AIAA Aviation Forum and ASCEND co-located Conference Proceedings
21 - 25 July 2025, Las Vegas, Nevada
AIAA AVIATION FORUM AND ASCEND 2025
10.2514/6.2025-3519

Prediction of Fuel Burn for Runway-to-Runway Commercial Flight Operations with Machine Learning

Hyungu Choi*, Ameya Behere†, Jirat Bhanpato‡, Michelle Kirby§ and Dimitri N. Mavris¶
Georgia Institute of Technology, Atlanta, GA, 30332, U.S.A.

The mitigation of aviation's environmental impacts is crucial in enabling robust long-term growth for commercial aviation. A key aspect of this is fuel consumption and carbon emissions. With open-source flight data increasingly available, a capability to model fuel burn associated with real-world operations is desired. Existing methods rely on physics-based aircraft performance models and are therefore difficult to scale up due to computation costs. In this study, a data-based approach is explored wherein machine learning techniques are utilized to develop a prediction model for fuel burn, given real-world flight data.

Two Machine Learning models, the decision tree and random forest, are trained using operational, weight, and en-route weather variables. The comparisons indicate that Random Forest consistently outperformed Decision Tree for each condition. Takeoff weight is an important predictor, and using only mean values from weather observations was sufficient. Overall, our results confirm that machine learning is a scalable tool to estimate fuel burn and the potential for more extensive integration of environmental metrics into analyses of the aviation industry.

I. Introduction

The commercial aviation industry is expected to witness robust long-term global growth over the next few decades [1, 2]. This growth will be manifested through a variety of airline measures, such as upgauging, creation of new routes, and increased frequency of operations on existing routes. With rising air traffic, the associated environmental impacts will need to be mitigated. The aircraft fuel burn is an important metric that has strong correlations with CO_2, NO_x, and other emissions. In order to accurately evaluate future policies and air traffic scenarios, it is important to have a robust fuel burn modeling capability.

Current methods for fuel burn estimation require resource and time-intensive modeling of the aircraft performance using physics-based models and equations. Such methods cannot be easily scaled up to compute large inventories associated with a large number of operations. Thus, a data-based approach is proposed in this paper that can rapidly and accurately compute fuel consumption using readily available flight data. The focus of the proposed model is to provide fuel burn estimates to researchers and industry practitioners while taking inputs which are readily available, such as ADS-B flight data.

II. Background

A. Existing methods for modeling fuel burn

Multiple studies have attempted to model the fuel burn during a flight, which can be grouped into two main categories. These are simulation tools based on physics and data-driven models that aim to identify the connection between flight inputs and fuel consumption.

---

*Graduate Research Assistant, Aerospace Systems Design Lab, School of Aerospace Engineering, Georgia Institute of Technology, AIAA Student Member

†Research Engineer II, Aerospace Systems Design Lab, School of Aerospace Engineering, Georgia Institute of Technology, AIAA Young Professional Member

‡Research Engineer I, Aerospace Systems Design Lab, School of Aerospace Engineering, Georgia Institute of Technology, AIAA Young Professional Member

§Senior Research Engineer, Aerospace Systems Design Lab, School of Aerospace Engineering, Georgia Institute of Technology, AIAA Member

¶S.P. Langley NIA Distinguished Regents Professor and Director of Aerospace Systems Design Laboratory, Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, AIAA Fellow

The first set of methods makes use of aircraft performance models and equations to calculate fuel burn. This calculation is typically done by dividing the entire flight into segments and computing the fuel flow rate and duration of each segment. However, to obtain the fuel flow rate, the engine deck and thrust data must be known. Unfortunately, thrust data is rarely available in real-world flight data. Instead, thrust models may be used to compute thrust using aircraft speed, trajectory, and aerodynamic characteristics. These additional considerations introduce their own complexity and associated errors, making it infeasible to use for large-scale studies. The FAA's Aviation Environmental Design Tool (AEDT) offers this capability through a "sensor path" modeling workflow [3, 4]. AEDT is used for comprehensive environmental impact analysis and modeling data from AEDT has been used in several machine learning applications such as prediction models and sensitivity analyses [5–7].

The second set of methods instead relies on data and surrogate models to estimate fuel burn associated with an entire flight or different phases of a flight. Huang and Cheng [8] estimated fuel consumption for general aviation aircraft by creating machine learning models of avionics system data. The researchers trained Classification and Regression Tree (CART) and Neural Network (NN) modeling methods on data collected from a Cessna Skyhawk 172 aircraft equipped with Garmin G1000 system and ADS-B Out transponder. Their findings exhibited encouraging performance in general aviation.

Seymour et al. [9] developed a two-component multi-fidelity approach which could predict aggregated fuel burn within 5% error using simple origin-destination pair and aircraft type inputs.

Horiguchi et al. [10] employed random forests, XGBoost, and deep neural networks for fuel consumption prediction, utilizing datasets comprising flight, passenger, and reservation information. The training dataset comprised 47,000 flights, with 7,000 allocated for testing. Their XGBoost regression model achieved an RMSE of 8.8% for individual flights, significantly outperforming human dispatchers by reducing prediction errors by 39.3%. Key features identified by XGBoost include the departure day of the year, the cosine of the departure day of the year, the number of passengers, the scheduled departure time, and the scheduled arrival time. Their research emphasizes the significance of operational and scheduling features in predicting fuel consumption.

Wang et al. [11] proposed the optimization of a support vector machine (SVM) model using a receiver operating characteristic (ROC) curve for predicting aircraft fuel consumption. The SVM model was trained using flight data from flight manuals or Digital Flight Data Recorders (DFDRs), with the ROC curve serving as a tool for performance evaluation and parameter optimization. Its higher Area Under Curve (AUC) value of 0.89 highlighted the SVM model's superior performance over a Radial Basis Function Network (RBFN) model, which had an AUC value of 0.83. However, the study relied on a single aircraft type, the Boeing 737-800 with CFM-56 engines, which limits the generalization of its findings.

Baumann and Klingauf [12] focused on using Machine Learning algorithms, such as neural networks and decision trees, to model fuel flow. Their study showed that neural networks were effective for modeling entire flights with high accuracy, although they required a substantial amount of training time. Conversely, decision trees trained much faster and performed better during specific flight phases, such as cruise. The authors utilized 30 correlated parameters to predict fuel flow, but the study did not account for weather parameters.

Pan et al. [13] estimated aircraft fuel consumption across five flight phases: takeoff, climb, cruise, descent, and landing. They used separate back-propagation (BP) neural networks for each phase. The study indicated that incorporating meteorological factors, such as wind direction and speed, and data on aircraft motion, including longitudinal and transverse acceleration and tilt, enhanced prediction accuracy.

Kang and Hansen [14] developed a method to improve furnace burn prediction accuracy using ensemble learning techniques. This approach significantly outperformed existing airline fuel planning systems and even the most effective individual machine learning algorithms. Specifically, their Lasso-based stacking method cut prediction error by 50% compared to current methods. However, the model has excluded information related to en route weather and traffic forecasts, which limits its operational applicability in dynamic airspace environments.

These studies in the second set of methods illustrate a broad range of data-driven modeling techniques, including decision trees, SVMs, neural networks, and ensemble methods, which have been applied to various flight segments and types of aircraft. Many models are limited by narrow aircraft types or by a lack of incorporating weather in the methods.

B. Sources of data

Two main datasets are used in this study: the Flight Operational Quality Assurance (FOQA) dataset and the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) dataset. FOQA offers detailed records of flight operations, whereas MERRA-2 provides global atmospheric variables.

FOQA is high-fidelity flight trajectory data recorded by airlines for operational purposes. FOQA provides a dense and rich database, including information like thrust settings, flap positions, and aircraft weight throughout each flight stage. This dataset is of high frequency, and captures time-series data from all flight phases between the departure and arrival gate. FOQA data has been used for various research applications in the literature, including for the indirect calculation of takeoff weight from representation as procedural profiles [15], direct takeoff weight calculations using machine learning [16], predicting aircraft configuration for arrival operations [17], and for other Machine Learning applications, such as clustering [18].

Table 1 Unique city pairs included in the experiment (bidirectional)

City Pair	Included Directions
Atlanta – New York (ATL – LGA/JFK)	ATL ↔ LGA, ATL ↔ JFK
Atlanta – Washington D.C. (ATL – DCA)	ATL ↔ DCA
Atlanta – Minneapolis (ATL – MSP)	ATL ↔ MSP
Atlanta – Las Vegas (ATL – LAS)	ATL ↔ LAS
Atlanta – Seattle (ATL – SEA)	ATL ↔ SEA
Atlanta – Salt Lake City (ATL – SLC)	ATL ↔ SLC
Las Vegas – Salt Lake City (LAS – SLC)	LAS ↔ SLC

Note: All pairs are considered bidirectional. The analysis included both directions for each city pair.

Table 1 presents the selected routes used in the fuel burn prediction experiment. These routes were selected from the original FOQA dataset based on their frequency of operation, typical route distance, and the aircraft utilized. Additional selection criteria were informed by the expertise of team members with deep knowledge of airline operations, fleet usage, and flight planning procedures. Each route is considered bidirectional, meaning that travel in both outbound and return directions is included in the study. This bidirectional approach ensures the model effectively captures asymmetries in operational conditions, such as variations in routing, altitude profiles, and wind conditions. The total number of flights was approximately 2,300.

MERRA-2 is the most recent atmospheric reanalysis of the modern satellite era from NASA's Global Modeling and Assimilation Office (GMAO) [19]. The MERRA-2 reanalysis data set is an extensive collection of reported meteorological data including temperature, humidity, and wind. The MERRA-2 data set and system is comprised of four dimensions: latitude, longitude, altitude and time. The resolution is approximately 0.5° by 0.625° and includes 72 levels hybrid-eta levels from the surface to 0.01 hPa [20]. Since the FOQA dataset used in this study consists only of domestic U.S. flights, the MERRA-2 dataset was filtered to include observations within the continental United States. From the available variables, eastward wind, northward wind, temperature, and specific humidity were selected due to their strong relationship with aircraft performance and fuel consumption.

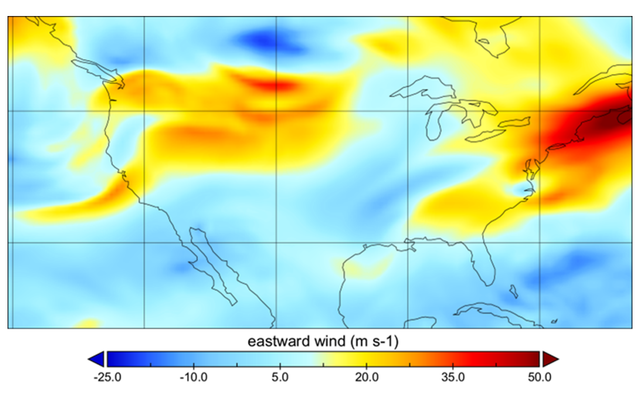

Fig. 1 Example of Eastward Wind (U-component) from Instantaneous MERRA-2 Data via Panoply Visualization Tool

Figure 1 displays the spatial distribution of instantaneous eastward wind (m/s) over North America, based on MERRA-2 reanalysis data. Positive values show eastward flow, while negative values indicate westward wind components.

C. Machine learning methods

To predict fuel burn for runway-to-runway operations, we applied Decision Tree (DT) and Random Forest (RF) algorithms, which are widely used machine learning algorithms for both classification and regression tasks [21]. In this section, we provide a brief overview of each algorithm.

Decision Trees (DTs) are a non-parametric supervised learning method structured as a hierarchical tree consisting of root nodes, branches, internal nodes, and leaf nodes [22]. It predicts the target variable by learning a series of simple decision rules derived from the input features. Decision Trees are easy to understand and interpret, and they are capable of handling both numerical and categorical data. However, without limiting their depth, Decision Trees can overfit the training data and perform poorly on unseen data [23]. This issue can be mitigated by restricting the size of the decision tree and pruning techniques.

To prevent overfitting and improve robustness in decision trees, Random Forest was introduced as an ensemble learning method that evaluates multiple decision trees and combines their output to make predictions[24]. In the algorithm, each tree in the forest is trained using a random sample of the training data. When the tree decides where to split the data, it only looks at a random set of features instead of all the features. For regression problems, the final result is the average of all three predictions.

III. Technical Approach

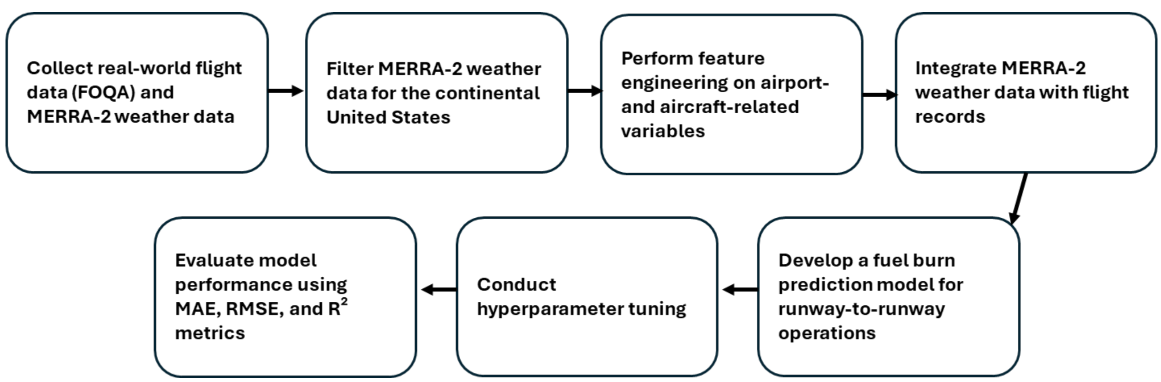

Fig. 2 Overview of the methodological process used in this study

Figure 2 illustrates the full processing pipeline. The procedure began with the collection of real-world FOQA flight data and instantaneous MERRA-2 weather data. The FOQA datasets were first cleaned by removing missing and erroneous values. Feature engineering was then conducted to extract variables related to airport operations and aircraft characteristics, including the departure and arrival airports and runways, takeoff and landing direction, aircraft type, engine type, and departure time. Each flight was also indexed by month to account for seasonal effects. Additionally, trajectory-derived features such as total flight distance, flight time, and total fuel burn, were included.

Following data preparation, relevant weather variables were extracted and also statistically summarized. Both operational and weather features were selected based on their relevance to fuel consumption. These selected features were then used to train predictive models, with total fuel burn serving as the target variable. The models were trained on 80% of the dataset and validated on the remaining 20%. Model performance was evaluated using standard metrics for regression: the Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). To minimize potential bias and ensure robustness, all evaluations were conducted using five-fold cross-validation to ensure robustness.

Finally, the results were analyzed to identify the most effective combinations of features and model configurations. This structured approach ensured a consistent methodology for building, validating, and interpreting the machine learning models used in this study.

A. Weather preprocessing

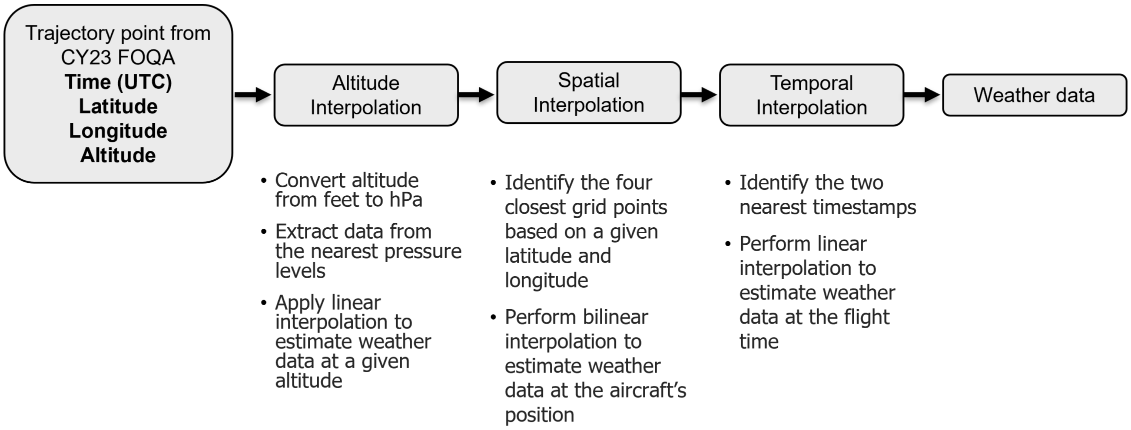

Fig. 3 Instantaneous MERRA-2 Weather Data Extraction Process

Figure 3 illustrates the process of extracting weather data along each flight trajectory, which includes timestamp, longitude, latitude, and altitude. Only en route segments above 10,000 feet Mean Sea Level (MSL) were considered for weather data extraction.

Since the en-route phase comprised the majority of both flight time and fuel consumption, this study concentrated solely on the impacts of weather during that phase. At cruise altitudes, broad atmospheric conditions like wind and temperature had the most significant effect on aircraft performance. By narrowing the analysis to en-route conditions, weather impacts could be assessed without interference from other operational complexities.

To obtain weather data corresponding to en route flight segments in the FOQA dataset, a three-step interpolation method was applied:

(1) Vertical interpolation: Altitude in feet was converted to pressure level (hPa), and weather values were interpolated linearly between the two nearest pressure levels.

(2) Spatial interpolation: Each trajectory point was constrained by four adjacent grid points in both latitude and longitude. Bilinear interpolation was employed to ascertain the meteorological values at the aircraft's location.

(3) Temporal interpolation: Two consecutive temporal intervals were identified, and linear interpolation was performed to compute weather conditions at the precise flight timestamp.

This sequence of vertical, spatial, and temporal interpolations ensured that weather conditions were accurately assigned to each trajectory point during the en route phase.

B. Model development

This section outlines the development and validation of the decision tree and random forest algorithms, for estimating fuel burn from runway to runway. A structured set of experiments was conducted to assess the influence of various combinations of input features and to determine the optimal configuration of the model.

The initial model was constructed using only operational features extracted from FOQA data. These features included airport and aircraft parameters such as including the departure and arrival airports and runways, takeoff and landing direction, aircraft type, engine type, and departure time. The purpose of this baseline model was to see how accurately fuel burn could be predicted using only operational data.

In the second phase, the aircraft takeoff weight was included as an additional feature. This variable demonstrated strong relevance in predicting fuel consumption, as noted in the prior literature, and was expected to further enhance prediction accuracy.

In the final phase, weather features were incorporated. Since the en-route phase constitutes the majority of flight time and fuel consumption, weather data were extracted exclusively for this segment. Instantaneous values of temperature, eastward wind, and northward wind were obtained from the MERRA-2 dataset for each trajectory point above 10,000 feet MSL. These variables were summarized using mean, maximum, and minimum values to represent weather conditions. Subsequently, comparative experiments were conducted to assess whether incorporating all three statistics provided a

Table 2 List of Features Used for Fuel Burn Prediction

Category	Features
Operational	Takeoff Airport, Takeoff Runway Landing Airport, Landing Runway Total Distance, Total Flight Time Takeoff Weight Aircraft Category, Airframe, Engine Flight Direction, Season, Flight Time Period
Weather	Mean Temperature (T_{mean}) Mean Eastward Wind (U_{mean}) Mean Northward Wind (V_{mean})
Target Variable	Total Fuel Burn

distinct advantage over using mean values alone. A total of six configurations were tested, each representing a different combination of operational, weather, and weight-related features, as summarized in Table 2.

Table 3 Random Forest Hyperparameter Grid

Parameter	Description	Candidate Values
n_estimators	Number of trees in the forest	100, 200, 300
max_depth	Max depth of each tree	5, 10, 15, Unlimited
min_samples_split	Minimum number of samples to split a node	2, 5, 10
min_samples_leaf	Minimum number of samples at a leaf node	1, 2, 4
max_features	Number of features to consider when looking for the next best split	sqrt, log2, Num_features

After identifying the feature set that delivered the highest performance, hyperparameter tuning was performed for the random forest model. Table 3 shows the hyperparameters used in the tuning process. A grid search method evaluated around 450 parameter combinations. The final model was chosen based on the trade-off between model fit and generalization, assessed using four regression metrics: MAE, MSE, RMSE, and R^2.

C. Evaluation Metrics

The performance of the fuel burn prediction model using the Decision Tree and Random Forest algorithms was evaluated using the most widely used evaluation metrics: MAE, MSE, RMSE, and R^2 [25]. While MAE, MSE, and RMSE focus on the magnitude of difference between predicted and actual values, R^2 measures how much of the total variability in the dependent variable is explained by the model [26, 27].Each of the four metrics is described in the following manner:

- **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Fuel\ Burn_{Actual}^{(i)} - Fuel\ Burn_{Predicted}^{(i)}\right| \tag{1}$$

  The MAE quantifies an average of absolute differences between actual values and the predicted values generated by the model. This metric signifies the actual magnitude of the error, regardless of whether the prediction surpasses or underestimates the actual value.

- **Mean Squared Error (MSE):**

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Fuel\ Burn_{Actual}^{(i)} - Fuel\ Burn_{Predicted}^{(i)})^2 \tag{2}$$

The MSE measures the average of the squared differences, instead of absolute differences, between actual and predicted values. It imposes greater penalties on larger errors compared to the MAE. A reduced MSE signifies enhanced predictive performance.

- **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Fuel\ Burn_{Actual}^{(i)} - Fuel\ Burn_{Predicted}^{(i)})^2} \tag{3}$$

The RMSE is obtained by taking the square root of the MSE. It provides an error measure in the same units as the predicted variable, highlighting large deviations more strongly.

- **Coefficient of Determination (R^2):**

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Fuel\ Burn_{Actual}^{(i)} - Fuel\ Burn_{Predicted}^{(i)})^2}{\sum_{i=1}^{n}(Fuel\ Burn_{Actual}^{(i)} - \overline{Fuel\ Burn}_{Actual})^2} \tag{4}$$

The R^2 score is used to assess how well a model fits the data and how well it can predict future outcomes. A value of $R^2 = 1$ indicates perfect prediction, while a value close to zero suggests that minimal variance is explained. Negative values may indicate that the model is performing worse than if it would have simply predicted the mean.

These metrics provide a comprehensive understanding of model performance, enabling a fair comparison between decision tree and random forest algorithms, as well as assessing whether integrating weather data enhances the model's performance. In the following section, these metrics will be used to evaluate the model's effectiveness under various experimental settings.

IV. Results and Discussions

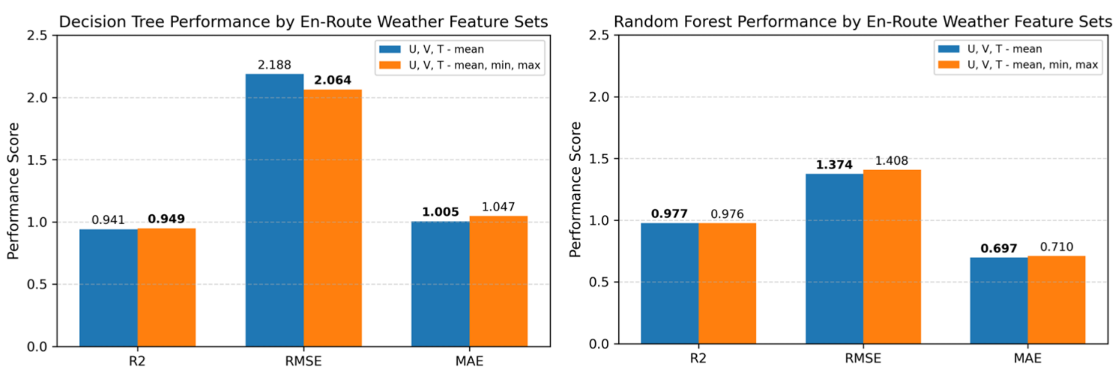

Fig. 4 Impact of Instantaneous MERRA-2 Weather Features on Decision Tree and Random Forest Performance

To examine the effect of the statistical representation of en-route weather variables on model performance, two feature sets were evaluated using the random forest algorithm. One feature set included only the means of eastward wind (U), northward wind (V), and temperature (T); the other feature set included means, minimums, and maximums of each of the weather variables to represent a broader scope of weather variability experienced by flights during the en-route phase.

Figure 4 illustrates the performance evaluation of the two feature sets across three metrics: R^2, RMSE, and MAE. Overall, the model that used only mean values of the weather variables showed slightly better performance across all three metrics, compared to the model that also included minimum and maximum values. Using only the mean weather values gave a slightly better R^2 score (0.977 vs 0.976) and lower errors: RMSE (1.374 vs 1.408) and MAE (0.697 vs 0.710) than when the minimum and maximum values were also included.

These results show that including the minimum and maximum values of weather data did not improve the model's accuracy. The model using only the mean values performed slightly better in all metrics. This means that using just the mean weather data is enough to make accurate predictions and is also simpler.

Table 4 Model Performance with Different Feature Combinations

Feature Set	Model	MAE	MSE	RMSE	R^2
Operational Features	Decision Tree	1.066	3.587	1.852	0.958
	Random Forest	0.800	2.244	1.442	0.975
Operational Features + Takeoff Weight	Decision Tree	0.917	2.820	1.664	0.968
	Random Forest	0.664	1.636	1.252	**0.982**
Operational Features + Weather	Decision Tree	1.005	5.112	2.188	0.941
	Random Forest	0.697	2.113	1.374	0.977
Operational Features + Weather + Takeoff Weight	Decision Tree	0.842	2.884	1.625	0.967
	Random Forest	**0.606**	**1.574**	**1.226**	**0.982**

The performance of machine learning models for predicting fuel burn was tested using four different feature sets, shown in Table 4. Each new set added more information to the previous one, including operational data, aircraft takeoff weight, and weather variables. This allowed us to see how each feature set helped improve model accuracy, both on its own and when combined.

The baseline model used only operational features. With random forest, it achieved an R^2 of 0.975 and an MAE of 0.800. When aircraft takeoff weight was added, the model improved significantly. The R^2 increased to 0.982, and the MAE dropped to 0.664, showing that takeoff weight is a strong factor in predicting fuel burn.

Adding weather features brought smaller improvements on their own, but when combined with takeoff weight and operational data, the model reached its best performance. The random forest model using all features gave the lowest errors: MAE of 0.606, RMSE of 1.226, and MSE of 1.574, with an R^2 of 0.982.

To better understand how the models behaved, Figure 5 shows scatter plots comparing predicted and actual fuel burn (top) and residuals (bottom) for both decision tree and random forest. Predictions from the random forest model closely follow the identity line, showing strong accuracy. In contrast, the decision tree model had wider spread and higher errors.

The residual plots reveal more detail. A good model should have residuals centered near zero. The random forest model had residuals that stayed close to zero across all fuel burn levels, meaning its predictions were consistent and less biased. The decision tree model, however, showed larger and more scattered errors, especially for flights with higher fuel use. These results confirm that random forest reduces overfitting and captures more complex relationships between inputs and fuel consumption.

Table 5 Top 3 Tuning Results for Random Forest

n_estimators	max_depth	min_samples_split	min_samples_leaf	max_features	r2	mae	mse	rmse
300	–	2	1	–	**0.982**	**0.602**	1.621	1.238
200	–	2	1	–	**0.982**	0.604	**1.612**	1.236
300	15	2	1	–	**0.982**	0.604	**1.612**	**1.235**

To better measure the accuracy of the model and see how hyperparameter choices affect performance, a grid search was used to find the best settings for the random forest model. The search tested different values for the number of trees, minimum samples to split a node, tree depth, minimum samples in a leaf, and the number of features used for each split. Table 5 lists the top three configurations, ranked by R^2 score.

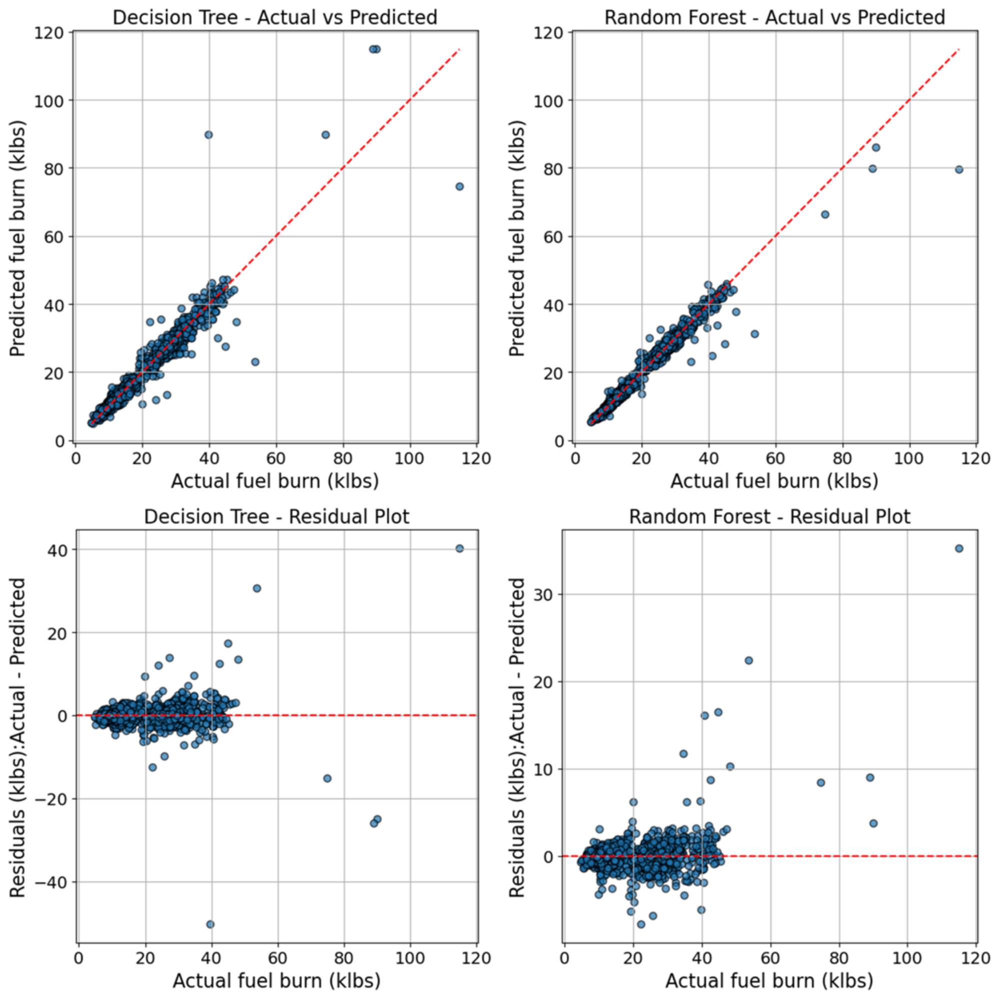

Fig. 5 Fuel burn prediction results using average en-route weather data

All three configurations had the same R^2 value of 0.982. This suggests that the model had already reached a high level of accuracy and further increases in complexity did not improve performance. In each of the top configurations, the minimum samples to split a node and to stay in a leaf were always 2 and 1. This shows that allowing deeper splits helped the model learn more detailed patterns in the data.

The number of trees used was either 200 or 300, which had very little effect on accuracy. The model with a depth of 15 gave the lowest RMSE and MSE values, although its MAE was slightly higher than the best MAE result. This means that limiting tree depth may help the model generalize better without increasing error. Overall, the high and consistent R^2 values show that random forest is reliable across different hyperparameter choices, and careful tuning can give slight improvements, especially when computational resources are limited.

V. Conclusions

This study developed a data-driven modeling framework aimed at estimating the total fuel consumption during commercial flights for runway-to-runway operations. The model incorporates Flight Operational Quality Assurance (FOQA) trajectory data alongside Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) reanalysis data, thus facilitating a more comprehensive representation of both operational and environmental inputs. The combination of FOQA trajectory data and MERRA-2 reanalysis weather data to represent both operational and environmental conditions. The focus was on the en route phase, during which fuel consumption is the highest. Decision tree and random forest algorithms were rigorously evaluated employing a variety of feature combinations to determine prediction accuracy.

The experimental results indicated that the Random Forest model consistently outperformed the Decision Tree model across all feature combinations. The most significant increase in prediction accuracy was noted when aircraft takeoff weight was included as a feature. This finding underscores the importance of aircraft weight as an operational factor in predicting fuel consumption. Additional tests investigated whether incorporating the maximum and minimum values of weather variables, along with the mean, would enhance accuracy. However, using only the mean values yielded similar results, implying that additional weather statistics were unnecessary.

Overall, this study shows that machine learning models, especially Random Forest, can accurately predict fuel consumption when given flight and weather data. The findings also highlight the importance of choosing the right input features and model. Future improvements could include incorporating dynamic weather data, expanding the model to cover more city pairs and international flights, and assessing its suitability for real-time operations.

References

[1] The Boeing Company, "Commercial Market Outlook 2023-2042," Online, 2023. URL `https://www.boeing.com/commercial/market/commercial-market-outlook#overview`.

[2] Airbus SAS, "Global Market Forecast 2023-2042," Online, 2023. URL `https://www.airbus.com/en/products-services/commercial-aircraft/market/global-market-forecast`.

[3] "Aviation Environmental Design Tool (AEDT) Version 3e User Manual," User manual, U.S. Department of Transportation Federal Aviation Administration, May 2022.

[4] "Aviation Environmental Design Tool (AEDT) Version 3e Technical Manual," Technical manual, U.S. Department of Transportation Federal Aviation Administration, May 2022.

[5] Behere, A., Kirby, M., and Mavris, D. N., "Relative Importance of Parameters in Departure Procedure Design for LTO Noise, Emission, and Fuel Burn Minimization," *AIAA AVIATION 2022 Forum*, 2022. https://doi.org/10.2514/6.2022-3916, URL https://arc.aiaa.org/doi/abs/10.2514/6.2022-3916.

[6] Behere, A., and Mavris, D. N., "Principal Component Analysis of Aviation Noise Grids for Dimensionality Reduction," *AIAA SCITECH 2023 Forum*, 2023. https://doi.org/10.2514/6.2023-0794, URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-0794.

[7] Behere, A., "A reduced order modeling methodology for the parametric estimation and optimization of aviation noise," Ph.D. thesis, Georgia Institute of Technology, 2022.

[8] Huang, C., and Cheng, X., "Estimation of aircraft fuel consumption by modeling flight data from avionics systems," *Journal of Air Transport Management*, Vol. 99, 2022, p. 102181. https://doi.org/https://doi.org/10.1016/j.jairtraman.2022.102181, URL https://www.sciencedirect.com/science/article/pii/S0969699722000023.

[9] Seymour, K., Held, M., Georges, G., and Boulouchos, K., "Fuel Estimation in Air Transportation: Modeling global fuel consumption for commercial aviation," *Transportation Research Part D: Transport and Environment*, Vol. 88, 2020, p. 102528. https://doi.org/https://doi.org/10.1016/j.trd.2020.102528, URL https://www.sciencedirect.com/science/article/pii/S136192092030715X.

[10] Horiguchi, Y., Baba, Y., Kashima, H., Suzuki, M., Kayahara, H., and Maeno, J., "Predicting fuel consumption and flight delays for low-cost airlines," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, 2017, pp. 4686–4693.

[11] Wang, X., Chen, X., and Bi, Z., "Support vector machine and ROC curves for modeling of aircraft fuel consumption," *Journal of Management Analytics*, Vol. 2, No. 1, 2015, pp. 22–34.

[12] Baumann, S., and Klingauf, U., "Modeling of aircraft fuel consumption using machine learning algorithms," *CEAS Aeronautical Journal*, Vol. 11, No. 1, 2020, pp. 277–287.

[13] Pan, Z., Chi, C., and Zhang, J., "A model of fuel consumption estimation and abnormality detection based on airplane flight data analysis," *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, IEEE, 2018, pp. 1–6.

[14] Kang, L., and Hansen, M., "Improving airline fuel efficiency via fuel burn prediction and uncertainty estimation," *Transportation Research Part C: Emerging Technologies*, Vol. 97, 2018, pp. 128–146.

[15] Behere, A., and Mavris, D. N., "A Method for the Parametric Representation of Take-off Time-Series Trajectory Data for Environmental Impact Assessment," *AIAA AVIATION 2023 Forum*, 2023. https://doi.org/10.2514/6.2023-4111, URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-4111.

[16] Daga, T., Bhanpato, J., Behere, A., and Mavris, D., "Aircraft Takeoff Weight Estimation for Real-World Flight Trajectory Data Using CNN-LSTM," *AIAA AVIATION FORUM AND ASCEND 2024*, 2024. https://doi.org/10.2514/6.2024-4291, URL https://arc.aiaa.org/doi/abs/10.2514/6.2024-4291.

[17] Bhanpato, J., Behere, A., Moharir, A., and Mavris, D. N., "Aircraft Configuration Prediction for Arrival Operations from Trajectory Data," *AIAA SCITECH 2024 Forum*, 2024. https://doi.org/10.2514/6.2024-2618, URL https://arc.aiaa.org/doi/abs/10.2514/6.2024-2618.

[18] Behere, A., Bhanpato, J., Puranik, T. G., Kirby, M., and Mavris, D. N., "Data-driven Approach to Environmental Impact Assessment of Real-World Operations," *AIAA Scitech 2021 Forum*, 2021. https://doi.org/10.2514/6.2021-0008, URL https://arc.aiaa.org/doi/abs/10.2514/6.2021-0008.

[19] Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al., "The modern-era retrospective analysis for research and applications, version 2 (MERRA-2)," *Journal of climate*, Vol. 30, No. 14, 2017, pp. 5419–5454.

[20] "Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) — nordata.physics.utoronto.ca," https://nordata.physics.utoronto.ca/en/annexes-datasets-descriptions/71-temperature/modern-era-retrospective-analysis-for-research-and-applications-version-2-merra-2/, 2025. [Accessed 14-05-2025].

[21] "1.10. Decision Trees — scikit-learn.org," https://scikit-learn.org/stable/modules/tree.html#tree, 2025. [Accessed 15-05-2025].

[22] "What is a Decision Tree? | IBM — ibm.com," https://www.ibm.com/think/topics/decision-trees, 2025. [Accessed 15-05-2025].

[23] Malakis, S., Psaros, P., Kontogiannis, T., and Malaki, C., "Classification of air traffic control scenarios using decision trees: insights from a field study in terminal approach radar environment," *Cognition, Technology & Work*, Vol. 22, No. 1, 2020, pp. 159–179.

[24] Breiman, L., "Random forests," *Machine learning*, Vol. 45, 2001, pp. 5–32.

[25] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.

[26] Khanna, N., "Understanding Regression Metrics in Machine Learning: A Comprehensive Guide," https://niharikakhanna.hashnode.dev/understanding-regression-metrics-in-machine-learning-a-comprehensive-guide, 2023. Accessed: YYYY-MM-DD.

[27] Brownlee, J., "Regression Metrics for Machine Learning," https://machinelearningmastery.com/regression-metrics-for-machine-learning/, 2021. Published: 2021-02-16, Accessed: 2025-05-09.