An Information-Theoretic Framework for Modeling Aircraft Actual Takeoff Weight (ATOW)

Emanuele Bossi* and Abd AlRahman R. AlMomani.†
Embry-Riddle Aeronautical University, Prescott, AZ 86301

I. Extended Abstract

In recent years, data-driven methods have become increasingly central to *aviation operations and research*, with machine learning (ML) offering powerful tools for performance modeling, anomaly detection, and prediction tasks. One critical parameter in aircraft performance modeling is the Actual Takeoff Weight. The *Actual Takeoff Weight (ATOW)* of an aircraft is the total mass of the plane at the start of the takeoff roll (including payload, fuel, and operating empty weight). This weight is a fundamental parameter in flight dynamics – it directly influences climb/descent rates, range, fuel burn, runway requirements, and overall aircraft performance.

Because ATOW is usually known *only* to the operating airline and is *not disclosed publicly*, trajectory-prediction and environmental models often must rely on assumed or estimated weight values. Inaccurate weight assumptions can lead to *significant errors* in fuel burn and trajectory forecasts. For example, previous research highlighted that using only the aircraft type (instead of the actual weight) in simulations can degrade the accuracy of flight path and fuel estimates.

Accurate ATOW estimates therefore improve safety margins (by ensuring climb/climb gradient predictions are valid), enable better fuel planning, and reduce uncertainty in emissions and noise modeling. However, estimating takeoff weight accurately and reliably from open data remains an open problem of both scientific and operational significance.

Traditional methods for estimating ATOW include *physics-based* models ([1]) that leverage aircraft design specifications, fuel planning assumptions, and simplified range or energy equations. Other methods *leveraged flight data* (radar, ADS-B, QAR/FDR records) to learn ATOW directly. One strategy is to infer weight from climb-phase performance: several studies use least-squares([2]) or adaptive algorithms ([3]) on trajectory speed/altitude profiles to compute an "equivalent TOW" that matches observed *energy rates*. More recently, statistical and machine learning models-ranging from Gaussian processes ([4]) and Bayesian inference ([5]) to ensemble regressors ([6]) and deep neural networks ([7])—have emerged that learn weight estimates directly from historical flight data, including trajectory recordings and weather conditions.

No current method perfectly predicts ATOW for every flight. Physics-based approaches suffer when flight-specific inputs are wrong or missing. For instance, weight breakdown models require accurate payload and fuel usage figures (often not public), and closed-form equations assume "textbook" performance (e.g. no engine derate, standard surface friction), which may not hold. Any inexact data (like extra fuel uplift or engine inefficiency) would lead to approximate weight outputs. Climb-based estimation can mispredict if the aircraft experienced unmodeled thrust changes or turbulence, and these methods are typically validated indirectly by how well they predict flight tracks, not by ground-truth weight measurements. Without true weight data, their absolute error remains unknown.

Modern ML methods have their own challenges as well. First, they depend heavily on the availability and quality of data. ADS-B surveillance data can be incomplete (poor coverage at low altitudes or in remote regions) and contain errors. Second, ML models may overfit or be biased by the training data: for example, if the dataset has mostly routine flights of certain aircraft types or average weather conditions, the model may perform poorly on unusual flights (heavy cargo, atypical routes, extreme weather). Third, most ML models are *black boxes*: the complex ensemble of trees or neural nets used in current research does not transparently explain why a weight estimate was made. This *lack of interpretability* can hinder trust in safety-critical settings.

This research addresses these challenges through a *three-pronged approach* represented in the methodology flowchart displayed in 1. First, we explored the information content of multiple feature sources, including *flight metadata* (departure and arrival airports, day of year, time of day, airline, aircraft type) and *ADS-B trajectory data*, to understand their relative contribution to weight estimation. Using techniques such as *singular value decomposition (SVD)* and *conditional mutual information* analysis, we quantified the statistical dependency between candidate features and takeoff weight, providing a principled foundation for model development. Second, recognizing that ADS-B data can be affected

*bossie@my.erau.edu
†Assistant Professor of Mathematics and Data Science, College of Arts & Sciences, almomana@erau.edu

by various measurement errors, we developed an algorithm to *detect and correct corrupted or implausible ADS-B measurements*, with the goal of improving data reliability before feeding it into machine learning pipelines. Finally, we built a machine learning model that is not only *accurate* but also *highly explainable*—capable of providing transparent justifications for each estimate it produces.

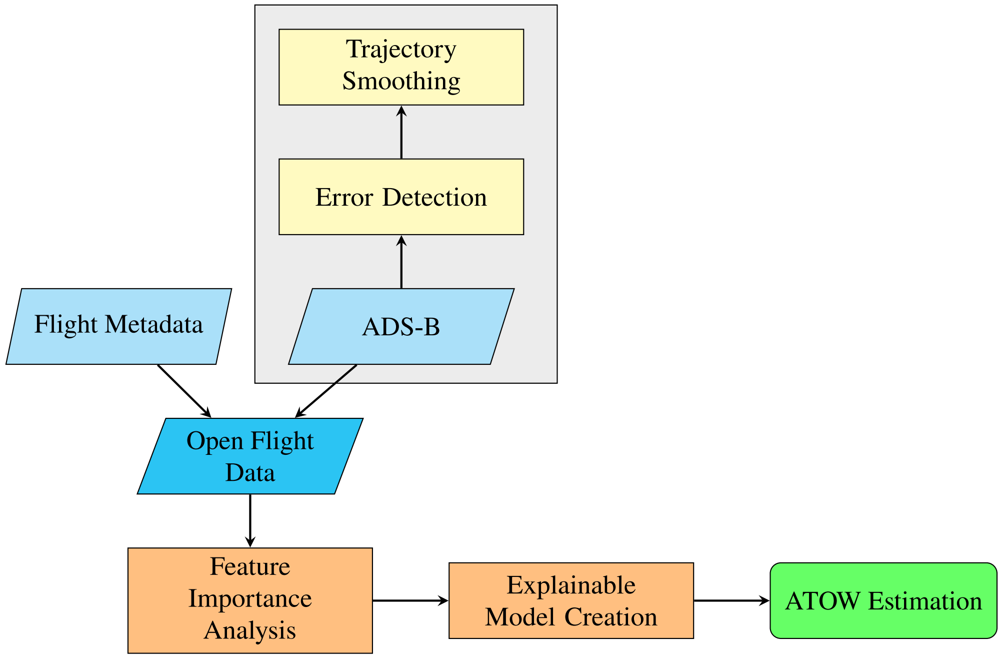

Fig. 1 Overview of the Three-Pronged Approach to Estimating Actual Takeoff Weight (ATOW).

Our long-term objective is to develop an *ATOW estimation model* that achieves a balance between predictive accuracy, robustness to real-world data imperfections, and interpretability suitable for aviation stakeholders. By combining rigorous feature analysis, data quality control, and interpretable machine learning, we hope to contribute a novel and practically viable approach to this critical problem in aviation analytics.

II. Preliminary Results

This section presents a summary of the preliminary results obtained thus far.

A. Data

Two interconnected datasets have been utilized in this study. The first dataset, hereafter referred to as Flight Metadata (as depicted in the methodology diagram above), contains general descriptive information about each flight, including:

- Flight Identification Number
- Date
- Callsign
- Departure Information (Airport ICAO code, Airport Name, Country, City)
- Arrival Information (Airport ICAO code, Airport Name, Country, City)
- Departure Time
- Arrival Time
- Aircraft Type Code
- Wake Turbolence Category (WTO)
- Aircraft Operator Code
- Flight Duration
- Taxi-Out Time
- Route Length
- Actual Takeoff Weight (TOW) (*ground truth*)

The second dataset comprises trajectory data derived from ADS-B position reports, collected and processed by the *OpenSky Network (OSN)* for each flight in the dataset. The trajectory data, recorded at one-second intervals, includes:

- Flight Identification Number
- Longitude
- Latitude
- Altitude
- Timestamp
- Ground speed
- Track Angle
- Vertical Rate of Climb/Descent
- Wind (u component of wind and v component of wind)
- Temperature

From a total of 369,013 flights provided by *Eurocontrol's Performance Review Commission (PRC)*, this study focuses on a subset of 19,929 flights from January 2022. This targeted selection was made to ensure a controlled and consistent dataset characterized by relatively stable seasonal and operational conditions. Focusing on a single month reduces variability caused by seasonal weather changes and operational disruptions, thereby allowing the model to better capture and explain the core relationships between flight parameters and takeoff weight without confounding factors. This focused approach supports the development of a more interpretable and explainable machine learning model, as it minimizes extraneous complexity that could obscure the underlying data patterns critical to accurate and transparent ATOW estimation.

B. Statistical Dependency Between Candidate Features and TOW

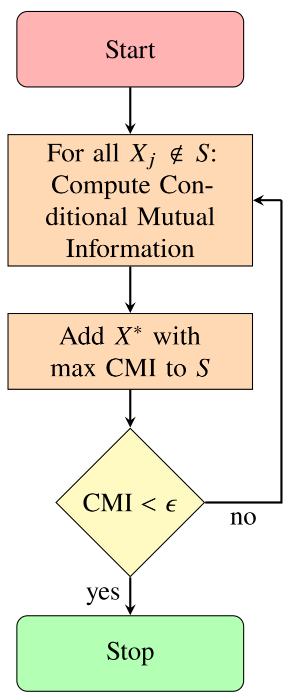

Fig. 2 Conditional Mutual Information Maximization (CMIM) Algorithm.

To identify the most informative predictors of ATOW, we conducted a rigorous analysis using Mutual Information (MI) and Conditional Mutual Information (CMI) across the set of candidate features from the Flight Metadata dataset. Given the continuous nature of TOW, we *discretized* it using three binning strategies: *equal-length*, *quantile-based*, and *clustering-based* (K-Means, Hierarchical, DBSCAN). Among these, quantile-based binning with 200 bins yielded the highest MI score (5.2640), outperforming equal-length (3.9362) and clustering-based (2.8887) approaches. This finding underscores the importance of *distribution-aware discretization methods* in entropy-based feature selection tasks ([8]).

Following binning, we evaluated the statistical dependency between each feature and ATOW using the *Conditional Mutual Information Maximization (CMIM)* algorithm introduced in [9] and depicted in the flowchart in 2. The analysis prioritized features that offered the highest information gain about ATOW, conditioned on already-selected features. Initially, mutual information was computed between each candidate feature and the quantile-binned ATOW. The top-scoring variable—aircraft type—was selected as the first feature due to its strong unconditioned dependency with ATOW, which aligns with domain knowledge about structural and operational limits associated with aircraft models.

Subsequent iterations calculated the CMI of remaining features, conditioning on the progressively expanding set of selected variables. The progression included:

1) Aircraft type
2) Destination airport
3) Day of the year
4) Flight duration
5) Departure time

As demonstrated in the conditional mutual information decay curve, the marginal information gain sharply diminished after the fifth feature, with CMI values falling below the threshold $\epsilon < 0.09$. This result corroborates the cumulative scree plot from *Singular Value Decomposition (SVD)*, which showed that the first four components explain nearly all of the variance in the input matrix—suggesting that dimensionality can be effectively reduced without significant information loss.

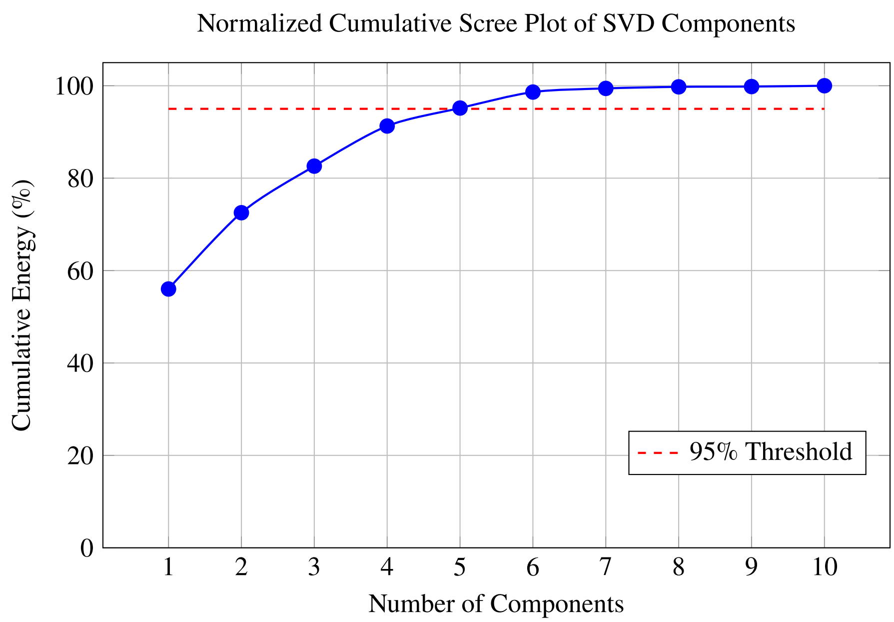

Fig. 3 Normalized Cumulative Scree Plot of Singular Values from SVD of TOW Dataset

The practical implication of these findings is twofold. First, it enables us to develop more *interpretable* and *computationally efficient* predictive models by reducing the input feature space. Second, it confirms that CMI is a robust criterion for *hierarchical feature selection*, particularly in systems where *multicollinearity* and *redundancy* are prevalent [10]. The method's robustness stems from its ability to isolate the unique contribution of each variable while adjusting for interdependencies.

This feature selection pipeline, integrating CMI with SVD-based validation, offers a statistically grounded framework for input variable reduction in aviation-related prediction tasks. These findings serve not only for predicting ATOW, but also as a foundation for future downstream modeling efforts that seek to optimize aircraft loading strategies in real-world operational settings.

C. Detection and Correction of Corrupted or Implausible ADS-B Measurements

Accurate feature extraction from ADS-B trajectories requires, as showed in 4, rigorous preprocessing to eliminate corrupted or implausible data points, particularly given the high sensitivity of ATOW prediction to trajectory-derived kinematic variables. Our approach integrates established smoothing and filtering methods with additional flight-phase-aware segmentation and transformation processes to robustly handle raw ADS-B data.

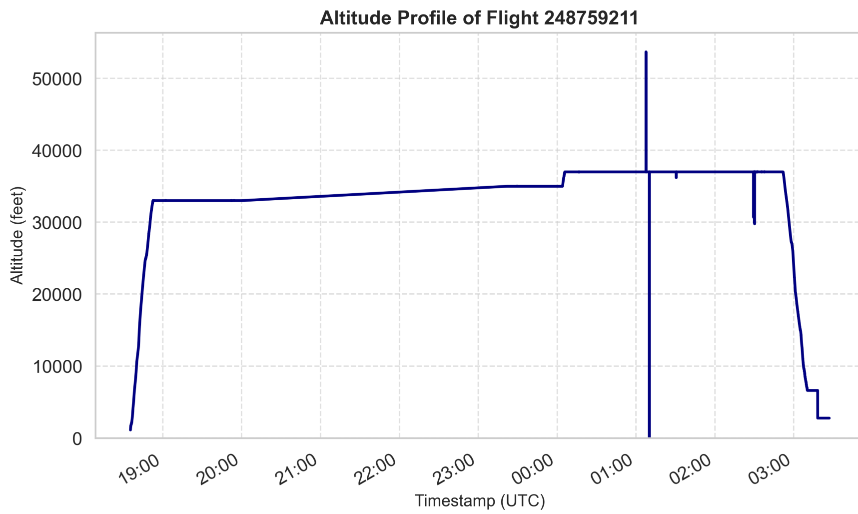

Fig. 4 Unfiltered ADS-B altitude data showing discontinuities and spurious variations.

The raw trajectory signals—comprised of positional, kinematic, and aerodynamic measurements—are first subjected to a multistage filtering framework designed to detect and remove observations that violate temporal continuity or physical realism. Measurements are flagged as redundant when multiple consecutive records exhibit invariant values across critical variables (e.g., latitude, longitude, altitude), suggesting either broadcast repetition or sensor staleness. These are discarded using a composite differencing criterion evaluated across successive time steps.

In addition, points are excluded if they are temporally isolated, defined as those separated from neighboring records by intervals exceeding a 20-second threshold. This prevents extrapolation across large temporal discontinuities. Further, measurements exhibiting anomalous first or second derivatives in positional or kinematic dimensions—indicating physically implausible accelerations or abrupt transitions—are also eliminated. This is determined by evaluating velocity and acceleration bounds grounded in known aircraft performance envelopes.

To reduce high-frequency noise without distorting valid physical dynamics, the retained trajectories are smoothed using spline-based techniques constrained by time-based proximity and local measurement density. This step is essential to maintain the fidelity of higher-order features such as vertical speed and energy rate, which are sensitive to minor fluctuations in raw altitude or groundspeed signals.

Observations where the reported geometric altitude falls outside the operational bounds of civil aircraft—specifically, below 0 feet or above 42,000 feet—are systematically removed. These thresholds are derived from operational regulations and verified through comparative analysis with domain literature on commercial flight envelopes.

To enable robust cross-comparison across heterogeneous flight durations, each trajectory is normalized onto a unit time scale based on actual airborne time. The normalized trajectory is then divided into ten consecutive, equal-duration partitions, each representing a decile of the total flight time. This segmentation allows for consistent extraction of phase-aware statistical features, independent of absolute flight length.

Within each of the ten temporal partitions, a comprehensive set of statistical features is computed to characterize the spatial and dynamic behavior of the aircraft:

- *Temporal Density*: For each partition, the total time spent (in seconds) is computed, providing a measure of data completeness and partition fidelity.
- *Spatial Statistics*: Minimum, maximum, and mean values of latitude, longitude, and altitude are recorded in each partition, yielding 90 spatial descriptors. These features are essential for capturing localized flight dynamics and identifying outlier trajectories.
- *Directional Speed Estimation*: The average speed in each cardinal direction is derived using a transformation

from geodetic coordinates to linear distances. Specifically, for latitude and longitude, angular displacements are converted to linear feet using the arc length formula, then normalized by the time duration of the partition. Altitude-based speed is computed as the vertical range divided by the same time duration.

- *Composite Speed Vector*: The total speed magnitude is calculated as the Euclidean norm of the three directional speed components, yielding a 3D velocity approximation that reflects total kinetic activity within each flight segment.

This set of 140 partition-based features provides a high-resolution temporal and spatial characterization of the trajectory, enabling the downstream modeling process to distinguish between typical and anomalous flight behaviors with respect to TOW-relevant performance indicators.

D. Expected Outcomes

We expect that the integration of information-theoretic feature selection—grounded in Mutual Information (MI) and Conditional Mutual Information (CMI)—will yield a model that not only achieves high predictive accuracy for ATOW, but also offers clear interpretability regarding the influence of individual flight parameters. By selecting a compact yet highly informative subset of features, our approach aims to mitigate overfitting, reduce model complexity, and enhance generalization across varied operational conditions. Early-stage results suggest that such dimensionality reduction does not degrade model performance; instead, it may improve it by eliminating redundant and noisy predictors. The anticipated outcome is a robust and explainable machine learning model that quantifies the relative contribution of each selected variable, thereby providing actionable insights for aviation practitioners and supporting transparent, data-driven decision-making in aircraft loading strategies. Given the maturity of our preprocessing and feature engineering pipeline, model development is already underway and will be completed in time for the final manuscript submission.

Acknowledgments

This work was supported by the Undergraduate Research Institute (URI) at Embry-Riddle Aeronautical University - Prescott [grant number: CC43121].

The authors would like to thank the Eurocontrol's Performance Review Commission (PRC) for providing information for 369013 flights flown all over Europe in 2022.

References

[1] Lee, H.-t., and Chatterji, G., "Closed-Form Takeoff Weight Estimation Model for Air Transportation Simulation," 2010. https://doi.org/10.2514/6.2010-9156.

[2] Alligier, R., Gianazza, D., and Durand, N., "Ground-based Estimation of the Aircraft Mass, Adaptive vs. Least Squares Method," 2013, pp. pp 1–10.

[3] Schultz, C., Thipphavong, D., and Erzberger, H., "Adaptive Trajectory Prediction Algorithm for Climbing Flights," 2012. https://doi.org/10.2514/6.2012-4931.

[4] Chati, Y., and Balakrishnan, H., "Statistical modeling of aircraft takeoff weight," 2017.

[5] Sun, J., Ellerbroek, J., and Hoekstra, J., "Aircraft initial mass estimation using Bayesian inference method," *Transportation Research Part C: Emerging Technologies*, Vol. 90, 2018, pp. 59 – 73. https://doi.org/10.1016/j.trc.2018.02.022.

[6] Murça, M., Maximo, M., Dantas, J., Szenczuk, J., Rutili de Lima, C., Carvalho, L., and Melo, G., "Open Machine Learning Models for Actual Takeoff Weight Prediction," *Journal of Open Aviation Science*, Vol. 3, 2025. https://doi.org/10.59490/joas.2025.7963.

[7] Daga, T., Bhanpato, J., Behere, A., and Mavris, D., "Aircraft Takeoff Weight Estimation for Real-World Flight Trajectory Data Using CNN-LSTM," 2024. https://doi.org/10.2514/6.2024-4291.

[8] Peng, H., Fuhu, L., and Chris, D., "Feature Selection Based on Mutual Information: Criteria of Max-Dependency," *Max-Relevance and Min-Redundancy J. Trans. on PAT A M/C Intell*, Vol. 27, 2005, pp. 1226–1238.

[9] Fleuret, F., "Fast Binary Feature Selection with Conditional Mutual Information," *Journal of Machine Learning Research*, Vol. 5, 2004, pp. 1531–1555.

[10] Brown, G., Pocock, A., Zhao, M.-J., and Luján, M., "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection," *The Journal of Machine Learning Research*, Vol. 13, 2012, pp. 27–66.