Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023)

Looking for the best flight efficiency indicators for arrivals

Pierrick Pasutto and Karim Zeghal
EUROCONTROL Innovation Hub
Brétigny-sur-Orge, France

***Abstract*—This study, conducted as part of an official working group with European air navigation service providers, explores various multi-dimensional flight efficiency indicators for arrivals. The objective is to find simple and intuitive indicators, easy to process and reproduce, that can best approximate fuel burn. We identified eleven indicators, covering the three dimensions (horizontal, vertical and speed), and decomposed them in two sub-indicators to capture permanent vs variable inefficiencies as proxies for airspace vs operations. We then considered all combinations and integrated each in a linear regression, calibrated per aircraft type on the top 30 European airports in the last 50NM, over 2.8 million flights. Predicting fuel burn in excess with a single indicator leads to diverse performances (mean absolute error ratio ranging from 25% to 66%). Combining indicators can improve performances (22% with two, 20% with three, 18% with four and 16% with the eleven) but adds complexity. The simplest indicators on the three dimensions (time, altitude and speed differences) show fair performances combined (24%). A good compromise may be with the two of the top 3 indicators (altitude average and time difference, 23% combined) possibly complemented by the corresponding speed indicator (speed average, 21% combined). Future work should investigate whether the prediction performance may be further improved.**

Keywords-component; flight efficiency, metrics and indicators, terminal area.

I. INTRODUCTION

Arrival flights are subject to significant inefficiencies in the terminal area. In a previous work [1], we estimated them to be in the order of 40% for the top 27 European airports[1]. These flight inefficiencies are caused by various factors (airspace complexity, traffic congestion, environmental constraints, …). Quantifying them and identifying their natures is essential to investigate if and how they may be attenuated. While fuel burn should be the final "judge", it has been documented that fuel models are sensitive to input data and show disparities in their estimations [2, 3]. In contrast, geometric indicators are easier to process and manipulate, enabling a form of transparency and reproducibility. They may also provide richer information such as revealing the most penalizing dimension (horizontal or vertical). Further to this, they can also be used during design (or re-design) phases, to help identify trade-offs and options when no traffic data is available for a fuel burn assessment. However, geometric indicators may not perfectly reflect all the inefficiencies and can only constitute proxies. Based on our experience in supporting air navigation service providers, we consider that fuel burn and geometrics indicators are complementary to cover all phases, from design to monitoring of operations.

The present study, conducted as part of an official working group with European air navigation service providers, explores various multi-dimensional flight efficiency indicators for arrivals and evaluates their accuracy against fuel burn. The objective is to offer a choice of simple and intuitive indicators that can best approximate fuel burn. These indicators may then be used for different purposes: to support the design of new procedures or operating methods, to identify inefficiencies and prioritize improvements, to assess and monitor the current situation. The paper is structured in three main parts, after the state of the art: (1) exploration following a form of systematic approach to find or develop (existing or new) indicators; (2) evaluation based on objective (fuel burn prediction) and subjective criteria (simplicity, intuitiveness); (3) application to the top 30 European airports within 50NM over 2.8 million flights (2019).

II. STATE OF THE ART

The assessment of flight efficiency requires, by definition, quantitative indicators. In 2003, Howell et al. [4] posed a ground principle by defining "en route inefficiency as the distance, flight time, or fuel consumption (...) in excess of what would occur if each sampled flight were the only aircraft in the system". At the same time, in Europe, Chesneau et al. [5] proposed route extension and extra fuel burn indicators resulting from the comparison of actual flights with direct or optimum reference profiles. In later works [6, 7], fuel burn indicator was generally not retained, due to its calculation complexity and inherent uncertainties. In 2009, Reynolds [2] quantify flight efficiency in different flight phases using track extensions and fuel metrics. More recently, Prats et al. proposed in [8] a new set of environmental performance indicators based on distance and

[1] Top 30 minus the three Turkish airports for which data is not available. In the following, we will consider the top 30 without the Turkish airports.

fuel, distinguishing different layers of inefficiencies (air traffic management, airspace users …). The influence of sub-optimal altitude or speed constraints was highlighted in both studies. On vertical dimension, Knorr et al. proposed in [9] an approach to identify the ATC constraints that impact the trajectory in the descent phase: vertical inefficiency is derived there from the distance flown on level segment and translated into excess fuel burn, assuming the shifting of level segments at cruise level. Such approach was applied and further developed through several works [10, 11] or used, in conjunction with time and fuel metrics to evaluate the impact of weather phenomena [12]. Many of those metrics are provided by the EUROCONTROL Performance Review Unit (PRU) to characterize air traffic management environmental performances at European level [13, 14]. In addition, fuel inefficiency metrics solely based on statistical principles were also developed to estimate excess fuel burn within the EUROCONTROL Network Manager (NM) area [15].

Other studies aim to establish a link between fuel (or excess fuel) consumption and a set of explanatory variables. In 2012, NATS presented a new metric called the 3Di Score developed as an option to enable assessment of fuel efficiency. This metric, described in [16] and further discussed in [17], is derived from factors relating to track extension and vertical inefficiency and calibrated using a fuel burn linear regression model, considering UK domestic traffic. In 2014, Ryerson et al., in [18] analyzed the actual fuel burn using econometric techniques to isolated different contribution (airborne and departure delay, terminal area inefficiencies…). Focusing on enroute part, Calvo et al. studied in [19] the correlation between usual horizontal efficiency metrics and estimated fuel efficiency to propose enhanced flight efficiency indicators. Using more sophisticated model, Jarry et al. [20] used aircraft data (including fuel burn) to develop environmental indicators.

In a previous work [21], we developed two flight efficiency indicators for arrivals, inspired from [13] and enabling the identification of airspace vs operations contributions. These indicators generated an interest from our stakeholders, however, the question that arose was: do they actually inform on fuel burn? Could they be considered as a proxy? In [1], we investigated whether a simple linear regression model with these two indicators, similarly to [16], could provide an acceptable prediction of the excess fuel burn. With a model calibrated on the top 27 European airports for the last 50NN, we obtained a fair correlation (R^2 0.89) and prediction (standard deviation of ±28kg for a median excess fuel burn of 40kg and 49kg for airspace and operations respectively). The question is now: could we find other indicators (existing or new) that would better approach fuel burn? A secondary question is: could we find a simpler vertical indicator? Indeed, the one developed was considered quite unusual as relying on a notion of surface (altitude difference × flight time). The present work may thus be seen as a form of generalization of [13] and [16].

III. Exploring Indicators

This section presents the principles of deviations from references and the definition of these reference, followed by a list of possible indicators to measure these deviations, and ends with an overview for the top 30 airports.

A. Deviation from references

Figure 1 below shows a typical arrival flow to a European airport, illustrating the two types of inefficiencies to be distinguished:

- **permanent**: any horizontal (e.g. path extension) or vertical restrictions (e.g. level-off), reflecting fixed procedures due to airspace (e.g. segregation of flows) or environmental constraints (e.g. non overfly areas, intercept altitude);
- **variable**: any tactical interventions (e.g. vectoring), reflecting the operations of sequencing and metering, and separation management, as instructed by controllers and executed by flight crews in given meteorological conditions.

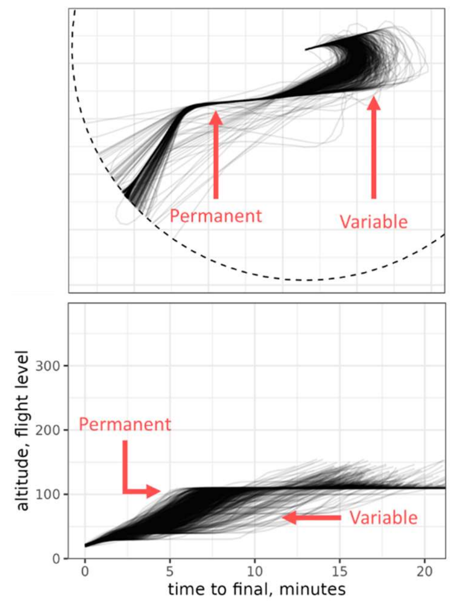


Figure 1. Permanent and variable inefficiencies

For this purpose, the idea is to consider two reference trajectories (Figure 2):

- **ideal unconstrained**: an ideal flyable trajectory with no constraints, defined by a common principle for all airports (e.g. direct from entry, continuous descent);
- **ideal constrained**: an ideal flyable trajectory incorporating any horizontal, vertical or speed restrictions applicable to arrivals to each airport.

These references are defined per airport, runway and flows.

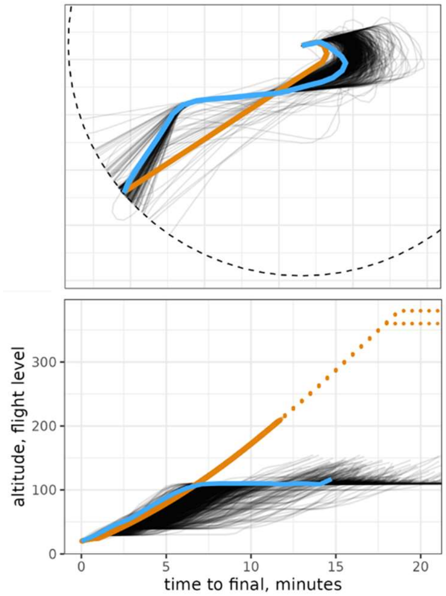


Figure 2. Ideal unconstrained (orange) and constrained (blue) references

The deviation from these references may then be expressed in different ways (e.g. time or altitude difference, fuel in excess) that will constitute the set of proposed indicators (section C). The deviation of a given trajectory from the ideal unconstrained reference captures a form of overall inefficiency, integrating both permanent and variable inefficiencies. It may be used for performance assessment. The permanent inefficiency, a proxy for **airspace** and environment, will be captured by the difference between ideal unconstrained and ideal constrained references (Figure 3). The variable inefficiency for a given trajectory, a proxy for **operations**, will be captured by the difference between the ideal constrained reference and this trajectory.

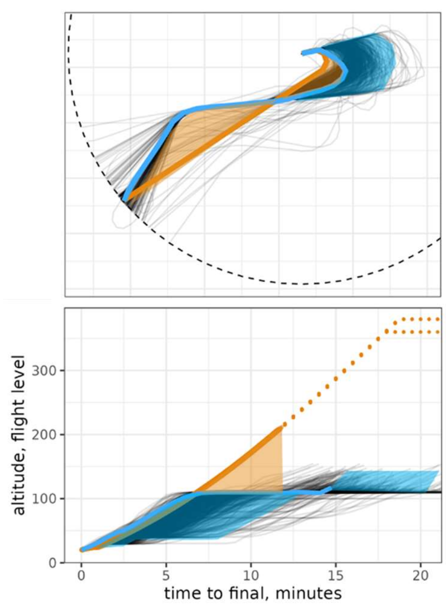


Figure 3. Airspace (orange) and operations (blue) related inefficiencies

B. *Defining references*

The **ideal unconstrained** reference is defined in the same way for all airports, based on simple considerations (other considerations may be added such as meteorological conditions). It is defined as the shortest direct route from entry (50NM) to the intercept point, with a final turn to intercept (shortest direction, constant rate), an intercept altitude at 2000ft (above ground level) followed by 30 seconds level-off and a straight segment until 5NM to threshold. The vertical profile until intercept is a continuous descent (standard 3-degree descent slope) starting at the highest flight level observed within 300NM (per aircraft type and city pair distance). The speed profile is defined as the maximum of the 90th percentiles of the speeds observed on all airports. Note: to measure a deviation from a given trajectory, the reference is extended until reaching the same flight time.

The **ideal constrained** reference could have been defined using published procedures. However, this would have been extremely effort demanding and uneasy for terminal areas relying on vectoring (sort of "free" area). We have thus decided to rely on a statistical approach, considering the flown trajectories (per arrival flow and aircraft type) and extract the "best" ones using a percentile approach (10th in the following). Taking the example of time, the 10th percentile yields to the blue trajectory shown above. This can be generalized: considering an indicator (e.g. time or altitude difference, fuel in excess), the ideal unconstrained reference for this indicator is defined (per flow and aircraft type) as a given percentile (10th) of the data sample.

For a given indicator, to generalize the principle introduced previously, the airspace related inefficiency is obtained by the difference of the indicator between unconstrained and constrained references; the operations related inefficiency by the difference of the indicator between the ideal unconstrained and the given trajectory (Figure 4).

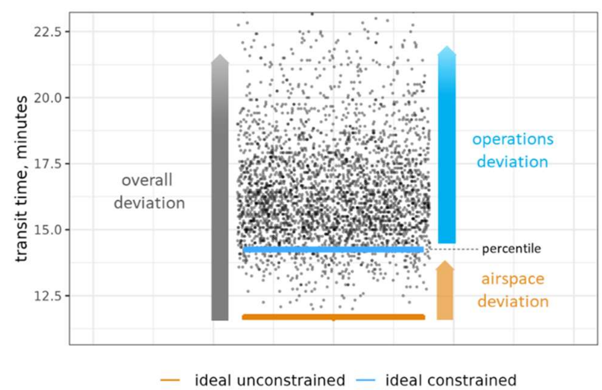


Figure 4. Airspace and operations

C. *Measuring deviation*

To identify potential indicators, we followed a form of systematic approach mixing dimensions and measurements. For the three dimensions, we considered: time or distance, altitude and speed. For time, we may consider the simplest measurement consisting of taking the difference of transit times between a given trajectory and the reference (ideal unconstrained). This measurement may be applied to distance, altitude and speed.

For altitude, in our previous analysis we considered the difference of surfaces along flight time (Figure 5) that brings a notion of cumulated deviation (integral). We may also consider the difference of average altitudes and the difference of slopes (notion of derivative), as well as the difference of flight times in level-off (Figure 5). These three measurements (surface, average and slope) may also be applied to speed[2].

We may note two types of indicators: one-dimensional as solely based on a difference of transit (time, distance, altitude and speed differences); two-dimensional as incorporating flight time (surface, average, slope, level-off).

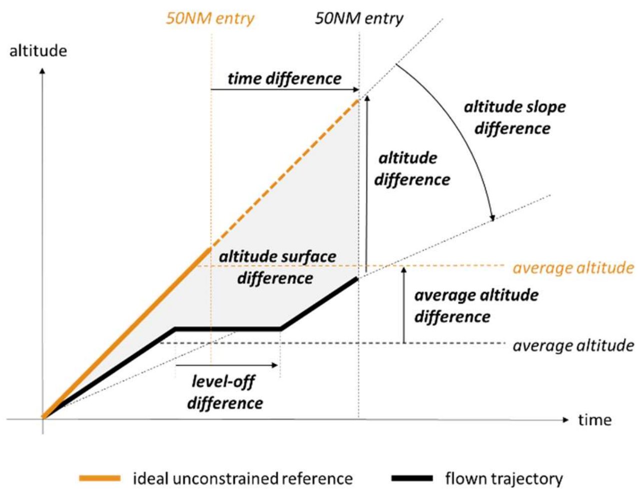


Figure 5. Illustration of time and altitude deviation indicators

We have in total eleven geometric indicators (Table I), some already existing and others new. In the following, the term "difference" will be omitted for average, surface, slope and level-off.

TABLE I. PROPOSED GEOMETRIC DEVIATION INDICATORS

	Time Distance	Altitude	Speed
Difference	✓	✓	✓
Average difference		✓	✓
Surface difference		✓	✓
Slope difference		✓	✓
Level-off difference		✓	

The indicators are then decomposed in two sub-indicators to capture airspace vs operations, as defined previously. The same principle is used to define the indicator of fuel burn in excess, using the ideal unconstrained and determining the ideal constrained as defined in section B.

We may note that the time difference for operations is similar to the ASMA additional time of the EUROCONTROL PRU that relies on the notion of unimpeded time [13, 14] (differences are 40NM horizon and 5th percentile which leads to lower unimpeded times hence higher additional times). The time level-off "overall" (i.e. merging airspace and operations) is similar to the continuous descent indicator of the PRU but with a different horizon (200NM and without counting cruise levels). The fuel burn in excess for operations is similar in principles (percentile approach) to the fuel burn indicator "wheels-off / wheels-on" developed by EUROCONTROL NM [15] but restricted in our study to 50NM.

D. *Data*

The data set initially contains more than 4.1 million flights arriving to the top 30 European airports in 2019. Data consists of position and altitude reports with an average update rate of 30 seconds. We consider an area within 50NM radius centered around each airport until 5NM to runway threshold. A preliminary data preparation allocates the landing runway and identifies the entry point in the 50NM using statistical clustering. Flights with data issues are filtered (e.g. gap between two tracks, 1.8%) as well as unusual flights (e.g. go-around or calibration, 0.4%). We focus on day time operations and exclude night time operations (9pm to 7am local time, 21.3%) that may rely on specific noise abatement procedures and would require a separate analysis. We also focus on the main flows and exclude the minor ones (less than 10% of the arrivals per runway, 9.3%). After filtering, the data set contains more than 2.8 million flights (67.2%).

The computation of fuel burn follows different steps. From the track data, we construct ground and vertical speed profiles. A filter attenuates the erroneous acceleration peaks due to surveillance issues that would lead to an overestimation of fuel burn. Then, we reconstruct a true airspeed profile, relying on wind archive data [22][3]. The fuel burn is estimated with a dedicated tool developed for BADA 4.2, considering the vertical and speed profiles, the aircraft model and the distance flown. A weight of 80% of the maximum landing weight is assumed when entering the 50NM (standard assumption from the BADA team). For each individual flight, the resulting fuel burn estimation on every point (30 seconds), is integrated to get the estimation of the total fuel burn within the 50NM.

E. *Overview*

To give an initial view of the orders of magnitude, we present average and median values of the fuel burn and the geometric indicators for the 30 airports globally (Figure 6 and Table II).

Looking at average values, the excess fuel burn (i.e. the fuel burn in excess compared to the ideal unconstrained reference) is 150kg, for a transit value in the 50NM area of 309kg (ideal unconstrained fuel burn: 159kg). The excess fuel burn may be decomposed in 67kg for airspace and 83kg for operations.

---

2 For time and distance, we only consider the difference of transits as we did not find an easy way to interpret the other measurements.

3 We are investigating whether we may directly rely on mode S data.

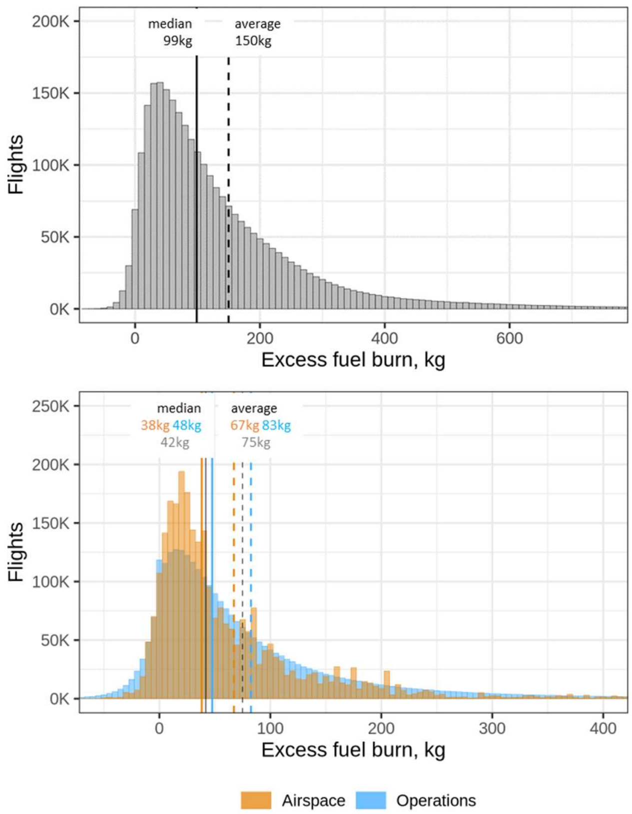


Figure 6. Excess fuel burn distribution
(overall top, airspace / operations bottom)

TABLE II. AVERAGE AND MEDIAN VALUES OF FUEL AND GEOMETRIC INDICATORS (TOP / BOTTOM)

Indicator	Transit	Deviation		
		Overall	**Airspace**	**Operations**
fuel, kg	309 247	150 99	67 38	83 48
time (difference), min	15.3 14.6	4.5 3.7	1.7 1.4	2.8 2.0
distance (difference), NM	67 64	12.2 8.4	3.8 2.1	8.4 4.9
altitude (difference), ft	15000 15000	10800 10000	5300 4400	5500 4600
altitude (average), ft	8800 8700	4800 4300	2000 1700	2800 2200
altitude (surface), ft×10min	10500 9800	8800 6400	2900 2200	5900 3600
altitude (slope), ft/min	1030 1020	660 680	370 340	290 270
time (level-off), min	2.1 1.1	1.6 0.6	0.0 -0.5	1.6 1.0
speed (difference), kn	198 198	71 72	21 17	50 48
speed (average), kn	261 261	79 77	43 39	36 33
speed (surface), kn×10min	165 158	89 71	17 10	79 55
speed (slope), kn/min	14 13	4.4 4.6	1.4 1.3	3.1 3.0

The time difference (i.e. the time in addition to the ideal unconstrained) is 4.5 minutes for a transit time of 15.3 minutes (ideal unconstrained transit time: 10.8 minutes). The altitude difference (altitude lower than the ideal unconstrained) is 10800ft for a transit altitude of 15000ft (ideal unconstrained transit altitude: 25800ft). Similar interpretations may be made for the other indicators. Overall, whatever the indicator considered, we may observe a significant ratio of inefficiency, with a contribution of operations higher than airspace.

IV. EVALUATING INDICATORS

This section presents the different steps followed to assess the prediction accuracy of each indicator, and how to select them. For the calibration, we retained the 10 most represented aircraft type (A320, B738, A319, A321, E190, A20N, CRJ9, DH8D, E195 and B737) accounting globally for more than 75% of the dataset (2.09 million flights). We observed varied traffic mix among the airports which may affect the correlation for those having a mix differing from the average.

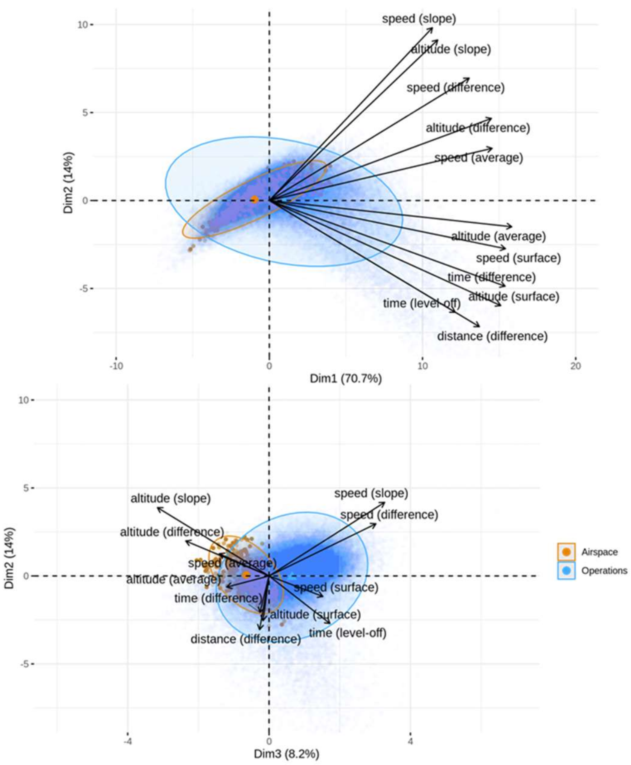
A. Correlation

To get an insight on the geometric indicators and how they relate to each other, we first present a correlation matrix (Table III). We observe strong correlations between some indicators, for instance time difference and distance difference (0.94) or time difference and altitude surface (0.96), speed difference and speed slope (0.96); medium or low correlations between others, generally with slope (altitude or speed), speed difference or level-off time. At this stage, is it unclear whether the diversity of correlations may enrich the prediction of fuel burn in excess.

TABLE III. CORRELATION MATRIX BETWEEN INDICATORS (A320)

	time (difference)	distance (difference)	altitude (difference)	altitude (average)	altitude (surface)	altitude (slope)	speed (difference)	speed (average)	speed (surface)	speed (slope)	level-off (time)
time (difference)	1	0.94	0.73	0.91	0.96	0.45	0.61	0.81	0.92	0.42	0.71
distance (difference)	0.94	1	0.62	0.81	0.89	0.34	0.47	0.63	0.82	0.29	0.70
altitude (difference)	0.73	0.62	1	0.87	0.71	0.91	0.67	0.85	0.69	0.58	0.51
altitude (average)	0.91	0.81	0.87	1	0.94	0.67	0.65	0.85	0.88	0.49	0.70
altitude (surface)	0.96	0.89	0.71	0.94	1	0.42	0.57	0.75	0.92	0.37	0.79
altitude (slope)	0.45	0.34	0.91	0.67	0.42	1	0.56	0.71	0.41	0.55	0.25
speed (difference)	0.61	0.47	0.67	0.65	0.57	0.56	1	0.68	0.76	0.96	0.50
speed (average)	0.81	0.63	0.85	0.85	0.75	0.71	0.68	1	0.79	0.57	0.46
speed (surface)	0.92	0.82	0.69	0.88	0.92	0.41	0.76	0.79	1	0.60	0.74
speed (slope)	0.42	0.29	0.58	0.49	0.37	0.55	0.96	0.57	0.60	1	0.35
time (level-off)	0.71	0.70	0.51	0.70	0.79	0.25	0.50	0.46	0.74	0.35	1

We may complement this view by a Principal Component Analysis (Figure 7) and could notice that 85% of the variability may be explained by two dimensions and 93% by three.



Figures 7. Principal component analysis (A320)

B. *Linear regression*

The objective is to evaluate the indicators in relation to their capacity to predict the fuel burn in excess. We consider two models: one to predict the overall excess fuel burn (i.e. with airspace and operations merged) and a second one to predict airspace and operations separately. The first one is to relate to other works, while the second is for the analysis. We consider a linear regression (per aircraft type), similarly to [16], as it is simple, explainable and easy to use when the model is calibrated. In short, this translates into finding the coefficients (for each aircraft type) minimizing the prediction error.

For overall excess fuel burn model:

$$\alpha \times \{indicator_{overall}\} = Predicted\ excess\ fuel_{\ overall}$$

For airspace / operation excess fuel burn model:

$$\alpha' \times \{indicator_{airspace}\} = Predicted\ excess\ fuel_{\ airspace}$$

and

$$\alpha' \times \{indicator_{operations}\} = Predicted\ excess\ fuel_{\ operations}$$

minimize the error with the excess fuel (BADA).

Figure 8 illustrates the prediction of both models for three indicators. We may already anticipate that these indicators will have varied prediction performances: good for time difference (alignment along the regression line), fair for altitude difference (alignment for low values but spread for large) and poor for time level-off (large spread).

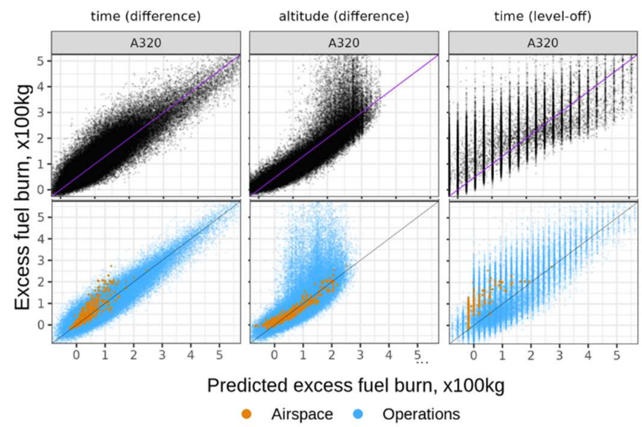


Figure 8. Predicted vs observed excess fuel burn, overall (top) and airspace/operations (bottom) for three indicators (A320)

To ensure the generalization of the models (i.e. estimate its performance when fed with new data), we split the sample in two sets: one training for calibration (80% random) and one testing to assess the prediction performances (20% remaining).

We use three performance indicators to evaluate the prediction. The first one is the **coefficient of determination** (R^2) which provides a goodness-of-fit measure for the linear regression. The second one is the **mean absolute error** (MAE) which enables the assessment of predicted versus observed values. The last one is the **ratio** between the mean absolute error and the mean predicted value (150kg for overall, 75kg for airspace and operations), to report the error in a relative form[4].

The performances of the eleven indicators are given in Table IV. We logically obtain better performances when predicting for airspace and operations merged (i.e. overall) than separately. The best indicator (altitude average) gives 0.94 (R^2) and 18% (MAE = 27kg) for overall, 0.91 and 25% (MAE = 19kg) for airspace / operations.

For both models (airspace and operations merged or split), considering R^2 and MAE, we observe varied performances, with a clear top 3 (altitude average and surface[5], time difference), complemented by a fourth indicator (altitude difference) having a lower R^2 for airspace / operations. This may suggest that time is the primary factor, either directly (time difference) or indirectly when integrated as flight time with altitude (altitude average and surface) which appears as the second key factor.

Distance shows a lower performance than time, probably because it does not compensate for wind. Speed shows poor performances, but slightly improved when used with surface and

[4] We did not use the mean absolute percentage error (MAPE) per flight due to many occurrences very close or equal to zero.

[5] We may note that altitude surface, developed in our previous work from a sort of intuition, although performing well, is beaten by the new altitude average which might be more intuitive.

average measurements, probably due to the indirect contribution of time. In contrast to average and surface, slope does not appear as a good measurement as its performance when associated to altitude and speed are the lowest. Finally, time level-off shows poor performances, probably as it only gives a partial view of the flight.

TABLE IV. PREDICTION PERFORMANCES (SINGLE INDICATOR)

Indicators	Overall			Airspace / Operations		
	R^2	*MAE*	*ratio*	R^2	*MAE* ↓	*ratio*
altitude (average)	0.94	27kg	18%	0.91	19kg	25%
altitude (surface)	0.94	28kg	18%	0.89	21kg	28%
time (difference)	0.92	33kg	22%	0.89	21kg	28%
altitude (difference)	0.88	35kg	23%	0.80	23kg	31%
distance (difference)	0.86	46kg	31%	0.82	29kg	38%
altitude (slope)	0.78	47kg	32%	0.63	30kg	40%
speed (surface)	0.86	44kg	29%	0.75	34kg	46%
speed (average)	0.83	49kg	33%	0.72	35kg	46%
speed (difference)	0.70	63kg	42%	0.52	44kg	59%
time (level-off)	0.72	66kg	44%	0.59	47kg	63%
speed (slope)	0.56	75kg	50%	0.36	49kg	66%

C. Combining indicators

We now investigate whether combining indicators may improve the prediction. This translates into finding the coefficients (for each aircraft type) minimizing the prediction error:

$$\alpha_1 \times \{indicator\ 1\} + \alpha_2 \times \{indicator\ 2\} + \alpha_3 \times \{indicator\ 3\} \ldots = Predicted\ excess\ fuel$$

minimize the error with the excess fuel (BADA).

Figure 9 shows the prediction performances (for airspace and operations separately) of the eleven indicators alone (this is a visual representation of the right part of Table IV); then of the combination of two, three and four indicators (55, 165 and 330 combinations respectively). In each graph, the black cross (top left) represents the performances of the eleven indicators combined.

With a single indicator, we may note the diverse performances already discussed (R^2 and error ratio ranging from 0.36 / 66% up to 0.91 / 25%). Combining indicators can improve performances (0.94 / 22% with two, 0.95 / 20% with three, 0.95 / 18% with four and up to 0.97 / 16% with the eleven indicators). The performances also become more consistent, with a large majority showing fair or good predictions.

The prediction performances for airspace and operations merged are logically higher (0.94 / 18 % with one indicator, 0.95 / 15% with two, 0.97 / 14% with three, 0.97 / 13% with four and 0.98 / 11% with the eleven) however at the expense of a loss of information.

Although combining many indicators provides better performances, it would be uneasy to understand and to handle. Further to this, some combinations, although showing good performances are not intuitive, for instance mixing time and distance or surface and slope.

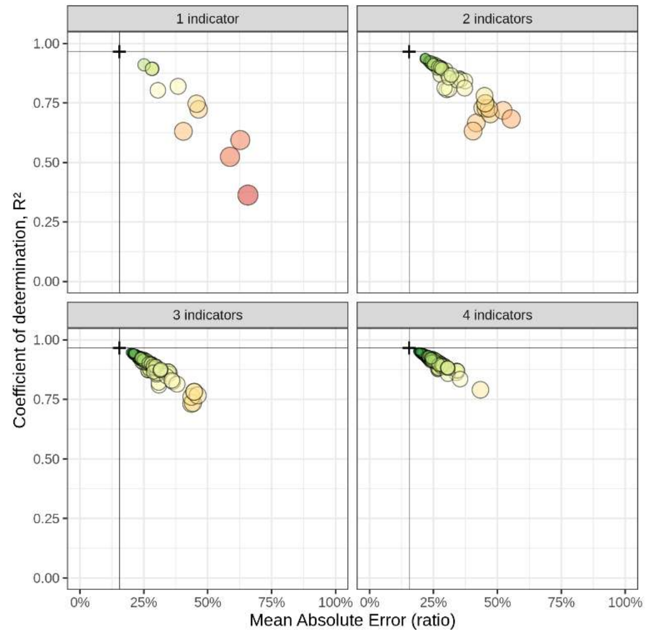


Figure 9. Prediction performances for one, two, three and four indicators (black cross: eleven indicators)

We have thus decided to introduce four empirical rules to select "valid" combinations: (1) the number of indicators is restricted to three and should reflect the three dimensions (horizontal, vertical and speed); (2) a combination with two (resp. three) indicators should be made of a valid combination with one (resp. two); (3) indicators should have good performances individually for their dimension; (4) a combination with two (resp. three) indicators should have higher performance than with only one (resp. two).

For individual indicators, we retain: time difference, altitude average, surface and difference, speed average and surface. We may also retain time level-off, although not a good proxy, it is rather common. With these four rules, possible combinations are presented in Table V.

The simplest horizontal and vertical indicators (time and altitude differences) show fair performances individually (0.89 / 28% and 0.80 / 31% respectively) and combined (0.92 / 25%). Adding the corresponding speed indicator (speed difference) slightly improves performances (0.93 / 24%). This combination may be retained as simple (one-dimensional) and intuitive, although it has a similar performance of the first indicator alone (altitude average, 0.91 / 25%).

The best "valid" combination is with the first two horizontal and vertical indicators (altitude average and time difference, 0.93 / 23% combined) complemented by the corresponding speed indicator (speed average, 0.94 / 21% combined, best with three indicators). The performances are just slightly below those of the best "non-valid" (0.94 / 22% with two, 0.95 / 20% with three) with 1kg difference on average.

TABLE V. PREDICTION PERFORMANCES (VALID COMBINATIONS, BEST-IN-CLASS "NON-VALID" IN GREY, SELECTION IN BOLD)

Indicators	Overall			Airspace / Operations		
	R^2	*MAE*	*ratio*	*R^2*	*MAE↓*	*ratio*
altitude (average)	0.94	27kg	18%	0.91	19kg	25%
time (difference)	0.92	33kg	22%	0.89	21kg	28%
speed (average)	0.83	49kg	33%	0.72	35kg	46%
altitude (difference) + altitude (surface)	0.96	23kg	15%	0.94	16kg	22%
time (difference) + altitude (average)	0.95	25kg	17%	0.93	17kg	23%
time (difference) + altitude (difference)	0.95	28kg	19%	0.92	19kg	25%
time (difference) + altitude (surface)	0.95	26kg	17%	0.91	19kg	25%
time (difference) + time (level-off)	0.96	25kg	17%	0.91	20kg	27%
altitude (difference) + altitude (surface) + speed (surface)	0.97	21kg	14%	0.95	15kg	20%
time (difference) + altitude (average) + speed (average)	0.96	22kg	15%	0.94	16kg	21%
time (difference) + altitude (surface) + speed (surface)	0.95	25kg	17%	0.93	18kg	23%
time (difference) + altitude (difference) + speed (difference)	0.95	26kg	18%	0.93	18kg	24%
all indicators	0.98	16kg	11%	0.97	12kg	16%

Tables VI and VII show the coefficients of the linear regression for a given aircraft type (A320), with one indicator and with the combination selected (time difference, altitude average and speed average).

TABLE VI. MODEL COEFFICIENTS (SINGLE INDICATOR, A320)

Aircraft type	Indicator	Coefficients
A320	time (difference)	28.2 kg.min^{-1}
	distance (difference)	8.1 kg.NM^{-1}
	altitude (difference)	12.5 kg.1000ft^{-1}
	altitude (average)	27.8 kg.1000ft^{-1}
	altitude (surface)	12.6 kg.(1000ft×10min)$^{-1}$
	time (level-off)	37.6 kg.min^{-1}

TABLE VII. MODEL COEFFICIENTS (THREE INDICATORS COMBINED, A320)

Aircraft type	time (difference) kg.min^{-1}	altitude (average) kg.1000ft^{-1}	speed (average) kg.10kn^{-1}
A320	12.3	20.7	-3.5

From Table VI, we may infer for instance that imposing a path elongation of 1 minute is equivalent to setting a vertical restriction that would lead to be 2200ft below a continuous descent. Table VII provides a more accurate and comprehensive view to select design options involving in particular both horizontal and vertical dimensions.

D. *Model for all airports vs per airport*

The performances of the selected combination (time difference, altitude average and speed average) per airport is shown in Figure 10. We may notice disparities among airports with R^2 and mean absolute error ratio ranging from 0.83 to 0.98 and from 12% to 67% for airspace and operations; from 0.85 to 0.99 and from 9% to 48% for overall. A calibration per airport slightly improves the performances and reduce the dispersion (R^2 ranging from to 0.90 to 0.98, mean average error ratio from 12% to 49% for airspace and operations; from 0.92 to 0.99 and from 8% to 36% for overall) however at the expense of a form of genericity.

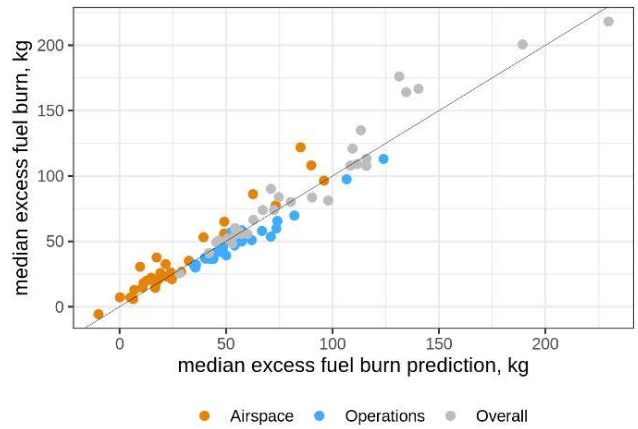


Figure 10. Prediction performances per airport (global model)

V. APPLICATION

In this section, to illustrate the use for flight efficiency assessment, we present the excess fuel and the selected horizontal and vertical indicators per airports (average values) for all aircraft types (2.8 million flights). For a given airport, a detailed view per runway and flow is available via an interactive dashboard [23].

A. *Excess fuel burn*

Figure 11 shows the fuel burn in excess per airports, with the split operations (x-axis) versus airspace (y-axis). The size of the dot reflects the number of movements.

We may observe a large disparity among airports with a range of more than 150kg on both axes, suggesting extremely varied situations. Some airports are subject to significant operations inefficiencies (e.g. EGLL and EGKK with more than 150kg), others to significant airspace inefficiencies (e.g. LFPG, EGLL and EDDF with more than 150kg), and a group showing operations and airspace inefficiencies below average. We may also note a ratio airspace / operations ranging between less than ½ to more than 2, with some airports mainly subject to operations inefficiencies (e.g. EIDW, LIMC) and others to airspace inefficiencies (e.g. LFPG, LFPO).

The significant operations inefficiencies materialize as frequent holding or extensive vectoring / tromboning at low altitude. They are due to the high traffic congestion compared to the arrival capacity and reflect the nature of the arrival management strategy (e.g. declared capacity higher than the real one, absence of arrival streaming). The significant airspace inefficiencies are caused by flow segregation taking the form of procedures incorporating level restrictions or route elongations, and may be aggravated by environmental constraints (e.g. high intercept altitudes imposing long distance for downwind flows). Airspace inefficiencies are typical of a complex terminal area with several airports in the vicinity with many arrival and departure flows interacting. We may note in particular EGLL, EGKK and EGSS for London, LFPG and LFPO for Paris (both having other airports not considered here). It may be observed that airspace and operations inefficiencies do not strictly depend on the number of movements.

We may look at the overall fuel burn in excess (operations + airspace) which varies considerably among airports (from 45 up to 395kg) with four airports inducing more than 200kg (EGLL, EGKK, LFPG and EDDF).

For the 30 airports globally, with a transit fuel burn in the 50NM of 309kg, the overall excess fuel burn[6] of 150kg corresponds to **a fuel inefficiency[7] of 49%**, with 22% for airspace and 27% for operations (these fuel burn values were already in Table II).

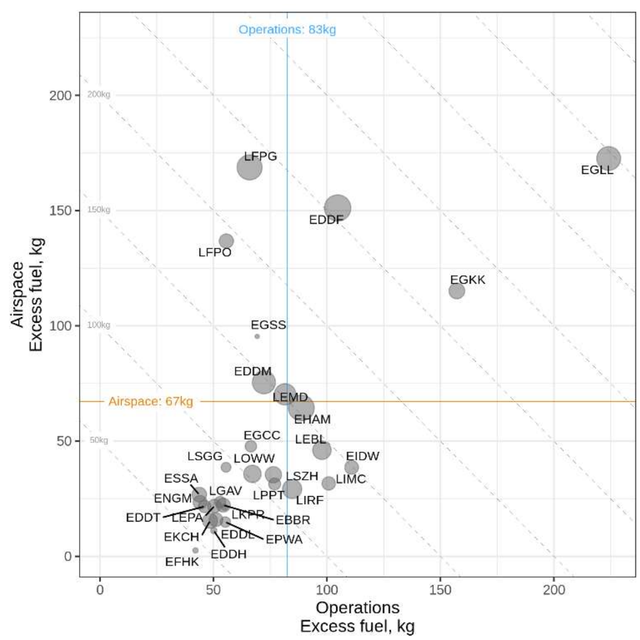


Figure 11. Excess fuel burn

The previous view with absolute fuel burn presents an information that depends on two factors: the level of inefficiency and the aircraft mix. Indeed, an airport with a majority of long-haul flights is more likely to have a higher fuel burn score than one with a majority of short / medium haul flights. Figure 12 presents the excess fuel burn as percentage of the total fuel burn for each flight. We may observe on average values a slightly different contribution for operations vs airspace compared with absolute values (operations 21% and airspace 18%).

More importantly, we clearly see the effect of the aircraft mix, for instance between EGLL and EGKK, or between LFPG and LFPO: while EGLL and LFPG (both with significant proportions of long haul flights) show a much higher inefficiency in absolute terms compared to respectively EGKK and LFPO (more short / medium haul flights), the difference is much reduced when considering percentage. Overall, all airports (except one) have an inefficiency higher than 20% and four higher than 50%.

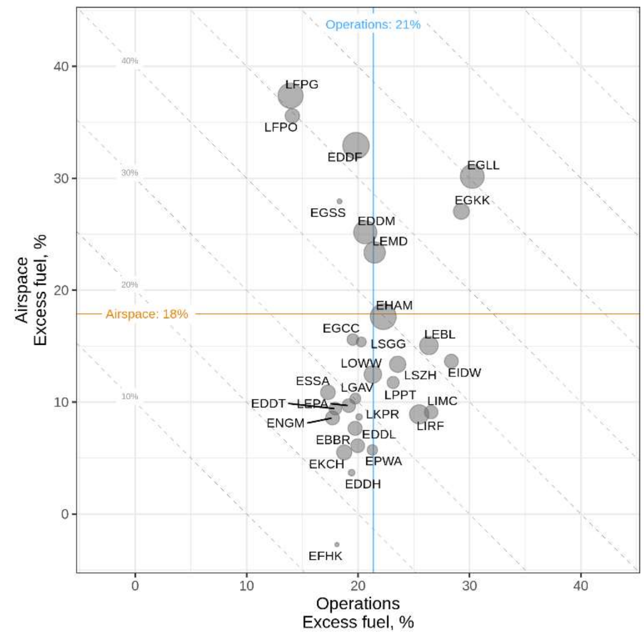


Figure 12. Excess fuel burn (percentage)

B. *Horizontal and vertical deviations*

Figures 13 and 14 show the horizontal and vertical inefficiencies per airport, using the indicators selected in section IV. Horizontally, the range is close to 3 minutes for both operations and airspace (7 minutes overall) and vertically it is more than 3500ft (7000ft overall). On both figures, we may identify the airports showing significant inefficiencies, possibly more clearly on the vertical dimension.

For the 30 airports, with an average transit time of 15.3 minutes, the overall horizontal deviation (4.5 minutes) represents an **horizontal inefficiency of 29%** (mean ratio per flight: 26%). Similarly, with an average transit altitude of 8800ft, the overall vertical deviation (4800ft) represents a **vertical inefficiency of 55%** (mean ratio per flight: 59%).

These two inefficiency values (29% and 55%) might suggest a higher impact of the vertical dimension than the horizontal one.

[6] It may be noted that the true fuel burn in excess may be higher, in particular for long haul flights, as any kilogram carried and burn when arriving implies an extra fuel at departure.

[7] Calculated as the ratio of mean values. We may also consider the mean of the fuel inefficiency ratio per flight, which would lead to 39%.

This should be further investigated and confirmed since the corresponding indicators as only proxies for excess fuel burn with a noticeable inaccuracy (cf. Table V).

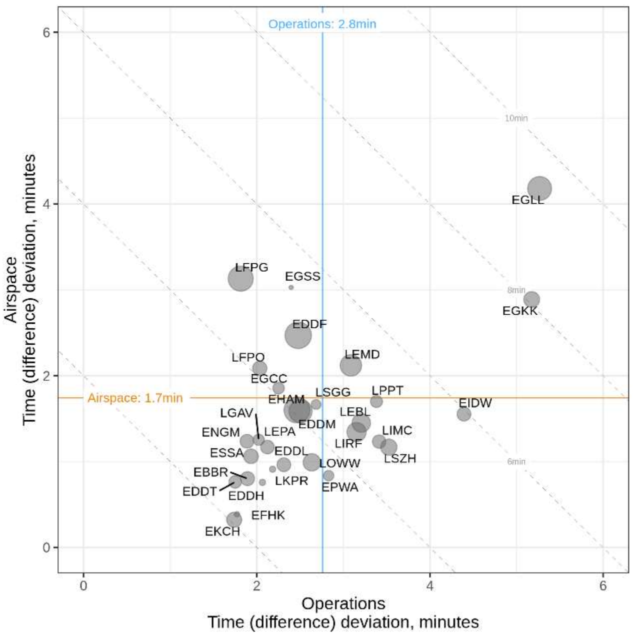


Figure 13. Horizontal deviations

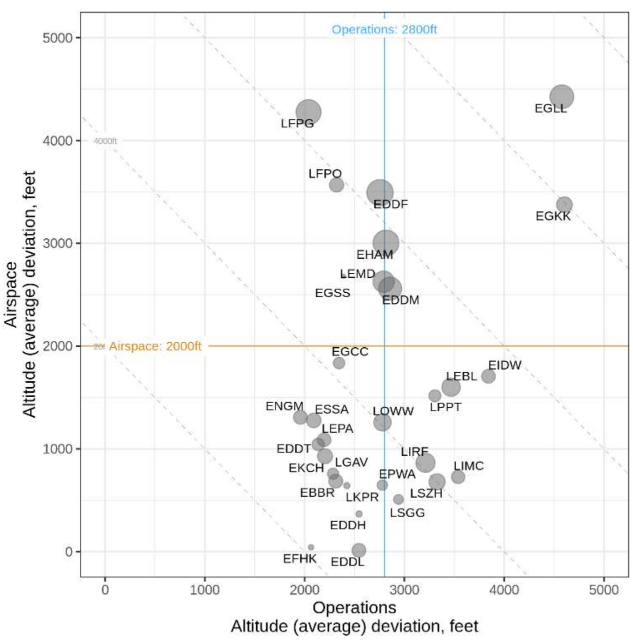


Figure 14. Vertical deviations

VI. CONCLUSION

The objective of this work was to explore and evaluate various flight efficiency indicators for arrivals. Following a form of systematic approach mixing dimensions (horizontal, vertical and speed) and measurements (difference, average, …), we identified eleven indicators, some existing and others new. We decomposed them in two sub-indicators to capture permanent vs variable inefficiencies as proxies for airspace vs operations. We then considered all combinations (with one to eleven indicators) and integrated each in a linear regression, calibrated per aircraft type on the top 30 European airports in the last 50NM, over 2.8 million flights.

Predicting fuel burn in excess for airspace and operations separately with a single indicator leads to diverse performances (mean absolute error ratio ranging from 25% to 66%) with a clear top 3 (altitude average, time difference and altitude surface). Combining indicators can improve performances (22% with two, 20% with three, 18% with four and 16% with the eleven) but adds complexity. Similar performances are observed for the top 10 when combining two or three indicators. Some combinations, although showing good performances are not intuitive and may not be retained. The simplest indicators (time and altitude differences) show fair performances individually (28% and 31%) and combined (25%). Adding the corresponding speed indicator (speed difference) slightly improves performances (24%) but with a minor gain compared to the top 1 alone (altitude average, 25%). A good compromise involving the three dimensions may be with two of the top 3 indicators (altitude average and time difference, 23% combined) complemented by the corresponding speed indicator (speed average, 21% combined). Predicting fuel burn in excess for airspace and operations merged leads to higher performances (mean absolute error ratio of 18% with one indicator, 16% with two, 14% with three and 11% with the eleven) however at the expense of a loss of information. A calibration per airport slightly improves performances at the expense of some form of genericity. However, disparity among airports remains and should be investigated.

The analysis of the 30 airports in 2019 reveals an average fuel burn in excess of 150kg (corresponding to an inefficiency of 49%), with 67kg due to airspace and 83kg to operations (22% and 27% inefficiency); an average time deviation of 1.7 and 2.8 minutes (11% and 18% inefficiency); and an altitude deviation of 2000 and 2800 feet (23% and 32% inefficiency). An interactive dashboard with the selected indicators has been developed, providing a detailed view per airport, runway and flow, for an easy identification of inefficiencies by our stakeholders.

Future work should investigate whether the prediction performance may be further improved with the same data source (radar tracks) by considering more sophisticated models or new indicators, or by constructing new data (e.g. vertical speed). It should also involve the extension to larger horizons (e.g. 100NM or 200NM) which may lead to a different selection of indicators, and the development of the airspace user perspective possibly with the use of aircraft derived data. The integration of departures is also an essential aspect to bring the complete picture.

Acknowledgment

This work is supported by the EUROCONTROL Aviation Sustainability Unit led by Marylin Bastin and constitutes a contribution to the ATM/ANS Environmental Transparency Working Group, part of the EUROCONTROL - EASA Joint Work Programme. We would like to thank our colleague Raphaël Christien for his technical advices.

References

[1] P. Pasutto and K. Zeghal, "Flight efficiency indicators in descent: a proxy for fuel burn?," in *AIAA AVIATION 2022 Forum*, USA, 2022.

[2] T. G. Reynolds, "Development of flight inefficiency metrics for environmental performance assessment of ATM," in *8th USA/Europe ATM Seminar*, Napa, California, 2009.

[3] G. Enea, H. Fricke, J. Bronsvoort and A. Ramadani, "Fuel burn estimation modeling for ATM benchmark applications," in *12th USA/Europe ATM Seminar*, Seattle, Washington, 2017.

[4] D. Howell, M. Bennett, J. Bonn and D. Knorr, "Estimating the En Route efficiency benefits pool," in *5th USA/Europe ATM Seminar*, Budapest, Hungary, 2003.

[5] S. Chesneau, I. Fuller and J.-C. Hustache, "ATM flight efficiency and its impact on the environment," EUROCONTROL, 2003.

[6] T. Kettunen, J.-C. Hustache, I. Fuller, D. Howell, J. Bonn and D. Knorr, "Flight efficiency studies in Europe and the United States," in *6th USA/Europe ATM Seminar*, Baltimore, Maryland, 2005.

[7] EUROCONTROL and FAA, " U.S./Europe comparaison of ATM-related operational performance," 2009.

[8] X. Prats, R. Dalmau and C. Barrado, "Identifying the Sources of Flight Inefficiency from Historical Aircraft Trajectories," in *13th USA/Europe ATM Seminar*, Vienna, Austria, 2019.

[9] D. Knorr, X. Chen, M. Rose, J. Gulding, P. Enaud and H. Hegendoerfer, "Estimating ATM efficiency pools in the descent phase of flight," in *9th USA/Europe ATM Seminar*, Berlin, Germany, 2011.

[10] S. Peeters, H. Koelman, R. Koelle, R. Galaviz, J. Gulding and M. Meekma, "Towards a common analysis of vertical flight efficiency," in *Integrated Communications Navigation and Surveillance Conference*, 2016.

[11] EUROCONTROL, "Analysis of vertical flight efficiency during climb and descent," 2017.

[12] A. Lemetti, T. Polishchuk, V. Polishchuk, R. Sáez and X. Prats, "Identification of significant impact factors on arrival flight efficiency within TMA," in *9th International Conference on Research in Air Transportation*, Tampa, Florida, 2020.

[13] EUROCONTROL, "Efficiency and environment," [Online]. Available: https://ansperformance.eu/efficiency/. [Accessed December 2022].

[14] EUROCONTROL Performance Review Commission, "Performance Review Report 2021, an assessment of Air Traffic Management in Europe," 2022.

[15] EUROCONTROL, "Environmental assessment: European ATM network fuel inefficiency study," 2020.

[16] C. Nutt, "NATS fuel efficiency metric," Technical Report, 2012.

[17] Q. McEnteggart and J. Whidborne, "A trajectory optimization based analysis of the 3Di flight efficiency metric," in *11th USA/Europe ATM Seminar*, Lisbon, Portugal, 2015.

[18] M. S. Ryerson, M. Hansen and J. Bonn, "Fuel consumption and operational performance," in *9th USA/Europe ATM Seminar*, Berlin, Germany, 2011.

[19] E. Calvo, J. M. Cordero, L. D'Alto, J. López-Leonés, M. Vilaplana and M. L. Civita, "A new method to validate the route extension metric against fuel efficiency," in *11th USA/Europe ATM Seminar*, Lisbon, Portugal, 2015.

[20] 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.

[21] P. Pasutto, E. Hoffman and K. Zeghal, "Vertical efficiency in descent. Assessing the potential for improvements at the top 30 European airports," in *AIAA Aviation Forum*, USA, 2020.

[22] "National Centers for Environmental Information, Global Forecast System," [Online]. Available: https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast.

[23] EUROCONTROL, "TMA Metric dashboard," December 2022. [Online]. Available on request : https://analytics1-test.eurocontrol.int/#/site/inolab/views/TMA_flight_efficiency_dashboard/Introduction.