Aircraft Take-off Weight Prediction with Operational Data and Supervised Learning

Gheorghe, A.I.; Sun, Junzi; Ribeiro, M.J.; Hop, Pascal; Cramet, Benjamin

Publication date
2024

Document Version
Final published version

Published in
14th SESAR Innovation Days, SIDS 2024

Aircraft Take-off Weight Prediction with Operational Data and Supervised Learning

Andrada Ioana Gheorghe, Junzi Sun, Marta Ribeiro
Air Transport & Operations, Faculty of Aerospace Engineering
Delft University of Technology, The Netherlands

Pascal Hop, Benjamin Cramet
EUROCONTROL
Brussels, Belgium

***Abstract*—Predicting aircraft Take-Off Weight (TOW) has been a long-standing goal for aviation stakeholders, especially for operational and regulatory bodies involved in flight planning. Accurate TOW values would enable better emissions computation, leading to more effective regulation of aviation's climate impact. However, aircraft operators prefer to keep TOWs confidential because they are sensitive to operational trends and cost indices. Consequently, many works have attempted to circumvent this gap by predicting TOW values. Unfortunately, limited success has been achieved primarily due to the lack of accurate real-world operational data. This study is unique in utilizing operational TOW data provided by airlines. We predict TOW before take-off based solely on Flight Plan and Terminal Aerodrome Forecast parameters, primarily focusing on flights at Amsterdam Airport Schiphol. The accuracy of several Machine Learning algorithms is directly compared. The best Mean Absolute Percentage Error of 2.17% on the Schiphol testing dataset is achieved. The model is further validated on flights at Paris - Charles de Gaulle Airport and Brussels South Charleroi Airport with errors of 4.07% and 3.41%. We found that the distribution of flights in the training dataset, particularly aircraft and airline types, significantly influenced the model's applicability. Recommendations are also made on how to improve the model further.**

***Keywords*—Aircraft Take-off Weight, Supervised Learning, Flight Plan, Terminal Aerodrome Forecast**

I. Introduction

The prediction of aircraft Take-Off Weight (TOW) has been a difficult problem to solve for many aviation stakeholders. More than just a safety-critical parameter for take-off performance, TOW impacts fuel consumption and plays an important role in trajectory prediction computations, especially during the climb phase. Most operational and regulatory organisations involved in flight planning and network operations aim to improve their flight planning and emissions calculations before take-off. However, without TOW data, the accuracy of such predictions cannot be guaranteed. Unfortunately, aircraft operators are generally not willing to share this data as it is used to calculate their cost index. The belief is that TOW may reveal sensitive information about airlines' operational trends, making them vulnerable to market competition or even penalties. However, predicting aircraft TOW could enable aviation authorities to better compute emissions and other climate-oriented parameters, improving regulations on aviation's climate impact.

The current state of the art, for the most part, has studied the estimation of aircraft TOW using supervised Machine Learning (ML) algorithms. However, these are highly dependent on data quality, quantity, and selection. Thus, the scarcity of TOW data makes training an ML algorithm a challenging task. For this reason, previous studies have relied on trajectory data - mostly sourced from Automatic Dependent Surveillance-Broadcast (ADS-B) - to build a training dataset by reverse engineering trajectories with a total energy model. However, these approaches often introduce a sequence of mass estimations for the climb profile, potentially leading to propagating errors. Additionally, synthetic data is often used to introduce certain assumptions into the data. Finally, and most notably, all studies involved post-flight computations, which is not practical for flight planning and operational applications prior to take-off.

This work aims to use Flight Plans (FPLs) and perform TOW predictions solely based on operational parameters known to air traffic controllers before take-off. Airlines fill their *operational* TOW in the FPL. As such, the data used in this study is the closest to real TOW data and provides the best achievable accuracy available for operations. The use of FPLs, provided by EUROCONTROL, captures airline preferences and enables a (pre-)tactical prediction horizon that is one to seven days prior to take-off, including the day of operations.

This paper is structured as follows. Section II highlights the main take-away points from previous studies, including potential research gaps covered by this work. Section III details the methodology of the developed model, including the ML algorithms and features selected. Next, Section IV describes the case studies and data selection procedure. The findings of the analysis are discussed in Section V, together with results from two validation activities treating the model's applicability. Finally, several points for improvement are discussed in Section VI, followed by the conclusions of the study in Section VII.

II. Literature Review

A crucial consideration regarding previous research is the lack or scarcity of operational TOW data. For this reason, previous studies have attempted to deduce aircraft mass via analytical calculations, focusing on estimating the parameter *after* the flight has taken place. The computations are usually based on flight trajectory data such as ADS-B or radar Correlated Position Reports (CPRs). Works such as [1]–[6]

have made use of The OpenSky Network [7], an open-source platform providing real-time and historical ADS-B data for research and academia, while [8], [9] based their work on Quick Access Recorder (QAR) data. The latter is an airborne flight data recorder designed to provide raw flight data and is mainly used by aircraft operators for routine monitoring of their fleet and flight crew [10], [11]. These data sources introduce constraints to the models' accuracy.

Approaching the problem backwards involves building training datasets containing *synthetic* TOW data, before applying ML methodologies. Following the sequential nature of the available data (i.e. trajectories), the current state-of-the-art approach opts for reverse engineering a sequence of aircraft masses using a total energy model, generally over the climb profile. These methods adjust the mass to fit observed values of energy variation. Note that although [3]–[6] used statistical methods instead of ML, they still take this approach for TOW and mass estimations. Not only does this introduce assumptions and errors, but it also limits the models' capability of estimating TOW to a post-operations time frame, having mostly trajectory parameters as input. This restricts the prediction horizon and hinders (pre-)tactical prediction capabilities. At most, the predictions are computed using past trajectory points and with a 10-minute prediction horizon [1].

Although the reverse engineering step is no longer needed when using QAR data, the prediction time frame issue persists. This is due to the capability of the flight data recorder itself, which provides real aircraft mass data at each point along the trajectory, yet only when the aircraft is airborne. Additionally, building a model on QAR data introduces limitations to its applicability due to the origin of such datasets. Certain studies [8], [9] use QAR data gathered from two airlines respectively; so although the data quality is improved, the predictions become airline-specific. While QAR data is not limited to airlines, it is typically regulated due to privacy and security, limiting its variability on aircraft types, airlines, and origin/destination pairs.

In previous research, there is a clear lack of FPL integration in the studies, as well as long-term prediction capabilities extended to at least a few hours before take-off. To the best of the author's knowledge, no successful study has been conducted to predict TOW solely based on FPL data with a (pre-)tactical prediction horizon. While [2], [12] have used some FPL information, they either did not have access to the entirety of the dataset or simply did not use it as training features for their models. Omitting FPLs removes the airlines' preferences from the analysis and makes the results purely trajectory-based without having intent or route planning information. Furthermore, it has no added value for Air Traffic Management (ATM) authorities, as the predictions cannot be applied prior to the flight's execution.

This study aims to better incorporate these features and increase the prediction horizon that is best suited for ATM flight planning applications. The goal of this research is to predict TOW *before* take-off, hence with parameters available in the FPL itself as well as Terminal Area (Aerodrome) Forecast (TAF) at the airport of destination to include weather impact. Furthermore, this paper focusses on the use of ML - two main algorithms will be explored: Gradient Boosting Decision Trees (GBDTs) and Random Forests. These proved to be the most effective and least error-inducing algorithms for predicting single variables.

III. Methodology

An overview of the steps taken and datasets used is shown in Figure 1, where each white box represents one dataset. In ML, data plays a crucial role in defining the capability and applicability of the model to predict the selected target parameter(s). The data considered for this study is described in Section III-A. Regarding the ML algorithms, several have been tested in order to select the best-performing one for this particular application. Their selection, as well as their basic working principle, are detailed in Section III-B.

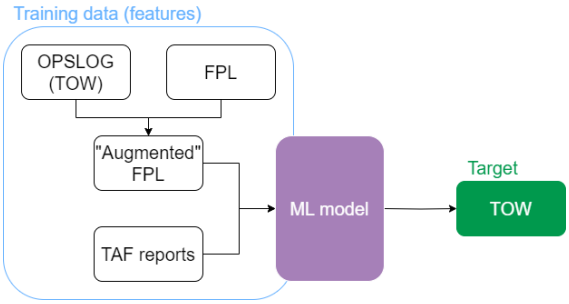

Figure 1. Methodology overview flowchart.

A. Datasets & Features Selection

This study has gathered data from different sources, as shown in Figure 1, all provided by EUROCONTROL. There are three main datasets: The Operational Logbook (OPSLOG), FPL, and TAF, each containing information for one flight prior to take-off. OPSLOG contains all the information about the flight in question, which is necessary for operational personnel in the execution of their duties. The aircraft TOW is an *optional* parameter which *can* be filed in the FPL, depending on the airline's willingness to share this data. Statistically, about 30% of the flights pertaining to EUROCONTROL's network have a TOW associated with their FPL. This corresponds to circa 3M flights in 2023 [13].

All training features are listed in Table I, together with the data type and unit used, when applicable, and the dataset they were extracted from. Note that some of these features are not taken directly from the dataset and were altered for different reasons. First, the departure, destination, and alternate aerodromes were not considered as categorical features themselves. Instead, for generalisation of the model and to make it independent of International Civil Aviation Organisation (ICAO) airport codes, the great circle distance between the aerodromes of departure and destination and between the destination and alternate aerodromes were used. Another altered parameter is the route available in the FPL and OPSLOG. Instead of considering the route itself, different

TABLE I. DESCRIPTION OF FEATURES USED FOR TRAINING.

Dataset	Feature	Description	Type	Units	Encoding
OPSLOG	great_circle_distance_ADEP_ADES	great circle distance between aerodromes of departure and destination	numerical	km	-
	great_circle_distance_ADES_ALTRNT1	great circle distance between aerodromes of destination and alternate	numerical	km	-
	AOARCID	aircraft operating agency ICAO ID	categorical	-	ordinal
	ARCTYP	aircraft type ICAO ID	categorical	-	ordinal
	EOBT	estimated off-block time	numerical	-	datetime cyclical
	TAXITIME	taxi time (taxi before take-off)	numerical	s	-
	TTLEET	total estimated elapsed time (flight duration)	numerical	min	-
	RFL	requested flight level	numerical	FL	-
	SPEED	requested speed	numerical	kts	-
FPL	flt_rvr_val	runway visibility range	numerical	m	-
	airac_cycl	AIRAC cycle	numerical	-	-
	flt_etot	estimated take-off time	numerical	-	datetime cyclical
	flt_eta	estimated time of arrival	numerical	-	datetime cyclical
	flt_f_rte_len	length of the route	numerical	nm	-
TAF	visibility_cavok	clouds and visibility ok	categorical	-	ordinal
	visibility_distance	visibility distance	numerical	m	-
	clouds_height	clouds ceiling height value	numerical	m	-
	clouds_amount	clouds amount	numerical	-	-
	wind_speed	mean wind speed	numerical	m/s	-
	wind_gust	wind gust speed	numerical	m/s	-
	wind_compass	mean wind direction	categorical	-	ordinal
	time	time of TAF report creation	numerical	-	-
	validity_start_time	start time of TAF report validity	numerical	-	-
	validity_end_time	end time of TAF report validity	numerical	-	-
	precipitation	presence of precipitation	categorical	-	one-hot
	obscuration	presence of obscuration	categorical	-	one-hot
	other	presence of extreme weather events (tornado, volcanic ash, etc.)	categorical	-	one-hot
	thunderstorms	presence of thunderstorms	categorical	-	one-hot
	freezing	presence of freezing	categorical	-	one-hot
	snow	presence of snow	categorical	-	one-hot
	clouds	presence of clouds	categorical	-	one-hot
	indicator	trend forecasts indicator	categorical	-	ordinal
	probability	trend forecasts associated probability	numerical	%	-

features were extracted from it, specifically the requested speed and flight level for the cruise. These will impact aircraft performance and are also linked to TOW. Note that adding the Standard Instrument Departure (SID) and Standard Arrival Route (STAR) was also considered. However, these are airport-dependent and may lead to overfitting or bias in the model predictions. Therefore, they have been left out of the features. Finally, it is important to state that when computing a flight's Total Estimated Elapsed Time (TTLEET), the airlines make use of weather predictions along the route, especially regarding head or tail winds. These may have a significant impact on the flight duration, so although cruise weather forecasts are not taken into consideration as separate features in this study, they are still accounted for via this feature.

B. Machine Learning Algorithms

A total of four algorithms were selected based on literature findings: Gradient Boosting Decision Trees (GBDTs) [14], LightGBM [15], Gradient Boosting Regressor (XGBoost) [16], and Random Forests [17]. Regardless of the algorithm, the goal in ML regression problems is to predict a target variable, in this case the aircraft TOW, from a vector of features listed in Table I. Table II lists the different hyperparameter values explored for each algorithm presented in this section. When the value is empty ('-'), the corresponding hyperparameter is not applicable to the algorithm in question.

IV. DESCRIPTION OF THE CASE STUDIES

Section IV-A presents the main case study along with the aircraft types considered. Finally, the validation datasets are detailed in Section IV-B.

A. Amsterdam Airport Schiphol (AMS)

The model first focuses on the flights departing and arriving at Amsterdam Airport Schiphol (AMS) to simplify the analysis and provide a first understanding of the results at one airport. In this way, potential lagging aspects of the model could be identified, especially regarding features considered for training. AMS is a large airport in terms of traffic volumes and passengers carried. Furthermore, it generally accommodates legacy carriers traffic such as KLM and Air France while also having a wide range of low cost aircraft operators, hence it provides a good mix of traffic types. Only the flights with the TOW information in their FPLs will be considered

TABLE II. HYPERPARAMETERS SEARCH SPACE FOR USED ALGORITHMS.

	Boosting stages	Max num trees	Number trees	Max depth tree	Learning rate	Max tree leaves	Min samples/leaf	Early stop rounds
BDTs	100 500 1000 2500 5000 7500	-	-	3	0.001 0.01 0.1 0.2	-	-	-
XGBoost	-	500	-	3	0.2	-	-	4
LightGBM	-	75 100 150 200 500	-	unconstrained	0.001 0.01 0.1 0.2	31	10 30	4
Random Forest	-	-	100 500 1000 2500 5000	6 17	-		10	-

since the latter provides the target output value for each flight. Furthermore, only those Scheduled (S) and following Instrument Flight Rules (IFR) are considered for these flights. Regarding time range, the oldest FPL in EUROCONTROL's database with associated TOW dates back to February 2022. For this reason, all flights scheduled starting February 2022 and up to the end of December 2023 are considered in this research, amounting to 122,379 flights at AMS. These were split using an 80-20% ratio between training and testing datasets, resulting in 97,639 and 24,740 flights, respectively. Note that this was not done randomly. The train-test split was conducted on a daily basis to guarantee robust training and to not omit potential cyclical patterns hidden behind 'datetime' features.

The distribution of flights across aircraft operators and aircraft types is shown in Figures 2 and 3 for the training set, respectively. Note that in Figure 2 only the top 10 airlines with the highest amount of flights are plotted. The main aircraft operator present in the training dataset is a low-cost carrier, precisely easyJet (EJU and EZY), amounting to almost 50% of the flights. The next airline in terms of flight count is Lufthansa, with almost 10% of flights from the training dataset, followed by TUI fly in The Netherlands with circa 7%. Interestingly, the second airline is a legacy carrier, while the third is a charter airline, giving a good variability for the training data despite a large number of low-cost carrier flights. Regarding Figure 3, almost 90% of most-flown aircraft types in the training dataset are classified as medium-range and narrow-body. As a consequence, the algorithms may have better accuracy for this type of aircraft.

B. Validation Datasets

While the ML model is built on AMS data, two more airports are considered for validation purposes: Paris - Charles de Gaulle Airport (CDG) and Brussels South Charleroi Airport (CRL). CDG was selected to test the trained model on another airport with similar traffic volumes and size, both being major international hubs and some of the busiest airports in Europe. Furthermore, both airports support a majority of legacy carrier operations, with KLM at AMS and Air France at CDG. On the other hand, CRL was chosen for its difference in size and operated flights, in order to analyse the model's applicability to a completely different traffic mix. CRL is known for its low-cost carrier operations, and Ryanair is one of the most important players.

These datasets contain flights departing from and arriving at the airports in question, for which TOW data is available. The CDG and CRL datasets amount to 320,032 flights and 54,788 flights, respectively. The distributions of aircraft operators are shown in Figures 4 and 5 for CDG and CRL, respectively. For CRL, Ryanair amounts to more than 80% of the flights. On the other hand, Figure 4 shows a promising distribution for CDG. As almost 70% of the flights are operated by Air France, a legacy carrier, this dataset will serve as a good baseline for validation, since most of the flights in the training dataset are

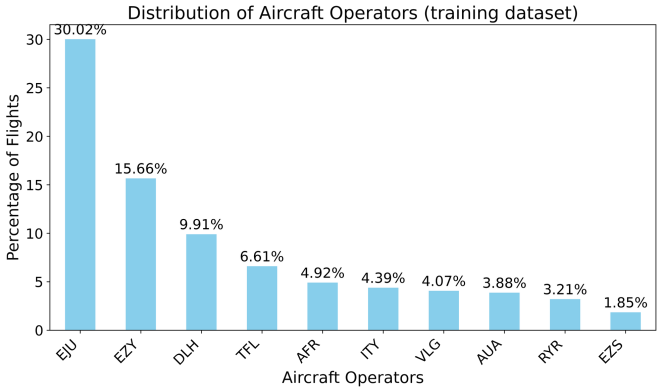

Figure 2. Distribution of aircraft operators - training dataset (AMS).

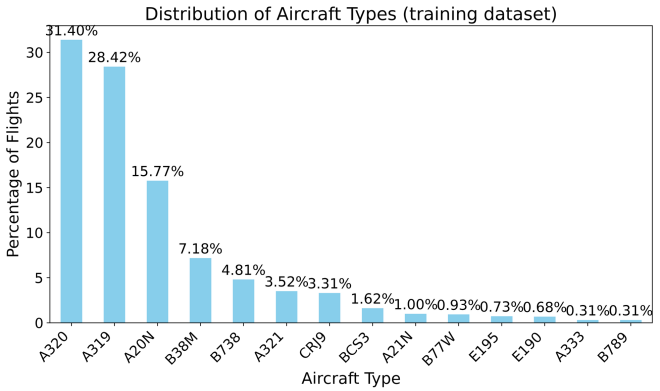

Figure 3. Distribution of aircraft types - training dataset (AMS).

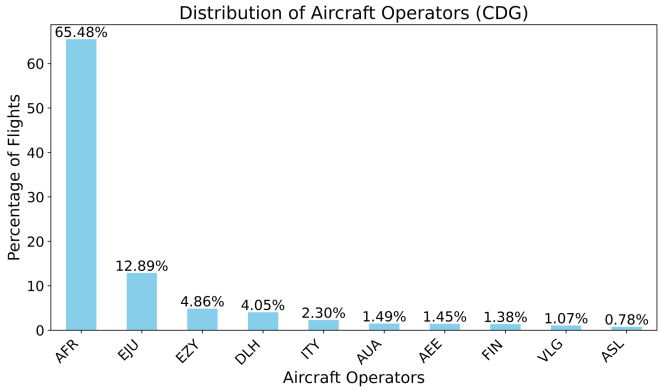

Figure 4. Distribution of aircraft operators - validation dataset (CDG).

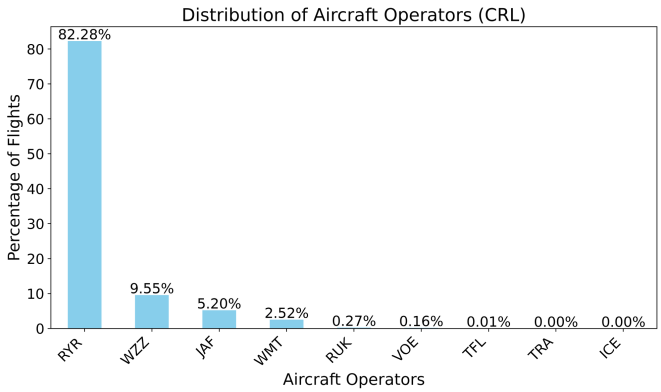

Figure 5. Distribution of aircraft operators - validation dataset (CRL).

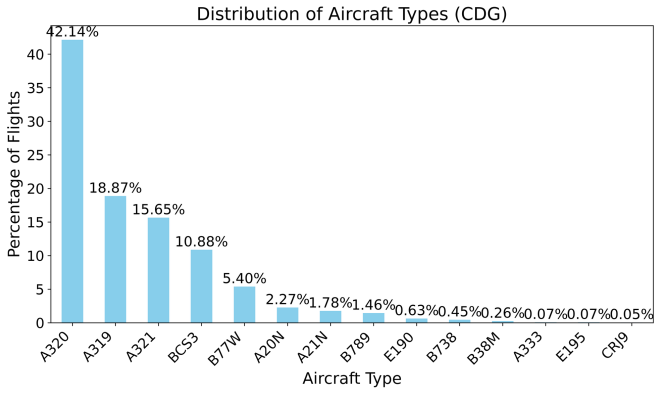

Figure 6. Distribution of aircraft types - validation dataset (CDG).

operated by a low-cost carrier (easyJet).

Finally, the most-flown aircraft types of each validation dataset are given in Figures 6 and 7. Although the aircraft type distribution of CDG may be similar to the training dataset, that of CRL is not. More than 70% of flown aircraft are B738, while these correspond to less than 5% in Figure 3. Consequently, CRL will serve as a good baseline for validation regarding the aircraft types feature, in the same way that CDG serves as a good validation baseline for aircraft operators.

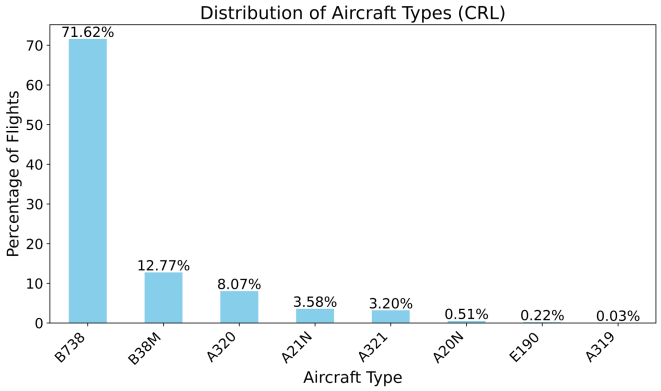

Figure 7. Distribution of aircraft types - validation dataset (CRL).

V. Results

This section presents the results of the analysis, starting with the selection of the most optimal ML algorithm trained with AMS data in Section V-A. Section V-B discusses the applicability of the best-performing model on CDG and CRL airports, as was described in Section IV-B.

A. AMS Case Study

The ML algorithms presented in Section III-B were trained with the same dataset from AMS, where a train-test split of 80-20% was followed. After training all the algorithms with 97,639 flights departing from and arriving at AMS, their performance could be analysed based on the testing dataset with 24,740 flights. Different error metrics were used to determine which ML algorithm performed best. These are listed in Table III along with the corresponding results. XGBoost and LightGBM greatly reduce training time compared to Random Forest and GBDTs. Nevertheless, the latter outperforms in terms of Mean Absolute Percentage Error (MAPE), the selected metric for this study. Based on this, the GBDTs model was selected for further analysis and validation activities. Therefore, from now on, prediction results and further details all refer to the GBDTs model.

The scatter plot of the regression achieved with GBDTs is shown in Figure 8, where the actual TOW values from the testing dataset are given on the vertical axis and the TOW predictions generated by the model are given on the horizontal axis. This graph gives a good illustration of the high R^2 score, with all data points located very close to the regression line and very few outliers.

The error distribution of the predictions done with the testing dataset are given in Figure 9. This plot defines the error

TABLE III. Error metrics overview across ML algorithms. Testing dataset used as reference.

Algorithm	Training time	MAPE (%)	MAE (kg)	R^2 score
XGBoost	57s	2.59	1,629	0.9877
Random Forest	6h 32m 30s	2.38	1,503	0.9851
GBDTs	12h 51m 36s	2.17	1,376	0.9907
LightGBM	4m 43s	2.18	1,373	0.9913

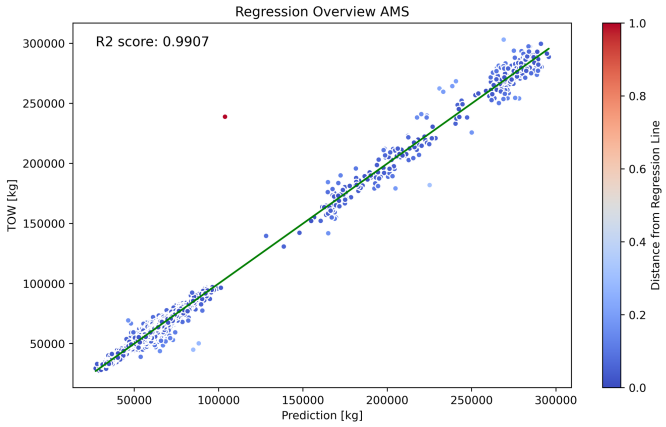

Figure 8. Scatter plot of GBDTs algorithm, testing dataset.

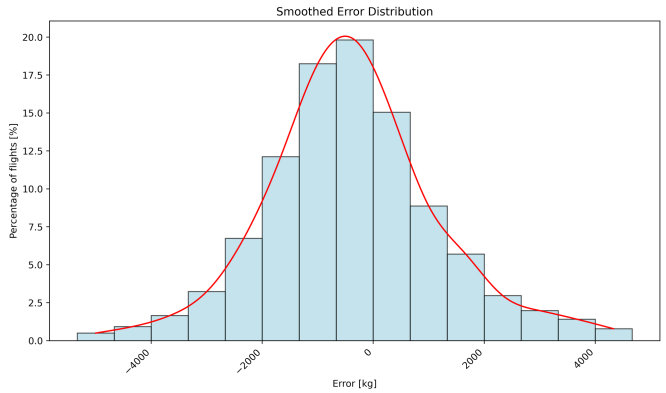

Figure 9. Error distribution of the testing dataset.

as the difference between predicted and actual values. The curve nicely depicts a normal distribution of the errors around 0, with a limited spread. The comparison between minimum and maximum errors suggests that the model may tend to underestimate the predictions. Indeed, the minimum TOW prediction of -4.7 tonnes is more than one tonne (absolute value) over the maximum of 3.8 tonnes. Overall, it was found that 53.19% of flights' TOWs were underestimated, and the remaining 46.81% were overestimated, which is not alarming.

TABLE IV. MAPE GROUPED BY AIRCRAFT TYPE, COMPARISON BETWEEN TRAINING AND TESTING DATASETS.

Aircraft Type (ICAO)	MAPE (%) - testing	MAPE (%) - training
A320	2.40	1.93
A319	1.83	1.56
A20N	2.04	1.63
B38M	2.23	1.52
B738	2.07	1.43
A321	2.52	2.11
CRJ9	2.18	1.72
BCS3	2.48	2.09
A21N	2.20	1.19
B77W	2.57	0.63
E195	3.33	2.25
E190	3.98	2.38
A333	3.15	1.21
B789	3.14	0.64

Finally, Table IV lists the average MAPE of the model for each aircraft type and for both training and testing dataset results. Comparing this table with the aircraft distribution of the training dataset from Figure 3, the results are consistent. The more the model is trained with a specific aircraft type, the better it predicts the TOW for that aircraft type. For example, E190 and E195 (together) account for circa 1.5% of flights in the training dataset, and they also have the highest MAPE.

The results of the feature importance analysis are given in Table V for the top eight most-used and influencing features during training. Note that the importance values are solely given to the best-performing model, namely GBDTs. The model's output is essentially dictated by the top three training features, that is requested speed, great circle distance between aerodromes of departure and destination, and aircraft type. The requested speed in cruise is the parameter which has the highest influence on TOW predictions for this case study. This can be surprising, as one may tend to hypothesise that great circle distance between airports of departure and destination could have more influence on TOW due to fuel carried. However, the requested speed at cruise affects the fuel consumption of the aircraft, so depending on this value, more or less fuel will be consumed. To reach higher cruise speeds, less fuel may be carried on board, affecting the overall value of TOW. Vice versa, when the requested speed is lower, the aircraft may accommodate a higher TOW. The great circle distance between the airports of departure and destination also influences TOW predictions. This suggests that there is a pattern between airport pairs and the fuel carried onboard to ensure that the aircraft reaches its destination. Finally, the aircraft type flown is an obvious factor, providing the model with a range of TOWs specific to each type.

A Shapley Additive Explanations (SHAP) overview is shown in Figure 10. 2k sample flights were extracted randomly for this calculation to reduce computational effort. SHAP capture the marginal contribution of each feature to the target output (TOW prediction). The top three features are identical, although a different ranking is suggested, with aircraft type being the most influential feature, followed by great circle distance between airports of departure and destination, and requested speed.

B. Extended Applications: CDG & CRL Airports

After performing the verification on both training and testing datasets and analysing the performance of the model, two

TABLE V. FEATURE IMPORTANCE ANALYSIS

Feature	Importance [%]
requested speed	42.38
great circle distance between aerodromes of departure and destination	33.72
aircraft type ICAO ID	19.71
length of the route	1.57
runway visibility range	0.81
requested flight level	0.54
aircraft operating agency ICAO ID	0.45
total estimated elapsed time (flight duration)	0.14

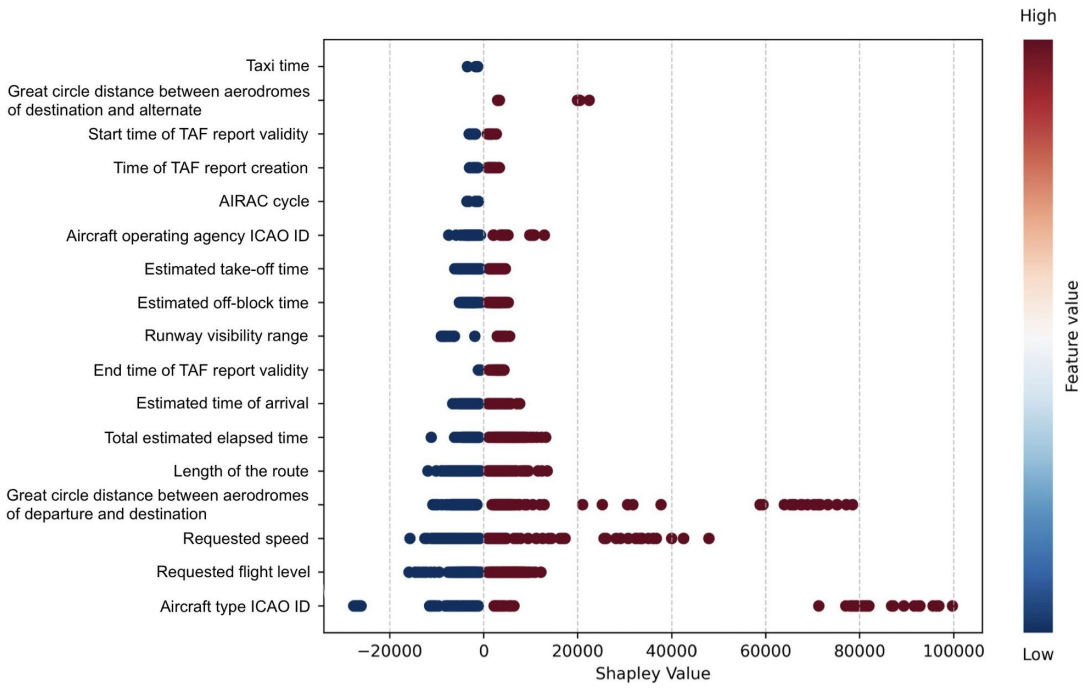

Figure 10. SHAP analysis results.

more datasets were tested for validation purposes. To ensure that the model can be applied to other airports, it was tested on CDG and CRL airports, as explained in Section IV-B. The error metrics are presented in Table VI.

TABLE VI. ERROR METRICS COMPARISON: CDG AND CRL DATASETS.

	MAPE [%]	MAE [kg]	R^2 score
CDG dataset	4.07	4,032	0.9722
CRL dataset	3.41	2,237	0.4344

As expected, the MAPEs for CDG and CRL are higher than that for the testing dataset. Most likely, this is due to the different flights being distributed in the dataset. Based on Table V, and Figures 2, 4, 5, and 10, these values can be explained. As the distribution of the training dataset consists mainly of low-cost carrier flights (easyJet), it makes sense that the errors are smaller for CRL than for CDG. The latter's traffic was mainly operated by Air France, a legacy carrier. Low-cost carriers tend to have lighter aircraft and, therefore, lower TOW values. This is due to limited fuel carried on-board for better aircraft performance and reduced costs, but also due to the constraints in luggage carried by passengers.

The score of the coefficient of determination (R^2 score) is positive for CDG, contrary to CRL. With Ryanair having the highest traffic slice for CRL, the model is exposed to a completely different distribution of the data, with an airline that is barely present in the training dataset. Furthermore, the distribution of aircraft types at CRL in Figure 7 shows that circa 70% of the flights are operated by B738 aircraft, while the training dataset only contains about 4.5% of its traffic with this aircraft type (see Figure 3), suggesting that the target output distribution in the CRL dataset does not match the training data distribution.

The MAPE grouped by aircraft type is given in Table VII for both CDG and CRL datasets. Note that the values missing ('-') for CRL are simply due to the aircraft types not being present in the dataset. The errors can be explained by comparing the distribution of aircraft types across flights in the training dataset, shown in Figure 3, with the same distributions of the validation datasets, given in Figures 6 and 7. The aircraft with which the model has been trained more (those ranked higher in Figure 3) are associated with lower MAPE in Table VII.

TABLE VII. MAPE GROUPED BY AIRCRAFT TYPE, COMPARISON BETWEEN CDG AND CRL DATASETS.

Aircraft Type (ICAO)	MAPE (%) - CDG	MAPE (%) - CRL
A320	3.14	5.33
A319	2.95	10.74
A20N	3.37	5.73
B38M	4.16	3.27
B738	2.77	2.93
A321	3.61	5.44
A333	4.22	-
B77W	11.16	-
B789	5.33	-
BCS3	5.22	-
CRJ9	5.23	-
E195	4.25	-
E190	31.99	36.06
A21N	3.83	5.15

VI. DISCUSSION

This section provides a discussion of the results. First, Section VI-A compares the GBDTs model with some of the

previous studies reviewed in II. Then, several conditions of the model's applicability are provided in Section VI-B along with potential improvements to consider for future work.

A. Comparison with Previous Studies

Work [1] focused on improving aircraft climb prediction by better estimating operational factors, specifically the mass and speed profiles during climb. As there was no access to FPL data, hence TOW, the total energy model was used to reverse engineer the flown trajectories and build a dataset containing aircraft mass sequences. The trajectory data was ADS-B data extracted from The OpenSky Network. A stochastic gradient boosting tree algorithm was trained to predict sequences of aircraft masses. Generally, work [1] proved to achieve lower Root Mean Squared Error (RMSE) per aircraft type despite the synthetic nature of the training data, suggesting that the use of trajectory data plays an important role in the quality of mass predictions. Even though TOW is a static parameter, it is part of a sequence of masses influenced by other trajectory factors, especially during the climb. Consequently, completely discarding trajectory features may not capture the entirety of the picture.

Works [3], [6] used take-off and initial climb ADS-B data to predict TOW. No quantitative error analysis was made due to the lack of validation data. Cruise data on speed and altitude is not used as the model focuses on the initial climb.

Work [9] predicted the initial-climb aircraft mass using a Multi-Layer Perceptron Neural Network (MLPNN) and QAR data. The MLPNN proved to outperform the GBDTs, with merely 0.61% MAPE on the testing dataset compared to 2.17%. Nevertheless, it is important to note that the QAR data used comes from one single airline, which has comparable trends and, especially, uses the same cost index. The reduced variability in training data is expected to lead to the lower encountered MAPE. Nevertheless, it would be worth exploring the capabilities of the MLPNN in the current study.

B. GBDTs Model Applicability and Improvements

A list of conditions for the applicability of the model can be drawn. The GBDTs model was essentially trained with narrow-body medium-range aircraft, with a majority of the flights operated by low-cost carriers, explaining its general tendency of underestimating TOW. The model's behaviour is found to be independent of aerodromes of departure and destination. Therefore, to apply the model to another dataset (e.g. another airport), it is essential to have sufficiently similar distributions of aircraft and aircraft operator types. The latter tend to significantly affect TOW due to luggage and fuel limitations (within safety bounds).

It is expected that incorporating more diverse data, especially in terms of aircraft and airline types, will improve the capabilities of the model and broaden its applicability. The most straightforward approach is increasing the training dataset to the entirety of the European network, for which EUROCONTROL is responsible. However, upsampling techniques could also be explored to synthetically balance the aircraft and airline types in the training dataset. Increasing the latter would enable the testing of neural network algorithms, which show better prediction accuracy for larger datasets.

Regarding TOW data accuracy, the model provides *operational* TOW estimations, which can sometimes deviate from the *actual* TOW depending on actual loading (passengers, fuel, etc.). While the TOW data used in this paper is the closest to *actual* TOW data and provides the best achievable accuracy available for operations, it is important to compare both to assess the precision of *operational* TOW data. However, this would require close collaboration with airlines as only they possess *actual* TOW information, and may quickly become a very demanding task (logistically) for a potentially minimal improvement in the training data.

Finally, because trajectory parameters are not considered among the features, neither the reduced thrust take-off and climb nor the corresponding cost index are captured by the model. For reference, the Flight and Flow Information for a Collaborative Environment (FF-ICE) format is expected to be implemented into operations by the end of 2025. FF-ICE will provide a speed schedule defined by (CAS_1, CAS_2, M), from which a more accurate requested cruise speed could be deduced. Alternatively, the reverse engineering approach could be used on provided FF-ICE climb trajectory predictions, to expand the training dataset to other aircraft operators that do not share TOW data.

VII. Conclusions

This study explored different supervised learning algorithms for the development of an ML-based TOW prediction tool at AMS. The model was trained solely on FPL and TAF data, reaching a MAPE of 2.17%. This proves that (pre-)tactical TOW prediction, solely based on features available prior to take-off, is possible and reliable. Feature importance revealed that the most influencing parameters, in order, were cruise requested speed, great circle distance between aerodromes of departure and destination, and aircraft type. Furthermore, the model proved to be independent of airports of departure and destination in terms of traffic volumes and passengers transported. Additionally, other flight-specific parameters had an impact.

The limitations of the model included its dependence on the distribution of AMS flights. When testing the model on CDG and CRL, similar- and different-sized airports, it was found that aircraft and airline *types* distribution influenced TOW predictions the most, but the size or similarity of the airport itself compared to AMS did not matter. Since the training was essentially conducted with medium-range aircraft and low-cost carriers, these categories showed better prediction accuracy, limiting the model's applicability to the distribution of flight types in the training dataset. Future work should focus on increasing the training dataset to provide a larger coverage of aircraft and airline types. Finally, speed profile and trajectory-based related parameters can potentially improve the current results of the model.

REFERENCES

[1] R. Alligier and D. Gianazza, "Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study," *Transportation research. Part C, Emerging technologies*, vol. 96, pp. 72–95, 2018.

[2] R. Alligier, D. Gianazza, and N. Durand, "Machine learning and mass estimation methods for ground-based aircraft climb prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3138–3149, 2015.

[3] J. Sun, J. Ellerbroek, and J. Hoekstra, "Modeling and inferring aircraft takeoff mass from runway ads-b data," 06 2016.

[4] ——, "Bayesian inference of aircraft initial mass," 06 2017.

[5] J. Sun, J. Ellerbroek, and J. M. Hoekstra, "Aircraft initial mass estimation using bayesian inference method," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 59–73, 2018.

[6] J. Sun, H. A. P. Blom, J. Ellerbroek, and J. M. Hoekstra, "Particle filter for aircraft mass estimation and uncertainty modeling," *Transportation Research Part C: Emerging Technologies*, 2019.

[7] OpenSky. (2018) Publication data.

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

[9] X. He, F. He, X. Zhu, and L. Li, "Data-driven method for estimating aircraft mass from quick access recorder using aircraft dynamics and multilayer perceptron neural network," *ArXiv*, vol. abs/2012.05907, 2020.

[10] flightrecorder.com. (2023) Quick access recorder (qar).

[11] F. A. Administration, "Advisory circular," 2004.

[12] G. A. Vouros, T. Tranos, K. Blekas, and G. Santipantakis, "Data-driven estimation of flights' hidden parameters," 2022.

[13] EUROCONTROL, "European aviation overview 2023," Jan 2024.

[14] J. Brownlee. (2016) A gentle introduction to the gradient boosting algorithm for machine learning.

[15] M. Corporation. (2023) Lightgbm documentation.

[16] X. Developers. (2022) Xgboost tutorials.

[17] A. Chakure. (2023) Random forest regression in python explained.