



HAL
open science

Towards aircraft generic Quick Access Recorder fuel flow regression models for ADS-B data

Gabriel Jarry, Daniel Delahaye, Christophe Hurter

► To cite this version:

Gabriel Jarry, Daniel Delahaye, Christophe Hurter. Towards aircraft generic Quick Access Recorder fuel flow regression models for ADS-B data. International Conference on Research in Air Transportation, Eurocontrol FAA NTU, Jul 2024, Singapour, Singapore. hal-04636710

HAL Id: hal-04636710

<https://enac.hal.science/hal-04636710>

Submitted on 5 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards aircraft generic Quick Access Recorder fuel flow regression models for ADS-B data

Gabriel Jarry*[†]

* Eurocontrol
Brétigny-Sur-orge France
gabriel.jarry@eurocontrol.int

[†] Direction des Services de la Navigation Aérienne
DGAC, DSNA, Mission Environnement
Toulouse, France

Daniel Delahaye[‡], Christophe Hurter[‡]

[‡] ENAC, Université de Toulouse
Toulouse, France
{daniel.delahaye, christophe.hurter}@enac.fr

Abstract—This paper presents an investigation into the development of generic Quick Access Recorder (QAR) fuel flow regression models applied to Automatic Dependent Surveillance-Broadcast (ADS-B) data, with the aim of improving the accuracy of fuel flow estimates for various aircraft operations. Given the critical need for accurate fuel consumption estimates to mitigate the environmental impact of aviation, this study explores a novel approach that integrates derivative features and aircraft-specific parameters into a unified model. This approach not only aims to generalise across different aircraft types, thereby providing scalability and flexibility for end users, but also demonstrates adaptability to missing parameters through data augmentation techniques. Using a dataset of QAR data from various aircraft type, this paper evaluates the performance of the model across different flight phases and aircraft types and compares it with other common fuel flow models.

Keywords—Fuel Flow Estimation, Aviation Sustainability, Machine Learning, Quick Access Recorder, Neural Network

I. INTRODUCTION

In the quest to reduce the environmental impact of aviation, accurate estimation of fuel flow during aircraft operations has been a long-standing research challenge. Responsible for a significant proportion of global CO₂ emissions [1], the aviation industry is at a critical moment where the need to reduce its carbon footprint cannot be overstated. This urgency is compounded by the increasing focus on the environmental impact of non-CO₂ effects such as contrails [2], which also contribute to global warming. In this context, the development of accurate and scalable models is essential to this environmental challenge.

The development of models for accurate estimation of fuel flow and other environmental metrics from aircraft operations is of paramount importance, not only for Air Navigation Service Providers (ANSPs) engaged in environmental monitoring, but also for the academic community focused on research in this area. These models offer a dual benefit: on the one hand, they enable ANSPs to accurately measure, evaluate and monitor environmental performance [3], which is crucial for achieving ambitious environmental goals such as those set by SESAR 2020 [4] in Europe and NextGen in the United States, which aim to significantly reduce CO₂ emissions and fuel consumption per flight by 2035. On the other hand, they serve as valuable tools

for academic researchers, providing a rich source of data for studying the environmental impact of aviation [5]–[7], exploring new ways to reduce emissions and developing innovative solutions to analyse [8] and improve [9], [10] the sustainability of air transportation.

II. STATE OF THE ART

A. Fuel estimation

In the field of air transportation research, understanding and predicting aircraft fuel consumption remains a critical challenge that is being addressed through various modelling approaches, ranging from physical models [11] to advanced neural network models [12] and other innovative methods [13].

Physical models: EUROCONTROL’s Base of Aircraft Data (BADA) is the basis for simulating aircraft trajectories for air traffic management, offering detailed kinetic modelling and improvements in its latest iterations [11], [14]. An open source model called OpenAP incorporates aircraft dynamics and fuel estimates, aiming for transparency and improved simulation accuracy [15]. The work of Poll and Schumann introduces a method for estimating cruise fuel burn and performance characteristics for turbofan aircraft, combining aerodynamic theory with empirical data for more accurate predictions [16], [17].

Neural network models: Using machine learning, Chati and Balakrishnan use Gaussian Process Regression (GPR) and operational data to outperform traditional models such as BADA, providing more accurate fuel consumption estimates [13]. Their additional research uses CART and Least Squares Boosting algorithms to predict fuel flow rates, improving the accuracy of the emissions inventory [18]. Baumann and Klingauf use machine learning algorithms that use full-flight sensor data to surpass the accuracy of traditional fuel flow models [19]. Baklacioglu’s work with advanced artificial neural networks, including multilayer perceptron and radial basis function networks, demonstrates high accuracy in predicting fuel flow rates [12].

Trani et al. present a neural network model validated against specific aircraft types, demonstrating its potential in fast time simulation models [20]. Li et al. use LSTM neural networks

to accurately predict performance-based contingency fuel, significantly improving operational efficiency [21]. Kayaalp et al. develop an LSTM model for predicting combustion efficiency and exhaust emissions, achieving high accuracy without extensive experimental testing [22]. Metlek presents a CNN-BiLSTM model that excels in predicting aircraft fuel consumption with minimal error, outperforming existing methods [23]. Baklacioglu’s Genetic Algorithm Optimised Neural Networks offer a novel approach to fuel consumption prediction, achieving remarkable accuracy during different flight phases [24]. Finally, a very promising paper proposes to add a physics-based loss to the training process to ensure the robustness of the model to parameter variations [25]. This is an interesting hybrid approach to increase the robustness of neural network models.

Other types of modelling have been proposed, such as Chatterji’s method for estimating jet fuel consumption, which uses real flight data to produce estimates within 1% of actual fuel consumption, validated against FAA data [26]. Senzig, Fleming and Iovinelli are refining terminal area fuel consumption modelling to an accuracy of $\pm 5\%$, supporting informed policy making [27]. Clemons and his team aim to improve the accuracy of aircraft fuel burn models on airport surfaces by addressing simplifications in existing tools and improving the estimation of baseline taxi fuel flow and taxi time [28].

In a previous work, we have benchmarked the use of neural networks to estimate on-board aircraft parameters during approach and landing, suggesting a promising direction for air traffic management systems to improve performance metrics such as fuel flow and flap configuration [29]. Furthermore, as part of the French ANSP’s ACROPOLE project, we have used machine learning to assess the environmental impact of air traffic, focusing on fuel consumption and CO2 emissions, thus providing a novel perspective from an air navigation service provider’s (ANSP’s) point of view [3].

B. Aircraft parameter estimation

More generally, several studies have made significant contributions to improving the accuracy and reliability of estimating critical parameters that are essential for safe and efficient operations. The studies by Delahaye and Puechmorel introduce innovative methods for wind estimation from radar data, which is crucial for accurate prediction of aircraft trajectories. Their work focuses on linear models and Kalman filtering to derive high quality wind estimates from radar track measurements and to estimate local wind conditions respectively, highlighting the importance of accurate environmental data for air traffic management [30] [31]. Further advancing the field, Alligier et al. evaluate and compare methods for estimating aircraft mass from ground-based observations. Their research demonstrates the effectiveness of adaptive and least squares methods, with a particular focus on improving the accuracy of climb prediction -

a key factor in optimising air traffic control operations [32] [33]. This information is a key feature as it is not normally available from ADS-B data.

Sun et al. propose a novel Bayesian inference method for estimating aircraft initial mass, using empirical knowledge and flight data to outperform previous techniques in terms of accuracy. This method shows promise for improving aircraft performance analysis and air traffic management by providing more accurate mass estimates, which are essential for various operational decisions [34].

A different approach introduced an interacting multiple model filtering technique for real-time identification of aircraft guidance modes using surveillance data to improve the accuracy of trajectory prediction. This methodology supports the advancement of trajectory-based operations in air traffic management by providing a reliable tool for guidance mode identification [35]. Askari and Cremaschi’s study presents a neural network-based approach to estimating aircraft departure performance metrics, illustrating the potential of machine learning applications in predicting critical flight parameters from radar and ADS-B data [36].

Finally, the research by Di Zhou et al. on hazard identification in civil aircraft auxiliary power units uses a hybrid deep neural network that combines convolutional neural networks and bidirectional long short-term memory models. This innovative approach aims to improve safety and maintenance efficiency by accurately identifying potential hazards, demonstrating the potential of deep learning in civil aviation safety applications [37].

C. Contribution

This paper presents several contributions that fill the gap in the state of the art in fuel flow estimation models. First, a single modelling approach using data augmentation is introduced to deal with missing parameters such as mass or true airspeed, allowing for easier use and scalability. Second, the previous network architectures [3], [29] are modified to integrate both derivative features instead of successive points, and aircraft fixed parameters to pave the way for a single generic model that generalises across aircraft. This allows flexibility for end-user systems and increases scalability. Finally, an open source library of the trained model is released as part of the Acropole project of the French ANSP [38].

III. DATA AND MODEL ARCHITECTURE

This section describes the neural network architecture and the performance achieved with our framework.

A. data description

The dataset consists of Quick Access Recorder (QAR) data from various airlines. The available dataset includes different

Aircraft	A320-200	A330-223	ATR72-600	B737-85P	B737-8GJ	B737-8K2	CRJ-1000	CRJ-700	E-170	E-190
Flights	16 453	186	2 605	8 744	2 995	21 226	29 422	17 234	30 462	36 287
Percent (%)	9.93	0.11	1.57	5.28	1.81	12.82	17.77	10.41	18.39	21.91

TABLE I
TRAJECTORIES NUMBER OF DIFFERENT AIRCRAFT MODELS.

aircraft types such as A319-211, A320-214, B737-8K2, A330-223, ATR72-600, E190, E170, CRJ100, CRJ700. The data set does not cover the entire fleet as each aircraft type accounts for a significant number of flights ranging from 186 to 30462. The number of flights per aircraft type is summarised in the table I. Each flight is gate-to-gate. The sampling rate is 1 second. As the data is unbalanced between aircraft types, the number of flights for training is limited to a maximum of 10,000 per aircraft type and a weight inversely proportional to the number of flights available is used during training to correct further unbalance.

The parameters used as input to the model are: altitude, ground speed, true airspeed, vertical speed, derivatives of ground speed and true airspeed, mass normalised by empty weight and maximum take-off weight, aircraft maximum operating speed and altitude, engine type (turboprop, jet). We apply min-max normalisation before feeding into the network. Except for mass, which is rarely available, and true airspeed, which may not be available without Mode S information or weather reanalysis data, all these parameters are available from ADS-B data and aircraft parameter tables. This makes the model applicable to ADS-B data during inference without access to QAR data. To evaluate the model, we also used these input parameters from QAR data, as the data are anonymised we cannot merge them with the corresponding ADS-B data. However, it is assumed that the difference between the QAR parameters and the corresponding ADS-B parameters will be negligible.

As usual in machine learning, the data set at the flight level is divided into training (71 623 flights), validation (16 557 flights) and test (16 557 flights) sets.

B. Model architecture, data augmentation and training

The chosen model architecture is a simple feed-forward neural network, as we have added derivative features as input to the model, time features are already included. The model consists of four fully connected layers with 50 neurons and relu activation, and the last layer is a sigmoid.

Indeed, as explained, one of the aims of our model is to pave the way towards a single generic modelling of engine fuel flow. This idea was proposed by Chati et al. [18] but never applied on a large scale. Therefore, to move in this direction, the last layer was designed to return the fuel flow of a single engine normalised by the take-off fuel flow given by the ICAO engine database [39].

Finally, to deal with unavailable parameters when applying to ADS-B data, we applied a data augmentation during training (training the model with predefined values when unavailable).

The default value for the mass parameter is set to -1.0. The idea here is to have a value outside the normalisation range (0.1) to ensure that the model learns without any mass assumption. For the true airspeed, we assume no wind hypothesis and fill the value with the ground speed. This aspect and possible improvements are discussed further in section V.

The learning phase was processed using the Adam optimiser [40] with decay, and the loss function used is the mean absolute percentage error. The learning rate is 10^{-3} and the decay is 10^{-9} . Each model is trained for 100 epochs and the best network over the validation set is retained using a checkpoint. Performance is then computed over the test set.

C. Model evaluation metrics and phases

Three different metrics are used to evaluate the performance of the model: the Mean Error (ME), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE).

Let \mathcal{D} be a set of input-output pairs (x, y) and h be a model to evaluate, the first three metrics are calculated as follows:

$$\text{ME}(h, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} h(x) - y \quad (1)$$

$$\text{MAE}(h, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} |h(x) - y| \quad (2)$$

$$\text{MAPE}(h, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} \frac{|h(x) - y|}{y} \quad (3)$$

For MAE and MAPE, the smaller the value, the more accurate the prediction. For ME, the closer to zero, the better.

We have also divided the data into Phases / Vertical Attitude as follows. If the ground speed is less than 65kt, the aircraft is considered to be on the GROUND. Then the split between CLIMB, LEVEL and DESCENT is assessed in terms of vertical speed. Between -150 ft/min and 150 ft/min (or below -150 ft/min, above 150 ft/min) the airplane is considered to be in LEVEL (or DESCENT, CLIMB).

IV. MODEL EVALUATION AND COMPARISON

This section evaluates the performance of our model. As a first illustration, Figure IV shows the prediction of the model on an A320 from the test set when mass and true airspeed are available.

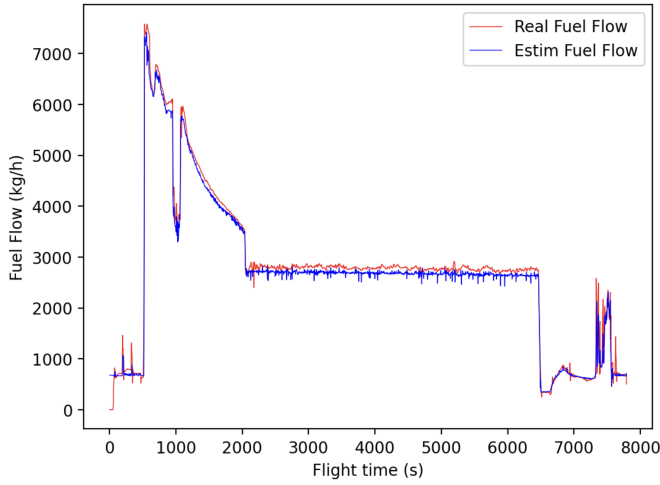


Figure 1. Example prediction of the model on an A320 flight when mass and true airspeed are available.

A. Performance with unavailable parameters

In this section we will illustrate how the model deals with unavailable parameters. To simplify the discussion, we will focus on the results for the A320 aircraft.

Table II shows a comparative analysis of performance metrics across different flight phases (GROUND, CLIMB, LEVEL, DESCENT and an overall assessment), highlighting the effects of removing aircraft mass data from the model. It shows that the inclusion of mass data generally improves prediction accuracy, as indicated by reduced MAPE and MAE values, with the effect being particularly noticeable during the climb phase, suggesting the critical role of mass data in accurately predicting fuel consumption. While the LEVEL phase shows a decrease in accuracy when mass data is excluded, the descent phase is identified as the most challenging for accurate predictions, with the highest errors observed. The analysis highlights the significant improvement in model performance across all flight phases when mass data is included, albeit with less consistent changes in ME, indicating a primary influence on accuracy rather than prediction bias.

Phase	MAPE (%)	MAE (kg/min)	ME (kg/min)	Samples #
GROUND	9.67 / 9.95	1.45 / 1.48	0.82 / 0.85	536 840
CLIMB	2.13 / 2.23	1.66 / 1.74	0.85 / 0.73	1 403 850
LEVEL	4.41 / 5.06	1.82 / 2.09	1.22 / 1.23	4 017 801
DESCENT	12.63 / 12.66	2.71 / 2.77	1.88 / 1.74	1 684 117
ALL	6.17 / 6.56	1.96 / 2.13	1.27 / 1.22	7 642 608

TABLE II

COMPARING THE MODEL WITH AND WITHOUT THE MASS PARAMETER (TRUE AIR SPEED AVAILABLE). THE FIRST VALUE REFERS TO THE MODEL ERROR WITH MASS AND THE SECOND WITHOUT THIS FEATURE.

Table III presents a detailed evaluation of the model per-

formance metrics, focusing on the influence of incorporating the true air speed parameter into the prediction model. This comparison shows that the inclusion of true air speed has a significant impact on the prediction accuracy, as evidenced by the variation in MAPE, MAE and ME values. In particular, the inclusion of true air speed improves the prediction accuracy most significantly during the CLIMB and LEVEL phases, highlighting the critical role of this parameter in accurate fuel consumption prediction.

Phase	MAPE (%)	MAE (kg/min)	ME (kg/min)	Samples #
GROUND	9.67 / 9.67	1.45 / 1.45	0.82 / 0.82	536 840
CLIMB	2.13 / 2.44	1.66 / 1.88	0.85 / 0.98	1 403 850
LEVEL	4.41 / 5.14	1.82 / 2.14	1.22 / 1.45	4 017 801
DESCENT	12.63 / 14.71	2.71 / 3.22	1.88 / 2.21	1 684 117
ALL	6.17 / 7.07	1.96 / 2.28	1.27 / 1.48	7 642 608

TABLE III

COMPARISON OF THE MODEL WITH AND WITHOUT THE TRUE AIRSPEED PARAMETER (MASS AVAILABLE). THE FIRST VALUE REFERS TO THE MODEL ERROR WITH TRUE AIRSPEED AND THE SECOND WITHOUT THIS FEATURE.

Finally, table IV highlights the compounded challenges when both mass and true airspeed are excluded. The comparison shows a more pronounced degradation in model accuracy across all phases, with the most significant increases in prediction errors observed during the CLIMB and LEVEL phases. This table illustrates the synergistic importance of both parameters in achieving high prediction accuracy and how their absence significantly reduces the model's ability to accurately predict fuel consumption.

Phase	MAPE (%)	MAE (kg/min)	ME (kg/min)	Samples #
GROUND	9.67 / 9.95	1.45 / 1.48	0.82 / 0.85	536 840
CLIMB	2.13 / 2.55	1.66 / 1.98	0.85 / 0.91	1 403 850
LEVEL	4.41 / 5.69	1.82 / 2.37	1.22 / 1.44	4 017 801
DESCENT	12.63 / 14.73	2.71 / 3.28	1.88 / 2.07	1 684 117
ALL	6.17 / 7.4	1.96 / 2.44	1.27 / 1.44	7 642 608

TABLE IV

COMPARISON OF THE MODEL WITH AND WITHOUT THE MASS AND TRUE AIRSPEED PARAMETERS. THE FIRST VALUE REFERS TO THE MODEL ERROR WITH MASS AND TRUE AIRSPEED AND THE SECOND WITHOUT THESE FEATURES.

B. Performance for different types of aircraft

In this section we will illustrate how the model performs for different types of aircraft.

The table V provides a detailed analysis of the performance of the prediction model for different aircraft types. Notably, the table includes a wide range of aircraft, from narrow-body jets such as the A320-214 and B737 variants, to regional jets such as the CRJ-700 and CRJ-1000, and even turboprops such as the ATR72. A number of key observations emerge from this analysis.

Phase	MAPE (%)	MAE (kg/min)	ME (kg/min)	Samples #
A320-214	6.17 / 7.4	1.96 / 2.44	1.27 / 1.44	7 642 608
A330-224	5.6 / 7.35	3.87 / 5.4	1.72 / 1.98	231 104
ATR72	7.83 / 8.55	0.48 / 0.55	0.19 / 0.18	1 058 484
B737-85P	7.38 / 8.96	2.37 / 2.78	0.08 / -0.48	1 361 170
B737-8GJ	6.44 / 7.48	2.2 / 2.49	0.29 / 0.0	330 212
B737-8K2	7.63 / 8.97	2.52 / 2.86	-1.24 / -1.44	2 540 614
CRJ-700	7.21 / 9.42	1.41 / 1.83	-0.07 / 0.02	4 122 830
CRJ-1000	8.01 / 10.32	1.54 / 1.99	0.1 / 0.02	4 085 026
E-170	8.96 / 110.39	1.81 / 2.36	1.14 / 1.18	4 422 710
E-190	6.27 / 8.77	1.46 / 2.13	0.29 / 0.19	4 645 568

TABLE V

COMPARISON OF MODEL PERFORMANCE BETWEEN AIRCRAFT TYPES FOR ALL TEST PLOTS WITH MASS AND TRUE AIRSPEED (FIRST VALUE), WITHOUT MASS AND TRUE AIR SPEED (SECOND VALUE).

a) *Variation in prediction accuracy:* The MAPE values, which indicate the percentage accuracy of the model's predictions, vary between aircraft types. The A330-224 shows the best performance with a MAPE of 5.6%, indicating a high level of accuracy in predicting fuel consumption for this aircraft type. In contrast, the E-170 has the highest MAPE at 8.96%, indicating less accurate predictions for this model.

b) *Differences in absolute errors:* The MAE values, which represent the average magnitude of error in fuel burn predictions (in kg/min), highlight the differences in absolute performance of the model. The A330-224 has a higher MAE of 3.87 kg/min, reflecting the larger size and fuel consumption patterns of wide-body aircraft compared to models such as the ATR72, which has the lowest MAE of 0.48 kg/min, probably due to its smaller size and efficiency on shorter routes.

c) *Bias in predictions:* The ME values provide an insight into the systematic over- or under-prediction bias of the model. For example, the B737-8K2 shows a significant negative ME (-1.24 kg/min), indicating a tendency to under-predict fuel consumption for this aircraft type. Conversely, the positive ME of the E-170 (1.14 kg/min) indicates an over-prediction bias.

The large number of samples for each aircraft type, ranging from over 7 million samples for the A320-214 to around 231,104 for the A330-224, provides a robust dataset for evaluating the model's performance. This analysis not only highlights the model's ability to adapt to different aircraft characteristics, but also underlines the challenges of achieving consistently high accuracy across different models, which can result from differences in aerodynamic properties, engine efficiency and operating profiles. This comprehensive performance comparison provides valuable insights for further refining the predictive model to better reflect the unique fuel consumption patterns of different aircraft types.

The table VI presents a detailed comparison of model performance metrics during different phases of flight (GROUND, CLIMB, LEVEL, DESCENT and ALL phases combined) for aircraft considering different configurations of input variables (TAS

Category	Phase	MAPE (%)	MAE (kg/min)	ME (kg/min)
TAS & MASS	GROUND	15.52 (6.43)	2.33 (1.75)	1.22 (1.31)
	CLIMB	3.38 (0.81)	2.47 (1.62)	0.26 (1.39)
	LEVEL	4.18 (0.76)	1.56 (0.68)	0.10 (0.87)
	DESCENT	11.45 (1.49)	2.38 (1.33)	0.83 (0.96)
TAS & No MASS	GROUND	15.84 (6.53)	2.36 (1.78)	1.12 (1.25)
	CLIMB	3.60 (0.99)	2.66 (1.79)	0.15 (1.29)
	LEVEL	4.76 (0.82)	1.84 (1.00)	-0.08 (0.95)
	DESCENT	11.88 (1.64)	2.47 (1.39)	0.56 (1.01)
No TAS & MASS	GROUND	7.60 (0.89)	2.15 (1.09)	0.21 (0.83)
	CLIMB	15.71 (6.54)	2.36 (1.78)	1.17 (1.30)
	CLIMB	3.59 (0.85)	2.60 (1.65)	0.25 (1.47)
	LEVEL	5.58 (1.36)	2.02 (0.71)	0.20 (0.97)
No TAS & No MASS	DESCENT	13.82 (1.89)	2.82 (1.43)	1.02 (1.08)
	ALL	8.40 (1.42)	2.28 (0.88)	1.48 (0.88)
	GROUND	16.07 (6.68)	2.38 (1.81)	1.06 (1.21)
	CLIMB	3.79 (1.02)	2.78 (1.84)	0.26 (1.49)
No TAS & No MASS	LEVEL	6.14 (1.30)	2.30 (1.07)	-0.01 (1.05)
	DESCENT	14.32 (1.88)	2.92 (1.47)	0.72 (1.12)
	ALL	8.86 (1.24)	2.48 (1.15)	0.31 (0.94)

TABLE VI

COMPARISON OF MODEL PERFORMANCE PER PHASE AVERAGE ACROSS AIRCRAFT TYPES FOR ALL TEST PLOTS WITH MASS AND TRUE AIRSPEED.

THE VALUE SHOWN IS THE AVERAGE BETWEEN AIRCRAFT AND THE STANDARD DEVIATION BETWEEN BRACKETS.

and MASS, TAS only, MASS only, neither TAS nor MASS). The performance metrics are reported together with their standard deviations within aircraft types (in brackets) to indicate the variability of model performance for different aircraft types. The configurations are designed to assess the effect of including true airspeed (TAS) and mass data on the prediction accuracy of the model. In fact, mass is usually not available from the ground side, the true air speed may be available with Mode S radar or by interpolation of reanalysis weather data.

Analysis of the data reveals the following trends and implications:

1) *Availability of parameters:*

- **TAS and MASS included:** this configuration generally produces highest MAPE values across all phases, indicating a reasonable level of accuracy in percentage terms. The MAE and ME values also suggest that, on average, the model is predicting with a good level of accuracy, particularly in the LEVEL phase where the errors are lowest. This suggests that the inclusion of both TAS and mass as inputs helps the model to accurately predict outcomes in different flight phases.
- **TAS Only (No MASS):** excluding mass slightly increases the MAPE and MAE values in most phases compared to including both TAS and mass, particularly noticeable in the LEVEL and DESCENT phases. However, we expected the model to show an even more significant drop in performance without the mass. As the model is a neural network, it is difficult to understand the reasons for this result. However, one possible explanation could be that the

model has somehow learnt statistics about the airline's preference/average mass and optimises the output fuel based on these statistics.

- MASS Only (No TAS): this configuration leads to a noticeable increase in errors in all phases, especially in the LEVEL and DESCENT phases where the MAPE and MAE are significantly higher than when TAS is included, highlighting the critical role of TAS in the predictive accuracy of the model.
- neither TAS nor MASS: excluding both TAS and Mass results in the highest errors at almost all stages, highlighting the importance of these variables in achieving accurate predictions.

2) *Specific Observations by Phase:*

- GROUND Phase: the GROUND phase consistently shows higher MAPE values compared to other phases, suggesting that this phase may be more difficult to model accurately, possibly due to the variability in ground operations.
- CLIMB and LEVEL phases: these phases generally benefit from the inclusion of TAS, as evidenced by lower errors when TAS is included. In particular, the LEVEL phase shows a significant improvement in model performance when both TAS and mass are considered.
- DESCENT Phase: The DESCENT phase shows a higher sensitivity to the exclusion of TAS, with a significant increase in errors when TAS is not included. This suggests that TAS is particularly important for accurate modelling of this phase.

C. *Comparison with other models*

Table VII gives a comprehensive comparison of the performance metrics for different fuel flow models (current modelisation multi aircraft, our last modelisation with a single aircraft [3], OpenAP [15], BADA [11], [14] and Poll-Schumann [16], [17]). The comparison is made on the 1000 test flights of A320-214 over different phases. It also shows the processing time in seconds to apply the model to 1000 flights.

To apply these models, sensitive features (vertical speed, true airspeed) are smoothed using the Savitzky-Golay filter [41]. Dedicated Python implementations are used: pybada available under the BADA 4.2 licence [11], [14], pycontrails [42] for the Poll-Schumann model and OpenAP [15]. To speed up the computation, we apply parallel processing to each trajectory. For all models we used the mass and true air speed from the QAR data and calculated the temperature using ISA conditions. As some models are unable to estimate fuel flow on the ground, we have removed this part of the trajectories for this evaluation.

Neural network based models (current our previous model) shows higher accuracy in all phases, with the lowest MAPE, MAE and ME values, indicating that it predicts aircraft performance with the least deviation from actual measurements. This

is particularly evident in the CLIMB phase, where its MAPE is significantly lower than the others, indicating a high reliability of predictions during the climb. BADA and Poll Schumann show slightly lowest accuracy, with Poll Schumann occasionally outperforming BADA, as seen in the LEVEL phase for the MAPE and MAE metrics; however, both models have higher errors compared to neural network based model, but perform better than OpenAP. The difference between a physical model such as BADA and a neural network based model could also be explained by the task for which they were designed. In particular, the BADA model is designed to ensure prediction accuracy in all areas of the flight envelope, whereas our model does not guarantee this property. Further discussion of parameter sensitivity is given in section V. Additionally, the data to train BADA were given by the manufacturer and correspond to performance of new aircraft. The model does not take into account engine aging which implies additional consumption. In the case of OpenAP, it has only been trained on open source data, which is not an easy task and may explain its performance. Finally, the distribution of the test set is likely to follow the same distribution as that of the train set, potentially facilitating the accuracy of our model.

If we compare this paper proposed model, with our previous model (single aircraft) [3], we can see that the new model with multiple aircraft has a lower performance. There could be several reasons for this. Firstly, the handling of multiple aircraft could require more parameters in the model architecture. Here, we kept the same architecture with the same number of layers and weights. Secondly, and more likely, the multi-aircraft model shows a higher bias (ME), which could be due to the fuel flow normalisation by take-off fuel flow that we introduced and the sigmoid output. The constraint could introduce a bias when training on multiple aircraft. This will be further investigated in future work to improve the performance of the multi-aircraft model. Finally, the decision to augment the data by filling in a default value when the data is not available could also introduce bias. To address this issue, further modelling could be added, for example by adding a dimension indicating whether the parameter is available or not.

The processing time data for the different models (Neural network based models, OpenAP, BADA and Poll-Schumann) shows a contrast in computational efficiency. Neural network based models stands out with a low processing time of just 3 seconds, indicating a highly efficient model that can provide fast predictions, making it ideal for applications requiring real-time data analysis. In contrast, BADA has the longest processing time at 474 seconds, suggesting a much slower computational process that could hinder its use in time-sensitive scenarios. OpenAP and Poll Schumann fall in between, with processing times of 284 seconds and 28 seconds respectively, indicating moderate efficiency. This results are consistent with neural networks,

Phase	Samples #	Model Metric	ACROPOLE Multi-Aircraft	Neural network Single Aircraft [3]	OpenAP [15]	BADA [11], [14]	Poll-Schumann [16], [17]
CLIMB	1 403 850	MAPE (%)	2.13	1.93	30.35	6.53	6.85
		MAE (kg/min)	1.66	1.42	25.81	5.53	5.65
		ME (kg/min)	0.85	-0.15	-25.66	-5.27	-4.62
LEVEL	4 017 801	MAPE (%)	4.41	3.23	18.59	7.01	4.84
		MAE (kg/min)	1.82	1.26	7.82	2.65	2.03
		ME (kg/min)	1.22	-0.33	-7.47	-1.43	-0.73
DESCENT	1 684 117	MAPE (%)	12.63	11.67	51.69	21.50	21.55
		MAE (kg/min)	2.71	2.27	8.62	3.71	4.71
		ME (kg/min)	1.88	-0.92	-1.75	-0.64	-3.67
ALL	7 105 768	MAPE (%)	5.91	4.97	27.60	9.84	8.61
		MAE (kg/min)	1.99	1.53	11.55	3.44	3.29
		ME (kg/min)	1.30	-0.43	-9.92	-2.03	-2.09
		Processing time (s)	3	3	284	474	28

TABLE VII

COMPARISON OF DIFFERENT MODEL PERFORMANCES PER PHASE FOR 1000 TEST FLIGHTS OF A320-214 AIRCRAFT USING REAL MASS AND TRUE AIRSPEED.

which are typically very fast at inference, and here benefit from GPU acceleration (Nvidia RTX A4500), which optimises the inference process for large batches that do not require per-trajectory processing. Whereas physical models are highly dependent on implementation optimisation and typically need to be processed per trajectory. here there is no GPU but parallel CPU processing with 36 chores. These variations in processing time are critical for practical applications; faster models such as neural network are more adapted for real-time performance analysis and decision making or Big-Data application such as at scale inventories, while slower models may be more suitable for in-depth, non-time critical analysis.

V. DISCUSSION

In this section we discuss the limitations and potential improvements of this work in terms of unavailable parameters, fleet coverage and generalisation, data imbalance in the training set, and model sensitivity to physical parameters.

A. Unavailable parameters

When dealing with the unavailability of critical parameters such as TAS (True Airspeed) and mass in predictive aircraft performance modelling, a practical approach could be to extend the model architecture to estimate these missing parameters. This can be achieved by incorporating sub-models or algorithms capable of inferring the missing data based on other available inputs and historical data patterns. In addition, the inclusion of a feature that explicitly indicates the absence of certain information could significantly improve the adaptability of the model. By flagging missing inputs, the model can adjust its predictions to account for the lack of specific data and potentially apply different weighting or inference mechanisms to compensate. This dual strategy not only mitigates the impact of missing data on model accuracy, but also enriches the model's understanding of the context of the data, leading to more robust and reliable predictions even in the face of incomplete information.

B. Coverage and generalisation

Our study has a limitation in terms of aircraft type coverage, in particular the sparse representation of wide-body aircraft, which are central to long-haul operations and have a profound impact on aviation's environmental footprint. The predominance of data from narrow-body jets, regional jets and turboprops in the dataset, while valuable, leaves a significant gap in understanding the fuel consumption patterns of wide-body aircraft. These heavy aircraft, due to their size, engine configurations and cruising altitudes, have different operational profiles, aerodynamics and fuel burn dynamics than the lighter aircraft types included in the study.

In addition, the ability of the model to generalise to different aircraft types not included in the original training dataset is not fully explored in this current work and requires further investigation. Specifically, future research could focus on training models with a subset of aircraft types while intentionally withholding data from certain aircraft models to serve as a hidden test set. This approach would allow for a more rigorous assessment of the model's predictive accuracy and generalisation across various aircraft types, including those not directly observed during the training phase. Such studies would be crucial for validating the model's utility in real-world applications, where the exact aircraft type may not be part of the historical dataset used for model training. By addressing this aspect, future research could significantly enhance the robustness and applicability of such fuel flow regression models, making them more versatile and effective tools for environmental monitoring and management in aviation.

Furthermore, the bias potentially introduced by the last layer (sigmoid multiplication by take-off fuel flow) also requires further analysis. A global level methodology should be sought. It could be composed of different sub-models: dedicated models for individual aircraft for which data are available, to maximise performance, and a generic model for unseen aircraft to cover the rest of the fleet.

C. Data imbalance

Another significant limitation of the study is the data imbalance observed between different flight phases within the dataset, resulting from the disproportionate representation of flight phases such as cruise, which could distort the model's learning process and potentially bias its predictions. In predictive modelling, a balanced representation of each class ensures that the model learns the characteristics of each phase accurately, leading to more reliable and generalisable predictions in different operational contexts. However, the predominance of certain phases over others in the training dataset can lead to an overfitting problem, where the model performs exceptionally well on the overrepresented phases but poorly on the underrepresented ones. To address the phase imbalance, an effective strategy could be either to rebalance the dataset by oversampling underrepresented flight phases, or to introduce a weighting scheme in the model training process to compensate for the imbalance.

D. Model sensitivity to physical parameters

Finally, it is important to ensure that sensitivity of the model to variations in key parameters such as aircraft mass and speed are physically realistic output. This sensitivity underscores the importance of not only accurately capturing but also effectively integrating these parameters into the modeling process. The problem of parameter sensitivity could be mitigated by incorporating a physical-based loss function, as suggested by Uzun et Al. [25]. Such an approach leverages domain knowledge to introduce constraints derived from physical principles of flight dynamics and fuel consumption, thereby guiding the model towards more physically plausible predictions. Integrating a physical-based loss function can help in aligning the model's predictions more closely with the underlying physical processes, reducing the impact of parameter sensitivity and improving the model's overall predictive performance.

VI. CONCLUSIONS

In this study, we investigated the development of generic Quick Access Recorder (QAR) fuel flow regression models to improve the accuracy of the fuel flow estimation using Automatic Dependent Surveillance-Broadcast (ADS-B) data during various aircraft operations. Through a comprehensive review of the state of the art, we identified and compared various modelling approaches, including physical models such as BADA [11], [14], Poli-Schuman [16], [17], and OpenAP [15], and advanced machine learning techniques, which have shown potential to outperform traditional methods in predicting fuel flow rates. We presented a novel approach that integrates derivative features and aircraft-specific parameters into a unified model, aiming to generalise across different aircraft types, thereby providing flexibility and scalability for end users. Despite not outperforming single aircraft neural net approach [3], this first attempt show promising results.

The dataset included QAR data from different aircraft, allowing us to train and validate our model across an extended fleet. We applied data augmentation techniques to deal with unavaialbe parameters such as mass or true airspeed. The evaluation of our model highlighted its consistent accuracy and efficiency in estimating fuel flow rates across different flight phases and aircraft types. Furthermore, the ability of our model to adapt to missing parameters demonstrates its potential for real-world applications where complete data may not always be available. We have compared our model with existing methods and demonstrated its competitive advantage not only in prediction accuracy but also in computational efficiency, which is particularly relevant for real-time applications where rapid data processing is essential. The global model is available in an open source library [38].

Nevertheless, several improvements and future work will be carried out to further refine the accuracy and robustness of the model. In particular, we plan to extend the aircraft type coverage and explore its generalisation capabilities to unseen aircraft types to ensure its broader applicability. We will investigate the integration of physically informed losses and architecture to improve the robustness of the model and perform a sensitivity analysis. Finally, we aim to analyse a pure physical model fit to the QAR data to assess the modelling error versus the error induced by the train data set difference.

REFERENCES

- [1] R. Teoh, Z. Engberg, U. Schumann, C. Voigt, M. Shapiro, S. Rohs, and M. Stettler, "Global aviation contrail climate effects from 2019 to 2021," *EGUsphere*, vol. 2023, pp. 1–32, 2023.
- [2] D. S. Lee, D. W. Fahey, A. Skowron, M. R. Allen, U. Burkhardt, Q. Chen, S. J. Doherty, S. Freeman, P. M. Forster, J. Fuglestvedt, *et al.*, "The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018," *Atmospheric Environment*, vol. 244, p. 117834, 2021.
- [3] G. Jarry and D. Delahaye, "Toward novel environmental impact assessment for ansps using machine learning," in *Climate change and the role of air traffic control Research Workshop, Vilnius*, 2021.
- [4] S. J. Undertaking, "European atm master plan. the roadmap for delivering high performing aviation for europe," *Brussels (Belgium), Tech. Rep.*, 2015.
- [5] M. E. Stettler, A. M. Boies, A. Petzold, and S. R. Barrett, "Global civil aviation black carbon emissions," *Environmental science & technology*, vol. 47, no. 18, pp. 10397–10404, 2013.
- [6] F. D. Quadros, M. Snellen, J. Sun, and I. C. Dedoussi, "Global civil aviation emissions estimates for 2017–2020 using ads-b data," *Journal of Aircraft*, vol. 59, no. 6, pp. 1394–1405, 2022.
- [7] R. Teoh, Z. Engberg, M. Shapiro, L. Dray, and M. E. Stettler, "The high-resolution global aviation emissions inventory based on ads-b (gaia) for 2019–2021," *Atmospheric Chemistry and Physics*, vol. 24, no. 1, pp. 725–744, 2024.
- [8] R. Dalmau, P. Very, and G. Jarry, "On the causes and environmental impact of airborne holdings at major european airports," *Journal of Open Aviation Science*, vol. 1, no. 2, 2023.
- [9] R. Dalmau and X. Prats, "Fuel and time savings by flying continuous cruise climbs: Estimating the benefit pools for maximum range operations," *Transportation Research Part D: Transport and Environment*, vol. 35, pp. 62–71, 2015.
- [10] R. P. Liem, G. K. Kenway, and J. R. Martins, "Multimission aircraft fuel-burn minimization via multipoint aerostructural optimization," *AIAA Journal*, vol. 53, no. 1, pp. 104–122, 2015.

- [11] A. Nuic, D. Poles, and V. Mouillet, "Bada: An advanced aircraft performance model for present and future atm systems," *International journal of adaptive control and signal processing*, vol. 24, no. 10, pp. 850–866, 2010.
- [12] T. Baklacioglu, "Predicting the fuel flow rate of commercial aircraft via multilayer perceptron, radial basis function and anfis artificial neural networks," *The Aeronautical Journal*, vol. 125, no. 1285, pp. 453–471, 2021.
- [13] Y. S. Chati and H. Balakrishnan, "A gaussian process regression approach to model aircraft engine fuel flow rate," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*, pp. 131–140, ACM, 2017.
- [14] A. Nuic and V. Mouillet, "User manual for the base of aircraft data (bada) revision 4," *EUROCONTROL Experimental Centre, EEC Technical/Scientific Report No. 12/11/22-58 v1.3*, 2016.
- [15] J. Sun, J. M. Hoekstra, and J. Ellerbroek, "Openap: An open-source aircraft performance model for air transportation studies and simulations," *Aerospace*, vol. 7, no. 8, p. 104, 2020.
- [16] D. Poll and U. Schumann, "An estimation method for the fuel burn and other performance characteristics of civil transport aircraft in the cruise. part 1 fundamental quantities and governing relations for a general atmosphere," *The Aeronautical Journal*, vol. 125, no. 1284, pp. 257–295, 2021.
- [17] D. Poll and U. Schumann, "An estimation method for the fuel burn and other performance characteristics of civil transport aircraft during cruise: part 2, determining the aircraft's characteristic parameters," *The Aeronautical Journal*, vol. 125, no. 1284, pp. 296–340, 2021.
- [18] Y. S. Chati and H. Balakrishnan, "Statistical modeling of aircraft engine fuel flow rate," in *30th Congress of the International Council of the Aeronautical Science*, 2016.
- [19] S. Baumann and U. Klingauf, "Modeling of aircraft fuel consumption using machine learning algorithms," *CEAS Aeronautical Journal*, vol. 11, no. 1, pp. 277–287, 2020.
- [20] A. A. Trani, F. Wing-Ho, G. Schilling, H. Baik, and A. Seshadri, "A Neural Network Model to Estimate Aircraft Fuel Consumption," 2004.
- [21] L. Li, S. Yuan, Y. Teng, and J. Shao, "A study on sustainable consumption of fuel—an estimation method of aircraft," *Energies*, vol. 14, no. 22, p. 7559, 2021.
- [22] K. Kayaalp, S. Metlek, S. Ekici, and Y. Şöhret, "Developing a model for prediction of the combustion performance and emissions of a turboprop engine using the long short-term memory method," *Fuel*, vol. 302, p. 121202, 2021.
- [23] S. Metlek, "A new proposal for the prediction of an aircraft engine fuel consumption: a novel cnn-bilstm deep neural network model," *Aircraft Engineering and Aerospace Technology*, vol. 95, no. 5, pp. 838–848, 2023.
- [24] T. Baklacioglu, "Modeling the fuel flow-rate of transport aircraft during flight phases using genetic algorithm-optimized neural networks," *Aerospace Science and Technology*, vol. 49, pp. 52–62, 2016.
- [25] M. Uzun, M. U. Demirezen, and G. Inalhan, "Physics guided deep learning for data-driven aircraft fuel consumption modeling," *Aerospace*, vol. 8, no. 2, p. 44, 2021.
- [26] G. B. Chatterji, "Fuel burn estimation using real track data," in *11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, including the AIAA Balloon Systems Conference and 19th AIAA Lighter-Than*, p. 6881, 2011.
- [27] D. A. Senzig, G. G. Fleming, and R. J. Iovinelli, "Modeling of Terminal-Area Airplane Fuel Consumption," *Journal of Aircraft*, vol. 46, no. 4, pp. 1089–1093, 2009.
- [28] E. Clemons, T. Reynolds, S. Badrinath, Y. Chati, and H. Balakrishnan, "Enhancing aircraft fuel burn modeling on the airport surface," in *2018 Aviation Technology, Integration, and Operations Conference*, p. 3991, 2018.
- [29] G. Jarry, D. Delahaye, and E. Feron, "Approach and landing aircraft on-board parameters estimation with lstm networks," in *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)*, pp. 1–6, IEEE, 2020.
- [30] D. Delahaye and S. Puechmorel, "Tas and wind estimation from radar data," in *2009 IEEE/AIAA 28th Digital Avionics Systems Conference*, pp. 2–B, IEEE, 2009.
- [31] D. Delahaye and S. Puechmorel, "Aircraft local wind estimation from radar tracker data," in *2008 10th International Conference on Control, Automation, Robotics and Vision*, pp. 1033–1038, IEEE, 2008.
- [32] R. Alligier, D. Gianazza, and N. Durand, "Ground-based estimation of aircraft mass, adaptive vs. least squares method," ATM Seminar, 2013.
- [33] R. Alligier, D. Gianazza, M. G. Hamed, and N. Durand, "Comparison of two ground-based mass estimation methods on real data," in *ICRAT 2014, 6th International Conference on Research in Air Transportation*, pp. pp–xxxx, 2014.
- [34] 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.
- [35] H. Khaledian, R. Sáez, J. Vilà-Valls, and X. Prats, "Interacting multiple model filtering for aircraft guidance modes identification from surveillance data," *Journal of Guidance, Control, and Dynamics*, pp. 1–16, 2023.
- [36] K. Askari and M. Cremaschi, "Simulation-based prediction of departure aircraft performance from surveillance-like data using neural networks," *Aerospace*, vol. 10, no. 6, p. 513, 2023.
- [37] Z. Di, X. Zhuang, and Z. Hongfu, "A hybrid deep neural network based on multi-time window convolutional bidirectional lstm for civil aircraft apu hazard identification," *Chinese Journal of Aeronautics*, vol. 35, no. 4, pp. 344–361, 2022.
- [38] DGAC, "Acropole library github." <https://github.com/DGAC/Acropole>.
- [39] ICAO, "Icao aircraft engine emissions databank." <https://www.easa.europa.eu/domains/environment/icao-aircraft-engine-emissions-databank>.
- [40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [41] W. H. Press and S. A. Teukolsky, "Savitzky-golay smoothing filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.
- [42] M. Shapiro, Z. Engberg, R. Teoh, and T. Dean, "pycontrails: Python library for modeling aviation climate impacts," Mar. 2023.