

# Data-Driven Trajectory Optimization on The CGK–UPG Route Using K-Means and Bluesky Stochastic Simulation

M.Efaldes Syahputra<sup>1,2</sup>, Toto Indriyanto<sup>2</sup>, Javensius Sembiring<sup>2</sup>

<sup>1</sup>Ministry of Transportation, Republic of Indonesia

<sup>2</sup>Faculty of Mechanical and Aerospace Engineering, Bandung Institute of Technology, Indonesia  
e-mail: efaldes123@gmail.com

Received: 12-12-2025. Accepted: 26-01-2026. Published: 12-02-2026

## Abstract

This study examines the operational efficiency of the Jakarta–Makassar (CGK–UPG) route, one of Indonesia’s busiest domestic corridors, by integrating historical ADS-B data analysis with stochastic simulation. Flight trajectories obtained from FlightAware were processed through data cleaning, standardization, Principal Component Analysis (PCA), and K-Means Clustering to extract a representative data-driven trajectory. This trajectory was evaluated alongside the RNAV T-5 and conventional VOR/NDB routes documented in AIRAC AIP AMDT 162. All routes were assessed using 300 Monte Carlo simulations in the BlueSky ATM Simulator, incorporating randomized variations in cruise speed and altitude to reflect operational uncertainty. Performance indicators included flight distance, travel time, fuel consumption, and CO<sub>2</sub> emissions. Results indicate that the data-driven trajectory reduces flight distance by approximately 36–51 km and shortens travel time by about 10–12 minutes compared to the procedural routes. These improvements correspond to fuel savings of approximately 148–209 kg per flight and CO<sub>2</sub> emission reductions of around 468–659 kg per flight. These findings demonstrate that trajectory extraction based on historical data can produce more efficient flight paths and support the implementation of Trajectory-Based Operations (TBO) within Indonesia’s domestic airspace. The approach also contributes to national sustainability efforts by enabling potential reductions in carbon emissions from the aviation sector.

**Keywords:** ADS-B Flight Trajectory, PCA, K-Means, BlueSky, TBO.

## 1. Introduction

The increasing need for efficient, predictable, and environmentally responsible air traffic operations has led to the growing importance of trajectory optimization in modern aviation. One of the domestic routes that reflects this operational complexity is the Jakarta–Makassar (CGK–UPG) corridor, which connects two major national hubs and represents one of the busiest east–west flows in Indonesian airspace. Although official procedures such as the RNAV T-5 route and the conventional VOR/NDB airway have long been used as the primary navigation structures, ADS-B observations indicate that actual flight trajectories along this corridor display significant variations in lateral paths, cruise levels, and speed profiles. These discrepancies arise because procedural routes are designed primarily for airspace structure and separation assurance, whereas actual flight operations are influenced by real-time air traffic management decisions, including tactical trajectory adjustments issued by air traffic control, prevailing traffic conditions, and weather avoidance. As a result, observed flight paths often deviate from the published procedures, highlighting the need for a more comprehensive assessment of trajectory behavior using data-driven techniques. This study also builds upon prior analyses of domestic route performance and contributes further insight into trajectory behavior on the CGK–UPG corridor.

In general, trajectory characterization and optimization rely on various methods. Traditional approaches often depend on statistical summaries, manual interpretation of radar tracks, or performance assumptions embedded in flight plan data. These methods, however, are limited in accuracy because they do not fully capture real-world flight variability or fine-scale trajectory deviations. Moreover, operational constraints and environmental factors—such as winds, traffic sequencing, or airspace restrictions—are not easily incorporated into such simplified models. The introduction of Automatic Dependent Surveillance–Broadcast (ADS-B) data provides significant improvements in trajectory observation, yet the high dimensionality of ADS-B datasets poses challenges for direct interpretation (Basora et al., 2017)(Mustaffa et al., 2018).

Recent developments have explored the use of dimensionality-reduction techniques such as Principal Component Analysis (PCA) and clustering algorithms such as K-Means to identify representative trajectory patterns within large ADS-B datasets. PCA helps reduce the complexity of multivariate trajectory variables, while K-Means has been widely adopted for grouping trajectories with similar geometric properties (Ding & He, 2004). Despite their advantages, these data-driven methods require further validation, especially in operational environments where fuel consumption, travel time, and CO<sub>2</sub> emissions must be evaluated quantitatively.

Simulation-based analysis has emerged as an alternative method for evaluating trajectory performance. Platforms such as the BlueSky ATM Simulator enable researchers to simulate aircraft behavior along different routes using stochastic variations in speed and altitude (Hoekstra & Ellerbroek, 2016)(Groot et al., 2024). While this method has been applied in several international studies, its use for validating Indonesian domestic routes particularly CGK–UPG remains limited. Previous Indonesian research has largely relied on theoretical modelling or deterministic calculations, which may not fully capture the operational variability present in real flights.

Based on these considerations, this study aims to investigate data-driven trajectory extraction for the CGK–UPG corridor using ADS-B data combined with PCA and K-Means clustering, followed by stochastic simulation using the BlueSky ATM Simulator. The research evaluates the performance of a data-driven mean trajectory relative to the RNAV T-5 and VOR/NDB routes published in AIRAC AIP AMDT 162. The analysis focuses on key performance indicators such as distance flown, travel time, fuel consumption, and CO<sub>2</sub> emissions. Ultimately, this study seeks to demonstrate how integrated data analytics and simulation can provide a cost-effective and operationally relevant approach to trajectory efficiency assessment, offering valuable insight for supporting Indonesia’s roadmap toward Trajectory-Based Operations.

## **2. Methodology**

This study consists of two main methodological components: processing ADS-B trajectory data using dimensionality-reduction, clustering techniques and validating the representative trajectory through stochastic simulation in the BlueSky ATM Simulator. Both components are designed to produce a data-driven trajectory model and evaluate its operational efficiency relative to official navigation procedures.

### **2.1. ADS-B Data Processing, PCA, and K-Means Clustering**

ADS-B historical data from the CGK–UPG corridor were pre-processed by removing duplicated rows, normalizing feature scales, and resampling each trajectory into 200 evenly distributed points to ensure temporal and spatial consistency across flights. Five dominant features were extracted for each trajectory: flight duration (s), path length (NM), lateral deviation

percentage, mean cruise speed (kt), and mean cruise altitude (ft). These features represent operational characteristics relevant to trajectory efficiency analysis.

In this study, Principal Component Analysis (PCA) is applied not merely as a generic dimensionality-reduction technique, but as a means to project heterogeneous trajectory efficiency features into a compact space that preserves dominant operational variability. This transformation enables consistent comparison and clustering of CGK–UPG flight trajectories while minimizing redundancy among correlated variables, such as flight duration, path length, and cruise characteristics.

Prior to PCA, all features were standardized using z-score normalization to ensure that each variable contributes equally to the analysis. The standardized feature matrix  $X$  represents individual CGK–UPG trajectories across the selected operational features, while the covariance matrix  $S$  quantifies the interdependencies among these efficiency-related variables. This step ensures that correlations, such as those between flight duration and path length, are properly accounted for and do not bias the subsequent distance-based clustering. Following the formulation described by Hastie et al. (2008), the covariance matrix is expressed as:

$$S = \frac{1}{N} x^T x \quad (2-1)$$

After constructing the covariance matrix, eigenvalue decomposition was applied to extract the principal components, which represent the dominant directions of variance in the data. Eigenvectors provide the weights of each feature in the principal components, while eigenvalues quantify the amount of variance captured along these directions.

For instance, the first principal component (PC1) is most strongly influenced by flight duration (0.548), path length (0.519), and lateral deviation percentage (0.522), capturing trajectory efficiency in terms of time and distance. The second principal component (PC2) is dominated by mean cruise altitude (0.978) and mean speed (0.116), representing vertical flight profile variations across trajectories. These values were directly obtained from the eigenvectors of the covariance matrix and used to interpret operational patterns. The computation of principal components follows the eigenvalue decomposition procedure described by Hastie et al. (2008), as expressed by:

$$X^T X = V D^2 V^T \quad (2-2)$$

where

- $V$  = eigenvector matrix (principal axes),
- $D^2 D^2 =$  diagonal matrix of eigenvalues.

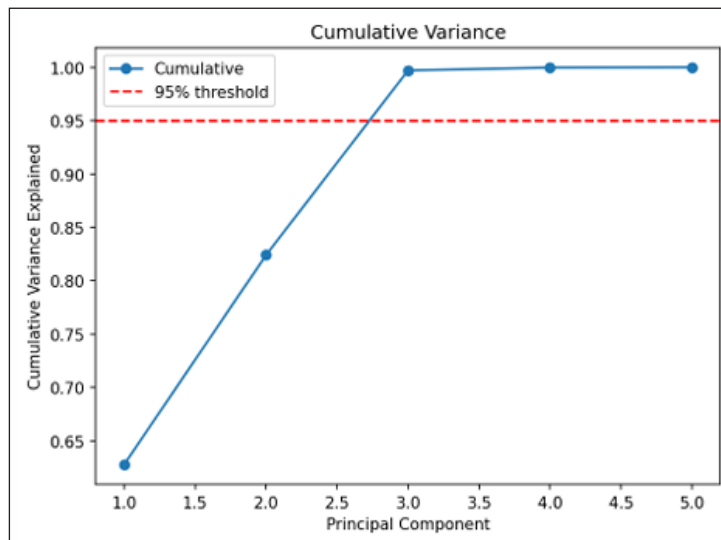
The explained variance ratio (EVR) for each principal component was quantified using the Proportion of Variance Explained (PVE), which measures how much of the total variation in the dataset is captured by each component. Components with higher PVE values represent a larger share of variance and are more informative for clustering, while components with negligible PVE contribute minimally and can be disregarded. According to Hastie et al. (2008), PVE is computed as:

$$PVE_m = \frac{\sum_{i=1}^n (\sum_{j=1}^p \phi_{jm} x_{ij})^2}{\sum_{j=1}^p \sum_{i=1}^n x_{ij}^2} \quad (2-3)$$

where,

- $\phi_{jm}$  = Loading for Feature  $j$  on principal component  $m$ ,
- $x_{ij}$  = value of variable  $j$  for sample  $i$  after z-score standardization,
- $n$  = total number of samples,
- $p$  = number of variables.

A higher PVE value indicates that the component accounts for a larger share of total variance, and components with negligible PVE contribute minimally to the representation of the dataset. Figure 2-1 shows the cumulative PVE for the five original features. The first two components jointly capture 82.4% of the total variance, exceeding the commonly applied thresholds of eigenvalue > 1 and cumulative EVR > 80%. Therefore, two principal components were retained as a compact yet sufficiently informative representation for the subsequent clustering stage. This 2D representation effectively summarizes operational variability while reducing dimensionality, making it suitable for subsequent K-Means clustering. The contribution of each original feature to the retained components was analyzed to interpret operational patterns: PC1 emphasizes flight duration, path length, and lateral deviation, representing overall trajectory efficiency in terms of distance and time, while PC2 primarily captures variation in mean cruise altitude, reflecting differences in vertical flight profiles across trajectories.



**Figure 2-1:** Cumulative variance explained by PCA components.

Next, K-Means clustering was applied to group CGK-UPG trajectories with similar operational characteristics in the reduced 2D space obtained from PCA. The goal of clustering was to identify representative patterns of trajectory efficiency while maintaining distinct separation among groups. As described by James et al. (2015), within-cluster variation for a cluster  $C_k$  is computed as:

$$W(C_k) = \frac{1}{|C_k|} \sum_{i \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \quad (2-4)$$

where,

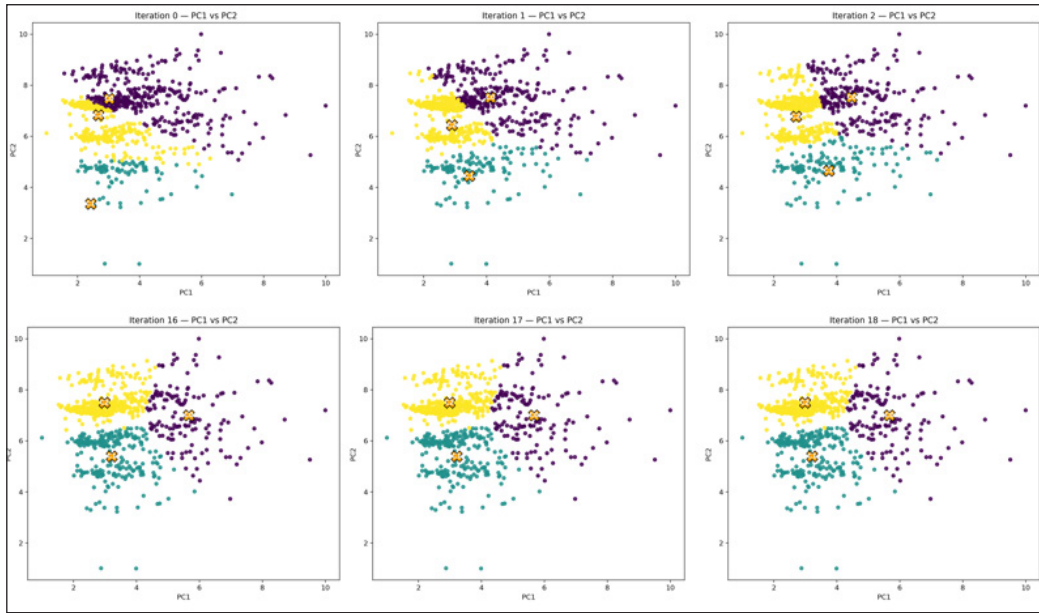
- $x_{ij}$  = value of feature j for sample i,
- $|C_k|$  = number of trajectories in cluster K,
- p = number of features.

K-Means aims to find a partition  $(C_1, \dots, C_K)$  that minimizes the total within-cluster variation:

$$\min_{C_1, \dots, C_K} \sum_{k=1}^K W(C_k) \quad (2-5)$$

In this study, K-Means clustering was initialized with K centroids randomly selected from the PCA 2D representation of the CGK-UPG trajectories. Each centroid acts as a provisional center for a potential cluster in the reduced feature space. During each iteration, every trajectory is assigned to the nearest centroid based on Euclidean distance, effectively grouping flights that share similar operational characteristics, such as flight duration, path length, and lateral deviation. After all trajectories are assigned, the centroid positions are recalculated as the mean of all points within the cluster, reflecting the updated cluster center. This assignment-update cycle continues iteratively, allowing the algorithm to progressively adjust the centroids so that clusters better fit the underlying structure of the trajectory data.

Convergence is reached when the changes in centroid positions between iterations become negligible, indicating that the cluster memberships have stabilized. In this research, the iterative process converged rapidly due to the compactness of the PCA-reduced 2D space (after 18 iterations). Figure 2-2 illustrates the convergence process, showing how trajectories gradually coalesce into three well-defined clusters. The final clustering reflects meaningful operational patterns is Cluster 3, exhibits the shortest path lengths, lowest lateral deviations, and shortest flight durations, making it the most efficient route and the basis for the subsequent stochastic simulation. This iterative fitting process demonstrates the ability of K-Means to accurately summarize the operational variability of CGK-UPG flights while maintaining clear separation between clusters.



**Figure 2-2:** Iterative Convergence of K-Means Clustering

To evaluate the optimal number of clusters  $K$ , the Elbow Method was employed by calculating the Sum of Squared Error (SSE) for cluster counts ranging from 2 to 10. following the formulation described by Rosyada & Utari (2024):

$$SSE = \sum_{k=1}^K \sum_{x_i} |x_i - C_k|^2 \quad (2-6)$$

where,

- $x_i$  = data point belonging to cluster  $C_k$
- $C_k$  = centroid of cluster  $K$
- $K$  = number of clusters evaluated
- $|x_i - C_k|^2$  = squared Euclidean distance between point  $x_i$  and its cluster centroid

SSE measures the compactness of each cluster, with lower values indicating tighter groups. When SSE is plotted for increasing values of  $K$ , the optimal cluster count is identified at the point where additional clusters no longer produce substantial reduction in SSE, this inflection point is known as the Elbow. In addition to SSE, the Silhouette Coefficient was used to evaluate clustering quality by measuring how well each trajectory fits within its assigned cluster relative to other clusters. Higher silhouette values indicate better cohesion and separation, meaning the clusters are more distinct and internally consistent

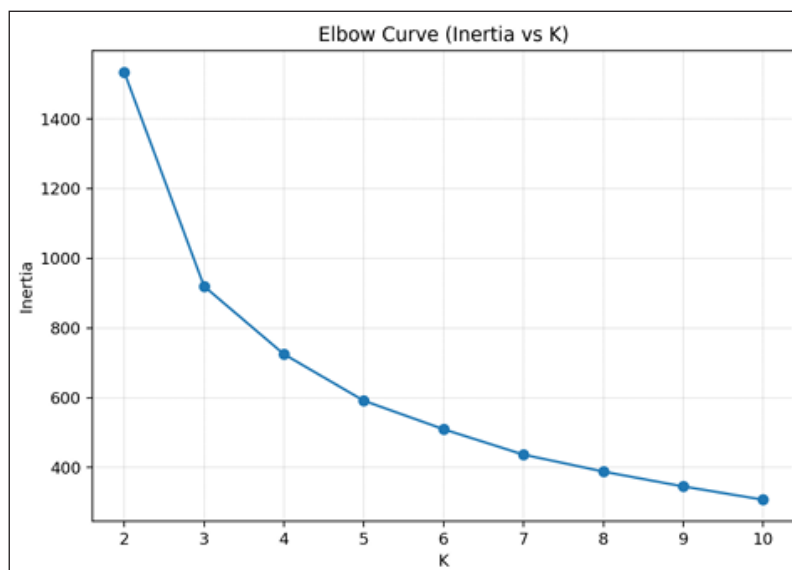
The optimal cluster count in this study was determined using both the Elbow Method and the Silhouette Coefficient. Both metrics consistently indicated that  $K=3$  provides the best partitioning of the data. The corresponding elbow curve is shown in Figure 2-3 and Table 2-1 presents the Inertia (SSE) and mean Silhouette Coefficient for  $K=2$  to 10. The inertia decreases sharply from  $K=2$  to  $K=3$  (1534.099 to 919.680), after which the rate of improvement diminishes, forming a clear elbow at  $K=3$ . The silhouette values show a consistent pattern, where the highest mean silhouette score also occurs at  $K=3$  (0.4321), indicating the best

balance between cohesion and separation. Silhouette scores for larger values of K fluctuate at lower levels and do not surpass the score at K=3, confirming that additional clusters do not improve the clustering structure.

Based on these complementary metrics, K=3 was selected as the optimal number of clusters. Among the three clusters, Cluster 3 demonstrates the shortest path length, lowest lateral deviation, and shortest flight duration. Its mean trajectory was therefore selected as the representative data-driven route for simulation.

**Table 2-1:** Inertia (SSE) and Mean Silhouette Coefficient for Different Values of K

K	Inertia	Silhouette Mean
2	1534.099	0.4075
3	919.680	0.4321
4	725.014	0.4009
5	591.659	0.3905
6	509.715	0.3884
7	436.980	0.3845
8	388.359	0.4029
9	345.649	0.4089
10	307.724	0.4177

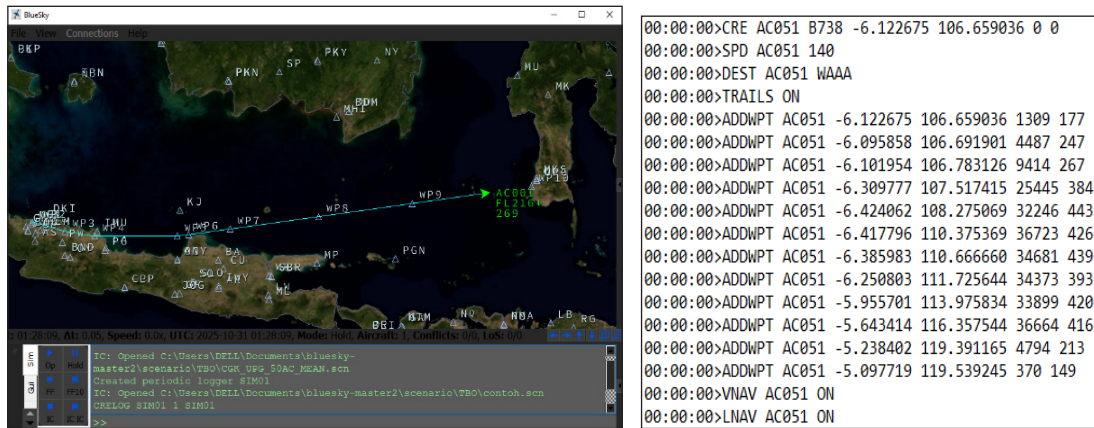


**Figure 2-3:** Elbow Curve for determining optimal number of clusters.

## 2.2. BlueSky Simulation Setup and Stochastic Trajectory Validation

The representative data-driven trajectory and the two procedural routes RNAV T-5 and the conventional VOR/NDB airway were evaluated using the BlueSky Air Traffic Management (ATM) Simulator. BlueSky provides an open-source, modular environment that supports waypoint-based navigation, aircraft performance modeling, and multi-scenario batch execution, making it suitable for analyzing operational variations across multiple simulated flights while maintaining consistency in aircraft behavior and route structure.

Each trajectory was translated into a BlueSky scenario by defining a sequential list of latitude–longitude waypoints that correspond to either the mean trajectory of Cluster 3 or the procedural airway coordinates. For all simulations, a Boeing 737-800 aircraft model was used to ensure comparability of operational performance. To capture realistic variations in flight operations, stochastic perturbations were introduced for cruise speed and cruise altitude, set at ±5% of the nominal values. This approach models typical operational deviations due to air traffic, weather, and pilot control inputs. For each route, a total of 300 Monte Carlo simulations were executed, generating a robust dataset for statistical evaluation. Figure 2-4 illustrates the simulation, including scenario construction and stochastic parameter assignment.



**Figure 2-4:** Simulation process and BlueSky scenario structure.

During execution, BlueSky generated log files containing simulation time ( $t$ ), aircraft identifier ( $AC\_ID$ ), distance flown ( $d$ ), and geographical coordinates (latitude and longitude). A Python-based parser was developed to process these logs: trajectories were separated by simulation run, reordered chronologically, and resampled to ensure consistent spatial and temporal resolution. Total flight distance and elapsed time were computed for each run, providing primary metrics for operational efficiency analysis. Based on Kühn & Scholz (2023), fuel consumption for each run was estimated using an empirical operational model defined as:

$$Fuel\ Consumption\ (kg) = C \times D \times N_{OEM} \tag{2-7}$$

where,

$C$  = Fuel Burn Rate (kg/km/seat),

$D$  = flight distance (km),

$N_{OEM}$  = effective seat count.

As defined by Hurtecant (2021), the effective number of seats was calculated using the following linear relationship:

$$N_{oem} = 0.6696 \times N_{max} + 22.858 \tag{2-8}$$

where,

$N_{max}$  = maximum seating configuration of the aircraft

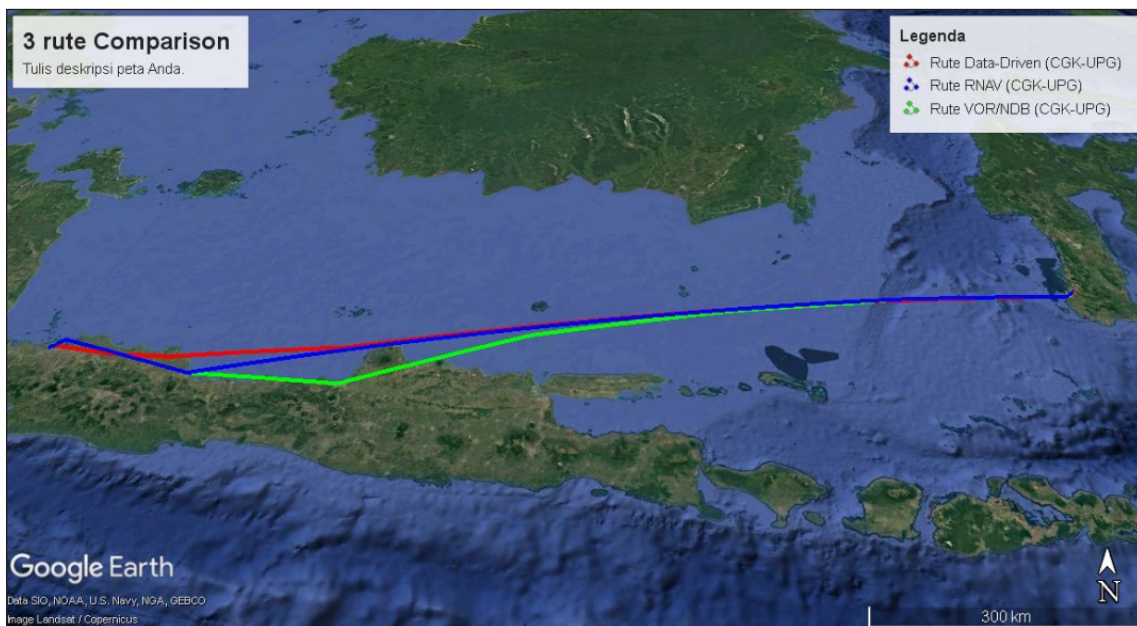
CO<sub>2</sub> emissions were computed using the standard conversion factor of 3.16 kg CO<sub>2</sub> per kilogram of fuel burned. These performance indicators—flight distance, travel time, fuel

consumption, and CO<sub>2</sub> emissions—provide a quantitative basis for comparing the efficiency of the three routes under randomized operational conditions.

### 3. Results and Analysis

The analysis begins with the extraction of trajectory patterns from the ADS-B dataset using PCA and K-Means clustering. The PCA results show that the first two principal components explain 82.4% of the total variance, allowing the CGK–UPG flight trajectories to be represented efficiently for clustering and subsequent analysis. Three clusters emerged from the K-Means algorithm. Cluster 1 represents flights with wider lateral deviations, Cluster 2 contains trajectories with moderate variations, and Cluster 3 forms a compact and stable group with minimal dispersion. Due to its geometric consistency, Cluster 3 was selected as the representative data-driven route for further evaluation.

A comparison of the Data-Driven, RNAV T-5, and VOR/NDB routes is shown in Figure 3-1. The RNAV route exhibits a longer lateral structure following fixed waypoint constraints, whereas the VOR/NDB airway shows sharper directional changes aligned with radio navigation aids. The Data-Driven route, in contrast, forms a smoother and more direct corridor between CGK and UPG, reflecting the natural flight behaviour captured from operational ADS-B data. These geometric differences suggest potential efficiency improvements when adopting data-driven structures rather than rigid procedural routes.



**Figure 3-1:** Visualization of Data-Driven, RNAV, and VOR/NDB routes.

It is important to note that the extracted data-driven trajectory is derived exclusively from historical ADS-B observations of routine commercial flights operating along the CGK–UPG corridor. Consequently, the resulting mean trajectory reflects airspace usage that has been operationally cleared and routinely managed by air traffic control. No systematic penetration of restricted or prohibited airspace is observed, and the trajectory remains consistent with existing airspace constraints encountered during day-to-day operations.

To assess performance, 300 Monte Carlo simulations were executed for each route using the BlueSky ATM Simulator with stochastic variations applied to cruise speed and altitude.

Logs containing total distance, flight time, and positional data were processed to compute fuel consumption and CO<sub>2</sub> emissions for each run.

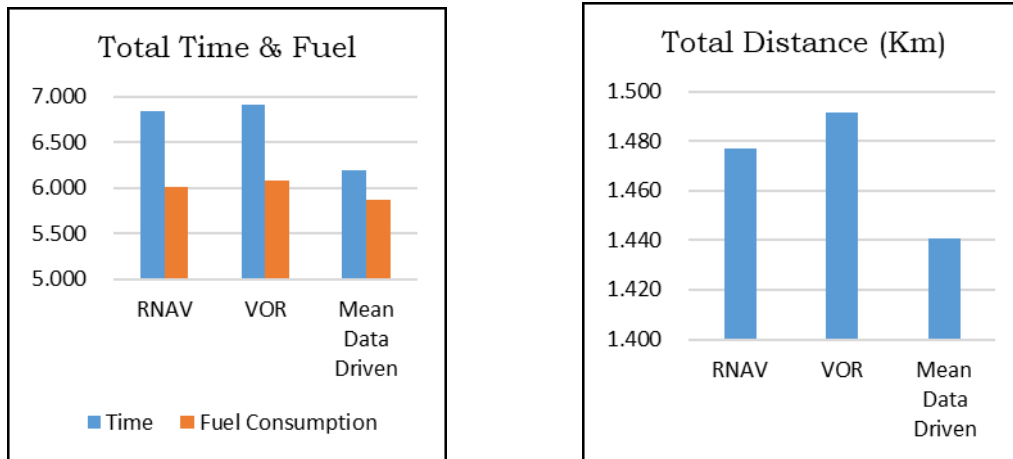
### 3.1 Flight Distance and Time Analysis

The performance comparison of flight distance and time is presented in Table 3-1. The Data-Driven route achieves the shortest average distance at 1,440.46 km, which is 36.40 km shorter than the RNAV route and 51.21 km shorter than the VOR/NDB airway. A similar pattern is observed in flight time, where the Data-Driven route records an average duration of 103.14 minutes, compared to 113.89 minutes for RNAV and 115.13 minutes for VOR/NDB. These differences correspond to reductions of 9.44% and 10.41%, respectively.

**Table 3-1:** Average Flight Distance and Time.

Route	Distance (km)	Difference from Data-Driven (km)	Time (min)	Difference from Data-Driven (min)
Data-Driven	1,440.46	–	103.14	–
RNAV T-5	1,476.85	36.40	113.89	10.75
VOR/NDB	1,491.67	51.21	115.13	11.98

These results demonstrate that the smoother geometry and reduced lateral deviation of the Data-Driven corridor directly contribute to shorter distances and faster flight times. As illustrated in Figure 3-2, the Data-Driven route consistently provides the most direct and efficient path among the three procedures, forming a strong basis for subsequent analysis in the following sections.



**Figure 3-2:** Bar chart comparing average flight distance and time

### 3.2 Fuel Consumption Analysis

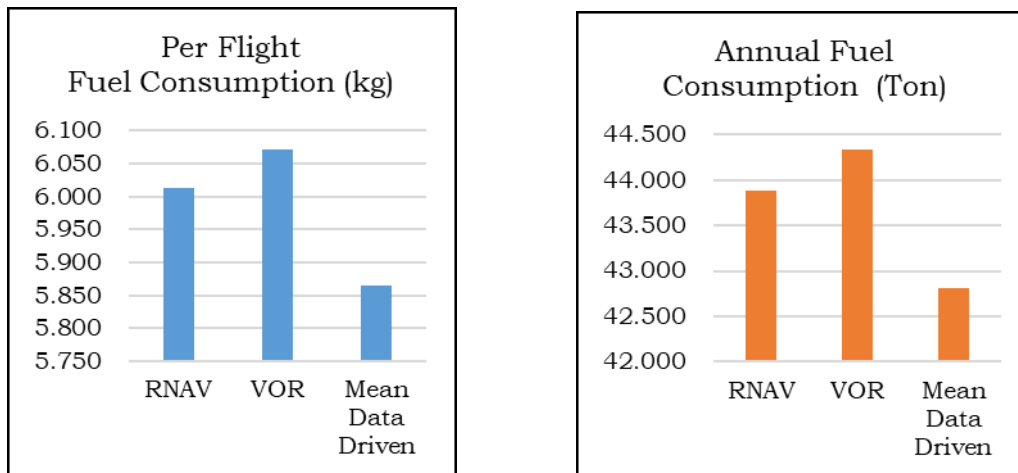
Fuel consumption was evaluated to determine how differences in flight distance and time influence the overall fuel requirements of each route. For every simulation run, fuel burn was

computed and then averaged across 300 Monte Carlo iterations for each route. The results, presented in Table 3-2 and illustrated in Figure 3-3, show clear distinctions in fuel demand among the three procedures.

The Data-Driven route demonstrates the lowest average fuel consumption at 5,864 kg per flight, making it the most efficient option. In comparison, the RNAV route requires an additional 148.17 kg of fuel per flight, while the VOR/NDB airway requires 208.49 kg more fuel than the Data-Driven trajectory. These values correspond to efficiency improvements of 2.46% over RNAV and 3.43% over VOR/NDB, aligning with the shorter distance and reduced flight time previously observed in Section 3.1.

**Table 3-2:** Average Fuel Consumption for the Three Routes.

Route	Fuel Consumption (Kg / flight)	Difference (Kg)	Fuel Consumption (Ton / year)	Difference (Ton / year)
Data-Driven	5,864	–	42,807.98	–
RNAV (AIP)	6,012	148.17	43,889.62	1,081.63
VOR/NDB	6,073	208.49	44,329.93	1,521.94



**Figure 3-3:** Bar chart comparing fuel consumption across the three routes

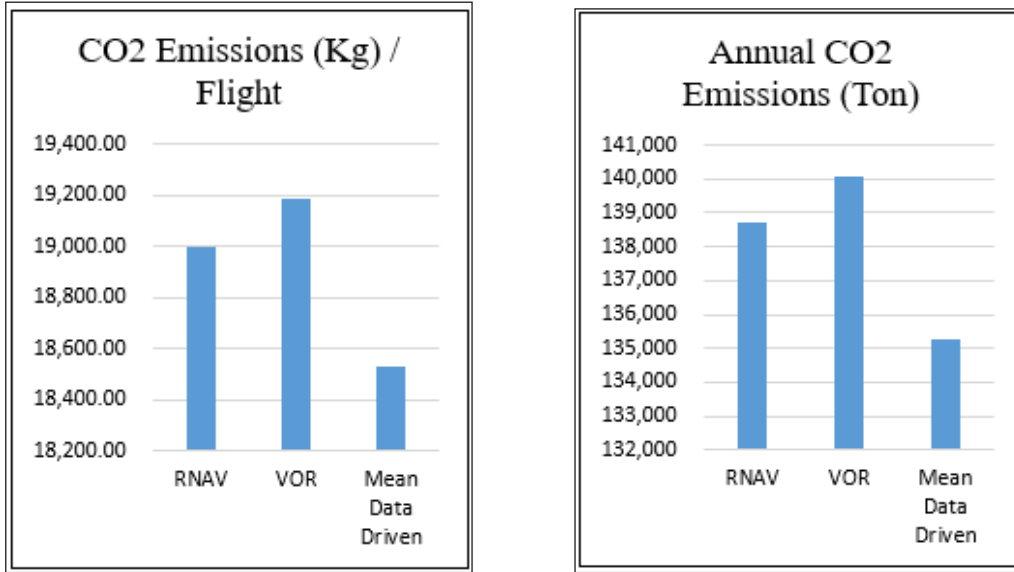
### 3.3 CO<sub>2</sub> Emissions Analysis

CO<sub>2</sub> emissions were analyzed to quantify the environmental impact of each route, using the fuel consumption results obtained from the 300 Monte Carlo simulations. Emissions were calculated by multiplying the fuel burn per flight by the standard emission factor of 3.16, following ICAO guidelines. The results show clear differences among the three procedures (Table 3-3).

The Data-Driven route records the lowest average CO<sub>2</sub> emission at 18,530.58 kg per flight, reflecting its reduced fuel consumption. In comparison, the RNAV route generates 18,998.79 kg of CO<sub>2</sub> per flight, while the VOR/NDB procedure produces 19,189.39 kg. This corresponds to additional emissions of 468.21 kg for RNAV and 658.81 kg for VOR/NDB relative to the Data-Driven route. These differences represent improvements of approximately 2.46% and 3.43%, respectively.

**Table 3-3:** Average CO<sub>2</sub> Emissions for the Three Routes.

Route	CO <sub>2</sub> Emissions (Kg / flight)	Difference (Kg)	CO <sub>2</sub> Emissions (Ton / year)	Difference (Ton / year)
Data-Driven	18,530.58	–	135,273.24	–
RNAV (AIP)	18,998.79	468.21	138,691.19	3,417.95
VOR/NDB	19,189.39	658.81	140,082.58	4,809.34



**Figure 3-4:** Bar chart comparing Carbon Dioxide Emissions across the three routes

### 3.4 Operational and Environmental Implications

The performance advantages of the Data-Driven route have several important implications for airline operations and air traffic management. From an operational standpoint, the reduction in fuel consumption contributes directly to lower operating costs. On the CGK–UPG corridor, the annual savings exceed 1,500 tons of fuel, which is a substantial amount for airlines operating on high-frequency domestic routes. Consistent savings of this scale enhance operational efficiency and support the financial sustainability of flight operations.

The environmental impact of the optimized route is equally significant. Reduced fuel burn results in lower CO<sub>2</sub> emissions, with annual reductions estimated between 3,400 and 4,800 tons when compared to the RNAV and VOR/NDB procedures. These improvements align with global aviation efforts to reduce greenhouse gas emissions and demonstrate how optimized en-route trajectories can contribute directly to national climate objectives. The reductions observed in this study also support Sustainable Development Goal (SDG) 13: Climate Action, which encourages immediate action to combat climate change by lowering emission sources in key sectors, including aviation.

From an air traffic management perspective, the Data-Driven route’s narrower dispersion and more predictable trajectory behavior provide additional benefits. The consistency of the corridor reduces uncertainty in aircraft movement, supports smoother sequencing, and simplifies conflict detection. This higher level of predictability is closely aligned with the

principles of Trajectory-Based Operations, which emphasize stable four-dimensional trajectories to improve the safety and efficiency of airspace operations.

The methodological framework employed in this study also demonstrates potential for broader implementation. The combination of trajectory clustering, corridor extraction, and stochastic performance validation can be applied to other domestic routes to identify more efficient en-route structures using real operational data. This provides a practical pathway for Indonesia to modernize its airway design while simultaneously supporting climate-oriented policies through measurable reductions in fuel use and emissions.

Overall, the Data-Driven route not only enhances flight efficiency and operational performance but also contributes to environmental sustainability in a way that is consistent with Indonesia's commitment to SDG 13 and the wider transition toward TBO-oriented airspace management

#### **4. Conclusions**

This study shows that a data-driven method supported by PCA-K-Means clustering and stochastic simulation can identify an en-route corridor that performs more efficiently than the existing RNAV and VOR/NDB structures on the CGK–UPG route. By deriving a representative trajectory from operational ADS-B data and testing it through 300 simulation runs, the Data-Driven route consistently produced shorter distances, shorter flight times, and lower fuel consumption.

The simulation results show that the Data-Driven trajectory provides the highest level of efficiency compared to the RNAV T-5 and VOR/NDB procedures. This route produces the shortest flight distance, the fastest travel time, and the lowest fuel consumption and CO<sub>2</sub> emissions. In contrast, the RNAV T-5 procedure results in a flight path approximately 36.4 km longer and 10.4 percent slower, while the VOR/NDB airway yields the longest distance and duration. Average fuel consumption on the RNAV route is about 148 kg higher than the Data-Driven trajectory, whereas the VOR/NDB route requires roughly 209 kg more fuel. CO<sub>2</sub> emissions follow the same pattern, with RNAV producing approximately 2.53% more emissions and VOR/NDB about 3.56 percent more compared to the Data-Driven route.

These per-flight improvements scale significantly at the operational level. When applied to the daily traffic level of the CGK–UPG corridor, the total annual savings amount to between 1,082 and 1,522 tons of fuel. Using Pertamina's Jet A-1 fuel price for Makassar, which is approximately IDR 15,326 per liter or around IDR 19,157.5 per kilogram, the financial impact assuming an average of 20 daily flights, reaches more than IDR 20 billion per year. This outcome shows that even small improvements at the flight level can generate significant operational benefits for airlines.

The reduction in fuel consumption is also reflected in lower carbon emissions. Annual CO<sub>2</sub> savings range from about 3,400 to 4,800 tons, which supports Indonesia's commitment to emission reduction and contributes to Sustainable Development Goal 13 on Climate Action. Since these improvements can be achieved without additional ground infrastructure, the approach offers a practical option for enhancing environmental performance in domestic airspace. The Data-Driven route also produced a more stable and predictable trajectory pattern. This consistency supports the requirements of Trajectory-Based Operations, particularly in relation to predictability and reduced variability. The methodology used in this study, which includes clustering historical flight data and validating the resulting route under stochastic conditions, can be adapted for other domestic corridors with similar traffic characteristics.

In conclusion, data-driven route extraction provides a realistic and effective way to improve operational efficiency while also delivering environmental benefits. The results offer clear value for airlines and air navigation service providers and support ongoing efforts to modernize Indonesia's airspace in preparation for wider TBO implementation

## Acknowledgements

The authors gratefully acknowledge the Ministry of Transportation of the Republic of Indonesia for the financial support and the opportunity provided to complete this study. The authors also acknowledge TU Delft for the development and open access of the BlueSky ATM Simulator, which was utilized for trajectory modelling and stochastic flight simulation.

## References

- Basora, L., Morio, J., & Mailhot, C. (2017). A trajectory clustering framework to analyze air traffic flows. *SESAR Innovation Days, November*.
- Blom, H. A. P., & Bakker, G. J. B. (2016). Agent-based modelling and simulation of trajectory based operations under very high traffic demand. *SESAR Innovation Days*.
- Dal Sasso, V., Djeumou Fomeni, F., Lulli, G., & Zografos, K. G. (2019). Planning efficient 4D trajectories in Air Traffic Flow Management. *European Journal of Operational Research*, 276(2), 676–687. <https://doi.org/10.1016/j.ejor.2019.01.039>
- Ding, C., & He, X. (2004). *K-means Clustering via Principal Component Analysis*. In *Proceedings of the 21st International Conference on Machine Learning (ICML 2004)*, Banff, Canada
- Directorate General of Civil Aviation. Indonesia Aeronautical Information Publication (AIP), Volume II. Directorate General of Civil Aviation, Ministry of Transportation of the Republic of Indonesia, Jakarta
- Enea, G., & Porretta, M. (2012). A comparison of 4D-Trajectory operations envisioned for NextGen and SESAR, some preliminary findings. *28th Congress of the International Council of the Aeronautical Sciences 2012, ICAS 2012*, 5, 4152–4165.
- EUROCONTROL. (2022). *Performance Review Report (PRR) 2021: An assessment of Air Traffic Management in Europe (v1.1)*. Performance Review Commission. European Organization for the Safety of Air Navigation
- FlightAware. (2025). Aircraft tracking and flight data. Retrieved from <https://www.flightaware.com>
- Groot, D. J., Leto, G., Vlaskin, A., Moec, A., & Ellerbroek, J. (2024). *BlueSky-Gym: Reinforcement Learning Environments for Air Traffic Applications*. *SESAR Innovation Days*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Hurtecant, D. (2021). *Launch of an ecolabel for passenger aircraft*. Master's thesis, Hochschule für Angewandte Wissenschaften Hamburg
- Hoekstra, J., & Ellerbroek, J. (2016). *BlueSky ATC Simulator Project: An Open Data and Open Source Approach*. In 7th International Conference on Research in Air Transportation: Philadelphia, USA.
- IATA. (2025). IATA - Sustainable Aviation Fuel (SAF). *International Air Transport Association*, 54(3), 1–19. <https://www.iata.org/en/programs/environment/sustainable-aviation-fuels/>
- ICAO & EUROCONTROL. (2024). *Asia Pacific TBO Pathfinder Project – Regional Perspectives*. Global TBO Symposium, Brussels, 5 June 2024.
- International Civil Aviation Organization. (2016). *ATM (Doc 4444) Issue November*. Montréal: ICAO
- IPCC. (2018). Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. In *Global Warming of 1.5°C*. <https://doi.org/10.1017/9781009157940.004>

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An Introduction to Statistical Learning: With Applications in R* (6th printing). Springer.
- Johnson, S. C., & Barmore, B. E. (2016). Nextgen far-term concept exploration for integrated gate-to-gate trajectory-based operations. *16th AIAA Aviation Technology, Integration, and Operations Conference*, 1–13. <https://doi.org/10.2514/6.2016-4355>
- Kim, D., Seth, A., & Liem, R. P. (2022). Data-enhanced dynamic flight simulations for flight performance analysis. *Aerospace Science and Technology*, 121. <https://doi.org/10.1016/j.ast.2022.107357>
- Kühn, M., & Scholz, D. (2023). Fuel Consumption of the 50 Most Used Passenger Aircraft. *Deutscher Luft-Und Raumfahrtkongress*, <https://www.fzt.haw-hamburg.de/pers/Scholz/arbeiten/TextKuehn.pdf>
- Li, F., Xu, X., Wang, R., Ma, M., & Dong, Z. (2025). Flight Trajectory Prediction Based on Automatic Dependent Surveillance-Broadcast Data Fusion with Interacting Multiple Model and Informer Framework. *Sensors*, 25(8), 1–22. <https://doi.org/10.3390/s25082531>
- Ministry of Transportation of the Republic of Indonesia. (2025). Air Transportation Statistics 2024. Directorate General of Civil Aviation.
- Mustaffa, M., Ahmad, S., Nor, R. M., & Zahari, N. H. B. (2018). Preliminary study on air traffic density of Peninsular Malaysia using visual flight path trajectories from Automatic Dependent Surveillance-Broadcast (ADS-B) data. *International Journal of Engineering and Technology (UAE)*, 7(4), 1–4. <https://doi.org/10.14419/ijet.v7i4.25.22238>
- Nations, U. (2024). The Sustainable Development Goals Report. *United Nations*, 26–27. <https://unstats.un.org/sdgs/report/2021>
- Nylin, M., Nilsson, J., Lundberg, J., & Bång, M. (2022). Atco Strategies in Tactical Tbo: a Simulator Study. *33rd Congress of the International Council of the Aeronautical Sciences, ICAS 2022*, 9, 6467–6479.
- Pamplona, D. A., de Barros, A. G., & Alves, C. J. P. (2021). Performance-Based Navigation Flight Path Analysis Using Fast-Time Simulation. *Energies*, 14(22), 7800. <https://doi.org/10.3390/en14227800>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rosyada, I. A., & Utari, D. T. (2024). Application of Principal Component Analysis for variable reduction in the K-Means clustering algorithm. *Jambura Journal of Probability and Statistics*, 5(1), 6–13. <https://doi.org/10.37905/jjps.v5i1.18733>
- SESAR Joint Undertaking. (2025). *European ATM Master Plan 2025*. 40.

