Evaluation of Flight Efficiency for Stockholm Arlanda Airport using OpenSky Network Data

Tatiana Polishchuk\textsuperscript{1}, Anastasia Lemetti\textsuperscript{1}, and Raúl Sáez\textsuperscript{2}

\textsuperscript{1} Communications and Transport Systems, ITN, Linköping University Norrköping, Sweden
\texttt{firstname.lastname@liu.se}
\textsuperscript{2} Department of Physics, Technical University of Catalonia (UPC), Castelldefels, Barcelona, Spain
\texttt{raul.saez.garcia@upc.edu}

\textbf{Abstract}

Identification of causes of the delays within transition airspace is an important step in evaluating performance of the Terminal Maneuvering Area (TMA) Air Navigation Services: without knowing the current performance levels, it is difficult to identify which areas could be improved. Inefficient vertical profiles within TMA and deviations from the optimal flight paths due to bad weather conditions are the main sources of performance decline. In this work, we analyse punctuality and vertical efficiency of Stockholm Arlanda airport arrivals, and seek to quantify the fuel consumption impact associated with the inefficient vertical flight profiles within the Terminal Maneuvering Area (TMA).

We use Opensky Network data for evaluation of the Stockholm Arlanda airport performance, comparing it to the DDR2 data provided by Eurocontol, outlining the advantages and disadvantages of both.

\section{1 Introduction}

Air transport punctuality in Europe is one of the major concerns for all the aviation stakeholders. Air traffic delays induce large tactical and strategic costs for airlines, airports and passengers. Delays may also cause environmental damage by increasing fuel consumption and gas emissions [24].

Determining causes of aviation delays as well as evaluation of the other aspects of the airport performance is the main purpose of the Impact of Fleet Diversity and Weather on Emissions, Noise and Predictability (IFWHEN) project supported by the Swedish Transport Agency (Transportstyrelsen) and in-kind participation of LFV (Swedish Air Navigation Service Provider). Within this project we evaluate the current state of Stockholm Arlanda airport performance, targeting to suggest the areas for further improvements in airport procedures. In particular, we analyse trade-offs between noise footprint, fuel consumption (which directly impacts emissions), flight time, resilience and other Key Performance Indicators (KPIs) for individual flight, taking into account weather impact on individual flights on short-time, tactical scale.

In this paper, we evaluate punctuality and vertical efficiency of Stockholm Arlanda airport arrivals. We seek to quantify the fuel consumption impact associated with deviations from the flight plans and inefficient vertical flight profiles within Terminal Maneuvering Area (TMA). We share our experience working with different data sources (OpenSky Network data and

\textsuperscript{This research is a part of the IFWHEN project supported by the Swedish Transport Agency (Transportstyrelsen) and in-kind participation of LFV. It is also supported by the SESAR Joint Undertaking under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 783287”.

\textsuperscript{1}}
Eurocontrol DDR2) and discuss their applicability for calculations inside TMA, outlining the advantages and disadvantages.

The rest of the paper is organized as follows. In Section 2 we review related work on the topic and provide background information on the methods we use for analysis of the performance of Stockholm Arlanda airport arrivals. We present the results of data analysis in Section 3 and summarize our findings in Section 4.

2 Background

This section reviews previous work and provides background information related to evaluation of flight efficiency of airport arrivals.

2.1 Related Work

Eurocontrol developed the methodology used by its Performance Review Unit (PRU) for the analysis of vertical flight efficiency during climb and descent [13]. Performance Review Commission of Eurocontrol made an assessment of air traffic management in Europe for the year 2018, where among other indicators reviewed air traffic punctuality and vertical flight inefficiency at the top 30 European airports, including Stockholm airport Arlanda [11]. In addition, Eurocontrol PRU develops and maintains open access cloud based data repositories to enable stakeholders to reproduce the performance review results [28].

In [24] fuel consumption is evaluated for terminal areas with a Terminal Inefficiency metric based on the variation in terminal area fuel consumed across flights, reported by a major U.S. airline. Using this metric they quantify the additional fuel burn caused by Air Traffic Management (ATM) delay and terminal inefficiencies. Furthermore, in [15] and [29], an analysis of fuel savings of the Continuous Descent Operations (CDO) with respect to conventional procedures is analyzed. A reduction in fuel consumption of around 25-40% by flying CDO was reported.

Estimation of the flight inefficiencies in terms of extra fuel burn calculated based on the algorithm proposed in [5] was considered in the scope of APACHE project (a SESAR 2020 exploratory research project) [22], [21], but mostly for en-route flight phase. In this work, we apply similar techniques to fuel estimation during the descent phase within TMA.

Classification and analysis of causes of airport delays was a topic of interest for many years. In early works [2], [4] weather uncertainties are mentioned as the main contributor to the deviations in airport schedules. Inability of ATM, which limits the airport capacity, to meet the ever raising demand is also one the most significant influencing factor resulting in airport delays [3].

Correia [7] analyzed vertical flight efficiency and additional fuel burn in details at Lisbon Airport for the years 2014-2017 and compared statistics for several European airports, including Stockholm Arlanda, for the year 2017. Evaluation of punctuality, vertical efficiency and additional fuel burn for Stockholm Arlanda airport arrivals was considered in [18].

Analysis of the impact of different weather phenomena on arrival punctuality and vertical flight efficiency of Stockholm Arlanda airport arrivals in 2018 was recently presented in [17].

2.2 KPIs

To capture inefficiencies within Stockholm TMA, in this work we use the following KPIs: average arrival delay, percent of delayed flights (delayed more than 15 minutes according to the
schedule), vertical flight efficiency (VFE) during descent phase (average time flown level, distance flown level, percent of level flights), additional time in TMA and fuel-based PIs.

Arrival delays can be a result of trajectories deviations from flight plans and could lead to extra expenses and additional fuel burn. To compute several KPIs we compare the actually flown flight trajectories against the corresponding ones of the flight plans. Flight plans are obtained from the Demand Data Repository (DDR2 [14], m1 file format) hosted by Eurocontrol. For the historical flight trajectories we use DDR2 (m3 file format) and the Historical Database of the OpenSky Network [20,27].

2.2.1 Fuel-Based PIs

Fuel-based PIs capture inefficiencies on tactical ATM layer in vertical domain as explained in [23]. The objective is to compare the fuel consumption of CDO trajectories with the actual flown trajectories. Fuel-based performance indicators are calculated using the 4.2 version of the Base of a Aircraft Data (BADA) [12].

The first expression used, known as the Total-Energy Model, equates the rate of work done by forces acting on the aircraft to the rate of increase in potential and kinetic energy, that is:

\[(T - D)V_{TAS} = mg \frac{dh}{dt} + mV_{TAS} \frac{dV_{TAS}}{dt}\]  \( (1) \)

Here \( T \) is the thrust acting parallel to the aircraft velocity vector, \( D \) is the aerodynamic drag, \( m \) is the aircraft mass, \( h \) is the geodetic altitude, \( g \) is the gravitational acceleration and \( V_{TAS} \) is the true airspeed.

The drag force is computed as follows:

\[ D = \frac{1}{2} \cdot \delta \cdot p_0 \cdot \kappa \cdot S \cdot M^2 \cdot C_D \] \( (2) \)

Here \( \delta \) is the pressure ratio, \( p_0 \) is the standard atmospheric pressure at mean sea level (MSL), \( \kappa \) is the adiabatic index of air, \( S \) is the wing reference area, \( M \) is the Mach number and \( C_D \) is the drag coefficient. BADA proposes equations for computing \( C_D \) depending on the aircraft configuration, and modelled as a polynomial of lift coefficient \( C_L \).

Three separate thrust models are proposed in BADA, depending on the engine type: turbofan, turboprop or piston. Each model includes the contribution from all engines and provides the thrust as a function of airspeed, throttle setting and atmospheric conditions. The general formula of the thrust force, \( T \), is:

\[ T = \delta \cdot W_{mref} \cdot C_T \] \( (3) \)

Here \( \delta \) is the pressure ratio, \( m_{mref} \) is the reference mass (obtained from the Propulsive Forces Model (PFM)), \( W_{mref} \) is the weight force at \( m_{mref} \) and \( C_T \) is the thrust coefficient, which is a function of Mach number.

For the three engine types, BADA proposes different equations to compute the thrust coefficient \( C_T \) depending on the engine rating: maximum climb, maximum cruise, idle and no rating (direct throttle parameter input).

For estimation of the fuel consumption, BADA proposes once again a different model depending on the engine type, and also depending on the engine rating. Each model includes the contribution from all engines and provides the fuel consumption as a function of airspeed,
throttle parameter and atmospheric conditions. The general formula for the fuel consumption, $F$, is:

$$F = \delta \cdot \theta^{\frac{1}{2}} \cdot W_{mref} \cdot a_0 \cdot L_{HV}^{-1} \cdot C_F$$

(4)

Here $\delta$ is the pressure ratio, $\theta$ is the temperature ratio, $a_0$ is the speed of sound at MSL in standard atmosphere, $L_{HV}$ is the fuel lower heating value (obtained from the PFM) and $C_F$ is the fuel coefficient, which depends on thrust for non-idle ratings. For each aircraft model, BADA provides an xml file with the corresponding aircraft performance data. For instance, the coefficients used to compute the thrust coefficient $C_T$ of the thrust equation (3) are in this file.

With the equations stated above, and the xml files for each aircraft, it is possible to compute the fuel consumption of a trajectory. The process followed is detailed below:

- **Thrust computation**: if the aircraft is climbing, max climb rating is chosen and the corresponding thrust formula (depending on the engine type) is applied. If the aircraft is descending, an idle rating is assumed. In level-offs, the total-energy model (equation (1)) is used in order to compute the corresponding aircraft thrust (drag is computed previously with equation (2)).

- **Fuel consumption computation**: for non-idle ratings, the thrust computed in the previous step is used to obtain the fuel coefficient $C_F$ used in equation (4). For descents, idle rating is assumed.

Wind was taken into consideration when computing the fuel consumption, and it was obtained from historical weather data from the National Oceanic and Atmospheric Administration (NOAA), as detailed in 2.3. Furthermore, a 90% of the maximum landing mass has been assumed at the destination airport for all aircraft.

**Generation of CDO trajectories**  In order to generate the CDO trajectories an optimal control problem has to be solved as explained in details in [26]. First, a state vector with the initial conditions is needed. In this paper, it has been chosen as $x = [v, h, s]$, where $v$ is the true airspeed, $h$ - the altitude of the aircraft, and $s$ - the distance to go. In order to obtain environmentally friendly trajectories, idle thrust is assumed and speed-brakes use is not allowed throughout the descent. In such conditions, the flight path angle is the only control variable in this problem ($u = [\gamma]$), which is used to manage the energy of the aircraft and achieve different times of arrival at the metering fix with minimum fuel consumption and noise nuisance.

The dynamics of $x$ are expressed by the following set of ordinary differential equations, considering a point-mass representation of the aircraft reduced to a "gamma-command" model, where vertical equilibrium is assumed (lift balances weight). In addition, the cross and vertical components of the wind are neglected, and the aerodynamic flight path angle is assumed to be small (i.e., $\sin \gamma \simeq \gamma$ and $\cos \gamma \simeq 1$):

$$\dot{x} = \begin{bmatrix} \dot{v} \\ \dot{h} \\ \dot{s} \end{bmatrix} = \begin{bmatrix} \frac{T_{idle} - D}{m} - g\gamma \\ \frac{v\gamma}{v + w} \\ v+w \end{bmatrix}$$

(5)

where $T_{idle} : \mathbb{R}^n \to \mathbb{R}$ is the idle thrust; $D : \mathbb{R}^{n_x} \times \mathbb{R}^n \to \mathbb{R}$ is the aerodynamic drag; $g$ is the gravity acceleration; $w$ is the wind and $m$ - the mass, which is assumed to be constant.
because the fuel consumption during an idle descent is a small fraction of the total $m$ [6]. The longitudinal component of the wind $w : \mathbb{R} \rightarrow \mathbb{R}$ is modelled by a smoothing spline [8]:

$$w(h) = \sum_{i=1}^{n_c} c_i B_i(h)$$

(B), $i = 1, \ldots, n_c$, are the B-spline basis functions and $c = [c_1, \ldots, c_{n_c}]$ are control points of the smoothing spline. It should be noted that the longitudinal wind has been modelled as a function of the altitude only, as done in similar works [9]. The control points of the spline approximating the longitudinal wind profile are obtained by fitting historical weather data [19].

In this paper, the trajectory is divided in two phases: the latter part of the cruise phase prior the ToD, and the idle descent down to the metering fix. Assuming that the original cruise speed will not be modified after the optimization process, the two-phases optimal control problem can be converted into a single-phase optimal control problem as follows:

$$J = \frac{f}{v_{cruise}} + \int_{t_0}^{t_f} (f_{idle} + CI) \, dt$$

where $f : \mathbb{R}^{n_x \times n_u} \rightarrow \mathbb{R}$ and $f_{idle} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ are the nominal and idle fuel flow, respectively; and CI is the cost index, which is a parameter chosen by the airspace user that reflects the relative importance of the cost of time with respect to fuel costs [1]. The CI is estimated by assuming that the aircraft was flying at the optimal speed in the cruise phase, as shown in [25].

To generate the optimum trajectories, five input parameters are used: aircraft model, cruise altitude, distance to go (i.e., the distance remaining to the metering fix by following a given route), speed (i.e., the true airspeed of the aircraft in cruise), and the cost index.

2.3 Data

We use two independent databases that provide air traffic data: the Demand Data Repository (DDR2) hosted by Eurocontrol [14] and the Historical Database of the Opensky Network [20,27], a crowd-sourced ground sensor network collecting air traffic data from transponder signals that are continuously transmitted by aircraft.

Eurocontrol offers data in SO6 format that is delimiter separated values files which store flight trajectories (the lists of waypoints containing aircraft position, barometric altitude and identity). We use DDR2 data in two formats: SO6 m1 and SO6 m3. The first one provides the last submitted flight plans, while the second one consists of the actual trajectories.

Opensky supports two types of data: state vectors and tracks. The aircraft state vector is a summary of all tracking information at a certain point in time. Opensky tracks do not contain all information about the flights the database keeps, but rather show the aircraft’s general movement pattern as a list of waypoints, similar to Eurocontrol’s DDR2; however, the choice of waypoints in Opensky and DDR2 does not match.

In this work we obtain flight plans from the SO6 m1 DDR2 files, but for the historical flight trajectories we use both DDR2 (SO6 m3 file format) and the Historical Database of the Opensky Network. To estimate the impact of vertical and horizontal flight inefficiencies on the fuel waste we use BADA v4 [12]. In addition, the historical weather data is taken from the National Oceanic and Atmospheric Administration (NOAA) through the National Operational Model Archive and Distribution System (NOMADS) [19].
3 Results

This section presents the results of the flight efficiency evaluation for the Stockholm Arlanda airport arrivals during the year 2018.

3.1 Vertical Flight Efficiency

Vertical inefficiencies during the descent phase result from the inability of flights to keep up CDO (Continuous Descent Operations). This type of operations enables the execution of a flight profile optimized to the operating capability of the aircraft, giving as a result optimal continuous engine-idle descents (without using speed-breaks) that reduce fuel consumption, gaseous emissions and noise nuisance. If the aircraft levels at intermediate altitudes before landing, this descent is considered as vertical inefficient.

For evaluation of VFE we use several KPIs described by ICAO [16]: average time flown level, average distance flown level per flight inside TMA, average percent of distance flown in level flight inside TMA, average percent of time flown in level flight inside TMA and percent of level flights. We calculate level flights inside TMA using the techniques proposed by Eurocontrol in [13] with small changes. We identify the point of the trajectory in which the aircraft enters the TMA and use it as a starting point for the calculations (instead of the Top of Descent (ToD), which may lie outside of TMA). A level segment is detected when the aircraft is flying with the vertical speed below the certain threshold. We use the value of 300 feet per minute for this threshold, the minimum time duration of the level flight is considered 30 seconds, and these 30 seconds are subtracted from each level duration as suggested in [13].

The results are summarized in Table 1.

<table>
<thead>
<tr>
<th>Months in 2018</th>
<th>$n_{arr}$</th>
<th>$P_{%}$</th>
<th>$L_{avg}$</th>
<th>$KPI_{19.1}$</th>
<th>$KPI_{19.1}$</th>
<th>$S_D$</th>
<th>$KPI_{19.2}$</th>
<th>$KPI_{19.2}$</th>
<th>$S_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>7721</td>
<td>55</td>
<td>1.00</td>
<td>2.80</td>
<td>5</td>
<td>4.62</td>
<td>0.72</td>
<td>5</td>
<td>1.11</td>
</tr>
<tr>
<td>02</td>
<td>7419</td>
<td>52</td>
<td>0.87</td>
<td>2.72</td>
<td>4</td>
<td>4.90</td>
<td>0.70</td>
<td>5</td>
<td>1.18</td>
</tr>
<tr>
<td>03</td>
<td>8131</td>
<td>56</td>
<td>1.08</td>
<td>2.99</td>
<td>5</td>
<td>4.79</td>
<td>0.77</td>
<td>6</td>
<td>1.17</td>
</tr>
<tr>
<td>04</td>
<td>8944</td>
<td>54</td>
<td>1.00</td>
<td>2.77</td>
<td>4</td>
<td>4.97</td>
<td>0.69</td>
<td>5</td>
<td>1.16</td>
</tr>
<tr>
<td>05</td>
<td>9552</td>
<td>50</td>
<td>0.88</td>
<td>2.22</td>
<td>4</td>
<td>4.73</td>
<td>0.52</td>
<td>4</td>
<td>1.01</td>
</tr>
<tr>
<td>06</td>
<td>8923</td>
<td>51</td>
<td>0.90</td>
<td>2.68</td>
<td>4</td>
<td>4.61</td>
<td>0.69</td>
<td>5</td>
<td>1.12</td>
</tr>
<tr>
<td>07</td>
<td>8426</td>
<td>46</td>
<td>0.68</td>
<td>2.38</td>
<td>4</td>
<td>4.21</td>
<td>0.61</td>
<td>4</td>
<td>1.02</td>
</tr>
<tr>
<td>08</td>
<td>8915</td>
<td>47</td>
<td>0.73</td>
<td>2.25</td>
<td>3</td>
<td>3.95</td>
<td>0.56</td>
<td>4</td>
<td>0.94</td>
</tr>
<tr>
<td>09</td>
<td>8779</td>
<td>48</td>
<td>0.79</td>
<td>2.28</td>
<td>3</td>
<td>4.03</td>
<td>0.59</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>9162</td>
<td>51</td>
<td>0.80</td>
<td>2.40</td>
<td>4</td>
<td>4.10</td>
<td>0.64</td>
<td>4</td>
<td>1.04</td>
</tr>
<tr>
<td>11</td>
<td>8558</td>
<td>41</td>
<td>0.59</td>
<td>1.91</td>
<td>3</td>
<td>3.71</td>
<td>0.49</td>
<td>3</td>
<td>0.91</td>
</tr>
<tr>
<td>12</td>
<td>6954</td>
<td>49</td>
<td>0.73</td>
<td>2.68</td>
<td>4</td>
<td>5.21</td>
<td>0.71</td>
<td>5</td>
<td>1.25</td>
</tr>
</tbody>
</table>

According to Eurocontrol’s Performance Review Report 2018 [11] average time flown level during descend is 1.1 minute at Arlanda in 2018, which does not contradict our calculations. The report [11] also states that vertical flight efficiency during climb and descent at the top 30 airports remained in 2018 at the same level as in 2017. From that we can conclude that the share of CDO during descent at Arlanda is approximately 50% ([10]), which coincides with our results.
We used OpenSky states data for the trajectories actually flown inside TMA. The OpenSky states data is very accurate as it provides altitudes for each second, while DDR2 provides much less waypoints for the flight trajectories. Figures 1 and 2 show an example of vertical and lateral profiles of two randomly-chosen flights (callsigns SAS410 and SAS964), obtained from DDR m3 data and OpenSky data (both tracks and states) for comparison.

The difference between the OpenSky states and tracks is clearly seen in Figure 1. We discovered some problems with the OpenSky tracks’ obtained using the techniques proposed by the OpenSky team. Although tracks’ points are supposed to constitute a subset of states’ points, it is clearly not the case in this example.

Figure 2 demonstrates non-optimal paths within TMA, as a result of bad weather conditions. OpenSky reflects holding patterns, while DDR only sketches the resulting path. Figure 3 confirms that OpenSky provides more accurate aircraft tracking: while DDR shows the aircraft landed on one runway, the last segments of OpenSky data correspond to the other runway in Arlanda airport.

We conclude that OpenSky states data is more suitable for reconstruction of the flight trajectories both in vertical and lateral domains. Even though the calculations based on such a dense data are very memory consuming; still, we managed to process OpenSky states inside TMA, as it only represents a short portion of the total flight. For the tasks requiring less number of waypoints, tracks would be a good data source, but current methods used for extracting tracks need to be refined.

![Figure 1: Vertical profile of the arrival flight with callsign SAS410 in Stockholm Arlanda airport TMA from DDR2 and OpenSky data on January 01, 2018.](image)

### 3.2 Additional Fuel Burn

Fuel-based PIs capture inefficiencies on tactical ATM layer in vertical domain as explained in [23]. The objective is to compare the fuel consumption of CDO trajectories within TMA with the actual flown trajectories. Fuel-based performance indicators are calculated using the 4.2 version of the Base of a Aircraft Data (BADA) [12].
We compare the fuel consumption of the actual trajectories with the CDO profiles. For the actual flown trajectories we use DDR2 M3 data, Opensky tracks and Opensky states. Then, in order to generate the CDO trajectories, an optimal control problem has to be solved as explained in detail in [26]. We calculate the additional fuel burn per day for all Arlanda airport arrivals during the month of February. Figure 4 illustrates the results.
While the fuel consumption is higher in Opensky tracks than in DDR (usually the difference is between 1% and 10%), the additional fuel burn with respect to CDO remains almost the same. However, if Opensky states are used, which is a more accurate data, the difference in fuel burn with respect to CDOs is at most 30%, as it can be observed in Figure 4.

![Figure 4: Additional fuel burn (in percent per day) due to inefficient vertical profiles, calculated as the difference between the actual flown trajectories (DDR m3, Opensky tracks or Opensky states) and the optimal trajectories, in total fuel consumption per day inside TMA for the month of February 2018.](image)

Absolute values for the fuel consumption are shown in Figure 5.

We observe that fuel consumption suffers significant changes throughout the month. The fuel calculation took into account wind components, which demonstrated the impact on the efficiency of vertical profiles, and our results confirm the decrease of the fuel consumption during the corresponding days with the low gust values reported, and increase of the fuel consumption during the days with the highest gust measurements (studied in detail in [17]).

The figures demonstrate that CDO provide a reduction of fuel consumption around 60-65% if actual flown trajectories are obtained from DDR2 or Opensky tracks, which constitutes a significant inefficiency of the vertical profiles actually flown. It is important to recall that we calculate the fuel inside TMA only; if the whole descent was compared, the difference would have been lower, as level-offs at lower altitudes are more detrimental for efficiency than those at higher altitudes. Regarding Opensky states, however, the reduction in fuel consumption is much less, from around a 5% up to a 30% depending on the day, which we believe is a more realistic result as the data obtained from Opensky states is more accurate (there is a data point each second inside TMA).

Regarding the data being used as actual flown trajectories, it could be affirmed that, on the one hand, DDR2 data is not very accurate in TMA, as it tends to provide very few segments (around 10 to 15 waypoints) if compared with Opensky data. On the other hand, Opensky
tracks are not as accurate as Opensky states. More levels can be found in tracks than in states (Figure 1), which increases the fuel consumption. However, tracks have a lower amount of data points, which reduces significantly the computation time. While the computation time for tracks could be around 10 seconds for 1 trajectory, the computation time for states could reach up to 2 minutes in some cases.

3.3 Punctuality of Arrivals

In addition, we compare two sets of actual arrival flight times: the one based on DDR2 data and the one provided by OpenSky Network, and conclude that due to incompleteness of the OpenSky data (OpenSky contains about 84% of DDR2 flights), DDR2 produces more accurate delay statistics. The resulting statistics for the Stockholm Arlanda airport arrivals during the year 2018 are presented in [18].

3.4 Discussion

Better data granularity of OpenSky makes it a better option to estimate fuel consumption inside TMA. While DDR usually provides only 10 to 15 waypoints inside TMA, in OpenSky tracks there are about 60-80 waypoints (depending on the trajectory), which makes it more reliable. However, there are also some errors in OpenSky data, which tend to cause very high values when computing the fuel consumption. Several data outliers were found in OpenSky tracks. For instance, in some trajectories there are repeated waypoints, even with the time advancing (which would mean the aircraft remains still, which is not possible). There are other situations where the latitude and longitude do not seem to correspond to the trajectory we are dealing with, and some of the speed values that could be extracted from Opensky tracks are wrong too. Another drawback of the OpenSky data is the lack of flight identifier, and aircraft type information, needed for the calculation of fuel consumption. We resolved this problem by merging the data with DDR flight plans (SO6 m1 files). Regarding Opensky states, there are even more data points, one each second, so the data is even more accurate. However, the time...
it takes to compute the fuel consumption is higher than in Opensky tracks, up to 2 minutes in some cases. As it happens with Opensky tracks, Opensky states contains some outliers that need to be filtered.

4 Conclusions

We evaluate the flight efficiency of Stockholm Arlanda airport arrivals using two different sources of historical data: Opensky Network data and Eurocontrol's DDR2. We share our experience working with OpenSky Network data and discuss its applicability for calculations inside TMA, outlining the advantages and disadvantages.

We conclude that due to incompleteness of the Opensky data, DDR2 is better suitable for delay statistics. Although Opensky has certain problems with data integrity and additional methods to detect and filter the different kinds of integrity breaches are needed, it provides the data of high density in form of state vectors with accurate three-dimensional aircraft positions along with precise timestamps for the signal arrivals. Thus, we recommend Opensky state vectors as a valuable data source for research with high accuracy demand, which however requires sufficient computational resources. We continue testing usability of the Opensky and DDR data within our research projects.

References


