

Duy Nguyen  
dnguyen44@seattleu.edu



SEATTLE  
UNIVERSITY

## Aircraft Fuel Efficiency Optimization: A Statistical Analysis of Cruise Phase Operations for Commercial Aviation

Seattle University DATA 5100, Group Project, December 2025

## Project Summary

This project investigates the determinants of fuel consumption during cruise-phase operations in commercial aviation using **1.88 million measurements** from 312 flights of Aircraft Tail 687. Through variance decomposition and sequential regression modeling, we quantify the relative importance of altitude, speed, engine performance, and meteorological conditions.

**Key Finding:** Engine fan speed accounts for **64.4%** of explainable variance in fuel consumption, approximately **2.2 times more** than altitude and speed combined (29.7%).

**Implication:** Airlines may achieve greater returns from engine health monitoring programs than from sophisticated flight planning optimization systems.

# Business Problem & Motivation

**Context:** Fuel costs represent approximately 30% of total operating expenses for commercial airlines

## Economic Stakes:

- ✦ Fuel cost per flight: \$15,000 - \$20,000
- ✦ A modest 5% reduction across 10,000 annual flights generates **\$15+ million** in savings
- ✦ Environmental co-benefit: Thousands of tons of CO<sub>2</sub> emissions avoided

**Research Question:** Which operational parameters have the greatest impact on fuel consumption, and how should airlines prioritize improvement initiatives?

**Scientific Aim:** Quantify the relative importance of operational parameters under pilot and airline control, decomposing total variance into components attributable to each predictor.

# Data Source & Characteristics

**Source:** NASA DASHlink Aviation Safety Reporting System (publicly available)

## Dataset Statistics:

- ❖ **Total Observations:** 1,878,441 measurements
- ❖ **Flights:** 312 flights (filtered from 651 original flights)
- ❖ **Flight Hours:** Approximately 130 hours of cruise operations
- ❖ **Aircraft:** Single aircraft type (Tail 687)
- ❖ **Sampling Rate:** 4 Hz (4 samples per second)

**Data Type:** Time-series sensor data from flight recorders during cruise phase

**Data Quality:** Zero missing values after filtering; all measurements verified within physical bounds; approximately 0.4% of observations removed as instrument errors

## Variables Measured

### Outcome Variable:

- ❖ **Total Fuel Flow:** Pounds per hour (lbs/hr), aggregated across all four engines

### Predictor Variables:

- ❖ **Barometric Altitude:** Feet (range: 20,000 - 40,000 ft)
- ❖ **Mach Number:** Speed relative to sound (range: 0.45 - 0.85)
- ❖ **Average Engine Fan Speed (N1):** % of maximum rated speed
- ❖ **Wind Direction/Speed:** Meteorological conditions

### Derived Variables:

- ❖ **Headwind Component:** Wind vector decomposed relative to aircraft track
- ❖ **Cumulative Fuel Burned:** Proxy for aircraft weight throughout flight

### Categorical Variables Constructed:

- ❖ **Altitude Band:** Low (25-29K), Medium (29-33K), High (33K+ ft)
- ❖ **Wind Category:** Strong Tailwind, Neutral, Strong Headwind ( $\pm 20$  knots)
- ❖ **Weight Category:** Heavy, Medium, Light (tertiles by cumulative fuel burned)

# Data Processing & Key Discovery

## Filtering Process:

1. Excluded climb/descent phases (vertical velocity  $> 500$  ft/min)
2. Applied physical bounds: Fuel flow 2,000-8,000 lbs/hr; Altitude 20,000-40,000 ft
3. Removed instrument anomalies ( $\sim 0.4\%$  of observations)
4. Synchronized all parameters at 4 Hz sampling rate

## Key Discovery: Slow Cruise Operations

During cleaning, we identified **5.4% of observations** exhibiting “slow cruise”, a distinct fuel-optimization strategy used under heavy weight conditions:

- ✦ Altitude: 28,739 ft (610 ft lower than normal cruise)
- ✦ Mach: 0.627 (0.08 slower than normal cruise)
- ✦ Fuel Flow: 4,519 lbs/hr (401 lbs/hr less than normal cruise)

**Insight:** Optimal flight parameters are **dynamic**, not static—validating weight variation as a critical predictor.

# Statistical Methodology

**Problem Type:** Supervised Learning — Regression

- ❖ **Outcome:** Total fuel flow (continuous, lbs/hr)
- ❖ **Predictors:** Altitude, Mach, N1, headwind component, weight proxy

**Analytical Approaches:**

1. **One-Way ANOVA:** Partition variance by categorical factors; report F-statistic and  $\eta^2$  (eta-squared) effect size
2. **Two-Way ANOVA:** Test altitude  $\times$  wind interaction effects
3. **Sequential Regression:** Add predictors in predetermined order to quantify incremental variance explained
4. **Nested F-Tests:** Compare successive models to assess if added complexity improves prediction
5. **Variance Decomposition:** Isolate unique contribution of each predictor set

**Practical Significance Criteria:** With  $n = 1.88$  million, we distinguish statistical from practical significance:

- ❖  $\Delta R^2 < 0.01\%$ : Trivial —  $0.01 - 0.1\%$ : Marginal —  $0.1 - 1\%$ : Moderate —  $> 1\%$ : Substantial

# Regression Model Specification

## Full Model Form:

$$\text{Fuel Flow} = \beta_0 + \beta_1(\text{Altitude}) + \beta_2(\text{Mach}) + \beta_3(\text{N1}) + \beta_4(\text{Headwind}) + \beta_5(\text{Weight})$$

+ Interactions +  $\varepsilon$

## Sequential Model Sequence:

- ❖ **Model 1:** Full model with all main effects + 3 two-way interactions (8 variables)
- ❖ **Model 2:** Remove altitude  $\times$  Mach interaction (7 variables)
- ❖ **Model 3:** Main effects only—no interactions (5 variables)
- ❖ **Model 4:** Remove average fan speed N1 (4 variables)
- ❖ **Model 5:** Core model—altitude and Mach only (2 variables)
- ❖ **Model 6:** Null model—intercept only (0 predictors)

## Diagnostic Procedures:

- ❖ Residual analysis: Mean, standard deviation, skewness, kurtosis
- ❖ D'Agostino-Pearson normality test; Breusch-Pagan heteroskedasticity test
- ❖ Emphasis on magnitude of deviations over formal p-values given large  $n$



# ANOVA Results

## One-Way ANOVA Results:

Factor	F-statistic	$\eta^2$	Variance Explained
Altitude Band	100,273	0.0965	<b>9.7%</b>
Weight Category	78,581	0.0772	7.7%
Wind Category	24,343	0.0253	2.5%

**Table:** All p-values < 0.001 (effectively zero with  $n = 1.88$  million)

## Two-Way Interaction (Altitude $\times$ Wind):

- ❖ F-statistic: 5,478 (statistically significant,  $p < 0.001$ )
- ❖ Additional variance explained: Only **1.0%** beyond main effects
- ❖ **Conclusion:** Limited practical importance despite statistical significance

**Interpretation:** Altitude selection explains  $\sim 10\%$  of fuel consumption variation, that's meaningful but not dominant. The remaining  $\sim 90\%$  stems from other factors.

## Sequential Regression Results

Model	Variables	$R^2$	Adj. $R^2$
M1: Full (interactions)	8	0.9586	0.9586
M2: $-\text{alt} \times \text{mach}$	7	0.9585	0.9585
<b>M3: Main effects only</b>	<b>5</b>	<b>0.9581</b>	<b>0.9581</b>
<b>M4: <math>-\text{N1}</math> (engine)</b>	<b>4</b>	<b>0.3404</b>	<b>0.3404</b>
M5: Core (alt, mach)	2	0.2850	0.2850
M6: Null (intercept)	0	0.0000	—

Table: Sequential model comparison showing dramatic drop when N1 is removed

### Critical Transition: Model 3 $\rightarrow$ Model 4

Removing engine fan speed (N1) causes  $R^2$  to **plummet from 0.958 to 0.340**, a decrease of **61.8 percentage points**. This establishes N1 as the dominant predictor.

## Nested F-Test Results

### Model Comparisons:

- ❖ **M1 vs M2:**  $F = 2,085$ ;  $\Delta R^2 = 0.0046\%$  (trivial)
- ❖ **M2 vs M3:**  $F = 10,766$ ;  $\Delta R^2 = 0.05\%$  (marginal)
- ❖ **M3 vs M4:**  $F = 27,700,000$ ;  $\Delta R^2 = 61.8\%$  (**massive**)
- ❖ **M4 vs M5:**  $F = 78,806$ ;  $\Delta R^2 = 5.5\%$  (substantial)
- ❖ **M5 vs M6:**  $F = 374,399$ ;  $\Delta R^2 = 28.5\%$  (substantial)

### Model Quality Assessment:

- ❖ Final Adjusted  $R^2 = 0.959$  — Model explains 95.9% of variance
- ❖ Residuals: Mean = 0.00, Std Dev = 86.6 lbs/hr ( $< 2\%$  of mean fuel flow)
- ❖ Minor normality deviations but robust given  $n = 1.88$  million

**Key Insight:** Interaction terms provide **negligible practical benefit** (0.05% total) despite statistical significance. Main effects model preferred for parsimony.

## Key Finding: Variance Decomposition

Decomposition of Explained Variance ( $R^2 = 95.86\%$ ):

Predictor Set	% Total Variance	% Explained Variance
<b>Engine Fan Speed (N1)</b>	<b>61.8%</b>	<b>64.4%</b>
Altitude + Mach Number	28.5%	29.7%
Headwind + Weight	5.5%	5.8%
Interaction Terms	0.05%	< 0.1%
Unexplained	4.1%	—

### The 2.2× Factor

Engine performance management offers approximately **2.2 times** the fuel savings potential of altitude-speed optimization:

$$\frac{61.8\%}{28.5\%} = 2.17$$

# Business Implications

## Three-Tier Priority Strategy:

1. **Priority 1 — Engine Performance (64% variance):**
  - Real-time N1/EGT monitoring systems
  - Predictive maintenance scheduling
  - Optimal power band operational protocols
2. **Priority 2 — Flight Planning (29% variance):**
  - Simplified altitude-speed guidelines
  - Treat factors as independent (no complex conditional logic)
  - Accessible guidance for flight crews
3. **Priority 3 — Environmental Factors (6% variance):**
  - Standard wind optimization practices
  - Weight management through existing procedures

**Challenges Conventional Wisdom:** Many carriers may be **overinvesting** in sophisticated flight management systems while **underinvesting** in engine condition monitoring and predictive maintenance programs.

# Study Limitations

## Important Caveats:

- ❖ **Single Aircraft Type:** Results specific to Aircraft Tail 687; generalization to heterogeneous fleets requires validation
- ❖ **Cruise-Only Focus:** Excludes climb and descent phases where different optimization strategies apply
- ❖ **Observational Design:** Identifies associations, not proven causal effects
- ❖ **Weight Proxy:** Cumulative fuel burned serves as imperfect proxy for actual aircraft weight
- ❖ **Engine Variance Interpretation:** The 61.8% engine contribution encompasses both power setting choices and engine health effects, decomposing these requires additional maintenance data

## Recommended Validation:

- ❖ Prospective intervention studies experimentally manipulating maintenance schedules
- ❖ Extension to multiple aircraft types and fleet compositions
- ❖ Integration of engine condition indicators and maintenance history

# Conclusions

## Key Findings:

1. Engine fan speed accounts for **64.4%** of explainable variance, approximately **2.2× more** than altitude and speed combined
2. Altitude and Mach contribute **29.7%** of explained variance—meaningful but secondary
3. Interaction effects contribute only **0.05%**—supporting treatment of altitude and speed as **independent factors**
4. Final model achieves **Adj.  $R^2 = 0.959$**  with robust residual diagnostics

## Answer to Research Question

Airlines seeking 5-10% fuel reductions should **prioritize engine condition monitoring and predictive maintenance** over sophisticated flight planning systems. Engine-related interventions deliver **2.2× the fuel savings potential** of altitude-speed optimization.

**Future Research:** Extend to heterogeneous fleets, incorporate climb/descent phases, develop predictive maintenance models linking engine condition to fuel consumption.

# Thank You

Questions?

Duy Nguyen  
dnguyen44@seattleu.edu  
Seattle University - DATA 5100

Data Source: NASA DASHlink Aviation Safety Reporting System  
[c3.ndc.nasa.gov/dashlink/resources/664/](https://c3.ndc.nasa.gov/dashlink/resources/664/)