

Fuel Burn Estimation Modeling for ATM Benchmark Applications

Perspectives from an International Collaboration

Gabriele Enea, Ph. D.
Engility Corporation
Billerica, MA, USA

Jesper Bronsvoort, Ph. D.
Airservices Australia
Melbourne, Australia

Hartmut Fricke, Ph. D.
Technische Universität Dresden
Dresden, Germany

Almira Ramadani
Federal Aviation Administration
Washington, DC, USA

Mike Paglione
Federal Aviation Administration
Atlantic City, NJ, USA

Christian Seiß & Judith Rosenow, Ph. D.
GfL-Gesellschaft für Luftverkehrsforschung
Dresden, Germany

Abstract—Fuel burn changes are often used as a proxy for benefits of new air transportation operational paradigms around the world. As phases of NextGen, SESAR and Australian AATMP come to maturity, the questions are pivoting from what benefits *will be* to what *have we achieved*. This pivot puts more pressure on fuel modeling being able to distinguish between the contribution of ATM and non-ATM factors to differences in fuel estimates. In addition, there are many fuel estimators used around the world, and their range of fidelity varies widely; in fact, fuel savings are frequently of the same magnitude of the error produced by the models used for their estimation. This paper summarizes an initial collaboration between researchers from several globally recognized institutions to address the question of fidelity of fuel estimation that may be required for different types of benefit assessments of Air Traffic Management improvements. Interviews were conducted initially to categorize common elements that typical ATM studies share. An international team of fuel modelers was assembled and participated by running their models on a common set of inputs. The outputs generated by these models, were categorized using metrics on empirical trajectories and other operational data, including predicted fuel burn. This provided a foundation for studying impacts of different fuel estimation approaches and assumptions, and how they relate to the analysis of fuel efficiency.

Keywords—component; Fuel Burn Modeling; ATM Performance; aircraft performance model;

I. INTRODUCTION

One of the main focuses of the modernization efforts of the air transportation systems around the world is to improve the flight efficiency of the aviation system, which is reflected largely by aircraft fuel burn. Moreover, the global warming

debate has focused the attention on aviation emissions including fuel pollutants such as CO₂, NO_x, SO_x, UHC, etc. Projects, such as NextGen in the United States [1], SESAR in Europe [2] and the Australian Air Traffic Management Plan (AATMP) in Australia [3], have been designed, and continue to be adjusted as required, to not just modernize the existing ATM system, but also to enable more efficient operations. To evaluate the effectiveness of these projects and estimate the benefits delivered, an accurate and reliable approach to calculate fuel consumption from recorded aircraft's four-dimensional trajectories is necessary. For example, this approach, applied to trajectories flown both under legacy and new operational paradigms, can be used to estimate the achieved fuel reduction benefits. The problem is that, per single trajectory, the benefits that are being measured are often small and sometimes of the same magnitude of the error in the estimation technique. Moreover, some of the new operations, such as Optimal Profile Descent (OPD) provide a tradeoff between better vertical profiles but sometimes at a cost of excess en-route distance to set up the procedure. For this purpose, a benefits' common denominator, such as fuel burn should be used.

There are many fuel burn models and analysis techniques available around the world. Most of them are based on similar complex aerodynamic or total energy equations. However, even if these models are built on similar physics with slightly different assumptions they may produce different outputs for the same set of inputs. This research effort¹ investigates such differences with the end goal of understanding fuel burn modeling capabilities, assumptions and inputs that are needed for accurate assessment of ATM improvements.

This effort was produced under contract to the Federal Aviation Administration while Dr. Enea was employed at Engility Corporation. He now works for The MITRE Corporation, McLean, VA.

To identify capability gaps in the fuel burn modeling techniques available, a preliminary literature review was performed. Once a subset of models was selected, a questionnaire was distributed to relevant institutions. The group included the FAA, Airservices Australia, and the Technische Universität of Dresden. Some of the questions that were addressed to the researchers developing and using fuel burn estimation models were:

- What is your institution’s need for fuel burn estimation models from recorded 4D trajectories?
- What type of studies do you perform with such models?
- What type of input/output data do you need to run these models?
- What are the challenges in using and validating these models?
- Are there any features that your model doesn’t currently have that would be desired to have?

This set of questions was useful to identify not only the data necessary for accurate fuel burn estimations from recorded tracks, but also gaps in capabilities that are currently present in the state of the art.

a. Literature Review

In addition to the models directly compared in this paper, which will be presented in the next section, many fuel burn estimation models have been used in the literature. Some of the studies that stand out are discussed here. A study from NASA Ames was presented by Chatterji in [4]. The fuel burn estimation model used the EUROCONTROL Base of Aircraft Data (BADA) 3 equations. Accuracy of 1% was reported with known initial TOW and validated with one flight whose Flight Data Recorder data were available. With recorded trajectory from surveillance, and unknown TOW, the model achieved 5.4% average error.

John Robinson, also of NASA Ames, presented results applied to a large set of flights in [5]. The Center-TRACON Automation System (CTAS) was used in conjunction with BADA 3.7 to estimate the benefits of continuous descent approaches (CDA) for a sample of more than 480,000 flights. CTAS’ trajectory synthesizer was used to create the CDA tracks to calculate the benefits of implementing these procedures into several airports in the NAS. These synthesized CDA tracks were compared to the actual recorded trajectories to evaluate the fuel burn savings of the new procedures. It is worth mentioning that adjustments to the BADA fuel flow model were performed to:

- Compensate for under-prediction of cruise fuel flow rate with respect to aircraft operating manual,
- Reduce the nominal airspeed to match typical speed profiles in congested terminal airspace, and
- Compensate for aircraft configurations using flaps at lower altitudes and slower airspeeds.

Trani et al. have explored the application of neural networks algorithms to fuel burn estimation modeling in [6]. Although they achieved very accurate results, the method requires detailed aircraft performance model information only available from manufacturers’ manuals that are seldom available. Belle and

Sherry have applied fuel burn modeling, using BADA equations, to estimate the benefits of introducing required navigation performance (RNP) procedures into the terminal area at Midway Airport (MDW) and presented the results in [7]. Although using only surface wind data from the ASPM data set, the model seems to provide a good proxy for variations in fuel burn at low altitudes in the terminal area for various arrival flows. In [8] the flight efficiency losses when speed control is applied by ATC were studied to correlate the effects of cost index selection onto fuel consumption. MITRE Center for Advanced Aviation System Development (CAASD) also developed an in-house fuel burn estimation model that uses fused surveillance data, generated by MITRE’s Threaded Tracks tool, to evaluate performance of various new operational concepts. This model is capable of analyzing large data sets [9].

b. Motivation

As illustrated above, there is a plethora of studies that report on results of fuel burn modeling, calculated over a sample scenario of recorded or simulated flights. In these studies, many variations of fuel burn models and metrics are used. Therefore, the results must be carefully assessed when comparing them to gain an overall perspective of the performance of different ATM concepts. This paper will compare the design, assumptions, and performance of different fuel burn models.

II. MODELING

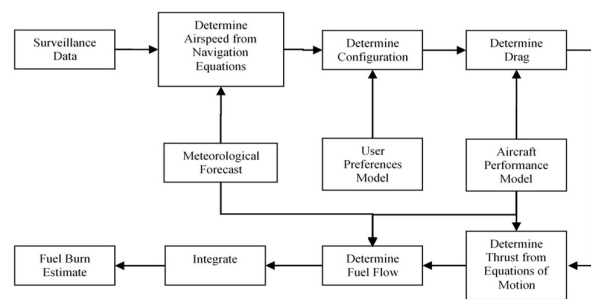


Figure 1: High level flow diagram of fuel burn estimation process.

When aiming to estimate fuel burn from surveillance data, all models apply, in principle, a similar process. A simplified high level flow diagram of this process is given in Figure 1, starting from surveillance data and clockwise illustrating the applied processing, finally resulting in the estimated fuel burn. In case the surveillance data provides no true airspeed (TAS) data, the TAS can be determined from the surveillance data using standard navigation equations and a meteorological forecast. At low speeds, flap configurations can be simulated accordingly to include the effect of additional drag. These flap retraction and deployment schedules can be derived from generic user operating models of the respective aircraft type. With an estimate of the drag and the aircraft’s attitude, the thrust is computed by evaluating the equations of motion. From this estimated thrust value and an aircraft performance model (like BADA), the fuel flow is calculated and subsequently integrated over the entire trajectory. Various modifications can be applied

TABLE 1 MODEL CHARACTERISTICS

Model		AFEST	ETAS	FAA NextGen	Dalí
Applicable phases of flight		All (excluding ground movement)	All (excluding ground movement)	All (excluding ground movement)	All (including ground movement)
APM	APM	BADA 3	BADA 3 & 4	BADA 3	BADA 3 & 4
	Type matching	Aircraft type matched through BADA 3 database of aircraft	Engine and aircraft type (for BADA4) is based on Global Aircraft Registry data base, based on registration	Aircraft type matched through BADA 3 database of aircraft	Engine and aircraft type (for BADA4) is based on aircraft data base, based on registration
MET	Meteorological model	NOAA Rapid Refresh (RAP) nowcast (40 km grid size)	NOAA Rapid Refresh (RAP) nowcast (13 km grid size)	NOAA Rapid Refresh (RAP) nowcast (40 km grid size)	World Area Forecast System (approximately 137 Km, but it depends on the latitude)
Integration	Equations of motion	BADA Total-Energy Model equations to calculate: lift and drag coefficients, thrust and fuel consumption	Kinetic equations of motion both time and distance integration applied	Distance-based integration using BADA lookups of nominal fuel burn rates as a function of flown distance and altitude, and reference TAS and climb/descent rates	Time integration and kinetic derivation of all state derivatives (e.g. vertical rate)
	Airspeed rate of change incl.?	Yes	Yes, kinetic and potential energy used for calculations	No	Yes, based on partial derivatives (none in level flight)
Input	TOW	Uses BADA nominal values along with the flight plan information, to iteratively simulate the required amount of fuel. Reserve fuel is added to the minimum amount to account for error, as well as the required 45-minute holding and diversion to next airport.	<ol style="list-style-type: none"> 1. Estimate the mass from the approach speed or, 2. Estimate the mass from the departure speed, or 3. Estimate over flight level and speed, else 4. Use standard values from BADA tables 	Estimated based on BADA nominal values and city-pair great circle distance	Estimated based on data in flight plan. If no flight plan is available, a more basic method is applied.
	Airspeed	Airspeed estimated from surveillance reported groundspeed and forecast	Airspeed estimated from surveillance reported groundspeed and forecast	Nominal TAS as a function of altitude and flight segment (climb, cruise, or descent)	Airspeed estimated from surveillance reported groundspeed and forecast

to this process depending on the required fidelity and/or available data.

The remainder of this section discusses the aircraft fuel burn estimation models used by Airservices Australia, the Technische Universität of Dresden and the FAA, whose level of complexity in the above diagram varies, and their respective applications. An overview of the different models and their main characteristics is provided in Table 1.

a. Airservices Dalí

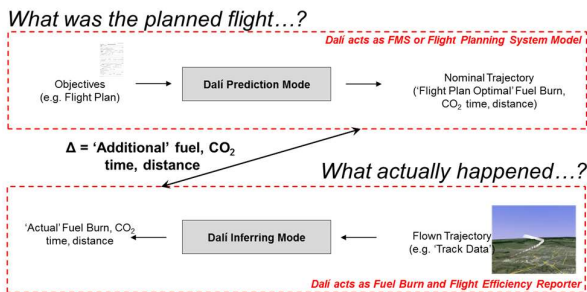


Figure 2 Inferring and prediction modes of Dalí

Airservices required an aircraft trajectory modeling capability to fulfill two major objectives [10]:

1. To develop a framework and capability to accurately measure and report flight efficiency of existing operations in Australian-administered Flight Information Regions (FIRs).
2. To develop a framework and capability to accurately model the impacts of proposed airspace changes and future ATM improvement initiatives on flight efficiency, and assess feasibility of a potential trajectory changes.

Dalí is an aircraft trajectory modeling toolbox developed in-house by Airservices, based on the concept of aircraft intent generation, where aircraft intent refers to the basic commands, guidance modes and control strategies available to an aircraft to control its trajectory. To model aircraft intent, Dalí applies the Aircraft Intent Description Language (AIDL) framework developed by Boeing Research & Technology Europe (BR&TE) [11]. Dalí has many ways of generating aircraft intent depending on the objectives of the modeling. First, it can take basic information from a filed flight plan and compute a high fidelity trajectory, essentially acting as the trajectory computation function within a Flight Management System (FMS) (prediction

mode). Second, based on provided surveillance data, Dalí can generate aircraft intent by simulating the aircraft following this trajectory, and performing fuel burn and emissions estimations (inferring mode). Combining these two models allows for assessment of the efficiency of the flight, by comparing what actually happened to what was planned, as illustrated in Figure 2, which corresponds with the first objective. Airservices is conducting this efficiency modelling on all IFR operations within the Australian flight information regions.

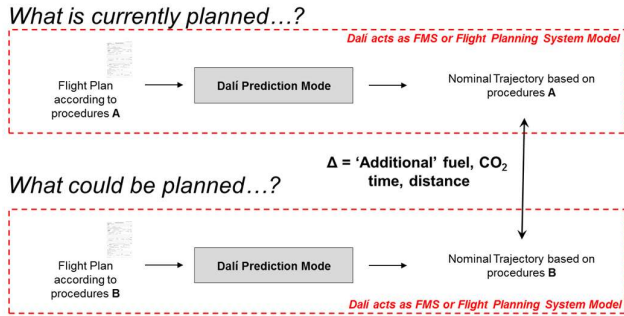


Figure 3 Different prediction modes of Dalí

In addition, comparing two trajectories computed in prediction mode using different procedures, allows for modeling the impacts of proposed airspace changes and future ATM improvement initiatives on flight efficiency (Figure 3), which corresponds with the second objective.

Airservices has been working with participating airline partners to validate the Dalí model. For a total of 300 flights, Quick Access Recorder data was provided to validate the fuel burn estimation from surveillance track data. An example of such analysis is shown in Figure 4 for a flight that was subject to significant intervention by ATC. Of particular interest is the similarity of the fuel flow in the lower plot estimated by Dalí (red) in comparison to the QAR recorded fuel burn (blue). The oscillations in the estimated fuel flow are a result of estimation noise added at each estimation step to infer fuel flow from surveillance data (Figure 2). While the resultant noise on the estimated fuel flow looks quite dramatic, the integrated effect is small. When referencing the estimated fuel flow with the airspeed data in the upper plot, it can be seen that the impact of

speed changes in fuel flow is accurately reflected. Also of interest, while the descent appears continuous in regards to the vertical profile (top plot), it was not fully performed at idle thrust. The momentary increases in thrust are detected by the model. The accuracy in modelling thus allows for detecting if a descent was both continuous *and* efficient (e.g. performed at idle-thrust).

Overall, for 136 of the analysed flights (45%), the estimated fuel burn error was within $\pm 2.5\%$, and for 253 of the analysed flights (84%), the error was within $\pm 5\%$.

b. *GfL/TU Dresden EJPM-based Trajectory Analysis Software (ETAS)*

The Enhanced Jet Performance Model (EJPM) was developed to provide a fast and highly precise fuel estimation model for post flight analyses. It relies on a six degree-of-freedom aircraft model. It is specifically capable of providing trajectory data for cruise [12] and (CDO)-descent [13] and climb with a position/fuel consumption error around 1% for deriving flight intent or target information on where the aircraft should fly to achieve its target state (e.g. minimum fuel/emissions). The biggest challenges in achieving that high accuracy are aircraft gross mass estimation and the limited quality of wind and temperature predictions in time and space. In [14], a novel, fast time capable mass estimation model, relying on the calculated approach ground (reference) speed v_{ref} and its correlation to the aircraft's landing mass was presented. For those trajectories not covering the landing phase, a second estimation model was added focusing on the take-off phase, relying on the safety climb out speed v_2 . Flap/slat configuration were assumed according to standard operating procedures. The mass estimation algorithm is presented in Figure 5.

Since October 2016, the ETAS is operationally used by the German ANSP Deutsche Flugsicherung (DFS) to estimate accurately the fuel consumption per flight of all operations executed by the six largest German airlines (LH, Condor, Airberlin, Germania, etc.) across Germany summing up to roughly 6.000 flights per day. DFS was chosen as “independent” ETAS host, providing analyses strictly confidential to each airline. DFS is further allowed to generate overall (cross airline)

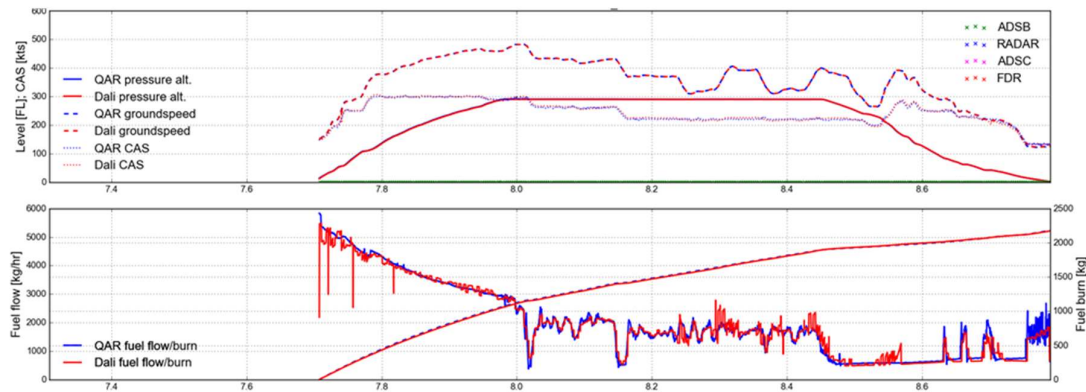


Figure 4 Analysis of actual flight data against Airservices Dalí model.

performance measurement data to estimate the national efficiency of existing operations feeding the performance indicators as set out by the SES Performance Scheme (Commission Regulation (EU) No 691/2010 of 29 July 2010).

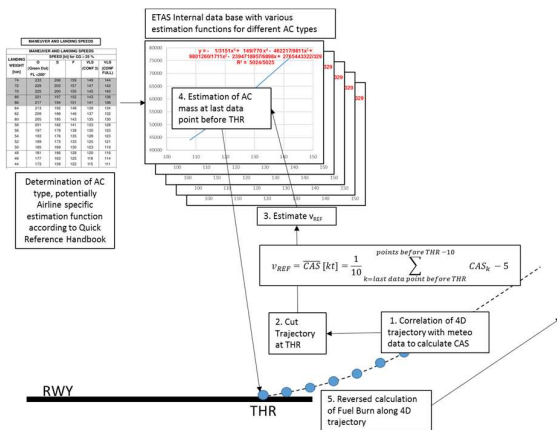


Figure 5 Take-Off oriented gross mass estimation algorithm of ETAS model.

Additionally, DFS constantly aims at improving cross border airspace structures as a member of the Functional Airspace Block-Europe central concept, FAB-EC. ETAS is intended to support these design activities by allowing the measurement of reachable ideal vertical flight efficiency (trajectory comparisons vs. individual maximum range and, just recently also minimum overall emission profiles).

Since then, the tool is used to validate new flight procedures (especially approach procedures into hub airports) and adjacent airspace structure design. Competing ideas for PBN approach transitions are being debated in Germany: the hitherto-favored trombone pattern (currently implemented for FRA, MUC, HAJ) versus point merge designs (LEJ). ETAS shall help identifying potential flight efficiency gains from the two alternatives.

Since January 2017, ETAS runs in an extended mode covering all types of relevant pollutants of aircraft engines for both complete and incomplete combustion processes. The APM was further extended at TUD to allow for a multi-criterion (fuel burn, emissions, radiation force) flight trajectory optimization process. This is set up through a user preference model, where competing target functions can be weighted by the users, and handled in a proportional–integral–derivative (PID) controller loop, interacting with the APM and a combustion model as depicted in Figure 6.

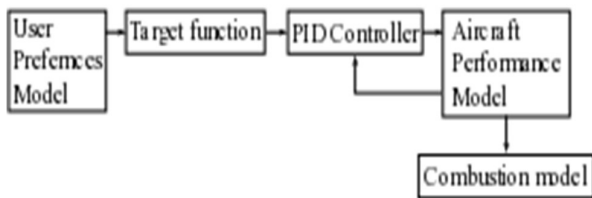


Figure 6 Multi-criteria trajectory optimization ETAS model considering fuel burn, emissions and radiation forcing.

These additional functions have not yet been considered within the model comparisons presented in the next section.

c. FAA AFEST

AFEST is developed by the Modeling and Simulation Branch in the NextGen Office at the FAA William J. Hughes Technical Center whose mission is to “validate new aviation concepts technologies, and system capacity issues, which evaluate the performance of both emerging and existing systems within the National Aviation System” [15]. The FAA group employs several COTS simulation models, most recently AirTop by AirTop Soft, to perform these studies, although these were not analyzed here. Often these tools have their own fuel burn models built in. The FAA prefers to use an in-house model to ensure consistency between studies and have an idea of its accuracy, given that some validation on the performance of the model has been conducted [16].

The model is typically used to help mature new NextGen concepts, and to help develop requirements and support the benefit and safety cases. In [17], the team estimated the operational impact that weather forecast performance has on the NAS to help define performance requirements for these weather products. In [18], the team estimated the impact that new entrant space vehicle operations can have on the NAS (today and in the future) and offered mitigation strategies and their benefit. In [19], the team estimated the cost/benefit for the implementation of a new NextGen concept.

d. FAA NextGen Office Model

It is challenging for the FAA to isolate changes in fuel efficiency that are solely driven by procedural and other improvements implemented over the years, even if consistent access to empirical fuel burn data of the required granularity is available. There are too many other contributors that also impact fuel consumption that would be difficult, if at all possible, to account for, including changes in winds, demand, fleet mix, aircraft weight and cost indices. Therefore, the FAA’s Office of NextGen has recently started working on a model that focuses on the correlation between differences in fuel burn and differences in typically flown horizontal and vertical profiles. By focusing on horizontal and vertical profile efficiency, the FAA’s goal is to develop an approach for isolating performance impacts that are directly instigated by NextGen improvements. This approach aims to not only help improve the reliability of post-implementation impacts and benefits, but also to enable better alignment between expectations and outcomes of these improvements.

The FAA NextGen Office Model uses a simplified fuel evaluation approach that applies BADA on full² or partial³ trajectories, and assumes aircraft take off at nominal weight and fly reference true airspeeds (TAS). The model considers trajectory segments specified by their horizontal and vertical profiles, meaning segments’ ground distance, and beginning and ending altitudes as shown in Figure 7. Since one of its key intended applications is to evaluate impacts from increased access to RNAV STARs with Optimal Profile Descent (OPD) across the NAS, the model adjusts fuel burn calculation to

² Runway-to-runway

³ For example, below top of descent, or within 300nm of the destination airport

account for differences between reference and actual descent rates. Also, since the NextGen Office’s goal is to isolate only the impacts caused by the improvement itself, as opposed to the impacts driven by other coinciding changes, the model does not account for winds, and does not make adjustments for differences between the reference and actual TAS. As a result, the model cannot capture impacts from any procedural speed restrictions that may be affecting aircraft behavior, but is also not sensitive to inaccuracies in surveillance data driven by sudden wind-bursts and other issues with merging and filtering raw data, especially difficult with overlapping data sources.

The FAA is testing and validating this model on several recent improvements, including North Texas Metroplex and OPD implementations at Boston Logan and Gary Indiana airports. Validation of the model is especially challenging because of the lack of empirical fuel burn data at the granularity that is required for such analysis; however, the FAA is closely collaborating with American Airlines and other airline partners in an effort to assure that the model produces both reasonable and meaningful outcomes.

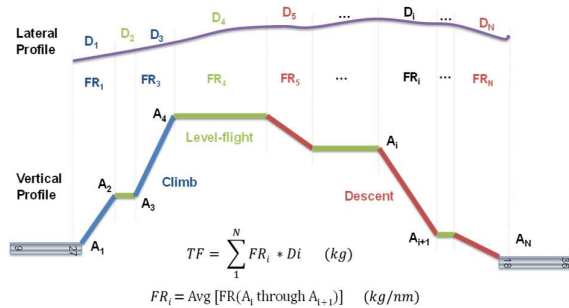


Figure 7 – FAA NextGen Office Model: Simplified Fuel Evaluation using BADA ptf-files and EmpiricalTrajectory Segments

III. APPLICATION AND DATA

a. Flights Sample

The set of flights analyzed for this study was recorded in June 2015 from various airports in the United States. There was a limitation on the number of flights that each of the participating organizations could run with their respective models, given the level of adaption required (especially for non-US participants). For this reason, an initial set of 111 flights was down-selected from the 38,000 flights that the FAA has recorded fuel and push-back weight for. The sample included flights with variable length, different O-D pairs, and various aircraft types and for different day of the month of June 2015. Flights from 19 different days, 65 O-D pairs and 16 different aircraft types were selected. Some of the fuel estimation models have complex procedures to load the weather files used for the calculation. Therefore, most of the flights were selected from June 29, 2015.

For various reasons, each fuel estimation model had issues running some of the 111 flights. The most common issue was a mismatch between the internal database of aircraft types and the data in the sample. Hence, each of the fuel estimation models produced meaningful results only for a subset of the 111 flights. To present a meaningful cross-comparison of the fuel burn

models evaluated, only the subset of common flights was used for this analysis. This subset represented the maximum number of flights with valid fuel and Take-Off Weight (TOW) predictions from all the models. The maximum set of common flights comprised 60 flights. The distribution of aircraft types in the sample is presented in Figure 8. The shortest flight in the sample was DCA-EWR, 220 nautical miles long. The longest flight was LAX-EWR, 2495 nautical miles long. 39 different origin-destination pairs were represented in the sample, with a maximum of 11 LAX-EWR flights.

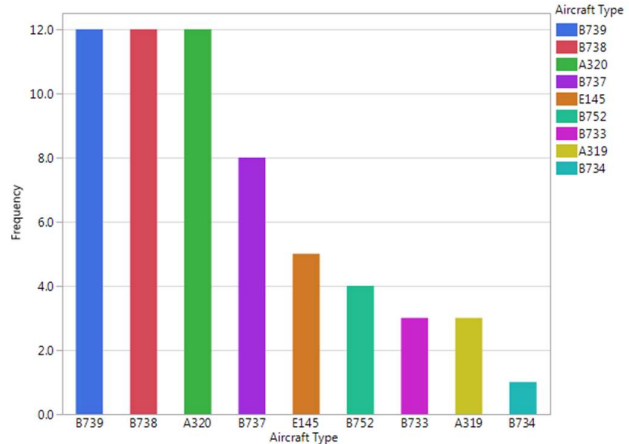


Figure 8 Aircraft type distribution in the sample of flights.

The raw trajectory data used as input by all the fuel burn models were provided by the FAA NextGen Office and consisted of National Offload Program (NOP) trajectories. These trajectories come from the blending of different surveillance sources but they do not include ground speed data. For this study, the ground speed was calculated from the succession of latitude, longitude and time data points, over variable intervals of multiple successive points. Smaller intervals were used for transition segments of climb and descent where larger variations in ground speed happen in shorter time compared to the cruise phase where a longer interval was chosen. A final smoothing of the resulting ground speed data was achieved by applying a 25-steps, 2-sided moving average. A side effect of this method was that 25 data points, both in the beginning and in the end of each trajectory, were lost. The benefit of using smoothed ground speed for the fuel estimation outweighed the loss of data.

All the models’ predictions in term of initial mass (TOW) and fuel burn were compared to the actual recorded values provided by Airlines for America (A4A). The recorded data set provides push-back weight, fuel burned on the ground and airborne. Unfortunately, there was only one number for each fuel burn category and not a continuous fuel flow record that would have provided for a more accurate validation of the models (like the example of Figure 4).

b. Data Results

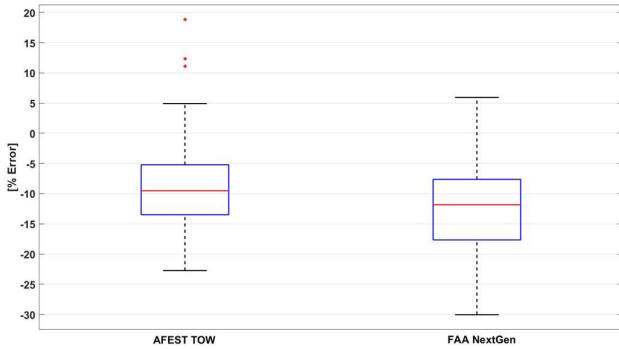


Figure 9 Fuel estimation models with known initial TOW errors.

As highlighted before, the point of this study was not to measure which model was the most accurate; this was infeasible with any statistical significance given the small sample of flights available.

The accuracy in prediction of both the TOW of the aircraft and the inflight fuel consumption will be compared across all models. The data will be divided into two sub-groups, one with the models that used the actual recorded TOW as initial input and the other with those models that did not use this information as input. A total of four different models and six total runs were collected and analyzed.

TABLE 2 FUEL ESTIMATION MODELS WITH KNOWN TOW STATISTICS

Fuel Burn Errors [%]	AFEST TOW	FAA NextGen
Minimum	-22.7	-30.1
25 th percentile	-13.5	-17.7
Median	-9.5	-11.8
75 th percentile	-5.2	-7.6
Maximum	18.8	5.9

1. Fuel burn results

Two models were run with known initial TOW, i.e. the AFEST model, from the FAA Tech Center, and the NextGen Office model. The AFEST model, significantly more complex than the NextGen Office one had a smaller median error, -9.5% versus -11.8%. AFEST had a larger maximum error, 18.8% versus 5.9%. The summary of the statistics is presented in Table 2 and a boxplot of the results in Figure 9.

Three models were run without initial known TOW; these were the AFEST model from the FAA Tech Center, ETAS from TU Dresden, and the Airservices Dalí, with BADA 3 and 4. These last two runs provided for an investigation of differences in outcomes produced by the two versions of the BADA performance model.

There was wide variability in the observed fuel burn error across these models, the smallest median error was obtained with the Dalí BADA 3 run with -3.9% while the largest median

error was observed for the AFEST run without known TOW, with -13.1%. Also the minimum and maximum error presented a wide variability. The summary of the statistics for these models is presented in Table 3.

TABLE 3 FUEL ESTIMATION MODELS WITH NO TOW STATISTICS

Fuel Burn Errors [%]	AFEST NO TOW	ETAS	Dalí	
			BADA 3	BADA 4
Min	-27.3	-31.2	-16.1	-13.8
25 th perc.	-17	-18.5	-7.8	-7.6
Median	-13.1	-11.8	-3.9	-4.6
75 th perc.	-8.5	-0.8	0.2	0.3
Max	13.3	46.9	26.5	32.7

A summary of the runs without known TOW is presented in Figure 10. From these runs it is worth noting how different models, with different level of complexity, present similar results. AFEST presented a median error of -13.1%. On the other end, the two runs with Dalí with different versions of BADA did not present major differences, BADA 3 performed slightly better, -3.9% versus -4.6% median error, and a smaller spread of the error. As mentioned previously, the sample is too small to draw any significant conclusions related to BADA 3 and 4. Airservices also indicated that during their own validation activities (see Section II.a), BADA 4 performed significantly better than BADA 3. ETAS performed in the middle with -11.8% median error but with a larger scatter.

All models performed worse than on previous studies and individual validation activities. Most likely, because of the quality of the surveillance tracks and the missing calibration for aircraft/airline specific data. In addition, only a single fuel burn figure was available that was recorded manually, likely containing some level of error. Ideally, a continuous fuel flow record should be available, as discussed earlier. Feedback from Airservices also indicated that the age of the fleet used in this sample, likely contributes to the low accuracy of the fuel estimation models (older airframes are generally less fuel efficient)

From Figure 10 it is clear that there is a general negative bias for all the model runs. This means that all the models were underestimating the fuel burn for the 60 flights in the sample. This fact was partially expected because of the missing portion (50 track points total) of the trajectories caused by the ground speed smoothing approach described in section III a. Since all the models take as input any piece of trajectory, they were predicting the fuel burn for a shorter trajectory that was actually flown. Only the runs in Dalí didn't present this issue, as this model adds nominal trajectory segments to the raw data to connect the initial and final portions of the trajectories to the origin and destination airports. This addition partially explains the smaller bias in their results.

It must be mentioned that with exception of the FAA NextGen Office experimental fuel estimator, each of the other models have been thoroughly tested, validated and accepted by the ATM community for use in their individual study objectives

(with better accuracy than reported here). The former model has been tested and validated to a lesser degree and it is still being improved. Its key goal is not to estimate fuel consumption for a specific flight, but to only account for the ATM-driven differences in fuel burn between flights with different vertical and horizontal profiles. Therefore, its inclusion into this comparison study will show more significant contribution in its next phase when the FAA will move the focus from studying accuracy in fuel estimation by flight to studying correlation between fuel estimates and trajectory characteristics.

2. TOW results

The statistics on the TOW predictions for all the models are presented in Table 4. Also for this error metric the results are variable across the models. In this case the Dalí BADA runs inverted their performance with BADA 4 being the more accurate with -0.4% median error versus -1.3% median error with BADA 3. This is interesting because the TOW error is considered to have a big impact on fuel burn estimation proving that there are multiple error components that affect this prediction. ETAS performs closely to Dalí with a median error of -3.5%, while AFEST presented a -8.1% median error. The TOW error statistics are presented in Figure 11.

TABLE 4 TOW ESTIMATION MODELS WITH NO TOW STATISTICS

Fuel Burn Errors [%]	AFEST NO TOW	ETAS	Dalí	
			BADA 3	BADA 4
Min	-16.1	-28.3	-8.5	-8.5
25 th perc.	-10.1	-8.2	-3.4	-2.5
Median	-8.1	-3.5	-1.3	-0.4
75 th perc.	-5.2	-0.2	1.28	1.8
Max	2.7	12.5	7.9	8.7

IV. DISCUSSION

In this paper a number of models were presented, each providing the ability to estimate the fuel burn of a flight from surveillance data. To fully support the design, implementation and review of any ATM performance improvement initiative or airspace change, three types of modelling are required:

- Baseline modelling to characterize current operations prior to any change with recorded actual tracks.

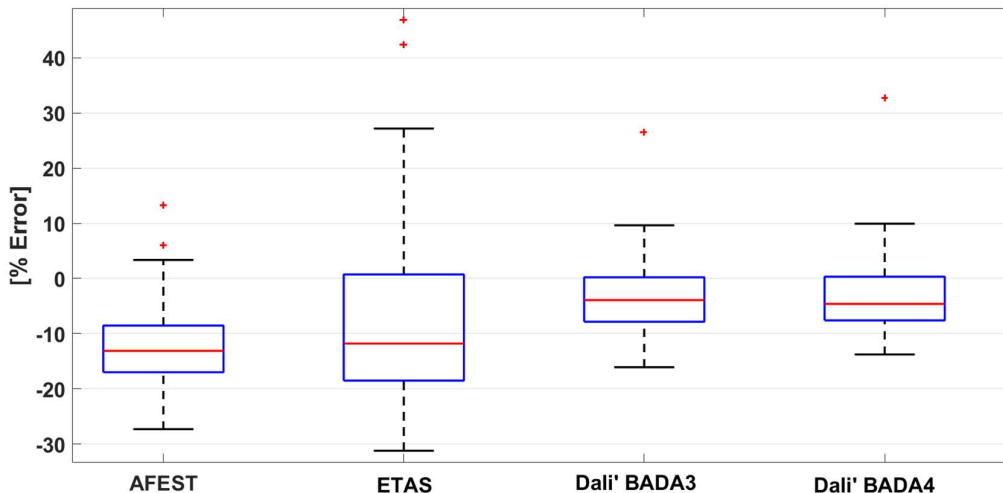


Figure 10 Fuel estimation models with no initial TOW boxplots.

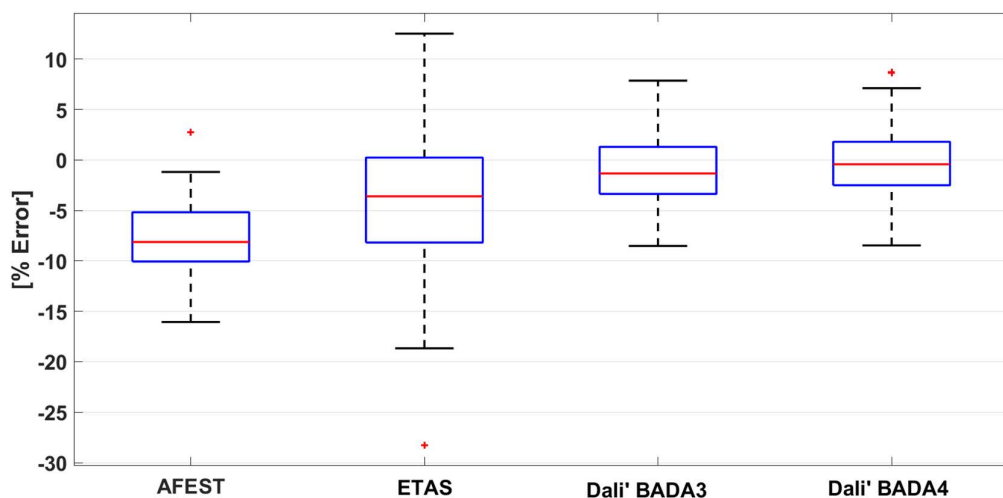


Figure 11 TOW estimations boxplots.

- Predictive modeling to provide a pre-implementation assessment of the benefit or impact of a proposed change with synthesized predicted tracks.
- Review (retrospective) modelling to provide a post-implementation assessment of the benefit or impact of an implemented change with recorded actual tracks.

Baseline and retrospective modelling can be performed on actual track data, but logically, predictive modelling needs to be based on the prediction of aircraft trajectories. These two types of modelling are subject to different errors.

In the case of modelling based on actual track data, there exists a large number of unknowns like aircraft weight and airframe specific characteristics, and error sources like meteorological forecast, aircraft performance, and noisy surveillance data etc. For example, assume an existing terminal area airspace (scenario A) is re-designed leading to shorter departure and arrival paths (scenario B). The estimated fuel burn over a period prior to the change is assessed, and compared to a similar period after the change. When subtracting the estimated fuel burn of one scenario from the other, the actual change in fuel burn is contaminated with a range of errors resulting from the models applied and by the actual changed conditions (e.g. actual winds/meteorology, potential airline fleet changes) between the times when the two scenarios were observed:

$$\begin{array}{l}
 \text{Estimated fuel burn scenario A -} \\
 \text{Estimated fuel burn scenario B} \\
 \hline
 \text{Fuel burn impact scenario change +} \\
 \text{Track data errors +} \\
 \text{Forecast data errors +} \\
 \text{Aircraft weight errors +} \\
 \text{Aircraft performance errors +} \\
 \text{Actual wind changes + ...}
 \end{array}$$

Because of the associated large spread in errors of the estimated fuel burn, as was illustrated in Section III, it will be extremely difficult to find a statistically significant change that can be confidently attributed to the change in scenarios being assessed. Instead, it could be better to assess the difference in flown track miles, flight time, or level-offs, which in general can be measured more accurately, and transform these to fuel savings. In fact, the FAA NextGen Office's fuel estimator has been developed with this purpose in mind. However, simplifying fuel estimation proved to be more complicated than anticipated; for instance, empirically observed differences in climb and descent rates may imply a different thrust setting and fuel flow, or may instead be executed at the same thrust setting but be driven by other factors such as increased aircraft weight or use of air brakes. Additional effort is still required to develop a more realistic version of simplified fuel estimation that is based solely on differences in empirical horizontal and vertical profiles.

In case of the predictive modelling, most of variables can be controlled (i.e. determined as part of the modelling process), meaning that when two predictive scenarios are being compared, most modelling errors cancel against one another (assuming a fairly linear process):

$$\begin{array}{l}
 \text{Estimated fuel burn scenario A -} \\
 \text{Estimated fuel burn scenario B} \\
 \hline
 \text{Fuel burn impact scenario change + ...}
 \end{array}$$

As a result, the difference can be confidentially attributed to the assessed change being modelled. The downside is that a predictive assessment is only valid if the modelled trajectories are representative of actual operations, which is often not the case due to the difficulty in modelling tactical ATC and flight crew behavior. The CDA benefit study of Robinson et al. [5] applied this technique. Comparing fuel burn estimated from non-CDA trajectories with modelled CDA trajectories would introduce noise as different errors affect the two types of modelling. Instead, the fuel burn along level segments during the descent was estimated assuming typical holding conditions. The same length of segment was then modelled in typical cruise conditions, with the difference leading to the CDA benefit estimate. A benefit of this approach is that the impact of assumptions on, for example, meteorological conditions and aircraft weight mostly cancel (constant components), but at the cost of simplification of the change being assessed. Depending on the accuracy of the models at hand, and the anticipated impact of the change being assessed, one method might be preferred over the other.

Another way to reduce the impact of the modelling error when estimating fuel burn from surveillance data, is through relative comparison with an appropriate reference. For example, the Dalí and the ETAS models compare the estimated fuel burn over a surveillance track, to a predicted reference trajectory for that flight based on the associated flight plan (see Figure 2) [10]. For both computations, the same forecast information, estimated take-off mass and aircraft performance models are used. By subtracting the reference fuel burn from the estimated fuel burn over the surveillance track, biases and constant components of the error cancel out, making this fuel burn relative to the reference an order of magnitude more accurate than the absolute figures. For example, by using the same estimated take-off mass for the estimated fuel burn on the flown trajectory and the reference trajectory, the strong dependence of the error on the take-off mass (as shown in the previous section of this paper), is accounted for. If the estimated actual fuel burn is subsequently expressed as a percentage of the reference fuel burn, some linear components of the error cancel out as well, provided the reference is adequate for the operation.

In summary, while some options exist to mitigate the impact of the fuel burn modelling error, consideration should be given if the estimated fuel burn is the most appropriate performance indicator for the objective being assessed. For an airline, both fuel burn and flight time are of the greatest interest, as both directly contribute to the operational cost of the flight. Additional distance or flight time can be an indicator of ATC intervention (e.g. radar vectoring and/or holding), and can be related to air traffic controller workload (i.e. greater controller workload results in more vectoring and holding, which in turn results in greater distances flown).

V. CONCLUSIONS

This paper discussed the necessity for fuel burn models from recorded 4D tracks to evaluate post operational changes in the

international air transportation system. It also showed that, to fully support the design, implementation and review of any ATM performance improvement initiative or airspace change, predictive modelling is also required to assess the impact of changes pre-implementation.

As it is clear from the cross-comparison of various models, with different levels of complexity, the accuracy of the models is not always adequate to the changes that model users want to observe. This is not just the limitation of the models itself, but also the accuracy of the input data. Even the highest fidelity model will significantly underperform if low quality input data is provided. Therefore, fuel burn metrics must be used carefully. As expected, one of the results of this analysis is that fuel burn estimation models with different level of complexity, present different level of accuracy performance. For example, models such as Dalí and ETAS that adapt the fuel burn to the aircraft tail number, generally perform better. Nonetheless the different level of complexity should be tailored to the objective of the analysis that is being performed.

Fuel burn estimation models are valuable to the ATM community, but as this paper has demonstrated, accuracy of these models depends on the quality of the inputs and is limited due to many assumptions made. Therefore, care should be taken in the application of these models. In particular,

- The limitations of these models and applied input data must be understood and transparent upon application,
- The model's level of complexity has to be adequate for the objective of the study being performed
- Fuel burn savings should be normalized to control for factors not related to procedural changes, and
- Fuel burn is not always the most appropriate performance indicator for flight efficiency.

In the next phase of this effort, the focus will shift from studying accuracy in fuel estimation by flight to studying correlation between fuel estimates and trajectory characteristics that isolate ATM-driven changes in operational performance. Most importantly, this study has opened an important international conversation about considerations and applicability of existing fuel modeling tools and approaches to assessment of trade-offs and benefits from ATM improvements and concepts.

ACKNOWLEDGMENT

The authors would like to acknowledge the contributions of Lakshmi Vempati (Harris Corp.); Paul Truong, Martin Durbin, James Bonn, Dave Knorr and Joe Post (FAA, NextGen Office); Bob Vivona, Howard Goldowsky and Lynne Oliver (Engility Corporation); and, Brian Schnitzer (General Dynamics Information Technology).

REFERENCES

- [1] Federal Aviation Administration (FAA), "NextGen Implementation Plan 2016," Washington, DC, June 2016.
- [2] SESAR Joint Undertaking (SJU), "Consolidated Annual Activity Report 2015".
- [3] Australian Government - Department of Infrastructure and Regional Development, "Australia's Air Traffic Management Plan - Draft", 2016.
- [4] Chatterji, G., "Fuel Burn Estimation Using Real Track Data", 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Virginia Beach, VA, 2011.
- [5] Robinson, III, J. E., and Kamgarpour, M., "Benefits of Continuous Descent Operations in High-Density Terminal Airspace Under Scheduling Constraints", 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Fort Worth, TX, 2010.
- [6] Trani, A., Wing-Ho, F., Shilling, G., Baik, H., Seshadri, A., "A Neural Network Model to Estimate Aircraft Fuel Consumption", 4th AIAA Aviation Technology, Integration and Operations (ATIO) Conference, Chicago, IL, 2004, AIAA 2004-6401.
- [7] Belle, A, Sherry, L., "Estimated Fuel Burn Performance for MDW Arrivals", AIAA Aviation Conference, Los Angeles, CA, 2013.
- [8] Rumler, W., Günther, T., Weißhaar, U., "Flight Profile Variations due to the Spreading Practice of Cost Index Based Flight Planning", 4th International Conference on Research in Air Transportation, 2010.
- [9] Amezcua, J., Hopper, J., "Estimating U.S. NAS-wide Fuel Consumption using BADA 3", presented at the BADA User Meeting, 2014.
- [10] Bronsvort, J., Zissermann, P., Barry, S.I., & McDonald, G.N. "A Framework for Assessing and Managing the Impact of ANSP Actions on Flight Efficiency". 11th USA/Europe ATM Research and Development Seminar, Lisbon, Portugal, 2015.
- [11] López Leonés, J., Vilaplana, M.A., Gallo, E., Navarro, F.A., & Querejeta, C., "The Aircraft Intent Description Language: A key enabler for Air-Ground synchronization in Trajectory-Based Operations". Proceedings of the 26th Digital Avionics Systems Conference, Dallas, TX, USA.
- [12] Kaiser, M., Rosenow, J., Fricke, H., Schultz, M., "Tradeoff between Optimum Altitude and Contrail Layer to Ensure Maximum Ecological En-Route Performance Using the Enhanced Trajectory Prediction Model (ETPM)", ATACCS 2012
- [13] Kaiser, M., Schultz, M., Fricke, H., "Automated 4D Descent Path Optimization using the Enhanced Trajectory Prediction Model (ETPM)", International Conference on Research in Airport Transportation, Berkeley, 2012
- [14] Fricke, H., Seiss, C., Herrmann, R., "Fuel and Energy Benchmark Analysis of Continuous Descent Operations", 11th USA/Europe ATM Research and Development Seminar, Lisbon, Portugal, 2015.
- [15] FAA, <https://acy.tc.faa.gov/>, accessed January 2017.
- [16] Oaks, R., Ryan, H., Paglione, M., "Prototype Implementation and Concept Validation of a 4-D Trajectory Fuel Burn Model Application", AIAA Guidance, Navigation, and Control Conference, Toronto, Canada, August 2010, AIAA 2010-8164.
- [17] Young, J., Schnitzer, B., Kee, M., "A Fast-Time Simulation to Measure the Effects of Temporally Inaccurate Convective Weather Forecasts in Terminal Airspace", DOT/FAA/TC-TN15/58, November 2015.
- [18] Young, J., Kee, M., Young, C., "Effects of Future Space Vehicle Operations on a Single Day in the National Airspace System: A Fast-Time Computer Simulation", DOT/FAA/TC-TN15/14 (updated version as of June 8, 2015).
- [19] Young, J., Bassett, P., Hailston, K., "Evaluating Integrated Arrival/Departure Control Services Using Fast-Time Simulation", DOT/FAA/TC-TN13/4, February 2013.