What makes this airplane burn more fuel? Quantifying drivers of fuel flow on Tail 687

Team Members:

1. Duy Nguyen,
2. Hemant Kumaar Aruljothi
3. Prithika Kandasamy

Problem Statement:

Using approximately 50 (for now, we can use more later) flights from the same 4-engine aircraft (≈ 950,672 rows × 33 columns, 4 Hz), we ask: Which cockpit controls, flight conditions and aircraft health (related to maintenance) are most associated with instantaneous fuel burn (dependent variable – per-engine **Fuel Flow** (FF_1…FF_4)), and by how much during cruise phase?

Why it matters: Fuel represents the largest operating cost for airlines and a primary source of aviation CO_2 emissions. The Boeing ecoDemonstrator program has demonstrated significant industry investment in operational efficiency, evaluating over 250 fuel-reducing technologies since 2012 with successes in flight optimization procedures that reduce fuel use, emissions, and community noise (Boeing Commercial Airplanes, 2024). Transparent, quantified relationships from flight data analysis can inform pilot technique, flight planning, and maintenance priorities, enabling cost reduction and emissions mitigation without new hardware investments.

These following questions would help us understand and guide us to break down our main research question by addressing nuances of our project statement, similar to the weather project:

1. Which engine and flight parameters have the strongest influence on fuel flow?
2. How much does fuel flow change with specific parameter adjustments? Particularly, What's the general pattern of pilot input vs autopilot that impacts fuel flow?
3. How do fuel consumption patterns differ across flight phases? (however, for simplicity, we plan to focus on cruising altitude only)
4. Are there any differences indicative of maintenance needs or potential efficiency improvements (evaluate engine health)?

Data Source:

- [NASA DASHlink – Tail 687]: 50 MATLAB flight files combined into a single 4 Hz sampling rate with these columns:
    - Engine Performance Parameters: Fuel Flow (FF_1, FF_2, FF_3, FF4), Fan Speed (N1_1, N1_2, N1_3, N1_4), Core Speed (N2_1, N2_2, N2_3, N2_4), Exhaust Gas Temperature (EGT_1, EGT_2, EGT_3, EGT_4)
    - Flight Control & Envelope: Throttle (PLA_1, PLA_2, PLA_3, PLA_4), Flight Conditions [ALT (altitude), MACH (Aircraft Mach number (speed relative to speed of sound)), CAS (calibrated speed), TAS (true airspeed)], Aircraft Attitude [AOAC (angle of attack)], Performance [ALTR (climb rate), WS (wind speed)], Engine Commands [N1T, N1C (target/command)], Lateral Load [LATG (lateral g-force)]
    - Other: Vibration (VIB_4), Added Column (flight_id)

- We will conduct a preprocessing pipeline, including converting .mat format to pandas DataFrames, filtering non-operational flight segments, and handling multicollinearity among flight parameters.
- After discussion with Prithika, our expert domain in Aeronautical Engineering, we believe these parameters are sufficient to address our research questions because they capture the complete operational picture for fuel consumption. For instance, throttle positions (PLA) and engine commands (N1T, N1C) reveal pilot versus autopilot patterns, while flight conditions (altitude, Mach, airspeed, wind) control for environmental factors, and engine performance metrics (N1, N2, EGT, vibration) across all four engines indicate health and efficiency differences. Together with fuel flow measurements across 50 flights, we can quantify which factors most influence fuel consumption during cruise phase.

Analytical Approach:

- We plan to use multiple regression to regress the relationship between FF (Fuel Flow) from operational variables like (PLA, N1, ALT, MACH, etc). The multiple regression will allow us to quantify how much each variable contributes to an increase or decrease to fuel flow, for instance, +10% N1 leads to +X lbs/hr FF.
- With multivariable regression models, we do not just focus on prediction but also on **interpreting the relationships between variables**. For instance, by examining the model's coefficients, we can ask:
    - Holding altitude and speed constant, how much does a 1% increase in N1 increase fuel flow?
    - Holding N1 and Mach constant, how does an increase of 10,000 feet in altitude affect fuel flow?

Solution Technologies:

- We will implement our analysis using Python with pandas and NumPy for data manipulation, SciPy for statistical computations, statsmodels and scikit-learn for regression modeling, and seaborn and matplotlib for visualization. Our code, version control, and other shared documents will be managed through a GitHub repository, which will also serve as the platform for presenting our final results.
- The initial dataset of approximately 950,000 rows (50 flights) is manageable on standard computing hardware using pandas and NumPy, requiring no specialized distributed computing resources. However, we anticipate the clean dataset prior to analysis will be substantially smaller after removing records with missing values, duplicate entries, unusual values, and data points outside acceptable aviation operational ranges. This reduction will further ensure efficient analysis on standard hardware.

Challenges:

- Understanding .mat file dataset from **NASA DASHlink**. We need to figure out how to turn .mat data structure to pandas dataframe for clean analysis.
- Aircraft during flight have parameters that possess potential risk for multicollinearity.
- Translating a relationship's association into actionable insights can be challenging since association isn’t causation. Further experiment or causation analysis from observational data might be required but it is out of the scope of this course.
- There is no phase label for segmentation and segmentation is out of the scope of this project, we plan to use an aeronautic expert domain for further research to estimate the altitude for cruise and only focus on cruise phase for our analysis.

Citation:

- Boeing Commercial Airplanes. (2024). The Boeing ecoDemonstrator Program [Backgrounder]. Boeing.
- Boeing. (2024). ecoDemonstrator Program. Boeing Sustainability. [Link]
- Matthews, B. (2012). Flight Data for Tail 687 [Dataset]. NASA DASHlink (C3).

Group Dynamics:

- We hold synchronous and asynchronous meetings twice a week (first meeting for planning and strategy discussion, second for coding sessions), with all meeting minutes and decisions documented in Google Docs shared with the instructor (email: fischer9@seattleu.edu). Our day-to-day coordination occurs through our GitHub repository.
- Next step, for this plan to move forward, we will divide tasks to each member to prioritize our three tasks:
  - **Task 1 (Hemant & Duy):** understand the `.mat` structure by loading a sample file, investigate the nested data structure (`data`, `Rate`, `Units`, `Description`), and exporting a CSV "data dictionary" of all parameters with descriptions and sampling rates
  - **Task 2 (Prithika & Duy):** Build a combined 4 Hz dataset across flights by (1) finalizing the parameter list, (2) iterating over all `.mat` files to extract only signals sampled at 4 Hz (with matching lengths per file) to avoid misalignment, (3) once aligned, stacking into a single Pandas DataFrame, and (4) adding a `flight_id` column for provenance.
  - **Task 3 (Collaborative – domain review by Prithika):** clean with aviation rules, filter non-operational segments (e.g., N1 < 15%), fix impossible values (e.g., N1 > 100%, negative PLA), remove invalid states (e.g., zero speed at high altitude), and filter (ground, climb, cruise, descent) to enable EDA and regression.