to RTK
Air Canada	Annual Report; CAPA	Fuel converted from million litres to tonnes; distance converted from miles to km
Ryanair	Annual Report; CAPA	Fuel converted from gallons to tonnes
Qantas	Annual Report; CAPA	Fuel converted from barrels to tonnes
EasyJet	Annual Report; CAPA	All operational KPIs directly reported
Air New Zealand	Annual Report; Climate Statement; CAPA	All operational KPIs directly reported

Appendix B:. Core analysis scripts in MATLAB

This appendix contains the core MATLAB scripts used for the main analytical stages of this study.

B.1 SBM-DDF-GML main function

The following script outlines the main function for the Stage 3 analysis, which calculates efficiency scores and the Global Malquist-

Luenberger (GML) index. It calls lower-level functions for the detailed DEA calculations.

```
function main_results = SBM_DDF_GML_analysis(data_processed, config)
% Main function for the SBM-DDF-GML analysis.
fprintf('Starting SBM-DDF-GML analysis...\n');
% Get data dimensions
n_total_obs = size(data_processed.inputs, 1);
n_dmu = n_total_obs / config.T;
fprintf('Data Check: Total Observations=%d, DMUs=%d, Periods=%d\n', n_total_obs, n_dmu, config.
    T);
% Initialize results structure
main_results = struct();
main_results.eff = zeros(n_total_obs, 1);
main_results.lambda = cell(n_total_obs, 1);
main_results.slacks = cell(n_total_obs, 1);
main_results.inefficiency = zeros(n_total_obs, 1);
try
% -- Step 1: Calculate SBM-DDF efficiency scores for each period --
for t = 1:config.T
period_indices = (t-1)*n_dmu + 1: t*n_dmu;
% Extract data for the current period
X_t = data_processed.inputs(period_indices,:);
Y_t = data_processed.outputs(period_indices,:);
Z_t = data_processed.undesirables(period_indices,:);
% Calculate efficiency scores by calling the solver function
[eff_t, lambda_t, slacks_t, ineff_t] = SBM_DDF_efficiency_solver(X_t, Y_t, Z_t, config);
main_results.eff(period_indices) = eff_t;
main_results.lambda(period_indices) = lambda_t;
main_results.slacks(period_indices) = slacks_t;
main_results.inefficiency(period_indices) = ineff_t;
end
% -- Step 2: Calculate GML, TC, and EC indices --
main_results.gml = zeros(n_dmu, config.T-1);
main_results.tc = zeros(n_dmu, config.T-1);
main_results.ec = zeros(n_dmu, config.T-1);
for t = 1:(config.T-1)
% GML components are calculated by calling a dedicated function
% that solves the four required distance functions.
[gml_t, tc_t, ec_t] = calculate_GML_components(data_for_period_t, data_for_period_t1, config);
main_results.gml(:, t) = gml_t;
main_results.tc(:, t) = tc_t;
main_results.ec(:, t) = ec_t;
end
% -- Step 3: Statistical Summary and Uncertainty Estimation --
main_results.summary = struct();
main_results.summary.mean_efficiency = mean(main_results.eff);
main_results.summary.std_efficiency = std(main_results.eff);
% Perform bootstrapping and jackknife analysis for robustness
main_results.bootstrap = bootstrap_analysis(data_processed, config);
main_results.jackknife = jackknife_analysis(data_processed, config);
fprintf('SBM-DDF-GML analysis completed successfully!\n');
catch ME
fprintf('Error during SBM-DDF-GML analysis: %s\n', ME.message);
% Create fallback results if an error occurs
main_results = create_fallback_results(n_total_obs, n_dmu, config);
end
end
```

B.2 Random Forest analysis

The following script details the implementation of the Random Forest model for feature importance analysis, including hyperparameter tuning and cross-validation.

```
% Random_Forest_Analysis.m
% This script performs Random Forest regression, including hyperparameter
% optimization, 5-fold cross-validation, and feature importance calculation.
```

(*continued on next page*)

(*continued*)

```
% Random_Forest_Analysis.m
% —— 1. Environment Setup ——
rng(42); % Set random seed for reproducibility
warning('off', 'all'); % Suppress warnings
% Load and normalize training and testing data (X_train, y_train, etc.)
% … (Data loading and normalization steps here) …
featureNames = final_feature_names; % Load feature names
% —— 2. Hyperparameter Optimization (Leaf Size) ——
fprintf('Optimizing leaf node size based on OOB error…\n');
leafSizesToTest = [5, 10, 20, 50, 100, 200, 500];
colors = lines(length(leafSizesToTest));
figure('Name', 'Leaf Node Size Optimization Analysis', 'Color', 'white');
hold on;
% Loop through different leaf sizes to find the optimal one
for i = 1:length(leafSizesToTest)
RF_model_tuning = TreeBagger(1000, X_train, y_train, …
'Method', 'regression', …
'OOBPrediction', 'on', …
'MinLeafSize', leafSizesToTest(i));
oobErr = oobError(RF_model_tuning);
plot(oobErr, 'Color', colors(i,:), 'LineWidth', 2);
end
xlabel('Number of Trees Grown');
ylabel('Out-of-Bag Mean Squared Error (OOB MSE)');
title('Effect of Different Leaf Node Sizes on Model Error');
legend(cellstr(num2str(leafSizesToTest')), 'Location', 'NorthEast', 'Title', 'Leaf Node Size');
grid on;
hold off;
% … (Code to automatically select bestLeafSize based on min OOB error) …
fprintf('Optimal leaf size selected: %d\n', bestLeafSize);
% —— 3. Performance Evaluation with 5-Fold Cross-Validation ——
fprintf('\nStarting 5-fold cross-validation…\n');
k = 5; % Number of folds
cv = cvpartition(length(y_train), 'KFold', k);
% Model parameters
numTrees = 500;
minLeaf = bestLeafSize; % Use the optimized leaf size
% Loop through each fold for training and testing
for i = 1:k
trainIdx = training(cv, i);
testIdx = test(cv, i);
% Train the Random Forest model for the current fold
RF_model_fold = TreeBagger(numTrees, X_train(trainIdx,:), y_train(trainIdx), …
'Method', 'regression', …
'MinLeafSize', minLeaf);
% Predict on the test set for this fold
predictions = predict(RF_model_fold, X_train(testIdx,:));
% … (Code to calculate and store performance metrics like R2, RMSE, MAE) …
end
% —— 4. Final Model Training and Feature Importance ——
fprintf('\nTraining final model on all data and calculating feature importance…\n');
% Train the final model on the entire training dataset
Final_RF_Model = TreeBagger(numTrees, X_train, y_train, …
'Method', 'regression', …
'OOBPredictorImportance', 'on', … % Use MDI importance
'MinLeafSize', minLeaf);
% Calculate permutation importance for robustness
permImportance = oobPermutedPredictorImportance(Final_RF_Model);
% Visualize feature importance
figure('Name', 'Feature Importance Analysis', 'Color', 'white');
[sorted_importance, sort_idx] = sort(permImportance, 'descend');
barh(sorted_importance);
yticks(1:length(featureNames));
yticklabels(featureNames(sort_idx));
xlabel('Feature Importance Score (OOB Permutation)');
title('Random Forest Feature Importance Ranking');
grid on;
fprintf('Random Forest analysis complete.\n');
```

Appendix C. . Returns-to-Scale robustness test output

This appendix presents the detailed computational results of the robustness test comparing constant returns to scale (CRS) and variable returns to scale (VRS) specifications within the Super-SBM-DDF-GML framework.

All analyses were conducted using the fully repaired algorithm to ensure convergence and feasibility across all decision-making units (DMUs).

C.1 Model configuration

- **Dataset:** 100 observations (20 DMUs × 5 years)
- **Baseline RTS assumption:** CRS (Constant Returns to Scale)
- **Alternative specification:** VRS (Variable Returns to Scale)
- **Orientation:** Non-oriented SBM-DDF model
- **Solver:** Fully repaired iterative algorithm with automatic slack correction
- **Infeasibility:** None detected across all iterations

C.2. Efficiency and GML results

Metric	CRS	VRS
Efficiency range	[0.3089, 0.9813]	[0.3059, 0.9908]
Mean efficiency	0.6603	0.6791
Standard deviation (efficiency)	0.1880	0.1960
Mean GML index	1.0014	0.9969
Standard deviation (GML)	0.0286	0.0306

C.3. CRS vs VRS comparative analysis

Statistic	Result	Interpretation
Efficiency correlation (CRS vs VRS)	−0.0467	Low correlation—efficiency magnitudes moderately sensitive to RTS assumption
GML correlation (CRS vs VRS)	−0.0188	Low correlation—minor effect on productivity decomposition
Mean efficiency difference	0.2216 (≈ 33.56 %)	Absolute efficiency levels differ, but ranking patterns remain stable
Mean GML difference	0.0348 (≈ 3.48 %)	GML trends highly consistent
Paired *t*-test p-value (efficiency)	$0.4996 > 0.05$	No significant difference between CRS and VRS efficiencies
Paired *t*-test p-value (GML)	$0.3389 > 0.05$	No significant difference between CRS and VRS productivity indices

C.4 Summary

The comparison between CRS and VRS specifications demonstrates that although the absolute magnitudes of efficiency scores vary with the assumed returns-to-scale, the overall efficiency ranking and productivity dynamics remain statistically consistent.

Both the efficiency and GML differences are statistically insignificant ($p > 0.05$).

These results confirm that the main CRS-based model provides a stable and representative benchmark for cross-airline performance comparison, while the VRS outcomes serve as a complementary robustness check.

Data availability

The authors do not have permission to share data.

References

Amankwah-Amoah, J., 2020. Note: Mayday, Mayday, Mayday! responding to environmental shocks: Insights on global airlines' responses to COVID-19. Transportation Research Part e: Logistics and Transportation Review 143, 102098. https://doi.org/10.1016/j.tre.2020.102098.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS.

Brueckner, J.K., Zhang, A., 2010. Airline emission charges: Effects on airfares, service quality, and aircraft design. Transp. Res. B Methodol. 44 (8–9), 960–971.

Chen, Y., Cheng, S., Zhu, Z., 2021. Exploring the operational and environmental performance of chinese airlines: a two-stage undesirable SBM-NDEA approach. J. Clean. Prod. 289, 125711. https://doi.org/10.1016/J.JCLEPRO.2020.125711.

Chen, R., Yang, H., Wang, K., Jiang, C., 2024. Impacts of a sustainable aviation fuel mandate on airline competition—Full-service carrier vs. low-cost carrier. Transp. Res. B Methodol. 190, 103098. https://doi.org/10.1016/j.trb.2024.103098.

Chen, Z., Wanke, P., Antunes, J.J.M., Zhang, N., 2017. Chinese airline efficiency under CO2 emissions and flight delays: a stochastic network DEA model. Energy Econ. 68, 89–108. https://doi.org/10.1016/J.ENECO.2017.09.015.

Chung, Y.H., Färe, R., Grosskopf, S., 1997. Productivity and Undesirable Outputs: a Directional Distance Function Approach. J. Environ. Manage. 51 (3), 229–240. https://doi.org/10.1006/JEMA.1997.0146.

Fukuyama, H., Weber, W.L., 2010. A slacks-based inefficiency measure for a two-stage system with bad outputs. Omega 38 (5), 398–409. https://doi.org/10.1016/J.OMEGA.2009.10.006.

Gualini, A., Zou, L., Dresner, M., 2023. Airline strategies during the pandemic: what worked? Transp. Res. A Policy Pract. 170, 103625. https://doi.org/10.1016/j.tra.2023.103625.

Hamdan, S., Jouini, O., Cheaitou, A., Jemai, Z., Granberg, T.A., Josefsson, B., 2022. Air traffic flow management under emission policies: Analyzing the impact of sustainable aviation fuel and different carbon prices. Transp. Res. A Policy Pract. 166, 14–40. https://doi.org/10.1016/j.tra.2022.09.013.

Huang, F., Zhang, T., Wang, Q., Zhou, D., 2023. CO2 emission change in China's aviation industry: a fleet-wide index decomposition and scenario analysis. Transp. Res. Part D: Transp. Environ. 119, 103743. https://doi.org/10.1016/J.TRD.2023.103743.

Huang, F., Zhou, D., Hu, J.L., Wang, Q., 2020. Integrated airline productivity performance evaluation with CO2 emissions and flight delays. J. Air Transp. Manag. 84. https://doi.org/10.1016/j.jairtraman.2020.101770.

Icao, 2016. *Resolution A39-3: Consolidated statement of continuing ICAO policies and practices related to environmental protection – Global Market-based measure (MBM) scheme*. International Civil Aviation. Organization. https://www.icao.int/environmental-protection/Pages/A39_CORSIA_FAQ2.aspx.

Icao, 2023. *Annual Report of the Council 2022 – Environmental Protection*. International Civil Aviation. Organization. https://www.icao.int/publications/Documents/AnnualReports/2022.

IEA. (2018). *The future of rail: Opportunities for energy and the environment*. International Energy Agency. https://www.iea.org/reports/the-future-of-rail.

Kaffash, S., Chomachaei, F., Aktas, E., 2024. Environmental sustainability performance of US airlines: implications of financial performance and technical efficiency. J. Oper. Res. Soc. https://doi.org/10.1080/01605682.2024.2429567.

Lee, D.S., Fahey, D.W., Forster, P.M., Newton, P.J., Wit, R.C.N., Lim, L.L., Sausen, R., 2009. Aviation and global climate change in the 21st century. Atmos. Environ. 43 (22–23), 3520–3537. https://doi.org/10.1016/j.atmosenv.2009.04.024.

Li, H., Fang, K., Yang, W., Wang, D., Hong, X., 2013. Regional environmental efficiency evaluation in China: Analysis based on the Super-SBM model with undesirable outputs. Math. Comput. Model. 58 (5–6), 1018–1031.

Li, H., Shi, J.F., 2014. Energy efficiency analysis on chinese industrial sectors: an improved Super-SBM model with undesirable outputs. J. Clean. Prod. 65, 97–107. https://doi.org/10.1016/J.JCLEPRO.2013.09.035.

Li, Y., Li, S., 2021. The Influence Study on Environmental Regulation and Green Total factor Productivity of China's Manufacturing Industry. Discret. Dyn. Nat. Soc. 2021 (1), 5580414. https://doi.org/10.1155/2021/5580414.

Liu, X., Hang, Y., Wang, Q., Zhou, D., 2020. Drivers of civil aviation carbon emission change: a two-stage efficiency-oriented decomposition approach. Transp. Res. Part D: Transp. Environ. 89, 102612. https://doi.org/10.1016/j.trd.2020.102612.

Liu, X., Jiang, P., 2023. How does civil aviation achieve sustainable low-carbon development? — an abatement–cost perspective. Heliyon 9 (10). https://doi.org/10.1016/j.heliyon.2023.e20821.

Louppe, G., Wehenkel, L., Sutera, A., Geurts, P., 2013. Understanding variable importances in forests of randomized trees. Adv. Neural Inf. Proces. Syst. 26.

Lu, M., Perez, E.J., Mason, K., Li, M.Z., 2025. A hierarchical spatial and temporal optimisation of the air-high speed rail intermodal network. J. Transp. Geogr. 123, 104085. https://doi.org/10.1016/J.JTRANGEO.2024.104085.

Nguyen, M.-A.-T., Yu, M.-M., Lirn, T.-C., 2022. Airlines' eco-productivity changes and the European Union Emissions Trading System. Transp. Res. Part D: Transp. Environ. 102, 103100. https://doi.org/10.1016/j.trd.2021.103100.

Oh, D. hyun. (2010). A global Malmquist-Luenberger productivity index. *Journal of Productivity Analysis*, *34*(3), 183–197. Doi: 10.1007/S11123-010-0178-Y/TABLES/8.

Owen, B., Lee, D.S., Lim, L.L., 2010. Flying into the future: Aviation emissions scenarios to 2050. Environ. Sci. Technol. 44 (7), 2255–2260. https://doi.org/10.1021/es902530z.

Pang, S., Chen, M.C., 2023. The EU emissions trading system and airline low-carbon transition: a game-theoretic approach. J. Air Transp. Manag. 110, 102407. https://doi.org/10.1016/J.JAIRTRAMAN.2023.102407.

See, K.F., Guo, Y., Yu, M.M., 2025. Measuring energy and CO2 efficiency of global airlines using dynamic network DEA under natural and managerial disposability framework. Energy Econ. 150, 108788. https://doi.org/10.1016/J.ENECO.2025.108788.

Sharma, A., Jakhar, S.K., Choi, T.M., 2021. Would CORSIA implementation bring carbon neutral growth in aviation? a case of US full service carriers. Transp. Res. Part D: Transp. Environ. 97, 102839. https://doi.org/10.1016/j.trd.2021.102839.

Strauss, J., Li, H., Cui, J., 2021. High-speed Rail's impact on airline demand and air carbon emissions in China. Transp. Policy 109, 85–97. https://doi.org/10.1016/j.tranpol.2021.05.019.

Su-ungkavatin, P., Tiruta-Barna, L., Hamelin, L., 2023. Methodological framework for Life Cycle Assessment of sustainable aviation (SA) systems. Sci. Total Environ. 885, 163881. https://doi.org/10.1016/J.SCITOTENV.2023.163881.

Tone, K., 2001. A slacks-based measure of efficiency in data envelopment analysis. Eur. J. Oper. Res. 130 (3), 498–509. https://doi.org/10.1016/S0377-2217(99)00407-5.

Yang, F., Yao, Y., 2025. Sustainable aviation fuel pathways: Emissions, costs and uncertainty. Resour. Conserv. Recycl. 215, 108124. https://doi.org/10.1016/j.resconrec.2025.108124.

Zhang, N., Choi, Y., 2014. A note on the evolution of directional distance function and its development in energy and environmental studies 1997–2013. Renew. Sustain. Energy Rev. 33, 50–59. https://doi.org/10.1016/J.RSER.2014.01.064.

Zhang, Y., Wang, K., 2024. Mitigation effect of the European Union Emission Trading System on aviation emissions. Transp. Res. Part D: Transp. Environ. 130, 104186. https://doi.org/10.1016/j.trd.2024.104186.

Zhang, Y., Law, C.C., Zhang, A., 2023. The efficiency performance of full-service carriers in east and southeast asia amid increasing air transport liberalization and LCC presence. In: *Airlines and Developing Countries*. Emerald Publishing Limited, pp. 137–151.

Zhao, Y., 2025. Sustainable Performance Evaluation and Prediction of the Banking Sector: opening the Black Box of DEA with Machine Learning and Explainable AI. In: *International Conference on Machine Learning and Computing*. Cham, Springer Nature Switzerland, pp. 795–809.

Zhou, Y., Cai, J., Xu, Y., Wang, Y., Jiang, C., Zhang, Q., 2021. Operation performance evaluation of green public buildings with AHP-fuzzy synthetic assessment method based on cloud model. Journal of Building Engineering 42, 102775. https://doi.org/10.1016/J.JOBE.2021.102775.



https://dspace.org/

A three-stage evaluation of airline low-carbon competitiveness

Lu, Mengyuan; Yu, Jinglei; Li, Fei; Zhang, Yue; Jia, Quanxing

2026-03
Elsevier
https://doi.org/10.1016/j.trd.2025.105173
Lu M, Yu J, Li F, et al., (2026) A three-stage evaluation of airline low-carbon competitiveness. Transportation Research Part D: Transport and Environment, Volume 152, March 2026, Article number 105173

Downloaded from a DSpace repository. https://dspace.org/