Approach and landing aircraft on-board parameters estimation with LSTM networks

Gabriel Jarry, Daniel Delahaye, Eric Féron

Approach and landing aircraft on-board parameters estimation with LSTM networks

Gabriel Jarry*, Daniel Delahaye* Eric Feron†

* ENAC, Université de Toulouse, 7 Avenue Edouard Belin, 31400 Toulouse
Email: {jarry.gabriel, daniel.delahaye}@enac.fr
† School of Aerospace Engineering, Georgia Institute of Technology Atlanta GA, USA
Email : feron@gatech.edu

Abstract—This paper addresses the problem of estimating aircraft on-board parameters using ground surveillance available parameters. The proposed methodology consists in training supervised Neural Networks with Flight Data Records to estimate target parameters. This paper investigates the learning process upon three case study parameters: the fuel flow rate, the flap configuration, and the landing gear position. Particular attention is directed to the generalization to different aircraft types and airport approaches.

From the Air Traffic Management point of view, these additional parameters enable a better understanding and awareness of aircraft behaviors. These estimations can be used to evaluate and enhance the air traffic management system performance in terms of safety and efficiency.

Index Terms—Machine Learning, Flight Data Monitoring, Neural Networks, Aircraft Engine Fuel Flow Rate, Supervised Learning, Learning Generalization, Long Short-Term Memory.

I. INTRODUCTION

Nowadays, air traffic controllers and air traffic managers still rely on ground accessible data to understand and analyze aircraft behaviors and to evaluate air traffic management system performance. Nevertheless, aircraft flight dynamic is defined with various on-board parameters. Therefore, the estimation of these parameters could be beneficial to enhance the air traffic system performance and evaluation.

Estimation, sometimes called regression, is a well known mathematical problem. It is divided into two main approaches: the physical model-driven estimation, which had been for a long time the most popular estimation process, and the data-driven regression, which has been more and more used during the last decade with the rise of machine learning and deep learning, showing equal or even more precise predictions.

In the context of Air Traffic Management, the available data are recorded from ground-based surveillance systems, sometimes enhanced with ADS-B data. However, accessing Flight Data Monitoring (FDM) parameters is usually difficult. This paper investigates estimating FDM parameters using ground available parameters with supervised learning models. Though the learning phase still needs the FDR parameters, the learned models can then be used in ground-operations with radar-based data.

In this paper, an estimation of on-board parameters such as fuel flow rate, flap and landing gear setting using a particular neural network architecture called Long Short Term Memory (LSTM) is proposed. Besides, this paper analyses the generalization to different aircraft types and airport approaches.

The study is divided into four parts. Firstly, the state of the art around parameter estimation is presented. Secondly, the learning methodology and data are explained. Thirdly, the trained model is evaluated on a test data set. Finally, the model generalization is discussed.

II. STATE OF THE ART

Physical model-driven parameter estimation consists in using physical equations and models to estimate target features. Usually, differential equation problems are solved or assumptions are taken to give simplified estimations. In aviation, these models have been studied for a long time in different areas such as aerodynamic [1], or fuel-efficiency [2]. These models may need important computational resource.

On the other hand, data-driven models try to estimate the underlying physical models using available data. The learning phase can be costly in computational resource, however, the prediction phase is then instantaneous, which presents interesting properties in real-time operations. Different works were led to compare physical models, such as the Eurocontrol aircraft performance model: BADA [3], to data-driven models [4]. Clemons et al. identified first order enhancements to airport surface fuel burn modeling [5]. Turgut et al. used the BADA model to estimate aircraft fuel flow for a three-degree flight-path-angle descent [6]. Chatter et al. used BADA model to give a fuel burn estimation [7]. Similarly Simone et al. presented a rapid estimation of global civil aviation emissions with uncertainty quantification [8]. Alligier et al. also used the BADA model with ground-based data to estimate the aircraft mass [9], [10].

Many studies have been conducted to predict on-board parameters using ground-based surveillance systems. Delahaye et al. proposed estimations of the True Air Speed and the wind using radar data [11], [12]. Sun et al. estimate the aircraft mass using Bayesian inference methods [13]. Chati et al. proposed different learning model types such as Gaussian Process regression [14] and tree-based classification [15] to predict the aircraft fuel flow rate. The proposed prediction models outperform the classical physical BADA models. Chati

et al. models only rely on ground accessible parameters, their results will be used as a benchmark in this paper.

Other algorithms such as neural networks have also been investigated to estimate the fuel flow rate but with access to many on-board parameters [16]. Peyada et al. presented a new filtering technique based upon a neural network and Gauss-Newton method to estimate aircraft parameters [17], [18]. A recent type of neural network called Long-Short Term Memory has proven to be very efficient on time series data to predict trajectories [19] or hard landings [20]. To the best of the authors' knowledge, LSTM neural networks have not already been used to estimate aircraft on-board parameters such as fuel flow rate or flap and landing gear settings.

III. METHODOLOGY AND DATA

This section presents the machine learning models chosen for the following experiments as well as the data considered.

A. Machine Learning and Multilayer Perceptron Neural Network

A learning process consists in using data analysis methods and artificial intelligence to predict the behavior of a system. The aim is to define a model that will fit as best as possible the system. Machine learning algorithms define learning models h_θ, with parameters θ, that approximate the system function. The learning process is done upon a finite training set $\mathcal{D}$, and aims at minimizing the empirical risk over the training set by tuning the parameters θ of the learning model. The empirical risk is evaluated with an error function, the least square error being the usual choice for regression. [21], [22], [23].

There are various learning models, and in this paper the choice of a particular neural network named Long Short-Term Memory (LSTM) is made and compared to a classical Multi-Layer Perceptron (MLP). LSTM networks were designed as an enhancement of Recurrent Neural Networks (RNN) to perform better supervised learning tasks on time series data [24], [25], [26]. LSTMs are capable of learning long-term dependencies, while simple RNN only learn short term dependencies. LSTMs use a cell state that keeps information from the past, and three gates that update the cell state and compute the prediction. First, the forget gate enables updating the cell state to forget information that is no longer relevant based on the current input. Then, the input gate enables saving in the cell state relevant information from the current input. Finally, the output gate computes the prediction using the updated cell state and the current input. A simplified illustration of an LSTM structure is depicted in Figure 1.

B. From data to features

The data set is composed of 14807 A320 approaches and landing Flight Data Records from two French airlines. This study focuses on predicting on-board parameters such as the flaps configuration, the fuel flow rate, or the landing gear position, knowing the last minute (with a four-second sampling rate) of radar equivalent available parameters: the ground speed, the vertical speed, and the altitude. Additional
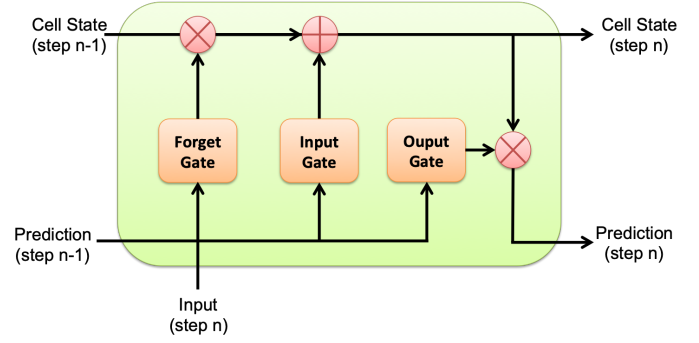

Figure 1: Simplified illustration of the structure of a LSTM cell

features are extracted from these parameters. First, the height is computed with the local QNH. Energy features such as specific energies per unit of mass (potential, kinetic, total) are then computed. In summary, the inputs of the neural network are one-minute time series of ground speed, vertical speed, altitude, height, specific potential energy, specific kinetic energy, and specific total energy.

The data set is first divided into two subsets: 10% for the testing set and 90% for the training set. Regarding the training set, 80% is used for actual training, and 20% for the validation.

Additionally, 200 A330 and 2000 B737 approach and landing Flight Data Records are used to assess the generalization of the model to other aircraft types.
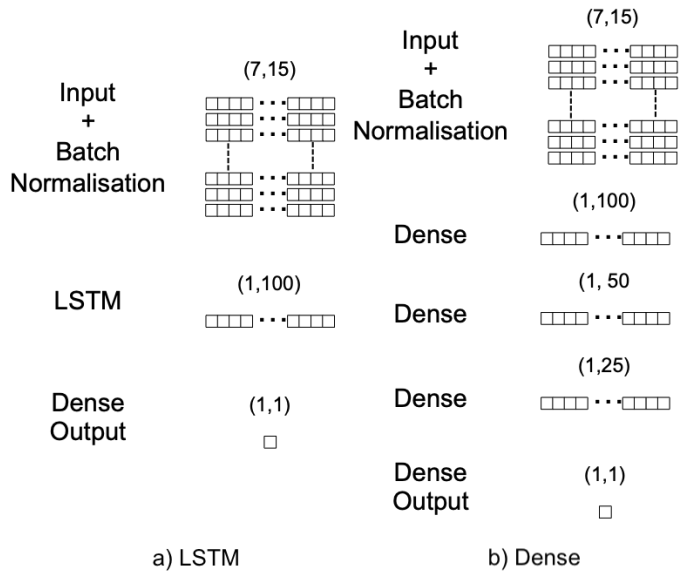
C. Network Architectures and Learning

Figure 2: Description of the neural network architectures. On the left, the LSTM architecture, on the right the MLP Dense architecture.

This section describes the neural network architecture used in the experiments. The LSTM neural network is composed

of two layers and a regression output layer. The first layer is a batch normalization. The second layer is an LSTM layer with 100 units and a ReLu activation function. The output layer is a single neuron dense layer. This architecture will be referred to as LSTM. The dense MLP is composed of one input batch normalization layer and three dense hidden layers with 100, 50 and 25 neurons with ReLu activation. The output layer is a single neuron dense layer. This architecture will be referred to as Dense. Figure 2 illustrates the networks' architecture.

The learning task is made using the Adam optimizer [27] with decay and the loss function used is the least square error. The learning rate is 10^{-3} and the decay is 10^{-9}. Each model is trained during 30 epochs several times on a multi-GPU cluster. The best network over the validation set is kept and analyzed in this paper. The cluster is composed of a dual ship Bi-Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz (80 Core) - with 8 GPU GeForce RTX 2080 Ti.

IV. RESULTS

A. Performance metrics

In order to compare the performance of the different models, four different indicators were used: the Mean Error (ME), the Mean-Absolute Error (MAE), the Normalized Root Mean Squared Error (NRMSE), and the Pearson correlation score. Let $\mathcal{D}$ be a set and h a function.

The mean error **ME** or mean absolute relative prediction error of h over $\mathcal{D}$ is computed using the following formula:

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

The mean-absolute error **MAE** is calculated using the following formula:

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

As for the **ME**, the smaller its value is, the more accurate the prediction is.

Regarding the Normalized Root Mean Squared Error **NRMSE**, similarly the smaller its value is, the more accurate the prediction is. It is calculated using the following formula

$$\mathbf{NRMSE}(h, \mathcal{D}) = \frac{1}{\sigma_{\mathcal{D}}(h)} \times \sqrt{\frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} (h(x) - y)^2} \quad (3)$$

where $\sigma_{\mathcal{D}}(h)$ is the standard deviation of the predicted value over $\mathcal{D}$

The Pearson correlation score indicates how well the curve of the predicted parameter follows the actual curve. The closer the score is to 1, the better the prediction is. It is calculated using the following equations:

$$r(h, \mathcal{D}) = \frac{\sum_{(x,y) \in \mathcal{D}} (h(x) - \bar{h})(y - \bar{y})}{\sqrt{\sum_{(x,y) \in \mathcal{D}} (h(x) - \bar{h})^2} \sqrt{\sum_{(x,y) \in \mathcal{D}} (y - \bar{y})^2}} \quad (4)$$

where $\bar{h}$ (resp. $\bar{y}$) is the average of $h(x)$ (resp. y) over $\mathcal{D}$.

In practice, to evaluate the fuel flow rate estimation, the **ME**, the **NRMSE** and the Pearson score are computed per trajectory and a global average score is given. Besides, the fuel consumption (in kg) is computed per trajectory and a **ME** score is given for all the trajectories. For the landing gear position and the flap setting, the remaining distance to the runway threshold is computed per trajectory when gear down (resp. flaps 1, 2, etc.) action is applied for both actual and predicted parameter. For example, if real flaps 1 appeared at 11NM and the model predict 10.5NM, the **MAE** error is 0.5NM. Then, the **MAE** score is averaged for all the trajectories to give an aggregated score. Therefore, it gives the average absolute distance difference between real and predicted values.

B. Airbus A320 - test set

In this section are presented the performance metrics on the A320 test set of 1410 flights.

1) Fuel flow rate: Both LSTM and MLP models show high scores on the fuel flow rate estimation. The average scores for all the 1410 flights are summarised in Table I. The high Pearson scores of 0.938 for LSTM model, and 0.927 for Dense model, indicate a high correlation between the real fuel flow and the prediction. This behavior is well illustrated in Figure 3. Indeed, the prediction seems to follow properly each variation of the real fuel flow rate. Even though the proposed models do not use any mass or take-off mass information, the **ME** and the **NRMSE** scores of the proposed models are similar to those described by Chati et Al. in [14], [15] for the descent and approach phases. The LSTM **ME** score is on average around 13.59% with a standard deviation of 3.26%, and the **NRMSE** is around 26.52% with a standard deviation of 8.41%. The fuel consumption **ME** for the test data set is around 4.2% (LSTM) and 4.5% (Dense), corresponding to an average absolute error of 8 kg. It seems to indicate that a performant estimation of the fuel flow rate can be given without any mass information. Furthermore, the LSTM network outperforms the Dense network.

Metric	LSTM	Dense
ME (%)	13.59 (3.26)	14.55 (3.97)
NRMSE (%)	26.52 (8.41)	27.97 (8.70)
Pearson	0.938 (0.046)	0.927 (0.065)
Fuel Consumption - **ME** (%)	4.2 (3.28)	4.52 (3.56)

Table I: Predictive performance of the LSTM and the Dense models on A320 test data for the fuel flow rate. Each entry shows the mean and the standard deviation (within parentheses) of the evaluation metric across all the 1410 flights in the test data set.

2) Landing gear and flap setting: In this section, the performance results of the landing gear and flap setting predictions are analyzed. Only the LSTM model performance metrics are presented here since it outperforms the Dense model. The average **MAE** performance and its standard deviation (within parentheses) are summarized in Table III. The landing
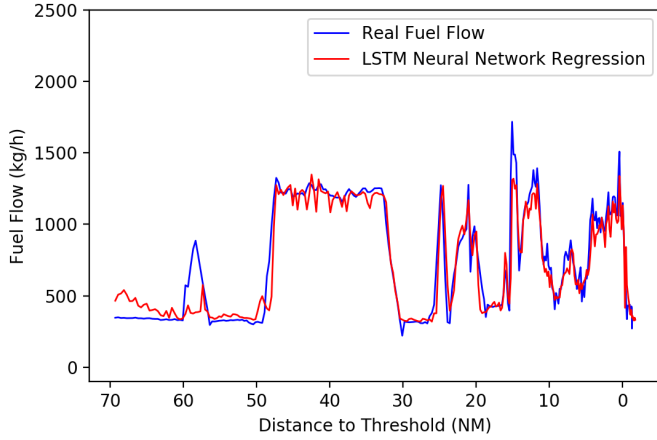
Figure 3: Illustration of the fuel flow rate prediction on a flight of the test data set
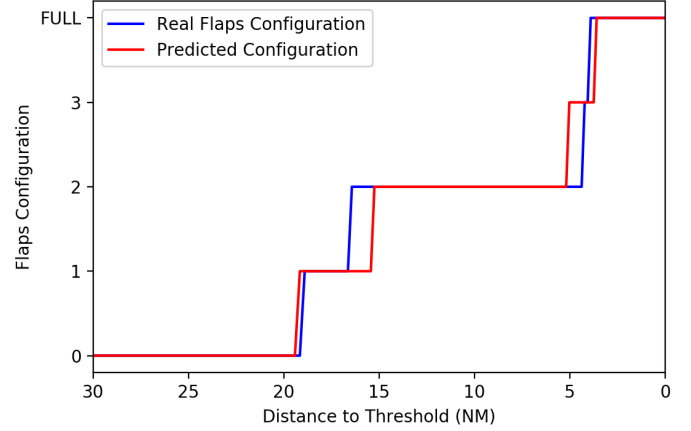

Figure 4: Illustration of the flap configuration prediction on a flight of the test data set

gear down action is well localized with an average **MAE** of 0.99NM. Regarding the flap setting, the average distance error goes from 2.27NM for flaps 1 to 0.88NM for flaps FULL. Examples of flap and landing gear setting prediction are illustrated in Figures 4 and 5. The configuration prediction might be enhanced if extra information such as the air speed (available with radar mode S) are added. Indeed, these parameters are directly linked to the aircraft flight dynamics and the airspeed.

Parameter	**MAE** (%)
Landing Gear	0.99 (2.38)
Flaps 1	2.27 (4.03)
Flaps 2	1.67 (2.35)
Flaps 3	0.97 (1.15)
Flaps FULL	0.88 (1.03)

Table II: Predictive performance of the LSTM models on A320 test data for the flap and landing gear setting. Each entry shows the mean and the standard deviation (within parentheses) of the evaluation metric across all the 1410 flights in the test data set.

V. GENERALIZATION

In this section, the generalization of the fuel flow rate and the landing gear configuration learning model to other aircraft types is investigated. Indeed, the model is built with A320 flight data records. The access to all the aircraft types might be very difficult. Therefore, the performance of the learning model on two other aircraft types (B737 and A330) is analyzed. The flap setting model generalization is not investigated since the flap configurations are different depending on the aircraft type.

A. Boeing B737

The B737 data set is composed of 2000 FDR approach and landing trajectories equally divided into two subsets. The first
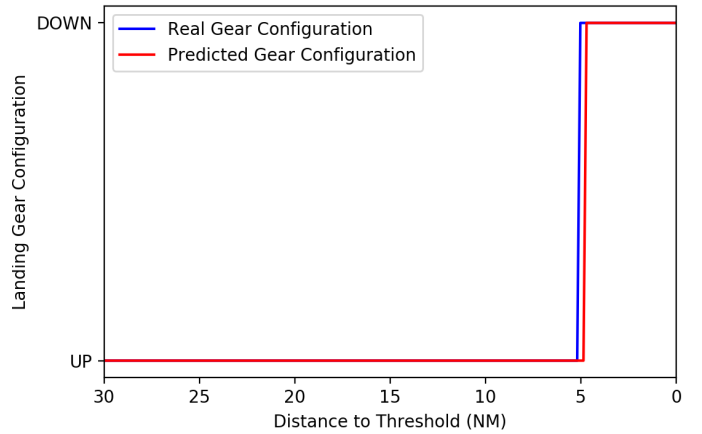

Figure 5: Illustration of the landing gear configuration prediction on a flight of the test data set

one is located at Paris-Orly Airport (LFPO), and the second at Madrid-Barajas Airport (GMAD).

For both airports the fuel flow rate model is still performing well with a correlation score of 0.917 (LFPO) and 0.921 (GMAD) as summarized in Table III. The **ME** (15.57% and 15.29%) and **NRMSE** metric 32.94% (LFPO) and 31.7% (GMAD) are not as good as the A320 test data set. Indeed, like in the example shown in Figure 6, there is a small offset between the real and the predicted fuel flow rate. Even though the A320 training set did not contain any flight at GMAD airport, the fuel flow rate model seems to perform well on this new platform. Moreover, the fuel consumption **ME** is on average of 4.35% (LFPO) and 4.85% (GMAD), which is similar to the mean error for the A320 test set. It confirms that the model is generic and can be used for any approach at any airport. Furthermore, the generalization to B737 seems to be relevant without any calibration. It is less precise than the
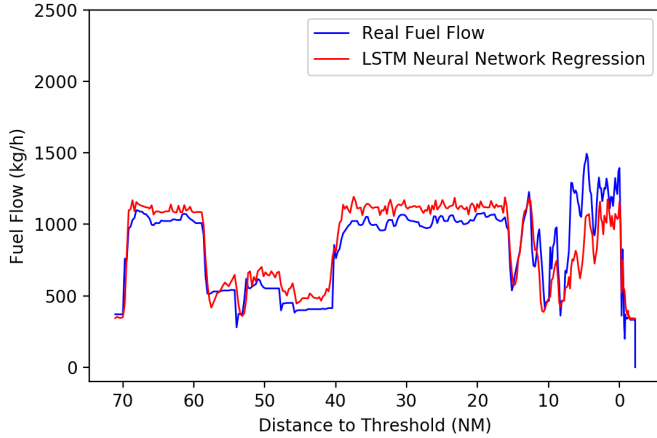
Figure 6: Illustration of the fuel flow rate prediction on a flight of the B737 test data set
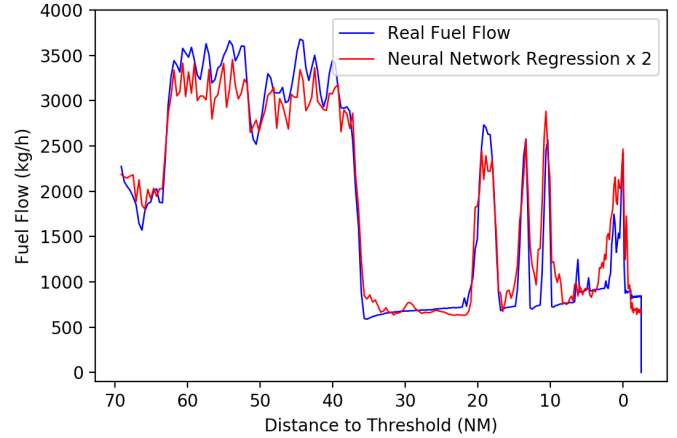

Figure 7: Illustration of the fuel flow rate prediction on a flight of the A330 test data set

A320 but depending on the use, the model might be efficient enough.

Regarding the landing gear configuration, the distance **MAE** is higher than for the A320 with 1.23NM and 1.92NM. The generalization process for the GMAD data set seems to be even more difficult.

Metric	B737 - LFPO	B737 - GMAD
ME (%)	15.57 (3.7)	15.29 (3.51)
NRMSE (%)	32.94 (10.25)	31.7 (8.87)
Pearson	0.917 (0.045)	0.921 (0.042)
Landing Gear - **MAE** (NM)	1.23 (2.21)	1.92 (3.52)
Fuel Consumption - **ME** (%)	4.35 (3.36)	4.86 (3.81)

Table III: Predictive performance of the LSTM model on B737 test data for the fuel flow rate and landing gear at LFPO and GMAD airport. Each entry shows the mean and the standard deviation (within parentheses) of the evaluation metric across all the flights in the test data set.

B. Airbus A330

The A330 data set is composed of 200 FDR approach and landing trajectories at Paris Orly Airport (LFPO).

The generalization of the fuel flow rate model is not as simple as the B737 since the fuel flow rate values are not in the same range. Therefore, a very simple optimization problem is solved to define a scaling coefficient: c. This method could have also been applied to the B737. With the A330, $c = 2$ is selected and the result of the prediction is illustrated in Figure 7. With the BADA model, it is possible to give a ratio of the fuel flow coefficient of the A330 over the A320. The coefficient is given per altitude and summarized in Table IV. The coefficients are similar (around 2 and increasing up to 2.5) to the coefficient obtained with the optimization problem. It is possible to enhance the optimization problem with a discretization per flight level. The problem will estimate different values minimizing the error. On the other hand, it

gives a possible solution without on-board real fuel flow rate for a particular aircraft type. One can refer to the coefficient ratio given by the BADA model to scale the A320 model and apply it to estimate the fuel flow rate to a new aircraft type.

Altitude (FL)	Coefficient ratio A330 / A320
0	1,9678
5	2,0333
10	2,0774
15	2,1309
20	2,1864
30	2,3018
40	2,4246
60	2,5184
80	2,5416
100	2,5422

Table IV: Fuel flow BADA coefficient ratio between A320 and A330 depending on the flight level

$$\underset{c \in \mathbb{R}^+}{\text{minimize}} \ \mathbf{ME}(c \times h, \mathcal{D}) \tag{5}$$

Moreover, the result of the performance metrics on the scaled LSTM model is summarized in Table V. The **ME** is worst than the B737 with on average 20,77%. However, the Pearson correlation score is high with 0.93 and the **NRMSE** is around 33.18 similar to the B737. The fuel consumption **ME** is on average 4.84%, it is still acceptable with a small increase compared the A320 test set. The generalization to the A330 seems to be relevant with the scaled model and acceptable depending on the use.

Regarding the landing gear setting, the distance **MAE** is higher than for the A320 and similar to the B737 data set with 1.63NM.

VI. DISCUSSION & CONCLUSION

This paper investigates the uses of LSTM neural networks in aircraft on-board parameter estimation with ground-based

Metric	A330 - LFPO
ME (%)	20.77 (4.4)
NRMSE (%)	33.18 (9.94)
Pearson	0.930 (0.041)
Landing Gear - **MAE** (NM)	1.63 (1.85)
Fuel Consumption - **ME** (%)	4.84 (3.53)

Table V: Predictive performance of the scaled LSTM model on A330 test data for the fuel flow rate at LFPO airport. Each entry shows the mean and the standard deviation (within parentheses) of the evaluation metric across all the flights in the test data set.

surveillance system features. Three parameters are estimated: the fuel flow rate, and the flap and landing gear configurations.

The study shows that even without any mass information, satisfying estimations of fuel flow rate can be computed. Estimation curves present high correlations with the real curves and the error seems to be non-significant. Moreover, the trained models can be generalized to different airport approaches. Furthermore, even if they are trained with only one aircraft type, generalization to other aircraft can be made with a scaling coefficient. The residual error seems to be acceptable depending on the use.

Regarding the flap and landing gear setting, adding extra information such as the air speed might enhance the estimation. The generalization task of the landing gear models presents higher difficulties for new airports and aircraft types.

Future works will focus on enhancing flap and landing gear setting prediction with airspeed information. Besides, air traffic performance metrics will be designed to use on-board parameter estimation models and studies will be conducted on how well these models enable enhancing the air traffic system performance evaluation.

ACKNOWLEDGMENT

REFERENCES

[1] R. V. Jategaonkar and F. Thielecke, "Aircraft parameter estimation—a tool for development of aerodynamic databases," *Sadhana*, vol. 25, no. 2, pp. 119–135, 2000.

[2] B. P. COLLINS, "Estimation of aircraft fuel consumption," *Journal of Aircraft*, vol. 19, no. 11, pp. 969–975, 1982.

[3] A. Nuic, "User manual for the base of aircraft data (bada) revision 3.7," *Atmosphere*, vol. 2010, p. 001, 2010.

[4] 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.

[5] 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*, 2018, p. 3991.

[6] E. T. Turgut, "Estimating aircraft fuel flow for a three-degree flight-path-angle descent," *Journal of Aircraft*, vol. 48, no. 3, pp. 1099–1106, 2011.

[7] 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*, 2011, p. 6881.

[8] N. W. Simone, M. E. Stettler, and S. R. Barrett, "Rapid estimation of global civil aviation emissions with uncertainty quantification," *Transportation Research Part D: Transport and Environment*, vol. 25, pp. 33–41, 2013.

[9] R. Alligier, D. Gianazza, and N. Durand, "Ground-based estimation of aircraft mass, adaptive vs. least squares method." ATM Seminar, 2013.

[10] 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*, 2014, pp. pp–xxxx.

[11] D. Delahaye and S. Puechmorel, "Tas and wind estimation from radar data," in *2009 IEEE/AIAA 28th Digital Avionics Systems Conference*. IEEE, 2009, pp. 2–B.

[12] ——, "Aircraft local wind estimation from radar tracker data," in *2008 10th International Conference on Control, Automation, Robotics and Vision*. IEEE, 2008, pp. 1033–1038.

[13] 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.

[14] 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*. ACM, 2017, pp. 131–140.

[15] ——, "Statistical modeling of aircraft engine fuel flow rate," in *30th Congress of the International Council of the Aeronautical Science*, 2016.

[16] A. A. Trani, F. Wing-Ho, G. Schilling, H. Baik, and A. Seshadri, "A Neural Network Model to Estimate Aircraft Fuel Consumption," 2004.

[17] N. Peyada and A. Ghosh, "Aircraft parameter estimation using a new filtering technique based upon a neural network and gauss-newton method," *The Aeronautical Journal*, vol. 113, no. 1142, pp. 243–252, 2009.

[18] ——, "Aircraft parameter estimation using neural network based algorithm," in *AIAA atmospheric flight mechanics conference*, 2009, p. 5941.

[19] Z. Shi, M. Xu, Q. Pan, B. Yan, and H. Zhang, "LSTM-based Flight Trajectory Prediction," Jul. 2018, pp. 1–8.

[20] H. Zhang and T. Zhu, "Aircraft Hard Landing Prediction Using LSTM Neural Network," in *Proceedings of the 2Nd International Symposium on Computer Science and Intelligent Control*, ser. ISCSIC '18. New York, NY, USA: ACM, 2018, pp. 28:1–28:5.

[21] V. Vapnik, *Statistical learning theory*, ser. Adaptive and learning systems for signal processing, communications, and control. Wiley, 1998. [Online]. Available: https://books.google.fr/books?id=GowoAQAAMAAJ

[22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer*, vol. 27, no. 2, pp. 83–85, 2005.

[23] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[24] Y. Bengio, P. Simard, P. Frasconi *et al.*, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[25] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.