Ref. Ares(2024)7985215 - 11/11/2024

Next generation data-driven reference European models and methods towards silent and green aircraft operations around airports

Horizon Europe | HORIZON-CL5-2022-D5-01-12

D2.2 – Advanced pre-processing algorithms and preliminary models and methods for aircraft operation reconstruction and statistical dispersion of flight operations

This project receives funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101095754 (NEEDED).

This publication reflects only the author's view and the European Climate, Infrastructure and Environment Executive Agency (CINEA) is not responsible for any use that may be made of the information it contains.

REVISION HISTORY

Deliverable No.	D2.2	
Deliverable Title	Advanced pre-processing algorithms and preliminary models and methods for aircraft operation reconstruction and statistical dispersion of flight operations	
Deliverable Type	Report	
Dissemination level	Public	
Written By	Lorenzo Dorbolò, Pietro Giannattasio, Marco Pretto (UNIUD); James Page, Alessandro Zanon (AIT); Laurent Cavadini (ECTL); Junzi Sun (TUDELFT).	14-06-2024
Checked by	Lorenzo Dorbolò (UNIUD), Laurent Cavadini (ECTL);	25-10-2024
Approved by	Alessandro Zanon (AIT)	11-11-2024
Status	Final	11-11-2024

TABLE OF CONTENTS

PROJECT ABSTRACT

NEEDED responds to the second and third bullets of the “expected outcome” of the HORIZON-CL5-2022-D5-01-12 topic, delivering the next generation data-driven reference European models and methods to estimate present and future aircraft emissions (pollutants and noise), achieving TRL 4 at the end of the project. To do so, NEEDED will advance the state of the art by:

- improving the accuracy of the reconstruction of aircraft operations by using real-world ADS-B data,
- advancing emission inventories for current and future aircraft technologies, while delivering more accurate pollution dispersion models,
- extending the applicability of the ECAC Doc.29 noise model towards future aircraft technologies,
- performing more accurate estimation of the number of people affected by local air transport operations by using dynamic population maps.

These activities are complemented by (i) local air quality (LAQ) and experimental noise measurements performed at Rotterdam The Hague Airport and Larnaca Airport, (ii) validation of the NEEDED toolchain in a 30-week pilot study involving three airports, and (iii) delivery of a methodology to optimize the flight patterns for minimum detrimental impact on the population in present and future scenarios. The project aims to function as a technology enabler, laying the methodological groundwork for facilitating the entry into service of transformative aircraft technologies while capitalizing on the potential of ADS-B data. The enabler role of NEEDED to the future Air Traffic Management (ATM) regulation and policies is facilitated by the direct involvement of EUROCONTROL.

The consortium combines a wide portfolio of competences from 11 partners from 8 different EU member states (Austria, Belgium, Italy, Sweden, The Netherlands, France, Spain and Cyprus) plus 1 non-EU Country and it is coordinated by AIT Austrian Institute of Technology. NEEDED is scheduled to run from January 1st, 2023, to December 31st, 2026, for a total duration of 48 months and has received funding from the European Union’s Horizon Europe research and innovation programme under Grant Agreement no. 101095754. A full list of partners and funding can be found at: https://cordis.europa.eu/project/id/101095754

LIST OF ABBREVIATIONS

Acronym / Short Name	Meaning
ADS-B	Automatic Dependent Surveillance – Broadcast
AGL	Above Ground Level
ARP	Airport Reference Point
ATM	Air Traffic Management
ECAC	European Civil Aviation Conference
FDR	Flight Data Recorder
FP	Flight profile
GT	Ground track
ICAO	International Civil Aviation Organization
LTO	Landing and Take-Off
METAR	Meteorological Aerodrome Report
MSL	Mean Sea Level
(M)TOW	(Maximum) Take-Off Weight
PDF	Probability Density Function
TMA	Terminal Manoeuvring Area

EXECUTIVE SUMMARY

This deliverable presents the work conducted by the partners of WP2, whose aim is delivering models and methods to collect/preprocess/enrich datasets of real-world aircraft trajectories, improve the reconstruction of near-airport aircraft performance, and obtain data-driven statistical dispersions concerning aircraft operations. This document is organized into three main chapters (Sections 1, 2 and 3), each one of them detailly describing the strategies developed to address the following Tasks:

- Task T2.1 – ADS-B and complementary data collection and pre-processing (ONERA, AIT, UNIUD, FR24, OSN).
- Task T2.2 - Aircraft operation reconstruction for individual flights (UNIUD, AIT, ECTL).
- Task T2.3 - Statistical dispersions for aircraft operations in absence of real-world data (TUDELFT, AIT, UNIUD).

The modelling strategies developed by the WP2's partners have been applied to the air traffic in March 2023 at four European airports: Schiphol Airport (EHAM), Dublin Airport (EIDW), Rotterdam The Hague Airport (EHRD) and Arlanda Stockholm Airport (ESSA).

1 TASK T2.1 – ADS-B AND COMPLEMENTARY DATA COLLECTION AND PRE-PROCESSING

1.1 INTRODUCTION

The methodology described in this section builds on the approach outlined in deliverable D2.1 and is applied to the dataset resulting from that processing (from now on “T2.1 dataset”). Such strategy, which is complementary to the one in the deliverable D2.1, consists of a pre-processing algorithm applied to each individual trajectory in the T2.1 dataset to check, correct, and enrich the data. The aim is to obtain datasets made of individual flights to which the aircraft operation reconstruction algorithm, detailly described in Section 2, can be applied. In fact, such reconstruction cannot depend solely on the T2.1 datasets, as ADS-B (and Mode-S) data typically lack several key pieces of information, including the aircraft model, its engine type, its weight, and a variety of aerodynamic and engine parameters. Therefore, to facilitate efficient and accurate estimation of the near-airport aircraft performance, external sub-models and databases must be included in the reconstruction methodology.

This part of the document is organized as follows: firstly, Section 1.2 presents the key steps of the pre-processing algorithm, focusing on the identification and elimination of incorrect flight movement, and assignment of model, engine, weather conditions and landing/take-off runways to each aircraft operation. Secondly, results are presented in Section 1.3 and conclusions are drawn in Section 1.4.

1.2 METHODOLOGY

ADS-B (and Mode-S) data from task T2.1 consist primarily in positional and kinematic information. However, these data alone do not allow to directly derive aircraft performance parameters such as engine thrust and fuel consumption, which are required to both noise and emissions of individual flight operations. As a result, several pre-processing operations are performed to enrich these datasets, enabling their utilization in the flight path reconstruction algorithm. For each T2.1 flight movement, the following operations are carried out:

1) identification of the type of flight operation, including its removal if key information is missing (Section 1.2.1);
2) assignment of a suitable aircraft airframe, from which an ANP proxy [1] and engine can be assigned (Section 1.2.2);
3) assignment of actual weather conditions (Section 1.2.3);
4) assignment of take-off and landing runway to each departure and arrival, respectively (Section 1.2.4).

1.2.1 Identification of valid and incorrect operations

The inaccuracies in individual flight operations stem from three primary causes:

a) the latitude and longitude data list only ‘NaN’ (not a number) values, making it impossible to determine the aircraft position over time;
b) the flight trajectory includes either only parts of the cruise phase, which takes place far away from the airport considered, or only the high-altitude portion of the take-off/landing procedure, lacking most of the positional and kinematic data below 5,000 ft AGL;

c) the data contain at most the low-altitude portion (1,500 ft or below) of the take-off/landing phase, in some cases containing only the taxi-out or taxi-in segments.

Examples of a number of operations which present type b) and c) issues are shown in Figure 1 for Schiphol Airport.

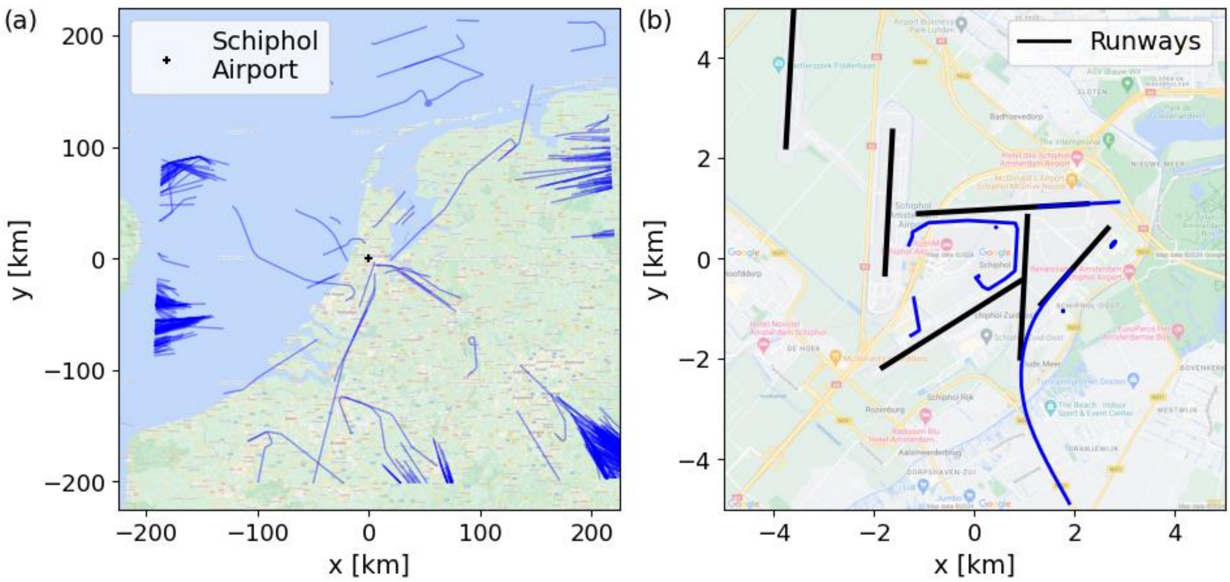

Figure 1 – At Schiphol Airport, high-altitude trajectories (a) and flight operations containing low-altitude portion of the take-off/landing phase or taxi in/out segments (b).

Operations falling into any of these categories are removed from the pool of usable flight operations.

1.2.2 ANP proxy assignment

The ANP proxies are reference airframe-engine combinations, each of which represents an actual aircraft commonly operated worldwide either currently or in the past. Manufacturers have been providing noise and performance data for such aircraft, and this information is stored in the ANP database [1]. Proxies, along with ECAC Doc.29 method [2] and ANP database, represent the backbone of the aircraft operation reconstruction outlined in Section 2. Therefore, the assignment of a suitable ANP proxy to each flight movement represent an essential operation to be carried out during the pre-processing routine and it is illustrated below.

ADS-B and Mode-S data do not provide explicit information on the type of aircraft that perform a given procedure. However, each individual operation in the T2.1 datasets has the ICAO24 identifier, which is a 24-bit unique number that is assigned to each vehicle or object that can transmit ADS-B messages. This identifier can be cross-referenced with the OpenSky aircraft database [3] to obtain the registration/tail number and four-character ICAO code of the aircraft: the structure of the above mentioned database is shown in Table 1. If no correspondence between ICAO24 and registration number is found, the flight is removed from the final dataset. To prevent these operations from being overlooked, one could manually add the missing ICAO24 identifiers into the OpenSky database. However, this handmade addition would need to be repeated each time the database is updated, hence it has not been implemented at this time.

Table 1 – Structure of the OpenSky aircraft database.

ICAO24	Registration	ICAO
aa3487	N757F	BE36
a4fa61	N42MH	PA31
a7a809	N5926K	AC90
391927	F-GGJH	DR40
…	…	…

Generally, the ICAO code is not a good enough indicator of the aircraft type, since it typically groups many different variants of the same aircraft family: for instance, the ICAO code for the Airbus 320 family (A320) may encompass various models such as the A320-211, A320-212 and A320-231, each one with significant differences in size, range, and performance characteristics. Instead, knowledge of the actual airframe-engine combination is mandatory for an effective flight operation reconstruction, and the registration number just retrieved, as well as Airlinerlist website [4], represent an excellent answer to this issue. In fact, Airlinerlist offers to the public 41 static databases, updated approximately every month, with data such as aircraft model and first flight date (accessible through the tail number, or registration) on more than 100,000 civil aircraft. These files were downloaded, opportunely processed to make the aircraft model names consistent with ANP syntax and merged together to generate a usable database that effectively acts as a lookup table. The structure of this database is shown in Table 2.

Table 2 – Structure of the Airlinerlist-based aircraft model database.

Registration	First Flight	Model	ICAO
B-HXL	21/12/2000	A340-313	A343
4R-ADG	21/12/2000	A340-313	A343
UP-A4001	21/12/2000	A340-312	A343
YK-AZA	21/12/2000	A340-312	A343
…	…	…	…

Once the aircraft model is retrieved, the last step consists in associating it with a suitable ANP proxy. In fact, the ANP database [1] provides performance and Noise Power Distance (NPD) data only for a restricted list of 150+ specific airframe-engine combinations (proxies), each of which is however very close to a variable number of actual combinations, so as to represent the vast majority of the aircraft types that make up the civil aircraft fleet of today with a reasonably limited number of database entries. As it is highly probable that a given aircraft model is not included in the ANP database, ANP offers also a way to map such a model to the most similar proxy in the form of two substitution tables. The first table (Table 3) is used to map a known aircraft model to a proxy, while the second one is used to select a proxy when the aircraft model is unknown and only the ICAO designator is provided.

Table 3 – Structure of the first ANP substitution table.

Model	ICAO	Proxy	N_{dep}	N_{app}	Engine
737-700	B737	737700	0.77	0.98	CFM56-7B22
737-8	B38M	7378MAX	0.77	0.96	LEAP-1B25
737-800	B738	737800	0.84	0.94	CFM56-7B26
737-8200	B38M	7378MAX	0.77	0.96	LEAP-1B27
…	…	…	…	…	…

However, the first ANP substitution table (Table 3) used here is a modified version of the original one. In fact, the original table requires the knowledge of both the aircraft model, which is known, and the specific engine mounted on it, which instead could not be determined from the information made available by Airlinerlist [4]. Thus, in the modified substitution table, for each aircraft model:

- the engine, whose name has been made compatible with the syntax of the ICAO AEED (aircraft engine emissions databank), corresponds to the most frequent one among the ones listed in ANP. After consultation of the ICAO AEED [5], it was found that the variation of rated thrust and pollutant emissions between engines of the same family is relatively small, suggesting that cumulative assessments of pollutant levels should not be significantly affected by this choice.
- the noise movement adjustment factors for departures N_{dep} and arrivals N_{app} [6] are the mean value of all the factors pertaining to different configurations of the same model in the original database. The decibel adjustment factors Δ_{dep} and Δ_{app} can be computed respectively from N_{dep} and N_{app} using Eqs.(1) and (2). About 95% of the aircraft models exhibit a maximum noise level deviation of 4 dB between the noisiest and the quietest configuration, and using averaged correction factors for a given model should lead to a maximum error of about 2 dB in the estimation of noise levels for a single operation. Again, this choice should have a modest impact on cumulative noise exposure assessments [7].

$$\Delta_{\mathrm{dep}} = 10 \cdot \log_{10} N_{dep} \qquad (1)$$

$$\Delta_{app} = 10 \cdot \log_{10} N_{app} \qquad (2)$$

The translation of the aircraft model found from Airlinerlist database into ANP proxy generally follows the pattern below:

- use the aircraft model in the first substitution table; if no proxy is found, use the ICAO designator in the second substitution table,
- if neither the aircraft model nor the ICAO code leads to an ANP proxy, discard the flight.

Overall, three types of failures can be identified if the algorithm is not able to find a suitable ANP proxy for the considered operation:

- type 1: the aircraft is a helicopter, which is not covered by ECAC Doc.29 and ANP;
- type 2: there is no match between the ICAO24 identifier and the registration inside the OpenSky aircraft database;
- type 3: no suitable ANP proxy could be identified, which can occur for small aircraft (e.g. general aviation types) without any model-proxy mapping in the ANP substitution tables.

Operations falling into any of these categories are removed from the pool of usable flight movements.

1.2.3 Actual weather conditions

For a more accurate flight operation reconstruction, the weather information provided by METARs and retrieved through the *traffic* Python library [8] is used. METARs are transmitted every 30 minutes by the airport weather station and, after being decoded, they can be stored as data lines in a database structured similarly to the one shown in Table 4.

Table 4 - Typical structure of weather reports decoded from METAR.

Time	Pressure [mbar]	Temperature [°C]	Wind Direction [°]	Wind Speed [kt]	Relative Humidity *RH* [%]
28/02/2023 23:25	1032	-1	350	2	85.79
28/02/2023 23:55	1032	-1	30	3	92.65
01/03/2023 00:25	1032	1	20	4	92.77
01/03/2023 00:55	1032	-1	20	5	100.00
…	…	…	…	…	…

Pressure and temperature are expressed in mbar and °C respectively, while wind direction and speed are in degrees and knots. Relative humidity (RH), in percentage, is not directly retrieved from METARs but it is evaluated through Eq. (3). The RH information is not required for the trajectory reconstruction (WP2), but relevant for noise (WP4) and emissions (WP3) calculations.

$$RH = 100 \cdot \frac{e^{17.625 \cdot T_{dew} / 234.04 + T_{dew}}}{e^{17.625 \cdot T / 234.04 + T}} \quad (3)$$

where T and T_{dew} are the temperature and dew point (present in METARs) expressed in °C.

For each flight operation, the first and last timestamps are retrieved from the ADS-B data. The METAR data before the first timestamp and after the last one are selected from the weather database, and all the meteorological information are obtained by simple linear interpolation and assigned to the correspondent flight.

1.2.4 Runway assignment

The assignment of the most appropriate runway for each flight operation, as well as the selection of a suitable take-off start point or landing end point, represents a crucial step for the subsequent aircraft operation reconstruction. For this assignment, two external databases are employed:

- the airport runway database, retrieved from the website OurAirports [9], whose main entries are shown in Table 5 for Schiphol Airport;
- the taxiway layout of the considered airport, available through OpenStreetMap.

Table 5 – Main entries of the runways database for Schiphol Airport.

Name	Latitude [deg]	Longitude [deg]	Elevation [ft]	Heading [deg]	Length [ft]
04	52.3004	4.78348	-13	41	6,627
06	52.2879	4.73402	-11	58	11,283
09	52.3166	4.74635	-12	87	11,329
18C	52.3314	4.74003	-13	183	10,826
18L	52.3213	4.77996	-12	183	11,155
18R	52.3627	4.71193	-13	183	12,467
22	52.314	4.80302	-14	221	6,627
24	52.3046	4.77752	-12	238	11,283
27	52.3184	4.79689	-13	267	11,329
36C	52.3018	4.7375	-12	3	10,826
36R	52.2908	4.77735	-11	3	11,155
36L	52.3286	4.70884	-12	3	12,467

Concerning the runway assignment algorithm, three main scenarios can occur:

a) on-ground ADS-B points are available: in this case the selection of the take-off/landing runway is straightforward. As illustrated in Figure 2, shapes are built around each airport runway, and the runway assigned is the one for which the on-ground points fall within its shape and have a heading angle compatible ($\pm 10°$) with its direction;

b) in the absence of on-ground data, the most probable runway is determined using geometrical consideration based on distances and heading angle differences between airport's runways and the first/last available ADS-B point P (departure/arrival). In particular, for each airport's runway the steps below are followed:
   i. evaluate the distance d between P and the runway midpoint;
   ii. compute the difference Δ_{ang} between P and runway heading angles;
   iii. compute the perpendicular distance Δ_{dist} between P and runway;
   iv. evaluate the dimensionless parameter $\varepsilon = {}^{\Delta_{ang}}/_{180} + {}^{\Delta_{dist}}/_{5000}$;
   v. if there are runways with $d \leq d_{th}$ (with d_{th} being a threshold value of 10 000 ft for departures and 25 000 ft for arrivals), assign the one with the lowest value of ε to the considered operation. Otherwise use method c) the assignment;

c) in the rare occasion in which neither method a) nor b) leads to a suitable runway, the runway is statistically assigned to the operation under consideration. For each combination of airport, date, and type of operation, the most frequently used runway is extracted for three separate time intervals (7:00-19:00, 19:00-23:00, and the remaining 8-hour period) and then assigned to the aircraft operation undergoing the pre-processing.

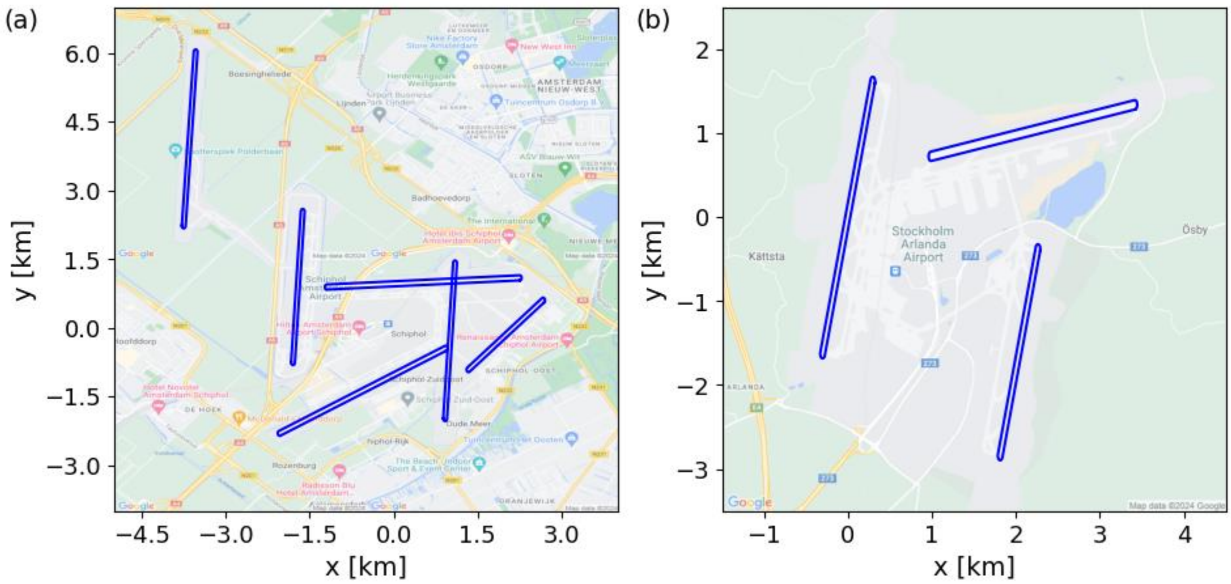

Figure 2 – Shapes surrounding the runways in Schiphol (a) and Stockholm Arlanda (b) airports.

After the runway has been selected, the assignment algorithm ensures that all departures and arrivals have a suitable start and end point, respectively. This is in general necessary since some aircraft operations lack most of the on-ground ADS-B data, having only airborne ADS-B positions which are not enough to fully identify the ground manoeuvres. In these cases, the algorithm artificially adds some points along the runway to guarantee a straight and realistic path on the ground. This addition, and especially the way the location of these points is determined, depends on the type of flight operation.

- For a departure, the taxiways of the considered airport are employed in such a way that depends on the presence of on-ground data.
    - On-ground data, specifically taxi-out points, are available. Initially, the taxi-out points are used to identify the taxiway that the aircraft is travelling on. Subsequently, the intersection between that taxiway and the runway is determined. Finally, both the distance d_1 between the beginning of the runway and this intersection, and the distance d_2 between the beginning of the runway and the first available ADS-B point on it, are calculated. If $d_1 < d_2$, the intersection is selected as the start-of-roll point. Vice versa, if $d_1 > d_2$, the start-of-roll is set at the first ADS-B data point on the runway.
    - On-ground data are not available. The first runway point is positioned at the end of the taxiway closest to the beginning of the runway assigned to the flight operation (example in Figure 3(a)).
- For an arrival, a very different approach is implemented, but it still depends on the availability of on-ground data.
    - On-ground data are available. Starting from the runway threshold, the position where the aircraft ends the deceleration procedure as prescribed by ECAC Doc.29 through the ANP database is evaluated. The distance d_1 between this point and the runway threshold is computed. The distance d_2 from the runway threshold to the last ADS-B point on the runway is also calculated. If $d_1 > d_2$, points are added to complete the ECAC Doc.29

procedure defining the end-of-roll; otherwise, the last point of the operation will be the actual last flight tracking point.

- In the absence of on-ground data, the end-of-roll point is placed at the position where the aircraft ends the deceleration procedure starting from the runway threshold, as prescribed by ECAC Doc.29 (example in Figure 3(b)).

Finally, at this preliminary stage of the methodology, emphasis has not been put on the actual reconstruction of the taxi-out (for departures) and taxi-in (for arrivals) phases. It is expected that this action will be undertaken by the next deliverable (D2.3), also in accordance with the methodological advancements and results achieved within T2.1.

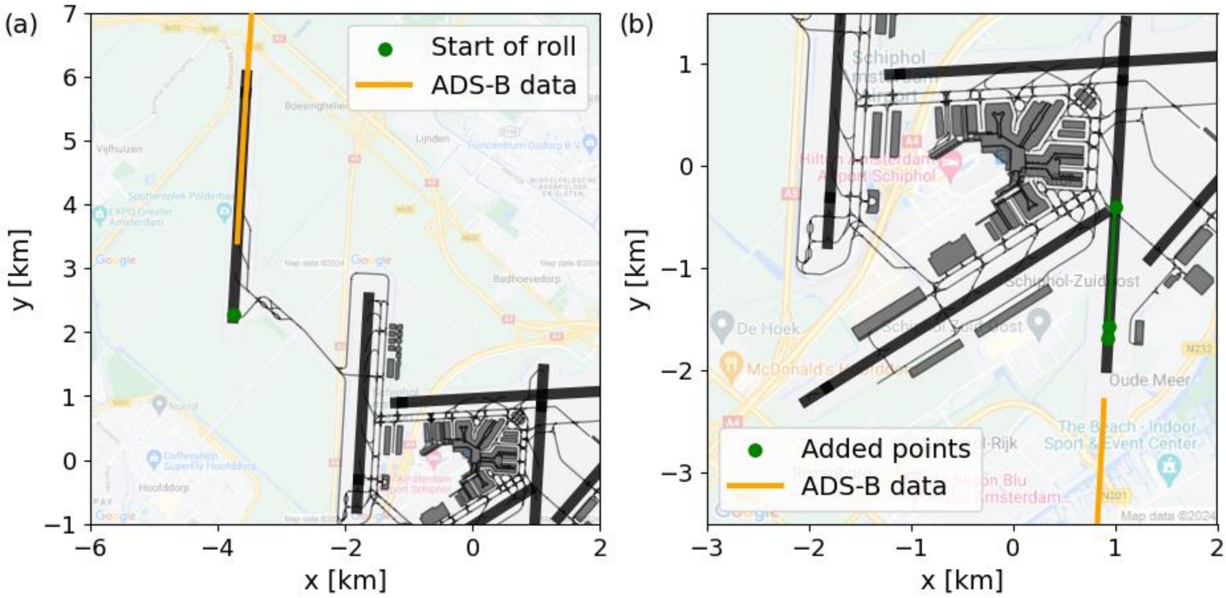

Figure 3 – Start of roll set at the lowest taxiway-runway crossing as taxi-out data are not available (a) and runway points addition according to ECAC Doc.29/ANP procedure for an arrival operation with bad ground coverage (b).

1.3 RESULTS

This section focuses on showing and examining the results obtained from the pre-processing routine. Section 1.3.1 describes the number of operations discarded due to incorrect data or the algorithm's failure to find a suitable ANP proxy. Meanwhile, Section 1.3.2 shows how information on runway usage and weather conditions can provide valuable insights into how air traffic in managed in at a particular airport.

1.3.1 Incorrect flight operations and ANP proxy identification failures

In Section 1.2.1 a number of issues were identified as the primary causes of inaccuracies in individual flight movements. The proportions of the inaccurate events relative to the total number of operations in the original T2.1 datasets are illustrated in Table 6 for the four airports under consideration. Additionally, the table provides the reasons for excluding these events (case a), b) or c) presented in Section 1.2.1), expressed as percentages of the total number of neglected operations.

Table 6 – Percentages of incorrect flights.

Airport	Total Discards [%]	Case a [%]	Case b [%]	Case c [%]
EHAM	4.25	47.72	51.31	0.97
EIDW	4.51	12.42	87.21	0.37
EHRD	0.75	43.75	31.25	25.00
ESSA	2.68	14.66	84.29	1.05

The percentage of incorrect flights is heavily dependent on the considered airport, and in general lies below the 5%. The distribution of the three types of failure is strongly related to the airport as well: type a) (absence of aircraft's position) and b) (high-altitude only flight trajectories) issues are mostly equivalent in both Schiphol and Rotterdam The Hague airports, while for Dublin and Stockholm Arlanda airports case b) is by far the predominant one. Finally, in all aerodromes apart from Rotterdam The Hague, case c) (low-altitude only flight trajectories) accounts for about 1% or less of the disregarded operations. Finally, all the aircraft operations are separated into departures and arrivals and sent to the following steps.

On the other hand, Section 1.2.2 outlined the steps followed by the ANP proxy assignment algorithm, highlighting also some situations in which such sub-routine is not able to identify a suitable proxy for the considered aircraft operation. The outcomes of this algorithm are presented in Table 7, where the percentages of discarded flight movements over the total number of operations at the considered airport as provided by the T2.1 datasets are displayed. Additionally, the table includes the reasons for the exclusion of these events, separated into the three types (1, 2 and 3, as per Section 1.2.2) and expressed as percentages of the total number of discarded operations.

Table 7 – Results of the ANP proxy assignment.

Airport	Total Discards [%]	Type 1 [%]	Type 2 [%]	Type 3 [%]
EHAM	2.36	67.54	28.73	3.73
EIDW	1.98	15.00	82.22	2.78
EHRD	14.93	56.78	4.42	38.80
ESSA	2.54	36.54	63.46	0.00

Results indicate that the percentage of discarded flights typically is typically around 2.5% of the total number of operations. However, for Rotterdam The Hague Airport this value is much higher at 14.40%. In this case, a noticeable fraction of the aircraft fleet is made of small and very small aircraft, for which the ANP substitution tables do not provide any mapping option.

The distribution of the three types of failures is strongly correlated with the airport. The type 1 failure (helicopters) is the most frequent at both Schiphol and Rotterdam The Hague airports, while it represents a sizeable but smaller fraction of the discarded operations at Stockholm Arlanda and Dublin airports. On the other hand, the type 2 failure (ICAO24 identifier not present in OpenSky aircraft database) is predominant at Dublin and Stockholm Arlanda, remains significant at Schiphol, but is not as prevalent at Rotterdam The Hague. Finally, the type 3 failure (general aviation aircraft types) is particularly significant only at Rotterdam The Hague Airport for the reason mentioned above.

1.3.2 Runway usage and weather conditions

During the pre-processing routine, each individual operation is assigned a range of information, including details about the take-off and landing runway, as well as the weather conditions. This data can be used, for example, to examine potential relationships between the runway usage and the wind direction. In this context, the daily runway usage statistics for the air traffic at Schiphol Airport in March 2023 are shown in Figure 4 alongside the daily median wind directions. It appears that the runway usage patterns tend to change based on the wind direction, and two patterns appear to be predominant. In the first one, when the wind is from the south/south-east, runways 18L and 24 for departures and runways 18R, 22, and 27 for arrivals are the preferred choices. In the second one, when the wind is from north-east/north-west, departures take place mostly on runways 09 and 36L, while arrivals are observed mainly on runways 06 and 36R.

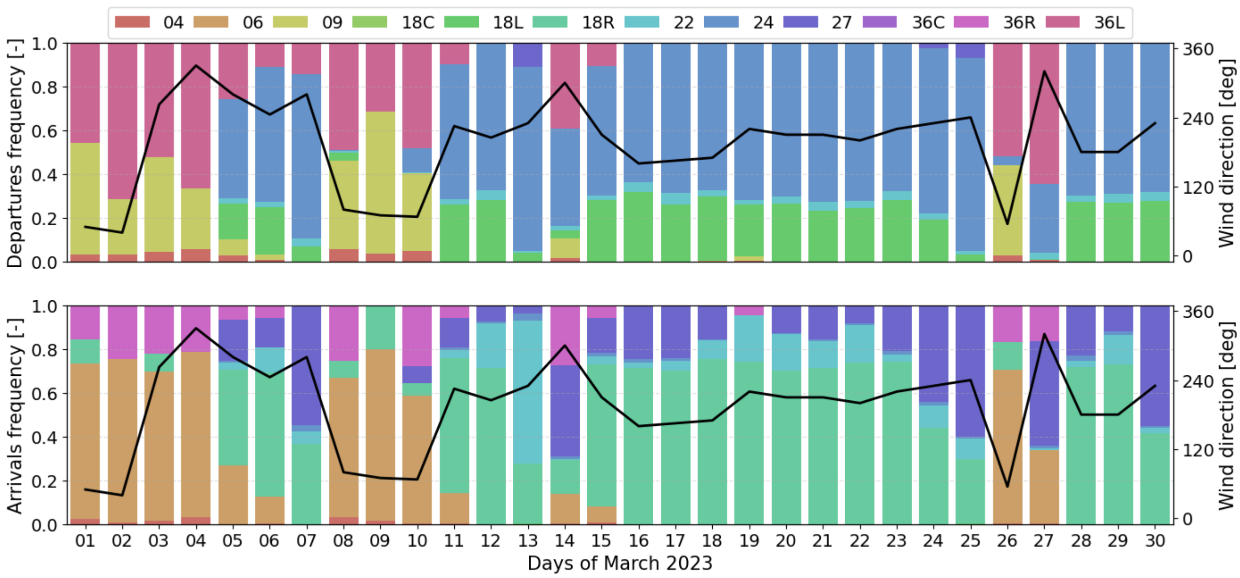

Figure 4 – Daily runways usage statistics for departure and arrival operations in Schiphol Airport in March 2023 and daily median wind direction.

1.4 CONCLUSIONS

Section 1 of the present deliverable D2.2, outlines the pre-processing methodology developed by UNIUD in task T2.1. Such routine complements the approach described in deliverable D2.1 and is aimed at checking, correcting and enriching the trajectories in the T2.1 dataset with additional information, mandatory for the subsequent aircraft operation reconstruction. For each operation, this information consists mainly of the aircraft configuration, the weather data, and its relationship with the airport layout in terms of both runway and taxiway(s) on which the operation took place.

As outlined in the relevant part of Section 1, the current preprocessing routine presents a number of limitations, but strategies are being formulated to address them. Particularly, the major drawback of the methodology is the current inability to associate each aircraft model with the exact engine type or variant mounted on it (see Section 1.2.2), which affects both noise and performance calculations, specifically the fuel flow. This issue could be mitigated by using information from the website *Rzjets* [10]: web scraping techniques can be employed to enrich the data in the *Airlinerlist* database (Section 1.2.2), enabling the mapping of specific aircraft to their corresponding engines whenever possible. Knowing the actual engine type will likely lead to better fuel flow estimation and enable the use of the original ANP substitution tables, thus retrieving more precise noise correction factors, N_{dep} and N_{app}. This is also likely to result in more accurate estimation of pollutant emissions and noise levels for single aircraft operations, thus providing a benefit also to the partners working in WP3 and WP4.

2 TASK T2.2 - AIRCRAFT OPERATION RECONSTRUCTION FOR INDIVIDUAL FLIGHTS

2.1 INTRODUCTION

The present section describes the preliminary methodology for the reconstruction of individual near-airport (TMA, or terminal manoeuvring area) aircraft operations on the basis of the datasets provided by task T2.1, properly pre-processed according to the guidelines outlined in Section 1. Such strategy builds upon elements of the aircraft performance calculation method described in the ECAC Doc.29 airport noise modelling guidance document [2], introducing complementary approaches aimed at refining the default calculation method of Doc. 29 according to the newly available sources of information. Specifically, the primary objective is to make the default departure and arrival procedures provided in the Aircraft Noise and Performance (ANP) database [1] accompanying ECAC Doc.29 more representative of the actual aircraft operations, thus allowing for a better estimation of both the aircraft trajectory profiles and associated flight performance parameters (e.g., thrust, weight, fuel flow, etc.).

Starting from the pre-processed T2.1 dataset, the present methodology consists of two major steps. Firstly, the ground tracks are computed for each flight operation using an algorithm that meets the IMPACT [6] / ECAC Doc.29 [2] requirement of trajectory smoothness. Secondly, the vertical profile component of the aircraft flight path is calculated using a mixed analysis-synthesis approach, according to which the synthetic profiles estimated on the basis of the Doc 29 methods are modified, through the use of multiple optimization variables, to follow more closely the analytical ADS-B-based aircraft trajectories. Finally, for each aircraft operation the ground track and the flight profile are merged, yielding the reconstructed near-airport flight paths of all operations.

This part of the document is organized as follows: Section 2.2 describes the reconstruction methodology, detailing the flight path reconstruction algorithm in Section 2.2.1.1 for the ground tracks and Section 2.2.1.2 for the vertical flight profile. Additionally, Section 2.2.2 describes the implementation of a variance-based sensitivity analysis designed to gain a better understanding of the relative importances of the optimization variables used in the mixed analysis-synthesis approach for reconstructing the vertical flight profile. Finally, the results of this methodology are presented in Section 2.3, while the conclusions are drawn in Section 2.4.

2.2 METHODOLOGY

An aircraft operation, alternatively named also flight operation, is represented by its flight path, which is composed by the ground track (GT), the projection of the aircraft trajectory on the ground, and the vertical flight profile (FP), that represents the aircraft motion along the GT. The present modelling methodology for the reconstruction of individual flight operations heavily relies on ECAC Doc.29 method [2] and the ANP database [1]. These resources provide all the necessary parameters for estimating the flight performance for 150+ reference airframe-engine combinations, also known as proxies, each of which represents an actual aircraft commonly operated worldwide either currently or in the past. However, the recommendations of ECAC Doc.29 essentially rely on the use of the default reference data of its accompanying ANP database and cannot account for the large variability of real-world operations over airports of any size, with a large number of aircraft with many airframe-engine variants, changing weather conditions, and flight plans tailored to both airport layout and aircraft payload. This variability is further shown by the ADS-B data, from which analytical flight paths can be extracted but performance parameters (particularly engine thrust and take-off weight) are not readily available. This issue is tackled by the present modelling approach, which combines the ECAC Doc.29 method with ADS-B (and Mode-S) tracking data and a number of support databases to identify actual flight operations and determine the aircraft trajectories. Flight identification and GT reconstruction are fully analytical, while the FP computation relies on the mixed analysis-synthesis approach, a methodology that uses both real data (analysis) and prescribed procedures (synthesis). The structure of the present modelling approach is shown in Figure 5**Error! Reference source not found.**, where three main stages can be identified:

1) pre-processing (see Section 1): starting from the daily air traffic datasets from T2.1, the identification of aircraft operations is performed using some support databases.
2) processing: this is the reconstruction of GT and FP of a given aircraft operation, from which the segmented flight path is obtained;
3) post-processing: this stage includes conducting several statistical evaluations to assess the quality and accuracy of the actions and computations carried out in 1) and 2).

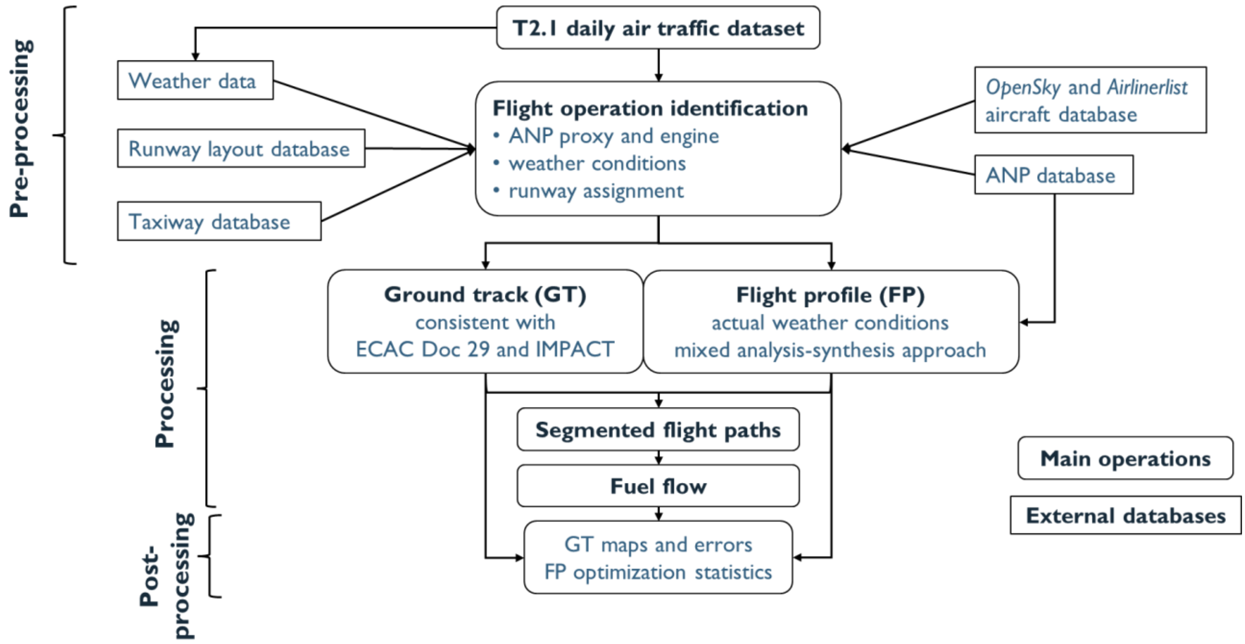

Figure 5 - Flowchart summarizing the present aircraft performance modelling methodology.

2.2.1 Processing: the flight path reconstruction

The processing of each aircraft operation involves the ground track (GT) reconstruction (Section 2.2.1.1), the vertical flight profile (FP) estimation through the mixed analysis-synthesis approach (Sections 2.2.1.2.1 and 2.2.1.2.2) and the flight path generation by merging GT and FP (Section 2.2.1.3). Finally, Section 2.2.1.4 illustrates a preliminary methodology for estimating the low-altitude fuel flow of turbofan-powered aircraft.

2.2.1.1 Ground track reconstruction

The approach described in this Section has been submitted as a short paper for the 11th OpenSky Symposium [11]. The main idea behind this algorithm is the generation of a ground track that is fully compatible with the IMPACT – ECAC Doc.29 compliant – modelling tool, meaning that two requirements are fulfilled: *i*) only segments and circular arcs are used, and *ii*) the smoothness condition or heading angle continuity, which effectively implies the tangency between consecutive arcs and segments. As mentioned at the top of this section, the GT reconstruction is a fully analytical process that relies only on the ADS-B positions. The GT reconstruction algorithm, applied separately to each aircraft operation, is illustrated here below.

1) A low-pass filter [12] is used for mitigating the data noise stemming from the highly time-resolved information available from the ADS-B reports. Latitude and longitude pairs are then converted to (x, y) Cartesian coordinates.
2) The *N* ADS-B points representing the flight operation, minimally displaced by the low-pass filtering, are connected by vectors $\boldsymbol{v}_i$, *i = (1, …, N-1)* in sequence. Value and sign of the angles α_i between consecutive vectors are computed and a threshold angle $\alpha_{th} = 0.15°$ is used to distinguish the points belonging to straight segments from those inside turns. The *i*-th point is deemed to be part of a turn if both the conditions below are met:

$$\begin{cases} |\alpha_i| \geq \alpha_{th} \\ \alpha_{i-1} \cdot \alpha_i > 0 \end{cases} \tag{4}$$

   As turns are identified through a non-zero threshold α_{th}, those with large radii (i.e. small α value) may not be recognized as actual turns, leading to inaccuracies in the ground track reconstruction.

   Finally turns longer than 70° are split into two or more sub-turns: in this way these turns can be modelled using multiple circular arcs with different radii, resulting in a reconstruction with higher accuracy.
3) Tentative GT nodes are identified as the first and last point of each turn resulting from step 2, plus the first and last ADS-B points.
4) The definitive GT nodes are identified by ensuring the trajectory smoothness (i.e. heading angle continuity), which is done by reconstructing each turn imposing three conditions:
   - tangency between circular turn and start-of-turn heading θ_S,
   - tangency between circular turn and end-of-turn heading θ_E,
   - circular turn passing through either the start or end node.

   The algorithm distinguishes between two cases:
   a) in the first and most common one (Figure 6(a)), the turn lies between two segments, and the turn-end node *E* is selected as the arc belonging node. Point *K* is then found according to equation (5) as the intersection between the extensions of the segments with headings θ_S and θ_E. Consequently, distance d_{EK} between *K* and *E* is computed and it is used to identify the turn-new-start node *S'* along the segment with heading θ_S.

$$\begin{cases} x_K = \dfrac{(y_E - y_S) \cdot \tan\theta_E \cdot \tan\theta_S + x_S \cdot \tan\theta_E - x_E \cdot \tan\theta_S}{\tan\theta_E - \tan\theta_S} \\ y_K = y_S + \dfrac{x_K - x_S}{\tan\theta_S} \end{cases} \tag{5}$$

   b) Instead, the second case (Figure 6(b)) occurs when turn subdivision mentioned in step 2). is necessary. In this case *S* is fixed, while *E* is the turn-node being moved. The (*x, y*) coordinates of the point *K* are calculated according to equation (6) and the distance d_{SK} between *K* and *S* is computed. *E* is then moved, with the same criteria described above, along the segment with heading θ_E in order to find the turn-new-end node *E'*.

$$\begin{cases} x_K = \dfrac{(y_S - y_E)\cdot \tan\theta_E \cdot \tan\theta_S + x_E \cdot \tan\theta_S - x_S \cdot \tan\theta_E}{\tan\theta_S - \tan\theta_E} \\ y_K = y_S + \dfrac{x_K - x_S}{\tan\theta_S} \end{cases} \quad (6)$$

This procedure is applied to all sub-turns after the first one, handled instead as in case a).

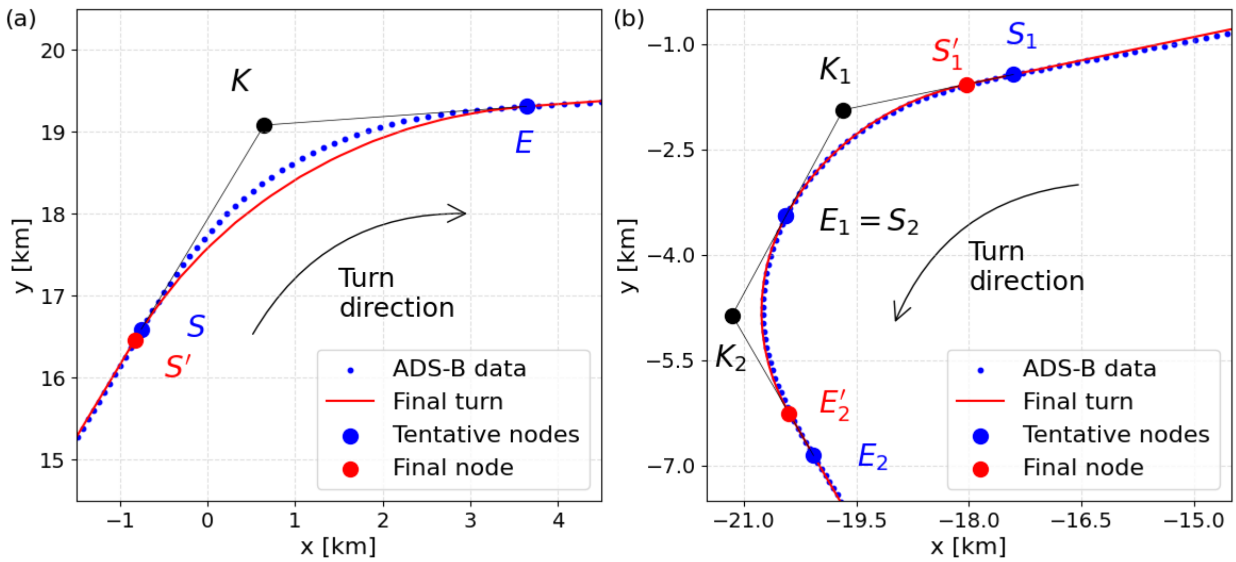

Figure 6 - GT algorithm applied to a single turn (a) and to an operation with two sub-turns (b).

5) After the identification of all the new ground tracks nodes as in step 4), for each turn, its radius and centre are calculated following the steps below:
   a) calculate the mid-point *M* between *S'* and *E* (or *S* and *E'*),
   b) calculate distances d_{KM}, $d_{S'M}$ (or $d_{E'M}$) and $d_{S'E}$ (or $d_{SE'}$),
   c) using the criteria for similarity of triangles, calculate radius *r* and the arc centre coordinates (x_C, y_C) according to Equation (7).

$$\begin{cases} r = {d_{S'M} \cdot d}/{d_{KM}} \\ x_C = x_M \pm \sqrt{r^2 - \left(\dfrac{d_{S'E}}{2}\right)^2} \cdot \dfrac{y_S - y_E}{d_{S'E}} \\ y_C = y_M \pm \sqrt{r^2 - \left(\dfrac{d_{S'E}}{2}\right)^2} \cdot \dfrac{x_E - x_S}{d_{S'E}} \end{cases} \quad (7)$$

The circular arc determined with this computation is split into a number of segments according to the requirements of the ECAC Doc.29 [2]. Each segment is the chord of a sub-arc that subtends a central angle $\Delta\theta$ defined by the user ($\Delta\theta \cong 5°$ in the present case).

As will be discussed in the sections dedicated to the results, the GT algorithm illustrated here is capable of satisfying ground tracks for the vast majority of aircraft operations, especially in the TMA. However, although rare, there are instances in which either the ground track cannot be built or the difference between the reconstructed GT and the actual positional data is too large. Typically, the root causes are an incorrect runway assignment and/or errors in the latitude/longitude values (which lead the low-pass filter to yield absurd results. In these cases, the considered operation is simply disregarded. The number of flight operations for which this occurrence was observed is reported in Table 8.

Table 8 – Number of operations discarded due to a failure in the ground track reconstruction.

Airport	Discarded operations
EHAM	6
EIDW	1
EHRD	0
ESSA	1

As a side note, it should be noted that IMPACT [6] can handle input GTs directly defined as a series of ordered, geo-referenced points. Therefore, a methodology could also be developed to enable the direct use (or nearly so) of ADS-B data, and this solution will likely be explored during the remaining period of Task T2.2. However, the current ground track modelling approach presents some useful advantages as illustrated below.

- Firstly, GTs resulting from this algorithm can be easily converted into vector tracks, i.e. ordered segments (straights and turns) to/from the airport.
- Secondly, since each reconstructed turn is modelled as an ideal circular arc, its segmentation, as described in ECAC Doc.29 [2], is straightforward. On the other hand, actual turns rarely represent perfect arcs, which complicates their segmentation into sub-turns and makes it more difficult to strictly adhere to the guidelines of Doc.29.
- Finally, the modelled GTs are perfectly smooth and do not exhibit the noise which commonly affects ADS-B points.

2.2.1.2 Aircraft performance estimation: the mixed analysis-synthesis approach

With the GT reconstructed for a given operation, the aircraft performance estimation is conducted by determining the vertical flight profile (FP) which consists in the evolution of aircraft altitude, speed, and engine thrust along the curvilinear coordinate s of the GT. This is conducted with the mixed analysis-synthesis approach formulated by UNIUD, which is available in its original form [13], but has undergone some improvements, currently under peer review, over the last year [14]. This approach consists in introducing, within the ANP procedures, appropriate degrees of freedom that are treated as optimization variables. Their values are then set by minimizing an objective function based on the difference between ADS-B data and ANP profile. Concerning the FP computation, the actual weather conditions retrieved as per Section 1.2.3 are used, and the vertical gradients of air temperature $T(Z)$ and pressure $p(Z)$ are built as alterations of ISA profiles according to Eqs. (8) and (9) respectively. The use of non-ISA atmospheres has been introduced to take into account the difference in temperature and pressure at MSL between a given non standard atmosphere and ISA.

$$T(Z) = T_{ARP} + L_b \cdot (Z - Z_{rnw}) \tag{8}$$

$$p(Z) = p_{ref} \cdot \left[\left(\frac{p_{MSL}}{p_{ref}} \right)^{1/5.256} + Z \cdot \frac{L_b}{T_{ref}} \right]^{5.256} \tag{9}$$

Being $p_{ref} = 101325\ Pa$, $T_{ref} = 288.15\ K$, and $L_b = 0.0065\ {}^{K}/_{m}$. T_{MSL} and p_{MSL} are the local temperature and MSL pressure expressed in K and Pa. Knowledge of these profiles allows for the calculation at any altitude of the atmospheric ratios $\theta(Z)$, $\delta(Z)$ and $\sigma(Z)$ (Eq. (10)), which are fundamental parameters for the aircraft performance evaluation.

$$\theta(z) = \frac{T(z)}{T_{ref}}, \qquad \delta = \frac{p(z)}{p_{ref}}, \qquad \sigma(z) = \frac{\delta(z)}{\theta(z)} \tag{10}$$

Consistently with the massive differences between departure and arrival operations, the formulation of the mixed approach is equally split into two and reported in the next two subsections. The reader is referred to the ECAC Doc.29 Vol 2 manual [2] for the details concerning the purely synthetic computation of the flight profiles, which are not reported in this document for the sake of brevity.

Finally, it should be noted that, since the GT and vertical FP are computed separately, the effect of the bank angle is neglected in the FP processing method. This modelling solution is in agreement with ECAC Doc.29 which in fact states that it is possible to disregard the effect of turns on vertical profiles to reduce the computational complexity.

2.2.1.2.1 Mixed analysis-synthesis approach for departures

The first requirement for applying the mixed approach to departures is the calculation of the ANP synthetic flight profile, which is done using simple flight mechanics equations in conjunction with the sequence of actions (i.e. take off, climb, or accelerate) that the pilot follows when departing. However, even for a fixed proxy aircraft and weight, there are up to three possible default sequences (or procedures) available to the pilot, denoted as DEFAULT, ICAO_A, and ICAO_B, which result in different noise exposure and are usually associated with the noise constraints at a specific airport [2]. These procedures differ mainly between 1,000 and 5,500 ft AGL, that is after the take-off roll and initial climb, but before the climb towards cruise following the switch of flap and engine settings, as shown by the example in Table 9.

Table 9 – Default (black) and ICAO_A (blue) procedural steps for an A320-211 departing at minimum ANP weight.

Step nr	Step	Flap settings	Engine settings	Height AGL h [ft]	Calibrated airspeed V_C [kt]	Climb rate [ft/min]	Energy share [%]
1	Take off	1+F	MaxTakeoff	-	-	-	-
	Take off	1+F	MaxTakeoff	-	-	-	-
2	Climb	1+F	MaxTakeoff	1,000	-	-	-
	Climb	1+F	MaxTakeoff	1,500	-	-	-
3	Accelerate	1+F	MaxTakeoff	-	186.2	1150.5	69.1
	Climb	1+F	MaxClimb	3,000	-	-	-
4	Accelerate	1	MaxTakeoff	-	208.1	1300.7	69.8
	Accelerate	1+F	MaxClimb	-	186.1	812.1	69.6
5	Climb	ZERO	MaxClimb	3,000	-	-	-
	Accelerate	1	MaxClimb	-	201.2	933.5	70.6
6	Accelerate	ZERO	MaxClimb	-	250.0	1230.7	69.0
	Accelerate	ZERO	MaxClimb	-	228.2	1119.7	69.9
7	Climb	ZERO	MaxClimb	5,500	-	-	-
	Accelerate	ZERO	MaxClimb	-	250.0	1240.5	69.6
8	Climb	ZERO	MaxClimb	7,500	-	-	-
	Climb	ZERO	MaxClimb	5,500	-	-	-
9	Climb	ZERO	MaxClimb	10,000	-	-	-
	Climb	ZERO	MaxClimb	7,500	-	-	-
10	Climb	ZERO	MaxClimb	10,000	-	-	-

Moreover, the ECAC Doc.29 calculation method and the ANP database assume normal (maximum) rated thrust (MaxTakeoff and MaxClimb) and a default take-off weight, TOW, tabulated as a function of flight distance bands ("stage length"), reaching the maximum value (MTOW) at the largest distance. These constraints may render the synthetic profile very different from the one reported by ADS-B, but such a mismatch can be heavily reduced by loosening these constraints and allowing the synthetic profile to follow more closely the ADS-B data.

For this reason, the mixed approach requires identifying a number of optimization variables whose variation is able to provide a good degree of operational flexibility. In this work seven variables have been selected, the first four of which are the same as in the original optimization [13]. These consist in the procedure type (DEFAULT, ICAO_A, ICAO_B), the TOW-to-MTOW ratio denoted as K_{MTOW}, and the take-off and climb thrust reduction coefficients, K_T and K_C respectively. The latter express the fraction of maximum corrected net thrust per engine available to the aircraft $(F_n/\delta)_{max}$, which is computed using the ECAC Doc.29 thrust model. This computation is conducted with Eq. (11), being K_{red} the generalized reduction coefficient and E to H the ANP thrust coefficients for either take-off or climb engine settings.

$$\left(\frac{F_n}{\delta}\right)_{red} = K_{red} \cdot \left(\frac{F_n}{\delta}\right)_{max} = K_{red} \cdot \left(E + F \cdot V_C + G_A \cdot Z + G_B \cdot Z^2 + H \cdot (T - 273.15)\right) \quad (11)$$

However, in the current method [14] three new variables have been added to enhance the flexibility within a given procedure: Δh_L, Δh_M and f_e. In particular, Δh_L and Δh_M are summed to the initial steps (before the first acceleration) and mid-climb steps (below 5,500 ft) respectively, while f_e is multiplied to all energy shares. Table 10 reports their ranges, also including the additional constraints applied to prevent calculation of unrealistic profiles for some proxy-profile pairs. Furthermore, K_C is decoupled from K_T, and, while $K_{MTOW,min}$ is computed following the original calculation [13] assuming 20% payload and minimum fuel to complete the flight, its upper bound is always 1.

Table 10 – Optimization variables of the mixed analysis-synthesis approach for departures.

Variable	Symbol	Type	Values/ranges	Purpose
Procedure type	P_{type}	Discrete	Default / ICAO_A / ICAO_B	Procedural flexibility
Take-off thrust reduction	K_T	Continuous	[0.75, 1]	Variable take-off thrust
Climb thrust reduction	K_C	Discrete	1 / 0.9 / 0.8	Variable climb thrust
Weight fraction	K_{MTOW}	Continuous	[$K_{MTOW,min}$, 1]	Variable weight
Initial climb height	Δh_L	Continuous	[-2000 ft, 500 ft], as long as $h_{initial\ climb}$ > 800 ft	Flexible initial climb step(s)
Mid-climb height	Δh_M	Continuous	[0 ft, 3000 ft], as long as $h_{mid-climb}$ < 5,500 ft	Flexible mid-climb step(s)
Energy share factor	f_e	Continuous	[0.7, 1.4]	Flexible acceleration step(s)

Having set the variables, the mixed approach proceeds by minimizing an objective function, OBF, which is composed of a measure of the distance between ADS-B data and synthetic profile, RMS_{ZV}, and a correction factor, CF, that acts as a penalty function, as illustrated in Eq. (12):

$$OBF = RMS_{ZV} + CF \quad (12)$$

Firstly, RMS_{ZV} is computed as shown by Eq. (13),

$$\begin{cases} RMS_{ZV} = RMS_Z + 25 \cdot RMS_V \\ RMS_Z = 20\sqrt{\frac{1}{N_L}\sum_{i=1}^{N_L}\left(Z_{FP,i} - Z_{ADSB,i}\right)^2} + 10\sqrt{\frac{1}{N_M}\sum_{i=1}^{N_M}\left(Z_{FP,i} - Z_{ADSB,i}\right)^2} + \sqrt{\frac{1}{N_H}\sum_{i=1}^{N_H}\left(Z_{FP,i} - Z_{ADSB,i}\right)^2} \\ RMS_V = 20\sqrt{\frac{1}{N_L}\sum_{i=1}^{N_L}\left(V_{FP,i} - V_{ADSB,i}\right)^2} + 10\sqrt{\frac{1}{N_M}\sum_{i=1}^{N_M}\left(V_{FP,i} - V_{ADSB,i}\right)^2} + \sqrt{\frac{1}{N_H}\sum_{i=1}^{N_H}\left(V_{FP,i} - V_{ADSB,i}\right)^2} \end{cases} \quad (13)$$

where Z and V denote altitude [ft] and speed [kt] respectively, while FP indicates the profile synthesized according to the optimization variables and $ADSB$ the recorded *i*-th data point at same GT coordinate s_i. Firstly, coefficient 25 ft/kt is used in RMS_{ZV} to weigh the speed component, RMS_V, against the altitude one, RMS_Z. This choice was made after careful analysis of ADS-B profiles and their impact on aircraft noise, and implies that a 1-knot error in the aircraft speed has the same impact as a 25-foot error in its altitude. This value is sensible considering the heights and speeds of typical departure manoeuvres (see Table 9). Secondly, each RMS is split into three components, which are the low-, mid-, and high-height ones, with each component having respectively subscripts $L,\ M,\ H$. The low-height component includes all N_L points below 1,500 ft AGL, the mid-height one all N_M points between 1,500 ft and 5,000 ft AGL, and the high-height one all N_H points above 5,000 ft. The weight coefficient of 20 was applied to the low-height component to give the most relevance to the take-off roll and initial climb, while the mid-height coefficient (10) is large enough to allow the mixed approach to detect the departure procedure used by the pilot, and the last one (1) is kept low since that portion of the climb affects noise levels and procedure detection the least. It should be noted that if the calibrated airspeed is available from Mode-S data, the algorithm uses it instead of the ground speed.

Secondly, concerning correction factor CF in Eq. (12), its role is to address some key issues of an insufficiently constrained optimization procedure. Primarily, these are the lack of tracking-based information on the aircraft TOW, and the reliance on a flight mechanics computation whose output mostly depends on the ratio between corrected net thrust per engine, F_n/δ, and aircraft TOW rather than on their separate values [2]. Leaving these elements unaddressed may lead to peculiar results, among which two are of high relevance:

1) very low TOW coupled with very low thrust;
2) very high TOW coupled with very low thrust.

However, analysis of FDR data from many aircraft and several aircraft models [15] show that both of these cases are very unlikely. For case 1), the reason is that aircraft tend to fly with as much payload (passengers or cargo) as possible, thus minimizing the chance of a low TOW. Additionally, heavy aircraft depart with high thrust to keep the take-off roll distance contained within the limits of runway length and associated safety margins, which makes case 2) a rare occurrence.

This issue is addressed in two steps. Firstly, the ECAC Doc.29 - based Eq. (14) is used to estimate the TOW, exploiting its relationship with take-off calibrated airspeed V_{CTO} through ANP-based lift coefficient C. This can be done because V_{CTO} can be extracted from the ADS-B lift-off speed, now identifiable from the high-resolution OSN data. This yields an estimate for K_{MTOW}, denoted as $K_{MTOW,est}$.

$$V_{CTO} = C\sqrt{TOW} \tag{14}$$

The second step is using this value in the correction factor. This is done defining CF as per Eq. (15):

$$CF = RMS_{ZV}\left[max(0, K_{MTOW} - K_T) + \left(exp\left(\left|K_{MTOW,est} - K_{MTOW}\right|\right) - 1\right)\right] \tag{15}$$

which acts as a two-pronged penalty function that multiplies the calculated RMS_{ZV}. The first part of the penalty is linear and is applied when $K_{MTOW} > K_T$, with the main goal of preventing case 2) above. Instead, the second part acts when K_{MTOW} differs from $K_{MTOW,est}$, but it is applied as a symmetrical exponential penalty that becomes particularly relevant only when $\left|K_{MTOW} - K_{MTOW,est}\right| > 0.05$. The reasons are a certain and unavoidable unreliability of the ADS-B data, and the fact that the actual take-off speed can be up to a few knots higher than the V_{CTO} from Eq. (14) [2]. These elements contribute to making $K_{MTOW,est}$ just a reference rather than a certainty, especially if the remainder of the profile gives better results with $K_{MTOW} \neq K_{MTOW,est}$.

Finally, the minimization of the objective function is conducted by means of a basin-hopping algorithm [16], which enables more effective identification of the global minimum in problems with multiple local minima such as this one, dominated by the thrust-to-weight ratio.

The mixed approach as illustrated here is applied to all pre-processed departure operations, and it leads to very good results in the vast majority of cases, that is when the ADS-B-derived profiles are at least somewhat similar to the ANP procedures. However, if the departure procedure is not covered by ANP, which is more likely to happen with small aircraft, the mixed approach provides only minor improvements. Examples of these opposite occurrences are provided in Figure 7.

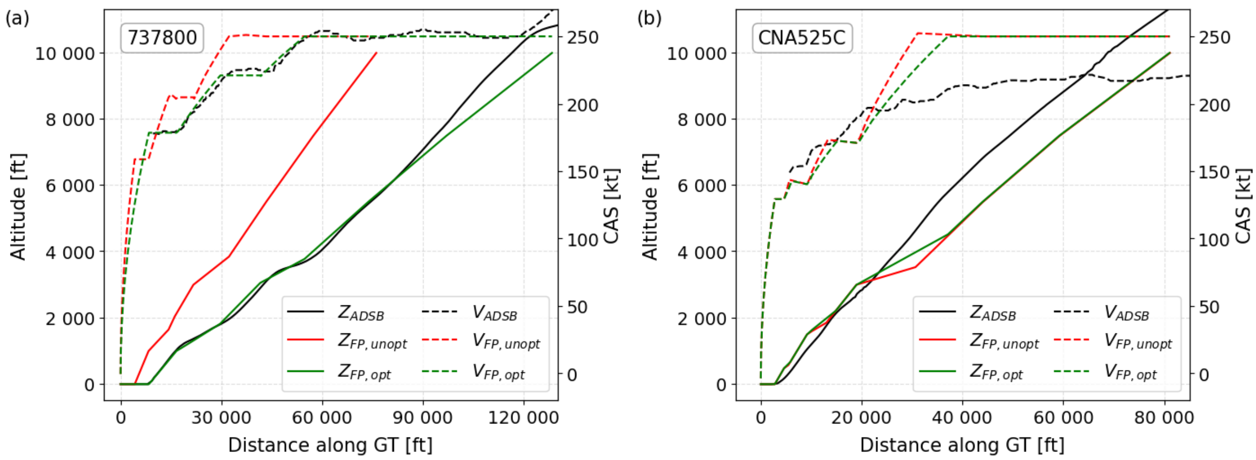

Figure 7 - At Schiphol Airport, (a) optimized departure FP when ADS-B data are similar to a procedure (here DEFAULT) covered by ANP, and (b) optimized departure FP in the case of a procedure outside the ANP coverage.

2.2.1.2.2 Mixed analysis-synthesis approach for arrivals

The mixed approach for arrivals is much simpler than that for departures, primarily because of the much higher degree of prescription in the ECAC Doc.29 calculation method. In particular, for each ANP aircraft type the landing weight is fixed, only a default profile is available, all descent angles are imposed rather than computed from energy shares or vertical rates, and the thrust outputs are mainly inferred from aerodynamic and kinematic quantities. Despite these constraints, during the final descent (generally below 1,500 ft AGL) the ANP profiles match quite well the ADS-B data, while more variability appears to be required above it. Additionally, observation of the ADS-B data showed that level-flight phases can be present for all aircraft models, but some ANP proxies (essentially old aircraft types) do not present such phases in their procedural steps. Hence, a calculation was performed to introduce an aerodynamic level phase for each one of the ANP proxies not having it in its profile, and this was done following the example provided in ECAC Doc.29 Vol. 2 for the Boeing 737-300 [2]. For this calculation, the level-flight phase is added maintaining consistency with the original procedural steps, that is the relationship between the flap setting used and the end-point calibrated airspeed present in the original ANP procedure is maintained in the modified one. An example showing an ANP approach profile without a level-flight phase and the same profile after the addition of such a phase is provided in Table 11.

Table 11 – Default (black) and modified (blue) approach procedural steps for a Boeing 737300.

Step nr	Step	Flap settings	Start Height AGL h [ft]	Start CAS V_C [kt]	Descent angle γ [deg]	Distance [ft]	Start Thrust [%]
1	Descend	ZERO	6,000	250	3		
	Descend	ZERO	6,000	250	3		
2	Descend	5	3,000	170	3		
	Level	5	3,000	250		21,000	
3	Descend	D-15	1500	148.6	3		
	Level	5	3,000	170		5,000	
4	Descend	D-30	1,000	139	3		
	Descend	D-15	3,000	148.6	3		
5	Land	D-30				316.8	
	Descend	D-30	2,500	139	3		
6	Decelerate	-		131.9		2,851.2	40
	Land	D-30				316.8	
7	Decelerate	-		30		0	10
	Decelerate	-		131.9		2,851.2	40
8	Decelerate	-		30		0	10

Similarly to departures, four optimization variables are introduced aiming to decrease the mismatch between ADS-B data and synthetic profile. These variables and their ranges are listed in Table 12.

Table 12 - Optimization variables of the mixed analysis-synthesis approach for arrivals

Variable	Symbol	Type	Values/ranges
Initial descent angle	γ_{in}	Continuous	[1.5 deg, 5 deg]
Initial descent calibrated airspeed	$V_{C,in}$	Continuous	[200 kt, 250 kt]
Length of level-flight phase	Δs_{lev}	Continuous (if present)	[40%, 250%] of ANP length, or removed
Height of level-flight phase	h_{lev}	Continuous (if present)	[1,500 ft, 4,500 ft] AGL, if level is present

Then, the optimization is conducted by minimizing the OBF of Eq. (12) with RMS_{ZV} computed for all N points via Eq. (16), that is without height components, not very useful due to the many constraints:

$$RMS_{ZV} = RMS_Z + 25 \cdot RMS_V = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Z_{FP,i} - Z_{ADSB,i}\right)^2} + 25\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(V_{FP,i} - V_{ADSB,i}\right)^2} \quad (16)$$

Moreover, $CF = 0$ since no penalty is justified when the landing weight is fixed and, more importantly, impossible to infer from the ADS-B data. As for departures, when the calibrated air speed is available from Mode-S data it is used instead of the ground speed.

The mixed approach as illustrated is applied to all pre-processed arrival operations, and similarly to its departure counterpart it yields satisfying results in the majority of cases, particularly when the ADS-B profiles are reasonably close to the pre-existing ANP procedures. However, problems arise if the actual approach procedure is not covered by ANP: this time it can happen also to larger aircraft, but the detrimental effects seem contained since the aircraft has to land on the runway assigned during the pre-processing. In particular, since the final approach is always conducted at about 3 degrees and 100 to 150 kt, depending mostly on the aircraft size and landing weight [2], the error between ADS-B data and reconstructed descent cannot be too high. Relevant examples referring to these opposite cases are provided in Figure 8.

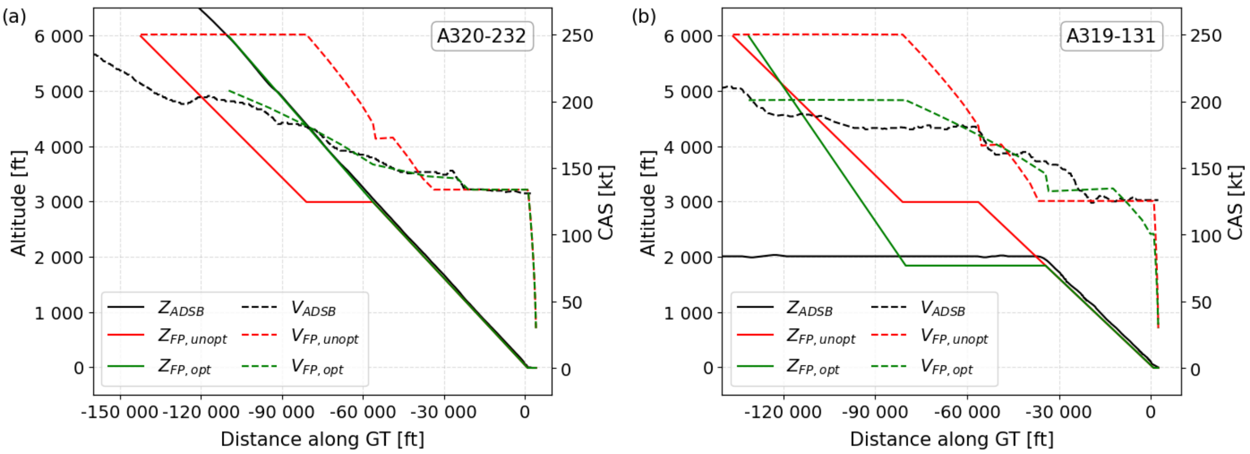

Figure 8 - At Schiphol Airport, (a) optimized arrival FP when ADS-B data are similar to a procedure covered by ANP, and (b) optimized arrival FP in the case of a procedure outside the ANP coverage.

2.2.1.3 Flight path generation and segmentation

The segmented flight path is the full representation of the whole 4D (3D over time) aircraft motion, and it is obtained by merging the ground track with the vertical flight profile. The merging is conducted exploiting the shared curvilinear coordinate *s* and following the guidelines provided by ECAC Doc.29 [2]. The result is a sequence of straight segments in space, at the end-point of which the following variables are known:

- aircraft 4D coordinates (coordinates on the ground plane, altitude, time elapsed);
- calibrated airspeed, true airspeed and ground speed;
- corrected net thrust of one engine (and hence of all engines, assuming equal thrust output for each of the aircraft engines);
- bank angle, computed after the actual ground track – vertical flight profile merging according to Eq. (17), where r is the turn radius [ft], V is the ground speed in kt and g is the gravitational acceleration in ft/s^2.

$$\varepsilon = \tan^{-1}\left(\frac{2.85 \cdot V^2}{r \cdot g}\right) \tag{17}$$

The number of flight path segments depends on ANP proxy, flight operation, and number of turns of the GT. The extension of the flight path is determined by the completion of the whole sequence of procedural steps that describe the TMA flight operation. Instead, the GT segments that lie beyond the last point of the flight profile (i.e., beyond 10,000 ft AGL for departures and 6,000 ft AGL for arrivals) currently do not contribute to the flight path, meaning that for these data the flight performance is not evaluated.

Although rare, there are instances where the length of the GT is shorter than that of the FP and, in such scenarios, it is not possible to project the FP onto the GT. To avoid removing these operations from the computation, the flight profile is shortened so that its extension along the curvilinear coordinate is lower or equal to that of the ground track, making the merging operation possible.

2.2.1.4 Fuel flow estimation

After determining the segmented flight path of a given aircraft operation, it is possible to use some of the resulting flight performance parameters (e.g., aircraft speed and altitude, net engine thrust) as the inputs of a fuel flow estimation methodology. However, in the time available before the present deliverable, the work focused primarily on the kinematic and dynamic aspects of the aircraft operation reconstruction, and the assessment of fuel flow estimation models started very recently (May 2024). As a result, in this document the methodology proposed is only preliminary, its outcomes in terms of fuel consumption are neither presented nor discussed, and relevant modifications and improvements are to be expected by the next T2.2 deliverable.

The preliminary fuel flow estimation methodology is based on the BFFM2 method [17] and the public data obtained from the ICAO AEED [5], which includes, for most turbofan engines, fuel flow values for each phase of the landing-and-take-off (LTO) cycle: take-off, climb-out, approach, and idle. These modes cover the usual pattern of near-airport aircraft operations, and each of them refers to a specific engine power setting. Following the mode order listed above, these settings are 100%, 85%, 30%, and 7% of the engine sea-level thrust rating, the latter also known as rated thrust.

To define a model capable of estimating fuel flow at any thrust level desired, the methodology detailed in ICAO Doc. 9889 [18] is employed. Firstly, the current engine power setting, $F_{\%}$ is calculated according to Eq. (18):

$$F_{\%} = \frac{F_n}{F_0} \tag{18}$$

where F_n is the net thrust obtained from the flight profile computation and F_0 is the rated thrust of the considered engine. The value of $F_{\%}$ is then used to derive a reference fuel flow value $\dot{m}_{ref}$ from the four data points in the ICAO databank through the twin quadratic interpolation underlined in ICAO Doc. 9889 [18]. The methodology is as follows:

a) $F_{\%} \leq 85\%$: the reference fuel flow at the desired thrust level $F_{\%}$ is computed using the quadratic equation (19) based on the 7%, 30% and 85% thrust and associated fuel flow points in the AEED databank.

b) $F_{\%} > 85\%$: the reference fuel flow at the desired thrust level $F_{\%}$ is computed using the quadratic equation (19) based on the 30%, 85% and 100% thrust and associated fuel flow points in the AEED databank.

$$\dot{m}_{ref} = Y \cdot \dot{m}_{ref,TO} = \left(A \cdot F_{\%}^2 + B \cdot F_{\%} + C\right) \cdot \dot{m}_{ref,TO} \tag{19}$$

In Eq. (19), $\dot{\mathrm{m}}_{\mathrm{ref,TO}}$ is the reference fuel flow at maximum rated thrust (i.e. take-off power setting), while coefficients A, B and C derive from the quadratic fitting of the three selected fuel flow-thrust level pairs in AEED database. Specifically, letting Y_i be the ratio between the *i*-th fuel flow value in

ICAO databank and $\dot{m}_{ref,TO}$, and $F_{\%,i}$ be the corresponding thrust level, coefficients A, B and C are determined by solving the system of equations (20).

$$\begin{cases} Y_1 = A \cdot F_{\%1}^2 + B \cdot F_{\%1} + C \\ Y_2 = A \cdot F_{\%2}^2 + B \cdot F_{\%2} + C \\ Y_3 = A \cdot F_{\%3}^2 + B \cdot F_{\%3} + C \end{cases} \tag{20}$$

An example of this operation for engine LEAP-1B28 is provided in Figure 9, where the input value of $F_{\%}$ is set at 0.75.

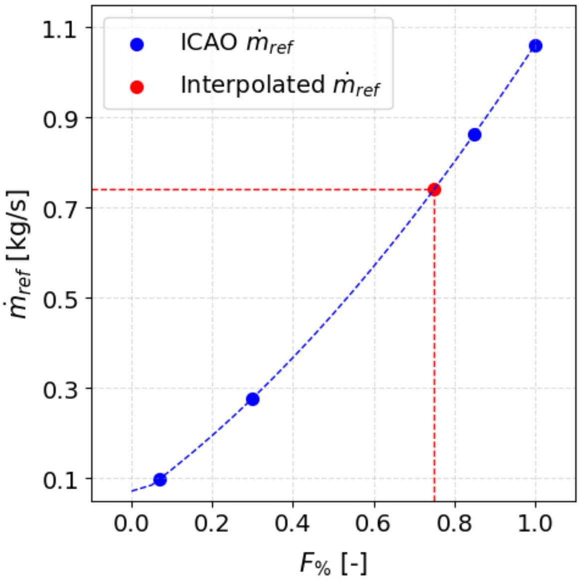

Figure 9 – Fuel flow interpolation from given $F_{\%}$ setting (here 0.75) for a LEAP-1B28 engine.

It should be noted that this approach allows for accurate estimation of fuel flow at reduced take-off thrust levels between 60% and 100% of the maximum rated thrust [18]. However, the same procedure is applied to find the flow-thrust relationship even for thrust ratios below 60%, with the understanding that the results may be less accurate. With this consideration firmly in mind, BFFM2 is employed to compute the non-reference fuel flow $\dot{m}$ for a single aircraft engine, as follows:

$$\dot{m} = \dot{m}_{ref} \cdot \frac{B_m \cdot \delta}{\theta^{3.8} \cdot e^{0.2 \cdot M^2}} \tag{21}$$

where δ is the air pressure ratio, θ is the air temperature ratio and M is the Mach number. Moreover, compared to static engine tests, actual aircraft operations entail the activation of the Environmental Control System (ECS), which requires some power extraction and some pressurized air to be bled from the engine compressor. This leads to a slightly higher fuel flow, which is modelled through factor B_m. Its value depends on the power setting and thus on the LTO mode, and the four values referring to the different engine modes are reported in Table 13.

Table 13 - AEED LTO modes and corresponding ECS adjustments.

LTO Mode	Power Setting	$B_m[-]$
Take-off	100%	1.010
Climb-out	85%	1.013
Approach	30%	1.020
Idle	7%	1.100

Finally, although this fuel flow model can be theoretically applied to all types of aircraft engines, the AEED does not provide information on non-turbofan engine models. Thus, the fuel flow was not estimated for flight operations performed by turboprop and piston-powered aircraft. However, the contribution in terms of fuel burned and emissions of these types of aircraft did not exceed 2% of the overall air traffic impact before 2010 [19]. Therefore, excluding them is unlikely to affect the results considerably. At the same time, efforts will be made by the next T2.2 report to evaluate any possibility concerning the fuel flow estimation for these aircraft, although a prominent challenge appears to be the lack of publicly available resources (e.g. an AEED-like database for turboprops), without which any effective modelling is unlikely to be implemented.

2.2.2 Flight path sensitivity analysis

In this Section, a sensitivity analysis has been designed to gain a better understanding of the relative importances of the optimization variables used in the mixed analysis-synthesis approach for reconstructing the vertical flight profile.

Variance-based sensitivity analysis has been selected to evaluate the sensitivity of the flight path altitude at specified points to the optimizable parameters. By decomposing the variance of the output into contributions from individual inputs and their interactions, variance-based sensitivity analysis provides a detailed understanding of the model's sensitivity.

The analysis is based on the following theory summarized below:

1. **Model definition**: Let $Y = f(X)$ be a model where Y is the output and $X = (X1, X2, \dots, Xk)$ are the input variables. Assume X is a vector of random variables with known distributions.

2. **The Law of Total Variance and variance decomposition**: Variance decomposition leverages the law of total variance, which states:

$$Var(Y) = E[(Y - E[Y])^2] = E[Var(Y \mid X)] + Var[E(Y \mid X)],$$

   where $E[Var(Y \mid X)]$ represents the variance within subgroups defined by different levels of X, and $Var[E(Y \mid X)]$ represents the variance due to differences in the expected values of Y across different levels of X.

3. **Additive decomposition**: For a model with independent inputs, the variance of the output can be decomposed additively:

$$Var(Y) = \sum_{i=1}^{k} Var(Y_{Xi}) + \sum_{i \neq j} Cov\left(Y_{Xi}, Y_{Xj}\right) + \cdots + Cov(Y_{X1}, \dots, Y_{Xk})$$

   where Y_{Xi} is the model output due to the input Xi, and the covariances capture interactions between different inputs.

4. **Sobol Sensitivity Indices:** Sobol sensitivity indices extend variance decomposition to measure the contributions of individual inputs and their interactions. First-order Sobol indices S_i:

$$S_i = \frac{Var[E[Y \mid X_i]]}{Var(Y)}$$

are the proportion of variance in Y directly attributable to the input Xi , while accounting for interactions with other inputs. Total Sobol indices ST_i:

$$ST_i = 1 - \frac{Var[E[Y \mid X_{-i}]]}{Var(Y)}$$

express the total effect of X_i, including interactions with all other input variables X_{-i}.

Sobol indices can be estimated via Monte Carlo simulations and polynomial chaos expansions, or via sampling of the parameter space.

Sobol sensitivity analysis considers both individual parameter and parameter interaction effects over the entire parameter space, making it particularly useful for more complex systems and for providing a global view of the sensitivities. This contrasts with local methods such as partial derivatives, which assess sensitivity at a specific point. For understanding of flight path sensitivities, it is considered important to consider the entire parameter space, as the full envelope is relevant in the context of noise emissions. The key drawback to the method is that it requires extensive model evaluations, which is computationally demanding. This was, however, assessed to be feasible with the flight path model. Hence the selection of the Sobol method for this work.

For this work, Saltelli sampling is used to sample the unit hypercube representing the normalized parameter space. Saltelli sampling is based on Sobol sequences, which are low-discrepancy sequences designed to offer more uniform coverage of a specified range than pure random sampling on the uniform distribution. The method is deterministic. Saltelli samples offer fairly uniform coverage and, thus, more efficient sampling in high-dimensional spaces, as well as reproducibility. This is illustrated in two dimensions in Figure 10 below.

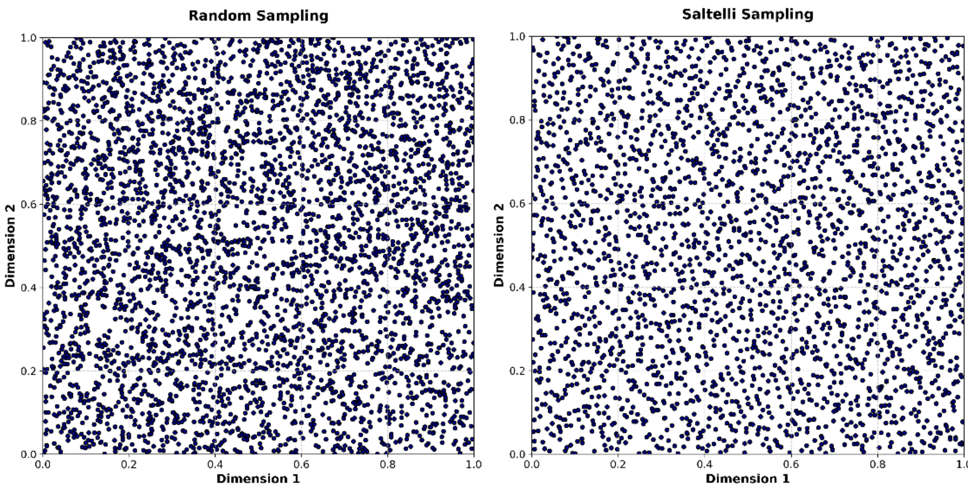

Figure 10 - Random uniform sampling (left) on a unit square versus Saltelli sampling (right) displaying a smoother distribution with the same number of points.

2.3 RESULTS

This section focuses on examining the results obtained from the aircraft operation reconstruction methodology illustrated in this document for Schiphol, Arlanda Stockholm, Dublin, and Rotterdam The Hague airports. Firstly, Section 2.3.1 describes the number of operations in both T2.1 and T2.2 datasets, comparing the outcomes with actual air traffic data. Section 2.3.2 includes the validation of the GT reconstruction algorithm. Secondly, Section 2.3.3 tackles the analysis of the flight profiles resulting from the application of the mixed analysis-synthesis approach, providing separate examination of departures and arrivals. Finally, Section 2.3.3.3 shows the outcomes of the variance based sensitivity analysis carried out on numerous departure vertical flight profiles.

2.3.1 Aircraft operations in the T2.2 datasets

As illustrated in Sections 1.2 and 2.2, aircraft operations can be removed from the analysis for several reasons, leading to the number of unique flight identifiers in the T2.2 dataset being lower than that in T2.1. Table 14 shows the comparison between T2.1 and T2.2 databases. Across the four airports considered, about 5-7% of the original datasets couldn't be included in the analysis. This percentage is higher (≅ 16.7%) for Rotterdam The Hague Airport for the reasons mentioned in Section 1.3.1.

Table 14 – Total number of operations present in T2.1 and T2.2 datasets.

Airport	Airport ICAO	T2.1 dataset	T2.2 dataset	Survival rate [%]
Schiphol Airport	EHAM	34,060 / 16,700 / 17,360	31,801 / 15,414 / 16,387	93.34
Dublin Airport	EIDW	18,219 / 9,449 / 8,770	17,033 / 8,422 / 8,611	93.49
Rotterdam The Hague Airport	EHRD	2,123 / 1,144 / 979	1,790 / 871 / 919	84.31
Stockholm Arlanda Airport	ESSA	14,346 / 7,370 / 6,976	13,597 / 6,733 / 6,864	94.78

Since some flights are not considered, it is crucial to quantify the number of remaining aircraft operations, and to do so Table 15 and Figure 11 are provided. Referring to Schiphol Airport, Figure 11 shows the comparison between T2.2 operations with the traffic data provided by EUROCONTROL [20]. Overall, the mean coverage over the month of March 2023 is 95.22% which is considered a satisfying result. Specifically, departures present a slightly worse coverage that arrivals (92.32% versus 98.13%). The relatively low coverage of take-offs is particularly affected by 11th and 13th of March which account for less than 90% of actual departures. Finally, Table 15 indicates the air traffic coverage at the other airports: while Dublin Airport shows results comparable to Schiphol Airport, significantly worse coverage is found in Stockholm and Rotterdam The Hague airports. In EHRD, these outcomes are primarily caused by a considerable number of smaller aircraft which cannot be mapped to any specific ANP proxy. On the other hand, at ESSA the number of flight operations was already low in the original T2.1 dataset, even considering those which could not be processed.

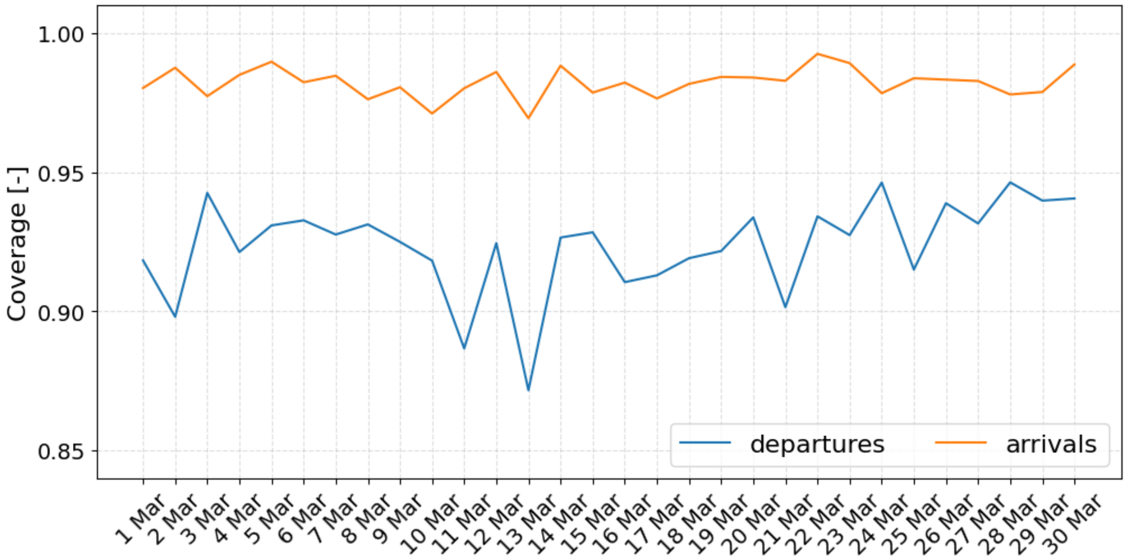

Figure 11 – At Schiphol Airport, T2.2 dataset air traffic coverage.

Table 15 – T2.2 dataset air traffic coverage across the four considered airports.

Airport	Mean coverage [%]		
EHAM	95.29	92.35	98.22
EIDW	95.37	94.21	96.54
EHRD	83.11	81.14	85.08
ESSA	88.63	87.69	89.57

2.3.2 Ground track reconstruction

After identifying the usable aircraft operations, the first result provided by the modelling methodology consists in the reconstructed ground tracks, which are reported in the GT maps of Figure 12. The maps show the preferential directions in the TMA, which typically depend on both airport traffic control orders and the weather conditions, particularly wind direction and speed.

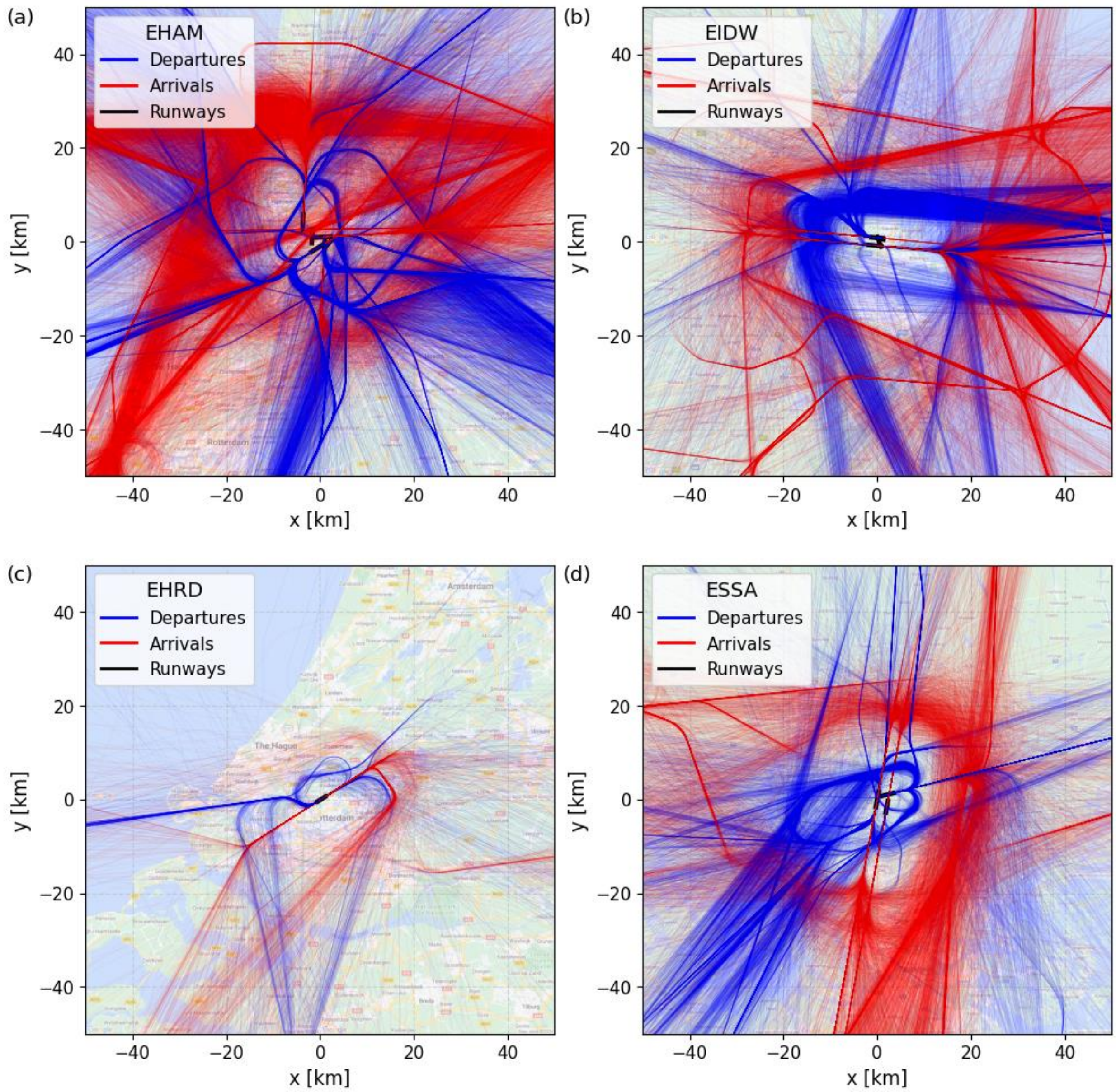

Figure 12 - Ground track map of air traffic around Schiphol (a), Dublin (b), Rotterdam The Hague (c), and Stockholm Arlanda (d) airports.

The validation of the GT reconstruction algorithm is conducted by examining the projection errors between ground tracks and ADS-B positions, with each error defined as the distance between ADS-B point and corresponding GT segment. The errors were computed for all ADS-B points of all operations, thus leading to several error-related PDFs. Several related outcomes for all four airports are reported in Table 16, while PDFs are shown only for Schiphol Airport in Figure 13.

Firstly, Figure 13(a) reports in a semi-logarithmic scale the global PDF of all GT errors, as well as the separate PDFs hosting the errors related only to straight segments and turns respectively, with their median values also added as dashed lines. The global median error is 13 m, and the number of points with error under 100 m is almost 90%, which can be considered satisfying given that the imposition of the tangency condition between segments and circular arcs is a strong constraint.

Furthermore, the median error for turns is just moderately higher than the one for straight segments (21 m vs 11 m), suggesting that aircraft turns can be modelled quite well using only circular arcs.

The only downside is the 0.99% of points with errors over 500 m, on which an examination is conducted in Figure 13(b) with a dedicated PDF. Analysis of this PDF indicates that the vast majority of such errors occurred more than 20 km away from the airport, with the root causes being both the Cartesian-geographic coordinate conversion, unsuited at large distances from the airport, and the inability of the algorithm to well capture large radii turns due to the non-zero threshold angle α_{th} (see Section 2.2.1.1). Therefore, these errors in the GT reconstruction are unlikely to affect significantly the aircraft flight path reconstruction.

Table 16 reports the main outcomes for all the airports considered, showing that the algorithm has consistent performance regardless of the airport. This result confirms the reliability of the present GT reconstruction algorithm when focus is put on the aircraft operations in the TMA.

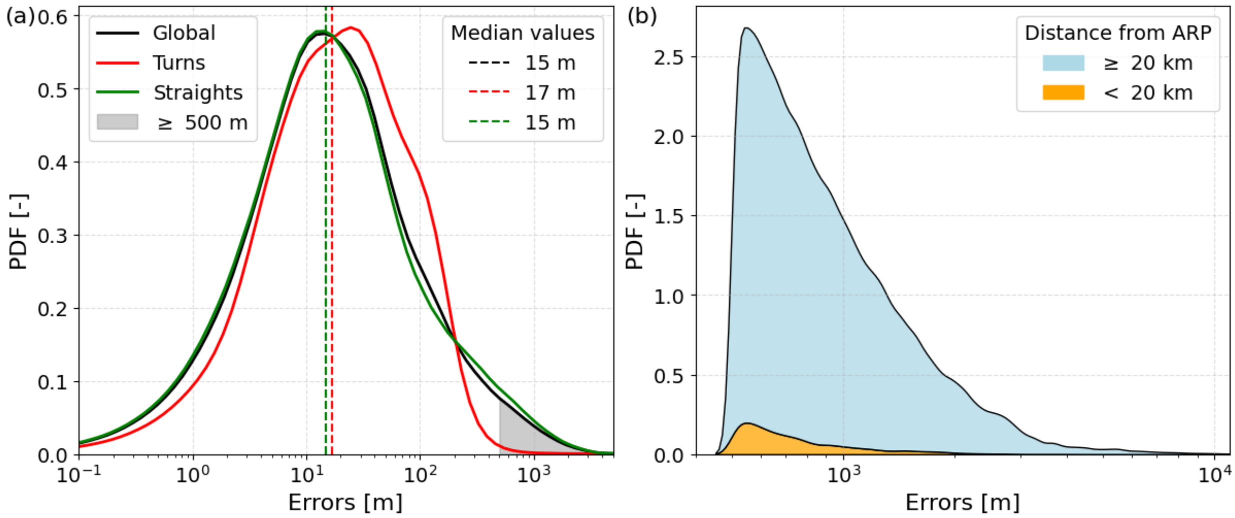

Figure 13 – At Schiphol Airport, (a) PDFs of GT errors accounting for all ADS-B points and separately for straight-flight segments and turns, and (b) PDFs of GT errors focusing on those over 500 m.

Table 16 – Outcomes of the GT reconstruction algorithm across all four considered airports.

Airport	Median errors [m]			Below 100 m [%]	Above 500 m [%]
EHAM	13	11	21	89.79	0.99
EIDW	14	13	19	91.25	0.73
EHRD	15	15	16	87.67	1.86
ESSA	13	12	19	88.16	1.27

2.3.3 Flight profile reconstruction with the mixed analysis-synthesis approach

The section is dedicated to the results obtained with the application of the mixed analysis-synthesis approach. The departures are discussed in Section 2.3.3.1, while the arrivals in Section 2.3.3.2.

2.3.3.1 Mixed approach applied to departure flight profiles

A quantification of the improvement obtained with the mixed approach applied to departures compared to the baseline ECAC Doc.29 method is provided in Figure 14 for Schiphol Airport. Firstly, Figure 14(a) illustrates the statistical distribution, together with its mean and median, of the ratio between optimized and non-optimized RMS_{ZV} for all the processed departures. The non-optimized value $RMS_{ZV,unopt}$ is computed using the default ANP procedure with maximum available thrust ($K_{red} = 1$) and flight-distance-dependent TOW as suggested by ECAC Doc.29. The distribution median is 0.31, meaning that the overall RMS_{ZV} reduction is close to 70%, a promising result considering that key elements such as ANP step sequences and flap settings were preserved. Moreover, around 94.8% of departures experience a 30% or larger reduction in RMS_{ZV}, while the remaining 5.2% with RMS_{ZV} ratio greater than 0.7 is generally performed by very small aircraft, for which the ANP procedures are lacking. This last consideration is made more evident in Figure 14(b), which shows the impact of the optimization, expressed through the distributions of the RMS_{ZV} ratio separately for each proxy, whose name is listed on the x-axis. For each proxy, the blue box indicates the 25th to 75th percentiles and the median, while each outlier is denoted as a black circle. This figure highlights that the proxies representing the most common aircraft (i.e. recent Airbus, Boeing, and Embraer models) have better outcomes compared to the ones mapped to smaller aircraft and private jets.

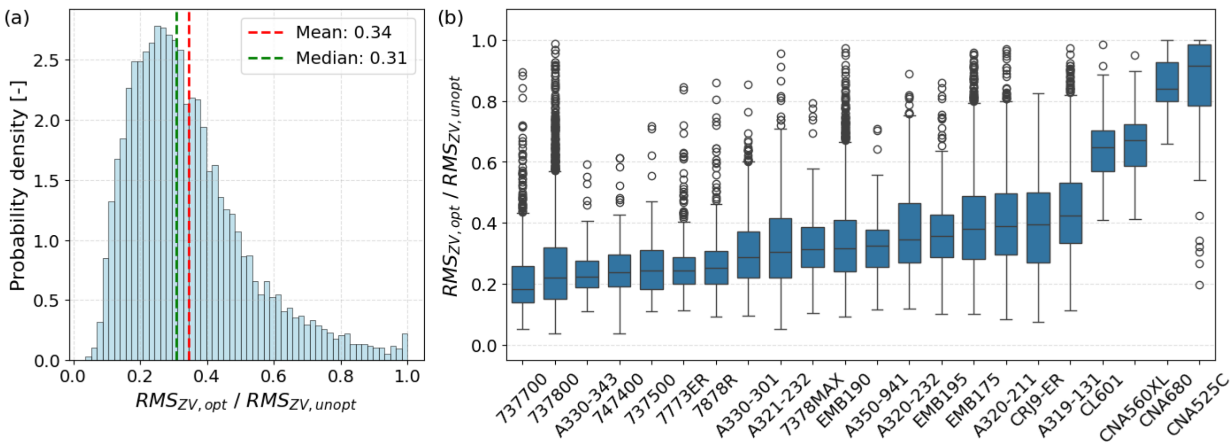

Figure 14 – At Schiphol Airport, (a) distribution of RMS_{ZV} ratio considering all departures, and (b) RMS_{ZV} ratio for each ANP proxy ordered by median.

Additionally, as the RMS_{ZV} ratio measures the improvement yielded by the mixed approach compared to the original ANP steps, a high value of $RMS_{ZV,\mathrm{opt}}/RMS_{ZV,\mathrm{unopt}}$ does not necessarily mean that the outcomes of this approach are bad. In fact, it could happen that the default (unmodified) ANP procedure is already close to the ADS-B data, and the mixed approach yields a result which is just slightly better the starting point. Two examples of this occurrence are given in Figure 15, which shows the results for two departure operations with RMS_{ZV} ratio higher than 0.9.

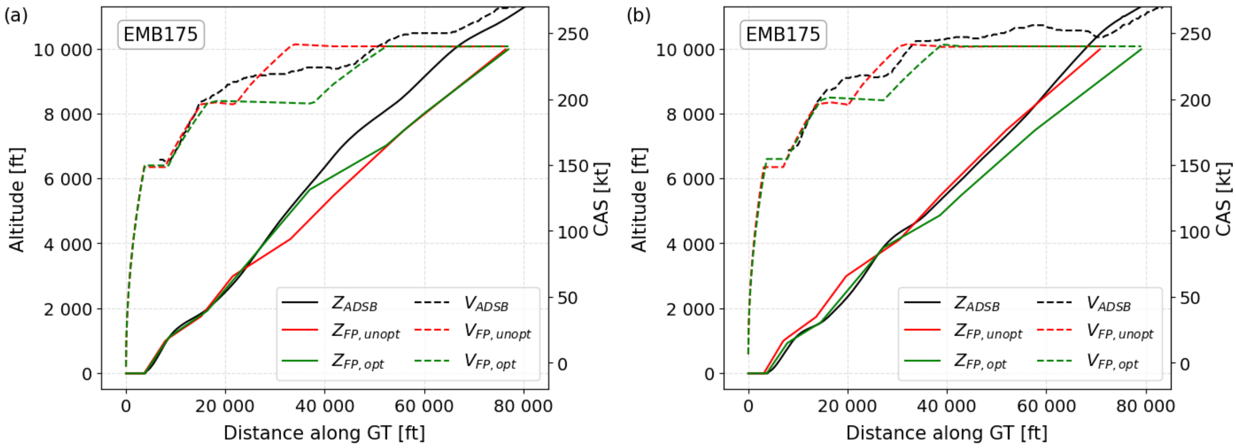

Figure 15 – At Schiphol Airport, flight profiles where the mixed approach yielded satisfying results despite the value of $RMS_{ZV,opt}/RMS_{ZV,unopt}$ being high (here ≥ 0.9).

Table 17 indicates that all these findings are quite consistent across all the considered airports, confirming the reliability of the mixed approach, but it is also observed that in Dublin, Stockholm and especially Rotterdam airports, both mean and median values of $RMS_{ZV,opt}/RMS_{ZV,unopt}$ are slightly higher than that in Schiphol Airport. In general, this happens because the fleet employed in those three airports is composed of a larger number of smaller aircraft and private jets, while the most common commercial airliners are comparatively fewer. Additionally, at Rotterdam The Hague Airport another key reason that contributes to the higher values of $RMS_{ZV,opt}/RMS_{ZV,unopt}$ was found: several ADS-B flight profiles show the presence of an acceleration-level phase, and a few of them are illustrated in Figure 16(a). As the ANP procedural steps for departures do not currently cover this phase type, in this scenario the mixed approach could not yield satisfactory results, at least regarding the altitude profiles (Figure 16(b)).

Table 17 – Overall results of the mixed approach applied to departure flight profiles across all considered airports.

Airport	$RMS_{ZV,opt}/RMS_{ZV,unopt}$	
	Mean	Median
EHAM	0.35	0.32
EIDW	0.38	0.34
EHRD	0.49	0.45
ESSA	0.43	0.42

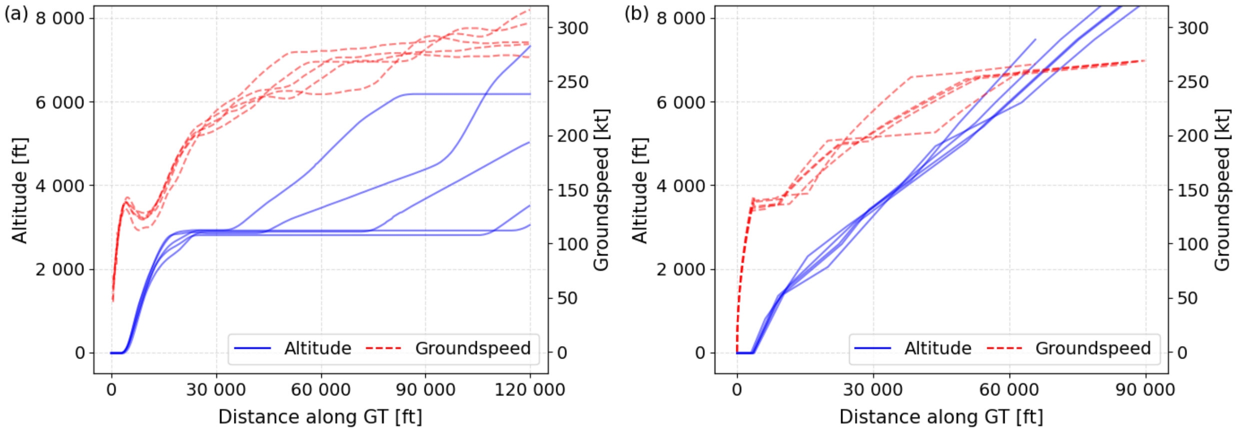

Figure 16 – At Rotterdam The Hague Airport, (a) examples of flight profiles which present acceleration-level phase, currently not covered by ANP procedural steps, and (b) the outcomes provided by the mixed approach.

A second type of analysis is conducted concerning the thrust-weight relationship, which consists in the behaviour of thrust reduction coefficient K_{red} (both K_T and K_C) compared to the TOW ratio K_{MTOW}. This is shown in the scatter plot in Figure 17, together with the PDFs of K_{MTOW} and K_T, respectively plotted along the x- and y-axes. Firstly, and most importantly, this scatter plot has a very similar appearance to those from literature [15] [21]. There is a clearly visible diagonal, where $K_T \cong K_{MTOW}$, which appears due to the penalty coefficient CF defined in Section 2.2.1.2.1. The majority of departures are distributed roughly along this diagonal, with some points to the right of it despite the penalty CF: this indicates that the solutions with $K_{MTOW} > K_T$ are still preferable to others, even with the increase in OBF caused by the penalty coefficient. Additionally, two accumulations are observed at $K_T \cong 1$ (larger) and $K_T \cong 0.75$ (smaller), and high K_C values are mostly associated with high K_T and K_{MTOW}, whereas low K_C with similarly low K_T and K_{MTOW}. Finally, the figure highlights two vertical lines corresponding to K_{MTOW} values approximately 0.875 and 0.95. The first accumulation of events, at $K_{MTOW} \cong 0.875$, is due to the Boeing 737-800 proxy, which presents a discontinuity in the ICAO_A procedural steps when transitioning from stage length 3 ($K_{MTOW} \cong 0.84$) to 4 ($K_{MTOW} \cong 0.90$). The second accumulation of events, at $K_{MTOW} \cong 0.95$, is caused by the Airbus A319-131 proxy, which presents a similar discontinuity in the default procedural steps between stage length 4 ($K_{MTOW} \cong 0.88$) and 5 ($K_{MTOW} = 1$).

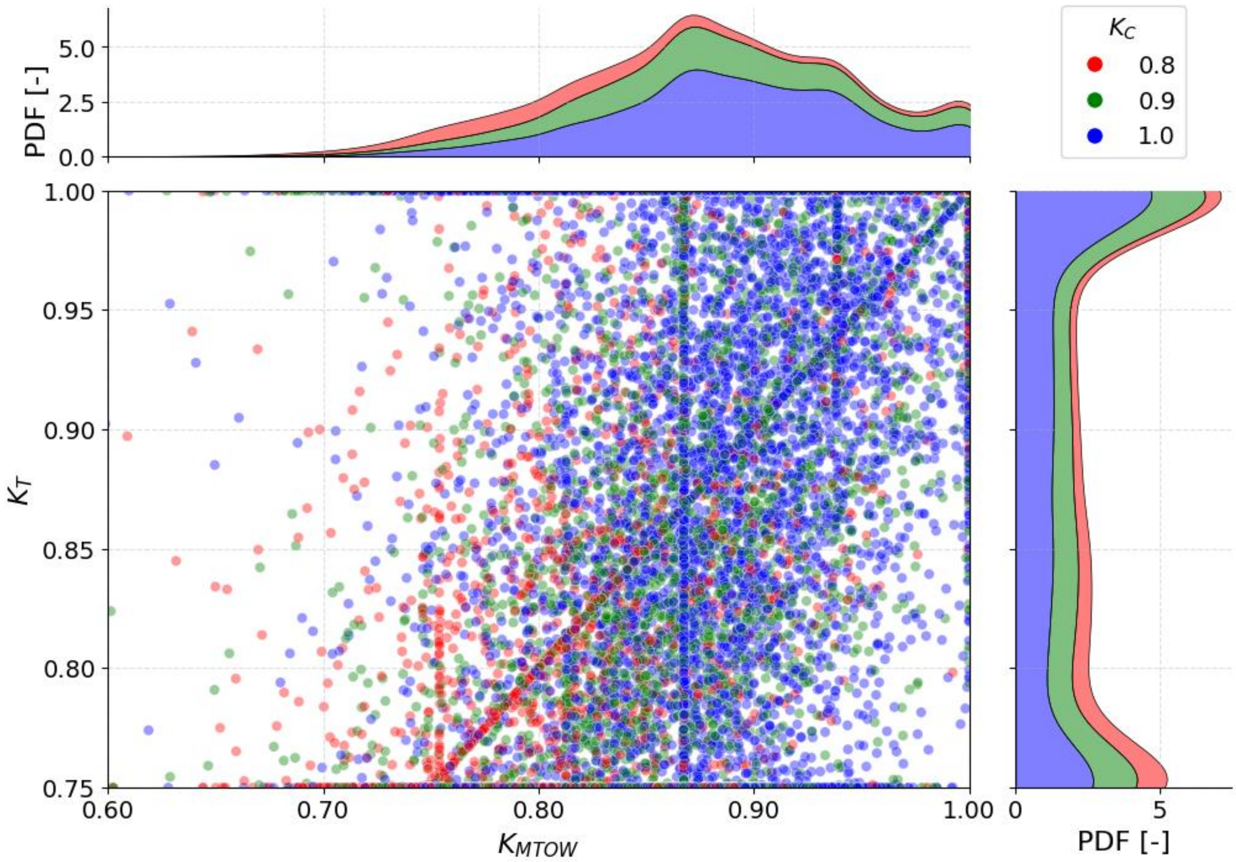

Figure 17 - $K_T - K_C - K_{MTOW}$ relationship, expressed as a scatter plot accompanied by PDFs of K_T and K_{MTOW} on the appropriate axes.

2.3.3.2 Mixed approach applied to arrival flight profiles

Similarly to what done for departures, the improvements achieved with the mixed approach applied to arrivals are presented as statistical distributions involving the ratio between optimized and non-optimized RMS_{ZV}. The non-optimized $RMS_{ZV,unopt}$ is computed relying on the unaltered ANP approach procedure. The outcomes are illustrated in Figure 18 for Schiphol Airport. Firstly, Figure 18(a) shows the distribution of RMS_{ZV} ratio for all arrivals, reporting also a distribution median of 0.23, which implies an overall improvement close to 80%. This value is markedly higher than that of departures, and this is mostly due to *i*) the fact that landing procedures are on average more constrained than departures, thus allowing the ANP profiles to better match the ADS-B data, and *ii*) the different expression for OBF employed compared to take-offs. On the other hand, Figure 18(b) illustrates the proxy-wise distribution of RMS_{ZV} ratio, showing that the median value is under 0.4 even for the worst proxy. In fact, the proxies that refer to the most common aircraft do not necessarily yield to better results compared to the smaller ones, and in general the differences between ANP proxies are now much less pronounced with respect to the outcomes registered for the departures.

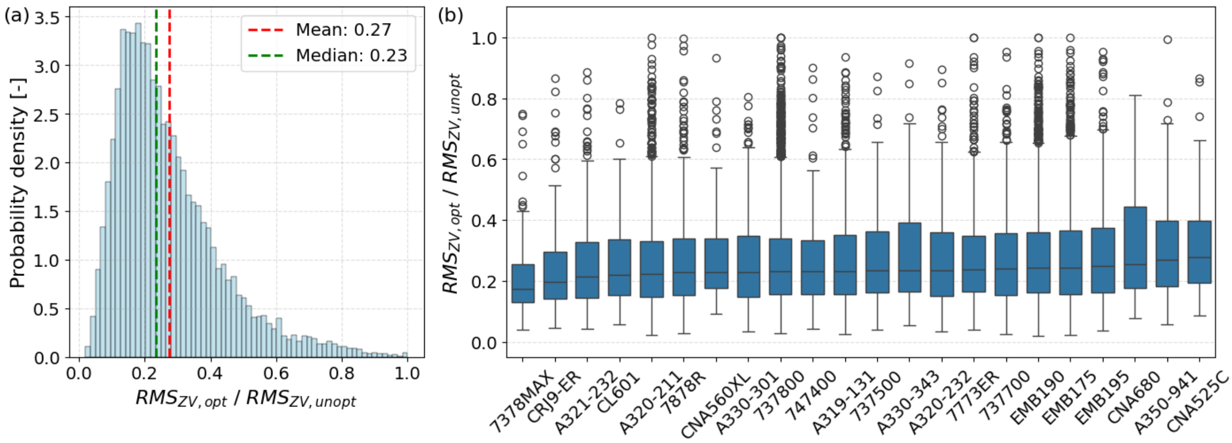

Figure 18 - At Schiphol Airport, (a) distribution of RMS_{ZV} ratio considering all arrivals, and (b) RMS_{ZV} ratio for each ANP proxy ordered by median.

Finally, Table 18 indicates that all these findings are consistent across all the considered airports, confirming the reliability of the mixed approach. The table makes no distinction between more common and smaller aircraft, and conclusions similar to those drawn for Schiphol Airport can be applied.

Table 18 – Overall results of the mixed approach applied to arrival flight profiles across all considered airports.

Airport	$RMS_{ZV,opt}/RMS_{ZV,unopt}$	
	Mean	Median
EHAM	0.27	0.23
EIDW	0.28	0.25
EHRD	0.34	0.30
ESSA	0.30	0.28

As a final remark, it is worth noting that the examination of the flight profile results that has been conducted in this subsection does not constitute full validation of the methodology. The latter, in fact, would be possible only if the actual engine thrust outputs were available for each of the aircraft operations processed, but these values are not publicly available. Therefore, other validation strategies must be employed concerning the thrust levels, and these usually involve referring to the environmental outputs, as done by UNIUD with airport noise levels for the previous [13] and current [14] versions of the mixed approach. However, although this strategy leads to generally satisfactory results, it is not definitive proof of a better modelling. In this regard, the availability of FDR data on an adequate number of aircraft operations from task T2.4 appears of paramount importance, so that by the next T2.2 deliverable a higher-quality validation can be achieved, concerning also the fuel consumption estimates.

2.3.3.3 Outcomes of the variance sensitivity analysis

Using the latest method of flight path optimization developed in NEEDED WP2, the flight paths for A320-211 proxy-related departure operations using the ICAO_A flight profile, on runway 36L at Amsterdam Schiphol airport over seven days from 2nd March 2023 to the 8th March 2023 have been calculated. Figure 19 shows the flight paths in terms of altitude versus ground track distance. Inspection of the results suggests considerable variation in both the altitudes at various fixed points and flight path shape within the confines of the ICAO_A profile.

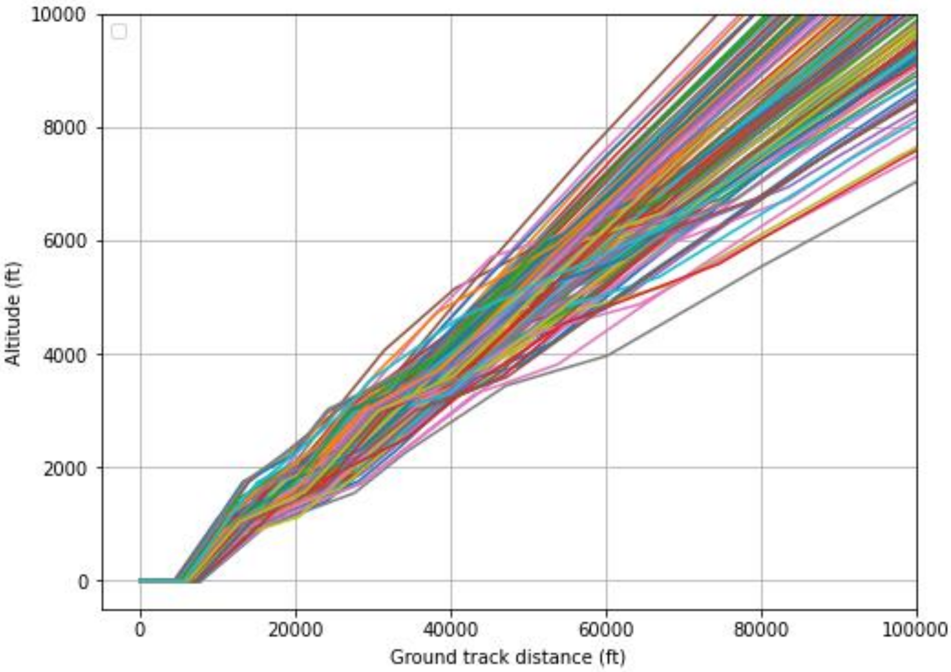

Figure 19 – Calculated flight paths for A320-211 proxy-related departure operations using the ICAO A flight profile, on runway 36L at Amsterdam Schiphol airport over seven days from 2nd March 2023 to the 8th March 2023

Results for a Sobol sensitivity analysis carried out on such vertical flight profiles are presented below. The sensitivity of the flight path altitudes at surface distances of 3km (10,000'), 6km (20,000'), 12km (40,000'), 18km (60,000'), and 24km (80,000') were calculated. The figures show the first-order Sobol sensitivity indices for the departure parameters (from left to right in the figures) take-off thrust reduction (k_takeoff), take-off weight (tow), energy share factor (acc_share), initial climb altitude change (alt_change_ini), mid-climb altitude change (alt_change_mid), and climb thrust reduction (k_climb). The code currently uses altitude to calculate the sensitivity indices. However it is easily extendable to use noise measurements. Such an extension will be considered if and when the capability becomes available.

Figure 20 to Figure 24 show the sensitivities for the DEFAULT departure profile. Figure 25 to Figure 29 show the sensitivities for the ICAO_A departure profile. Sample sizes of 80,000 flight paths for each departure profile were considered. The confidence intervals for each index are superimposed over their respective bars. The following observations hold for both profiles:

1. The take-off weight (tow) is the dominant parameter throughout the surface distance range.
2. Take-off thrust reduction (k_takeoff) appears to exert considerable influence at up to 12km and maintains some non-zero influence beyond.
3. Energy share factor (acc_share) and initial climb altitude change (alt_change_ini) show a spike in influence at around 12km.
4. Climb thrust reduction (k_climb) becomes influential beyond 12 km.

In terms of differences between the profiles:

1. Take-off thrust reduction (k_takeoff) maintains influence a little further with the DEFAULT profile.
2. Climb thrust reduction (k_climb) becomes more influential further out with the ICAO_A profile.

The indications suggest that in calculating flight profiles the estimation of take-off weight is of relatively high significance in achieving altitude accuracy throughout the range. For flight path calculation accuracy and, as a dual purpose, design of noise sensitive operation procedures, the indication is that take-off thrust reduction, energy share factor and initial climb altitude change, and climb thrust reduction could be worth closer consideration with respect to the distance ranges over which they exert greater influence on altitude, as listed above.

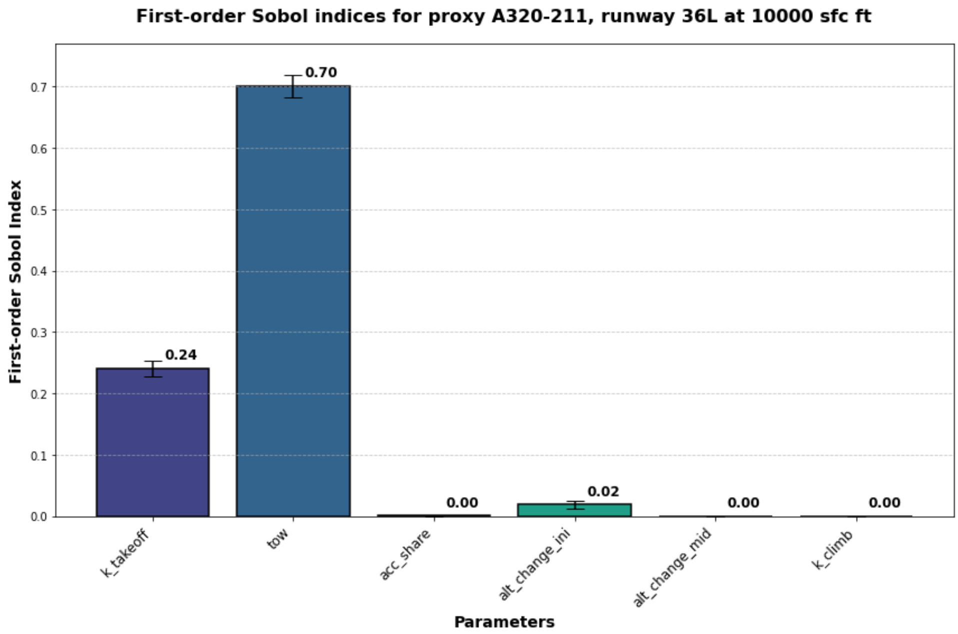

Figure 20 - Altitude sensitivities to flight path parameters at 3km ground track surface distance for aircraft with the A320-211 proxy using the DEFAULT departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

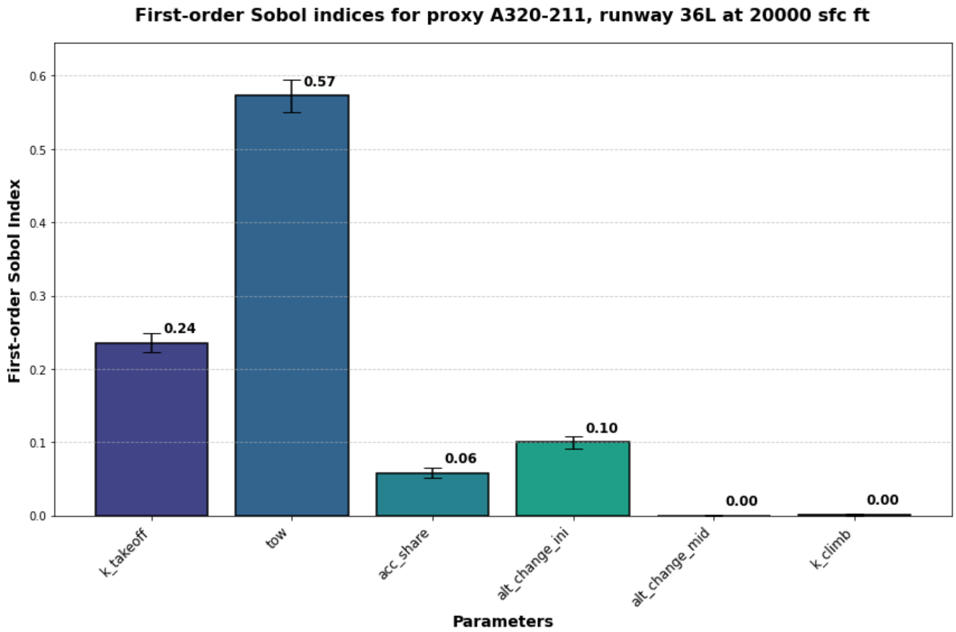

Figure 21 - Altitude sensitivities to flight path parameters at 6km ground track surface distance for aircraft with the A320-211 proxy using the DEFAULT departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

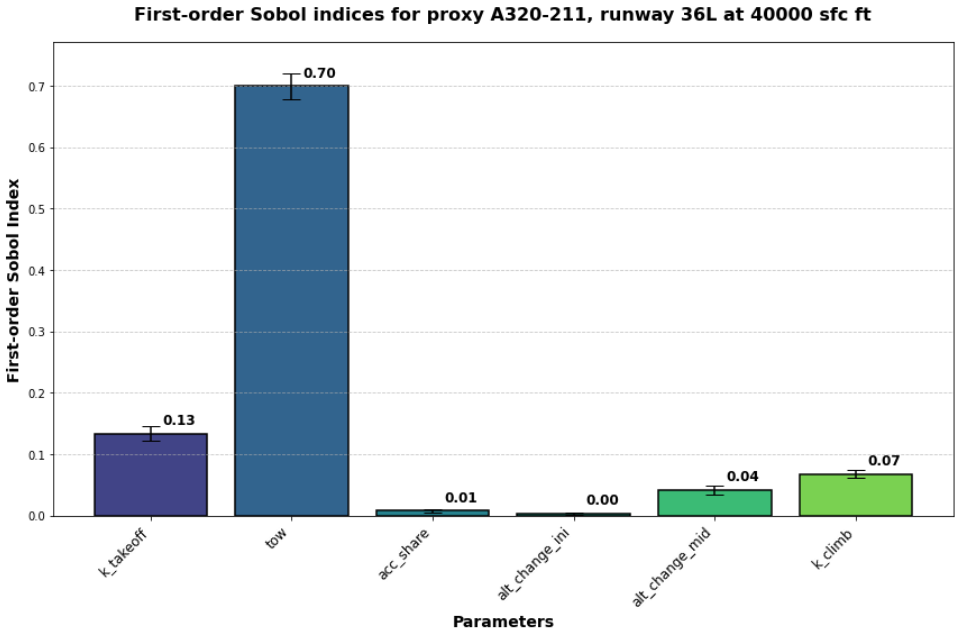

Figure 22 - Altitude sensitivities to flight path parameters at 12km ground track surface distance for aircraft with the A320-211 proxy using the DEFAULT departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

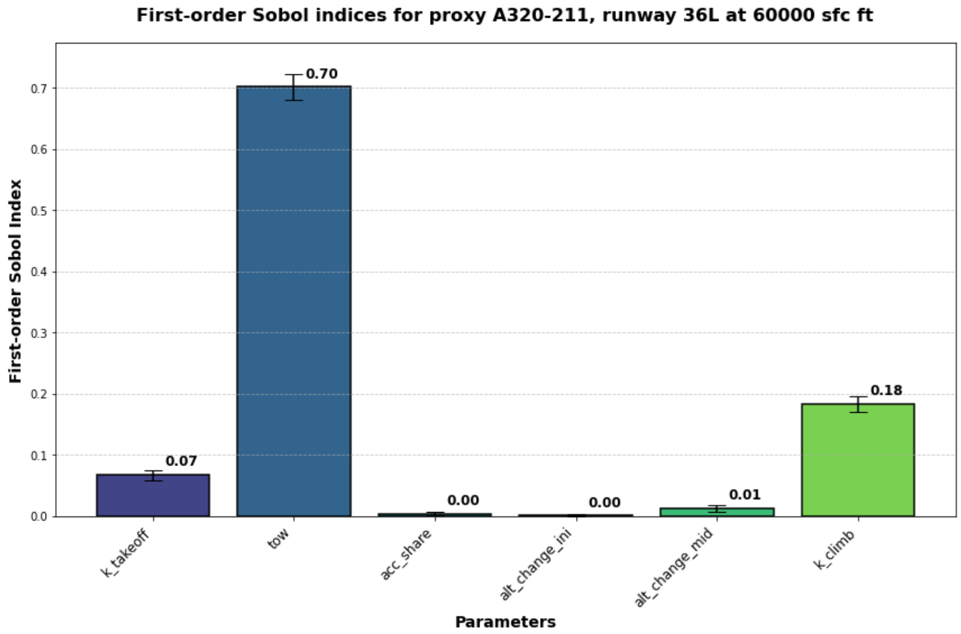

Figure 23 - Altitude sensitivities to flight path parameters at 18km ground track surface distance for aircraft with the A320-211 proxy using the DEFAULT departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

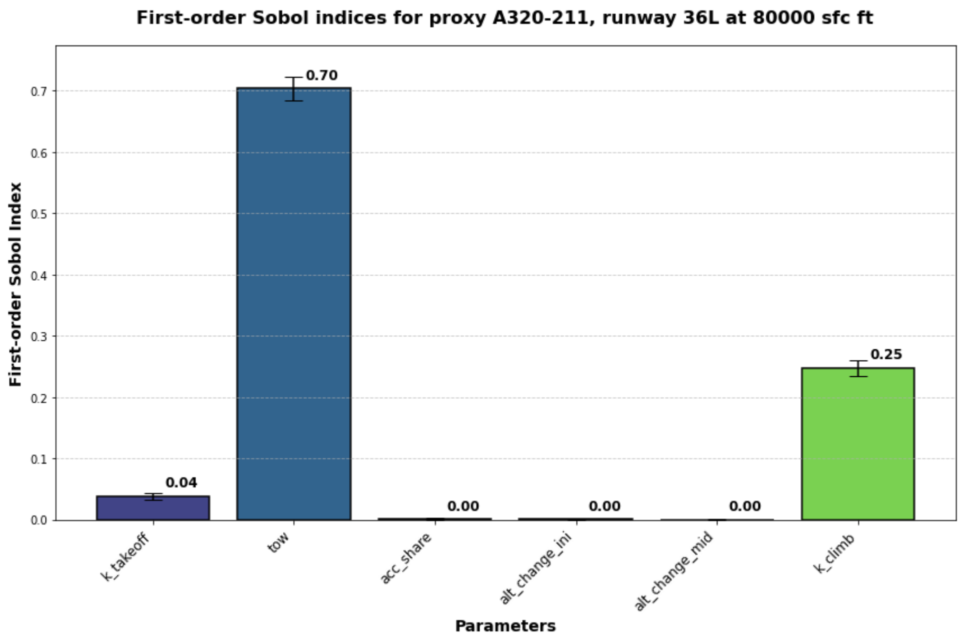

Figure 24 - Altitude sensitivities to flight path parameters at 24km ground track surface distance for aircraft with the A320-211 proxy using the DEFAULT departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

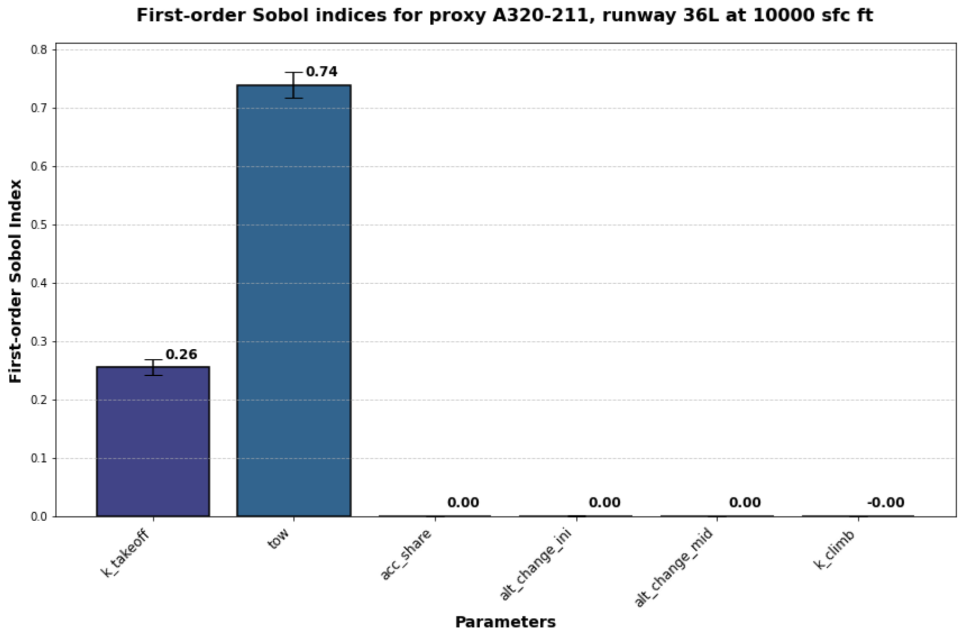

Figure 25 - Altitude sensitivities to flight path parameters at 3km ground track surface distance for aircraft with the A320-211 proxy using the ICAO_A departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

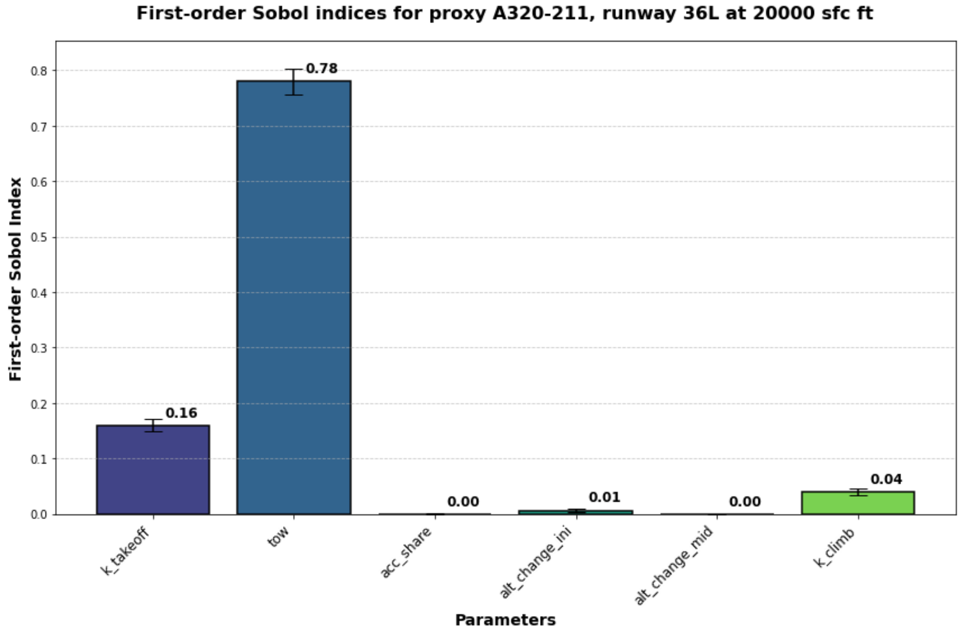

Figure 26 - Altitude sensitivities to flight path parameters at 6km ground track surface distance for aircraft with the A320-211 proxy using the ICAO_A departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

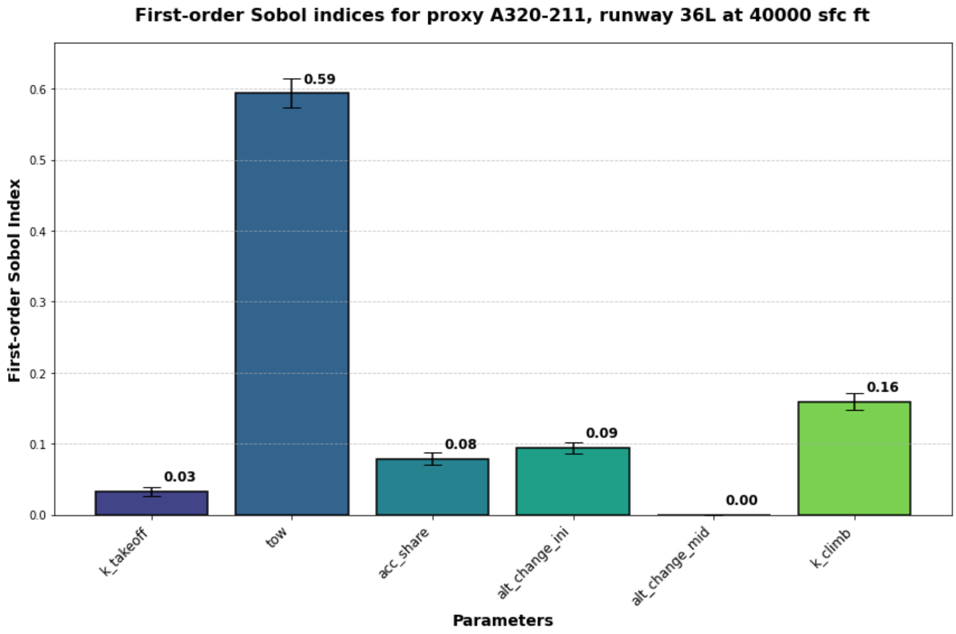

Figure 27 - Altitude sensitivities to flight path parameters at 12km ground track surface distance for aircraft with the A320-211 proxy using the ICAO_A departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

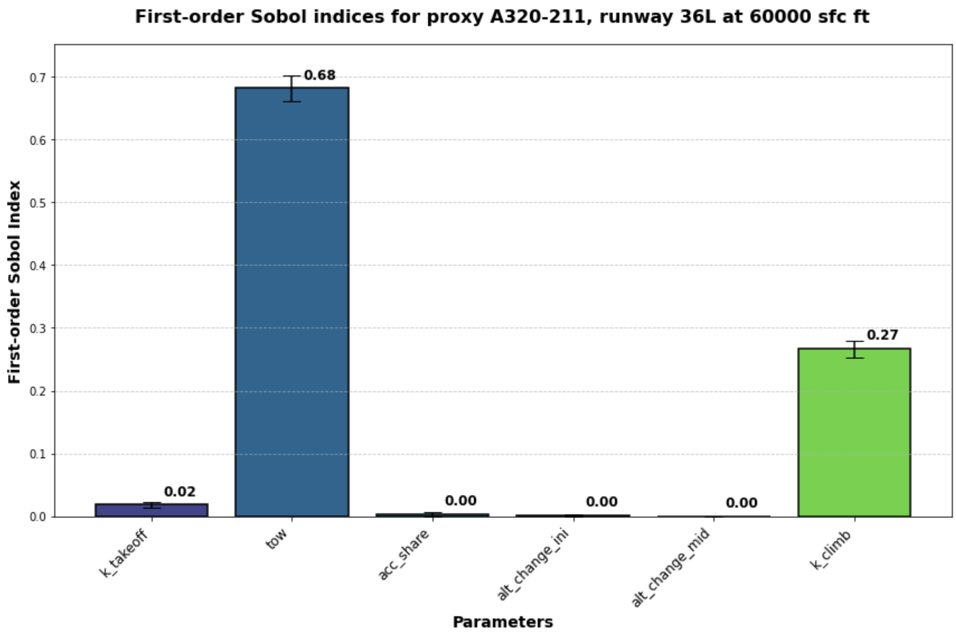

Figure 28 - Altitude sensitivities to flight path parameters at 18km ground track surface distance for aircraft with the A320-211 proxy using the ICAO_A departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

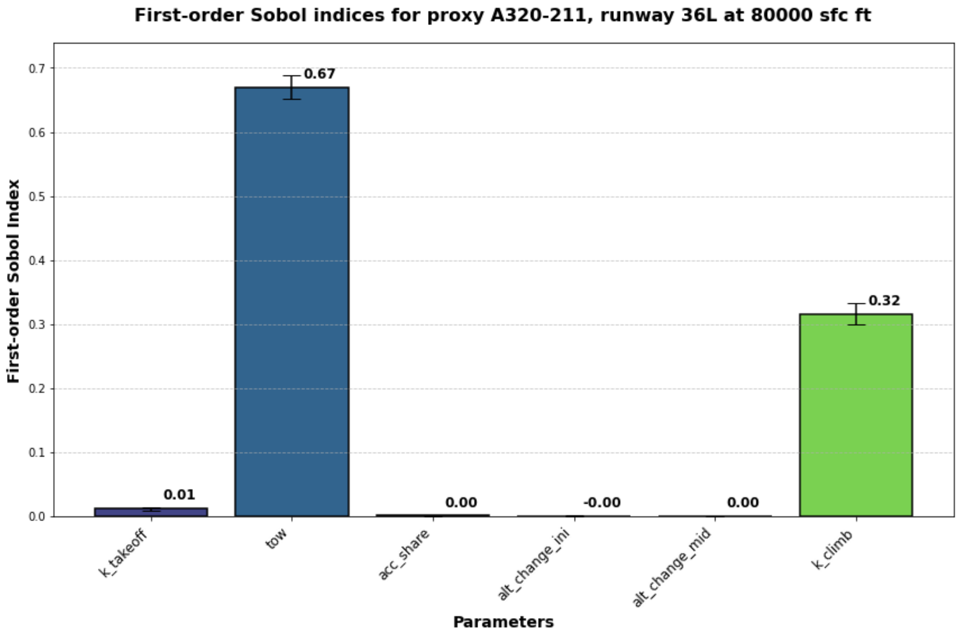

Figure 29 - Altitude sensitivities to flight path parameters at 24km ground track surface distance for aircraft with the A320-211 proxy using the ICAO_A departure profile on runway 36L at Amsterdam Schiphol, based on real world data from 02/03/2023 – 08/03/2023.

2.4 CONCLUSIONS

Section 2 of the present deliverable D2.2, has illustrated the methodology developed in task T2.2 as of June 2024 (M18 in NEEDED) for the reconstruction of aircraft operations in the TMA on the basis of the flight traffic T2.1 datasets enriched with the metadata coming from the pre-processing as per Section 1. Although this methodology is still in a preliminary stage, coherently with the duration of T2.2 (M5 to M28) and the mid-term placement of this deliverable, many of its key elements have already been defined. The final output of this task will consist in a large number of reconstructed aircraft operations that will include the parameters and metadata required for the estimation of pollutant emissions (WP3) and noise levels (WP4) in airport areas.

A summary of the key steps of the aircraft operation reconstruction methodology is given in this paragraph. The reconstruction of each aircraft operation is carried out by determining its flight path, which is built by merging the ground track and the flight profile, each of which is determined with specific modelling solutions. In particular, the ground track is computed with a dedicated GT reconstruction algorithm that yields smooth sequences of segments and circular arcs using only the operation flight tracking data, whereas the vertical flight profile, which hosts the key aircraft performance parameters (e.g. engine thrust), is computed with a mixed analysis-synthesis approach that uses the flight tracking information to improve the flight procedures prescribed by ANP and the calculation methods provided by ECAC Doc.29. A very preliminary fuel flow estimation method is also described that is based on publicly available databases and modelling methodologies.

Concerning the validation, the operation identification and ground track reconstruction are shown to be quite satisfying, especially when focusing on the TMA, while the flight profile reconstruction, although promising as shown here, can be assessed only indirectly, and this was done in journal submissions [13] [14] exploiting the official noise levels and contour maps made available by some European airports. In this regard, the FDR (flight data recorder) data expected from Task T2.4 should allow a much more direct validation of the flight profile reconstruction, with particular emphasis on aircraft weight, engine thrust levels, and fuel flow, and these outcomes will be very valuable for better identifying the most critical elements and proposing higher-quality reconstruction strategies.

As outlined in the relevant sections of this document, the current aircraft operation reconstruction methodology presents a number of limitations, but strategies are being formulated to address them. A list of the current main limitations and the possible countermeasures to be implemented in future developments is provided below.

- Firstly, it would be ideal to increase the accuracy of the ground track reconstruction algorithm, further reducing the error between the ADS-B data and the reconstructed ground trajectories: a viable solution could be to use a ‘dynamic’ threshold angle α_{th} (see Section 2.2.1.1) to better identify turns, even the ones with large radii, from the aircraft positions. The possibility of using directly (or nearly so) ADS-B data will also be explored.
- Developments are being considered regarding the mixed analysis-synthesis approach as well. Concerning this approach, two improvements could be implemented: 1) further increasing the number of degrees of freedom attributed to ANP procedural steps, and 2) conducting statistical analysis of the actual flight profiles, particularly those that were not well optimized through the mixed approach. In particular, the second improvement could provide key information on the procedures actually followed at an airport, fostering the definition of new procedural steps that may represent more accurately the actual flight trajectories, especially in terms of altitude and speed. For instance, an attempt could be made to introduce a pure acceleration-level phase in the ANP departure steps to capture better some of the take-off and climb procedures often observed at Rotterdam The Hague Airport (Section 2.3.3.1).
- Another aspect of the present modelling tool that requires attention is the fuel flow calculation for thrust levels lower that 60% of the maximum rated thrust, which mainly pertains to landing operations. Other calculation methods will be investigated and considered, but nevertheless only a comparison between FDR data and model outcomes is likely to provide the necessary information to validate the solutions implemented.
- Finally, the current methodology does not account for the taxi-out and taxi-in phases. Effort will be put into investigating suitable thrust and fuel flow models to include such phases in the future datasets whenever feasible. However, issues persist concerning the availability of an adequate amount of on-ground ADS-B data, which is highly dependent on number and location of the receivers within the airport area, airport infrastructure, terrain morphology, presence of obstacles impeding the data transmission/reception, and weather conditions. The contributions coming from T2.1 on this aspect will be closely monitored.

3 TASK T2.3 - STATISTICAL DISPERSIONS FOR AIRCRAFT OPERATIONS IN ABSENCE OF REAL-WORLD DATA

3.1 INTRODUCTION

The increasing complexity of modern air traffic management (ATM) systems, coupled with the need for accurate environmental assessments, necessitates the development of advanced models to predict noise and local air quality (LAQ) impacts under various scenarios, including those where real-world data are unavailable. This is particularly critical in the context of future scenarios where current data cannot be directly applied. To address these challenges, a comprehensive data-driven approach is required to model statistical dispersion in aircraft operations, which can then be used to simulate noise and LAQ impacts.

This task aims to deliver advanced statistical dispersion models and associated parameters, tailored for scenarios where empirical data may not be readily accessible. The primary goal is to develop robust models that can predict aircraft noise and LAQ under a variety of conditions, providing critical insights for future air traffic scenarios.

The task will be executed in multiple stages, each building upon the outputs of previous tasks. We use flight data from T2.1 and T2.2 to identify groups of flight trajectories that exhibit similar dispersion patterns. These trajectories will be clustered using unsupervised machine learning techniques. Clustering enables the categorization of flight trajectories into distinct groups, each representing a unique dispersion pattern, which is crucial for modelling.

The clustered trajectory data will then be modelled using two distinct approaches, including simple statistical methods and advanced machine learning techniques.

3.2 METHODOLOGY

In the context of analysing aircraft trajectories, clustering is a crucial technique used to group similar flight paths, which can then be studied collectively to identify common patterns, assess environmental impacts, or optimize air traffic management strategies. One of the most effective and widely used clustering algorithms for this purpose is DBSCAN (Density-Based Spatial Clustering of Applications with Noise). DBSCAN is particularly well-suited for trajectory clustering because it does not require the user to specify the number of clusters a priori, and it is robust to noise, making it ideal for handling real-world trajectory data, which often includes outliers and irregular patterns.

3.2.1 Data exploration

In this section, we describe the process of exploring a month's worth of aircraft trajectory data obtained from the OpenSky Network, focusing on arrivals and departures at Rotterdam The Hague Airport (EHRD). This analysis serves as a preliminary step for subsequent tasks, including providing insight on sensor placement experiments and the optimization of air traffic procedures. By examining the patterns in the collected data, we aim to provide insights that will contribute to the objectives of Work Package 4 (WP4), particularly in the position of placing noise monitoring stations around the airport.

3.2.1.1 Data collection and preprocessing

The dataset used for this analysis was sourced from the OpenSky Network, a collaborative platform that aggregates real-time and historical air traffic data from multiple sensors around the world. For this exploration, one month of flight data covering all arrivals and departures at Rotterdam The Hague Airport (EHRD) was downloaded.

Before data can be used for detailed analysis, the raw data was pre-processed to ensure that it included only the relevant information. The preprocessing steps included:

- Filtering for Relevant Flights: The dataset was filtered to include only flights that either landed at or took off from Rotterdam The Hague Airport (EHRD). This ensures that the analysis is focused on the traffic patterns specific to this airport.
- Time and Date Normalization: The timestamps associated with each flight were normalized to a common format, ensuring consistency across the dataset.
- Trajectory Data Cleaning: Outlier removal was conducted to eliminate any erroneous data points that could skew the analysis. This includes handling missing values, duplicate entries, and any flights with incomplete trajectory information.

3.2.1.2 Exploratory data analysis

In T2.3, we first conduct an exploratory data analysis of flight trajectories focusing on departures and arrivals at EHRD.

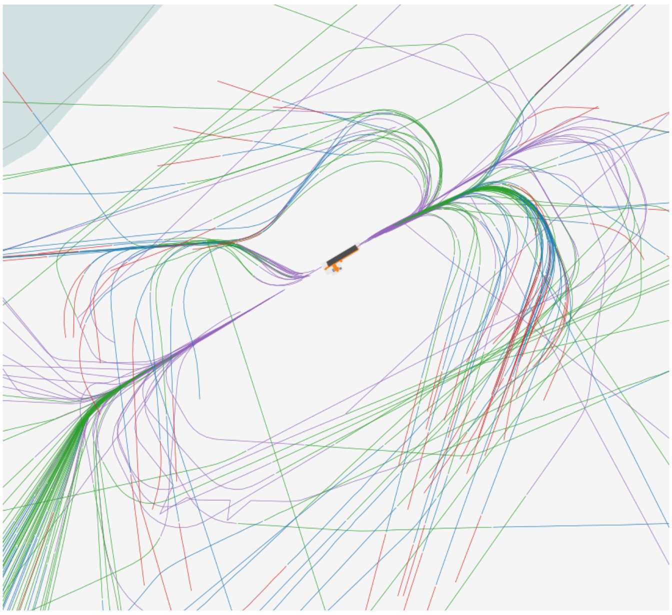

Figure 30 - Example of arrival and departure flights at EHRD, flight levels are coded with colours. Purple: 1000-2500 ft, Green: 2500-4500 ft, Blue: 4500-6500 ft, Red: 6500-8500 ft.

Flight paths for departures and arrivals were visualized to identify common routes and any deviations from standard procedures. The analysis showed that most flights adhered closely to the established arrival and departure paths, with only minor deviations in certain areas. This consistency provides a solid basis for predicting noise exposure areas.

The data was further segmented into different altitude bands to align with the typical departure and arrival procedures at EHRD. Each band reflects a distinct phase of flight, allowing for a detailed examination of how noise propagates as aircraft ascend or descend.

The segmented data was compared against standard departure and arrival procedures at EHRD. The analysis confirmed that aircraft consistently follow established procedures, with predictable patterns emerging in both ascent and descent phases. These findings are critical for identifying areas where noise is most likely to be concentrated.

The analysis also provides inputs regarding the optimal placement of microphones for noise measurements, contributing to the objectives of Work Package 4 (WP4). The data has been segmented into specific altitude bands, reflecting different stages of aircraft departure and arrival procedures, allowing for a targeted examination of noise exposure at various heights.

3.2.2 Clustering of flight trajectories

In this task, we employ the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) as the clustering algorithm for group trajectories with similar departure or arrival patterns. DBSCAN is a robust and widely used clustering algorithm in the field of data mining and machine learning. Introduced by Ester et al. [22], DBSCAN is designed to identify clusters of varying shapes and sizes in a dataset, while effectively distinguishing noise points. Unlike partitioning methods like k-means, DBSCAN does not require the number of clusters to be predefined, making it particularly useful for exploratory data analysis.

DBSCAN is built on several key concepts, each playing a crucial role in how the algorithm operates and defines clusters within a dataset. In the following figure, the key parameters are show. Note that each point in the figure represents a multi-denominational time series flight trajectory data.

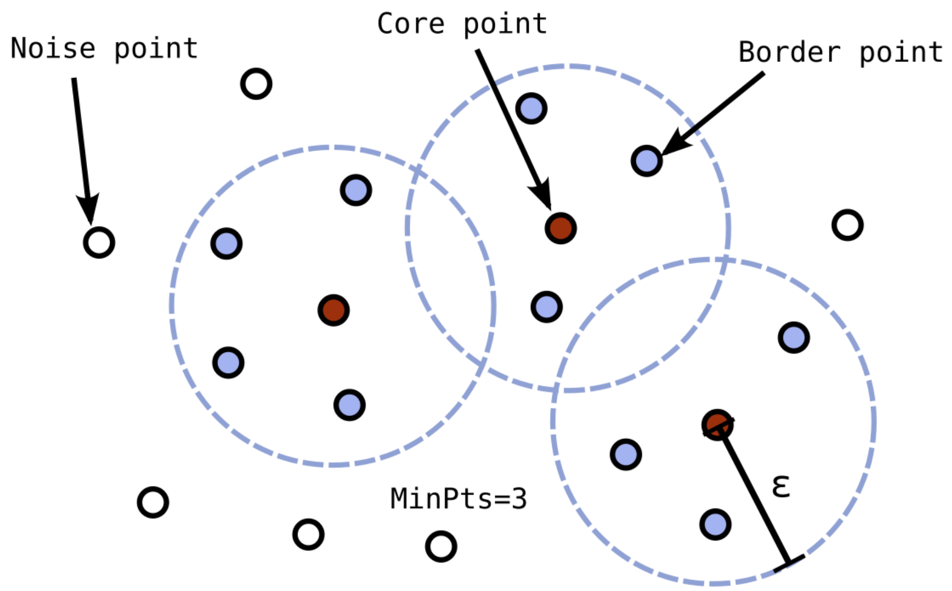

Figure 31 - The key parameters of the DBSCAN algorithm for flight trajectory optimization.

In the following part of this section, we explain these concepts in greater detail.

3.2.2.1 ε (Epsilon): The Neighbourhood Radius

The ε parameter is fundamental to the DBSCAN algorithm as it defines the radius around each point within which other points must fall to be considered part of the same cluster. This radius, often referred to as the neighbourhood radius, controls the local density criterion that DBSCAN uses to identify clusters.

- Small ε Value: If ε is set too small, many points will not have enough neighbours within this radius, resulting in many small, fragmented clusters or points being classified as noise.
- Large ε Value: Conversely, if ε is set too large, distinct clusters may be merged together into a single cluster, losing the algorithm's ability to differentiate between densely packed regions.

Choosing an appropriate value for ε is critical and is often done by examining a k-distance plot (typically for k equal to MinPts), which shows the distance to the k-th nearest neighbour for each point.

3.2.2.2 MinPts: The Minimum Number of Points

MinPts, the minimum number of points required within an ε-neighbourhood to form a dense region, is the second key parameter of DBSCAN. This parameter helps determine whether a region is dense enough to be considered a cluster.

- Low MinPts: Setting MinPts too low may result in many small clusters, capturing noise as part of clusters.
- High MinPts: Setting MinPts too high might miss smaller clusters and result in more points being classified as noise.

3.2.2.3 Core Points, Border Points, and Noise

DBSCAN categorizes points into three types, which determine how they contribute to cluster formation:

- Core Points: These are points that have at least MinPts within their ε-neighbourhood. Core points are the central elements of clusters, around which other points are grouped.
- Border Points: Border points are those that have fewer than MinPts within their ε-neighbourhood but are within the ε-neighbourhood of a core point. Border points are included in clusters, but they do not serve as seeds for expanding the cluster.
- Noise Points: Noise points are those that are neither core nor border points. These points do not belong to any cluster and are considered outliers.

In the following figure, we show an overview of the clustering results based on the dataset on EHRD provided by Task 2.1

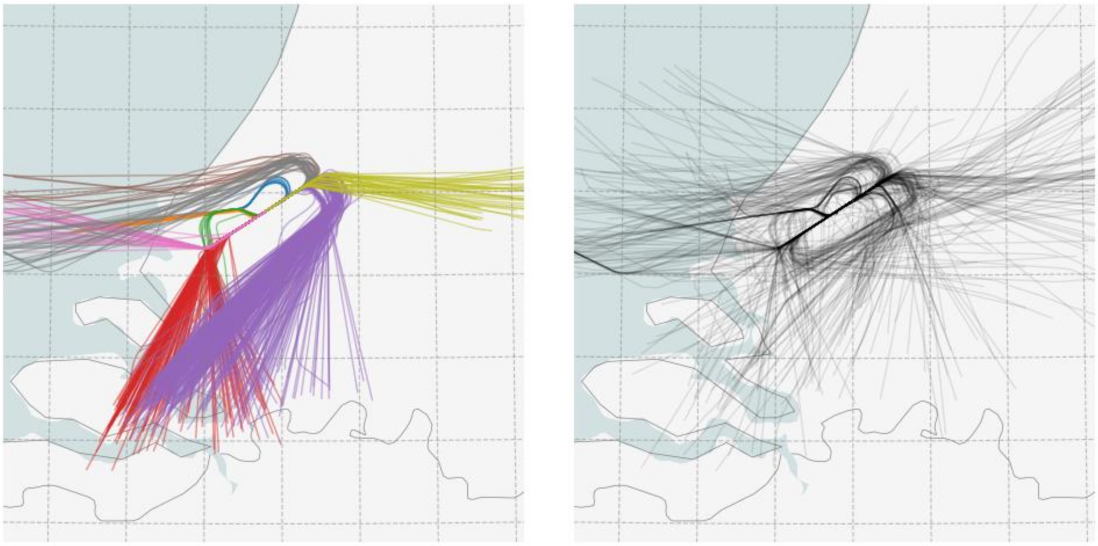

Figure 32 - An overview of the trajectory clusters and noise-point trajectories for arrivals and departures at EHRD.

3.2.3 Analysing the statistical dispersion key parameters in clusters

Furthermore, we analyse the statistical dispersion of two key parameters, airspeed and vertical rate, at different altitudes within each cluster. By examining these parameters across different altitude bands segmented in increments of 1,000 feet, we can model how airspeed and vertical rate vary during different phases of flight. This analysis is crucial for understanding the consistency of flight operations and identifying the range of parameters that could impact noise and air quality assessments.

3.3 RESULTS

3.3.1 Exploratory data analysis

The following figure shows the separation and illustration of arrival and departure flights at the EHRD, using a small test dataset directly downloaded from the OpenSky Network.

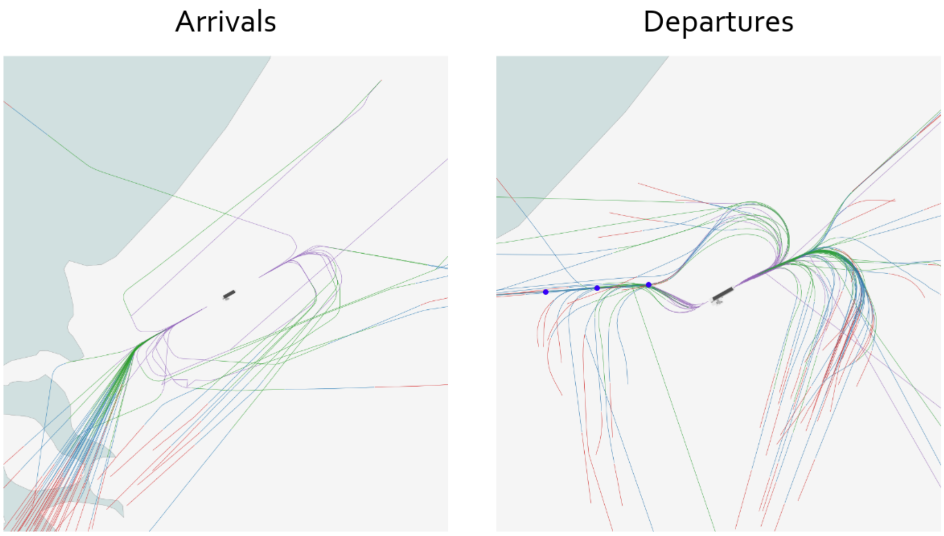

Figure 33 - The arrival and departure flights at EHRD.

We explored how the departures match with the departure procedures. The figure below illustrates three different departure procedures for three runway configurations at EHRD. We can see the alignments and mis-alignments between the flight trajectories and the standard instrument departure procedures.

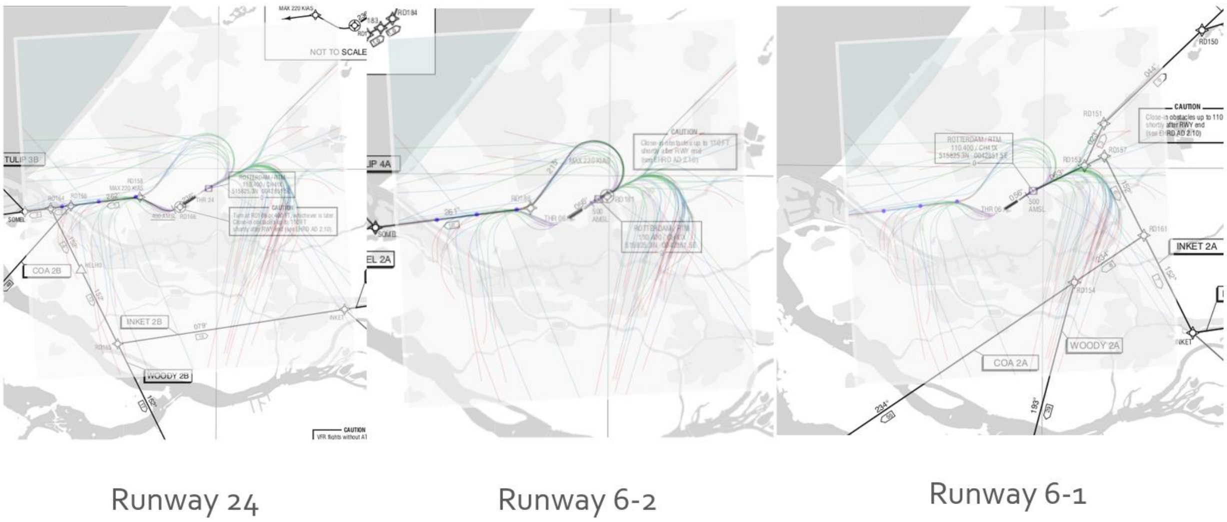

Figure 34 - Matching the departure flights with the standard instrument departure procedures at EHRD.

3.3.2 Clustering of flight trajectories

In this section, we show the clustering results generated from all the airports.

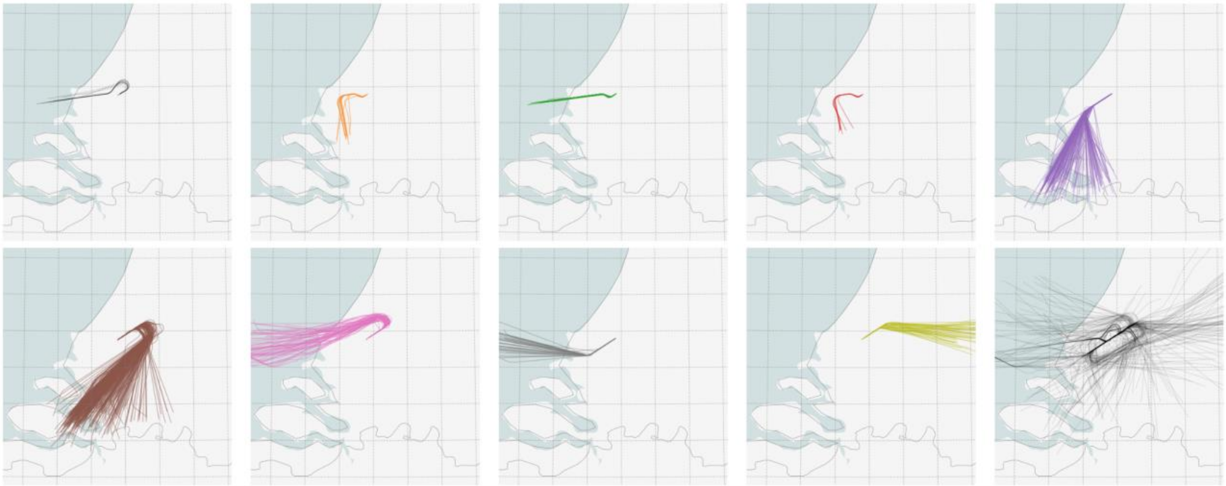

Figure 35 - Clusters extracted from the EHRD dataset.

Figure 36 - Clusters extracted from the EHAM dataset. We can observe that the vectoring is very widely adopted at EHAM due to the large dispersion of the ground track, especially for the arrival flights.

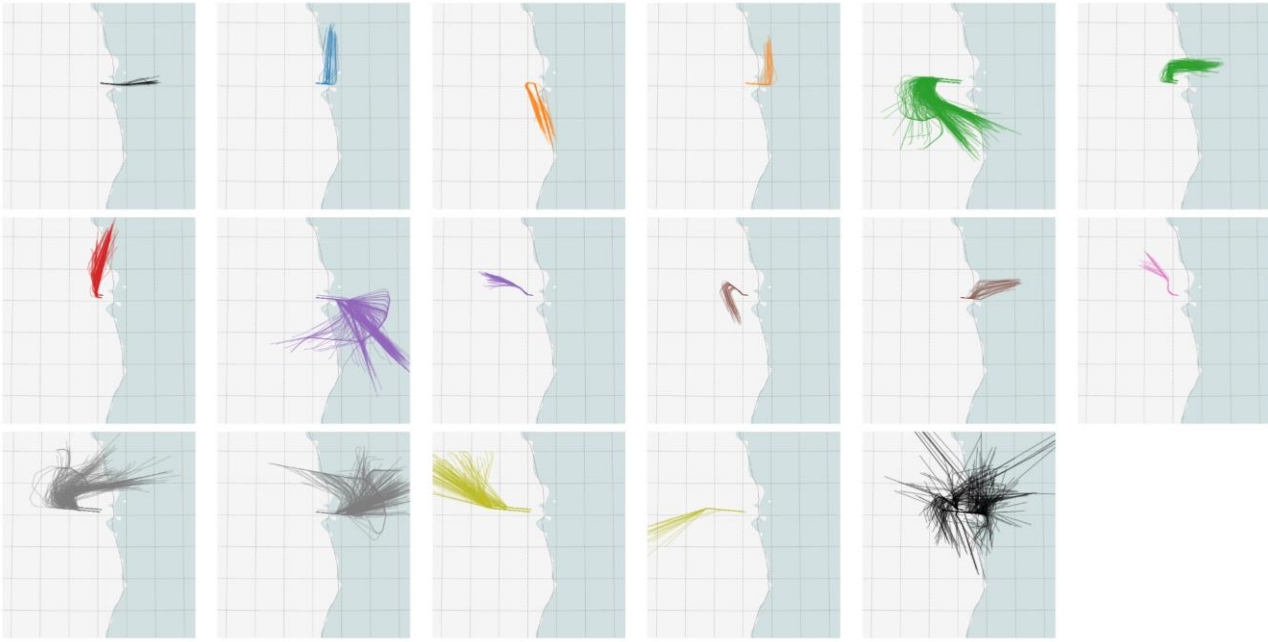

Figure 37 - Clusters extracted from the EIDW dataset. We can observe point-merge patterns for the arrival flights at EIDW.

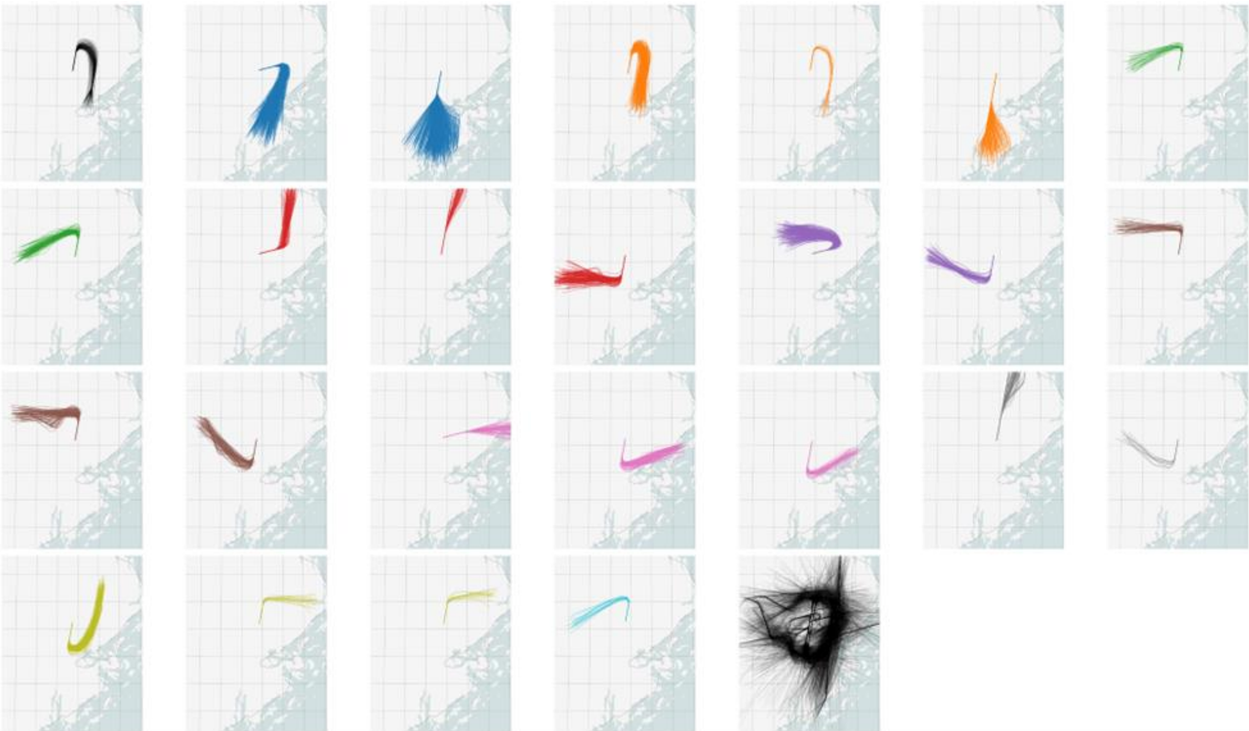

Figure 38 - Clusters extracted from the ESSA dataset.

Figure 39 - Reduced clusters extracted from the ESSA dataset. In this figure, we have reduced the number of clusters, where arrivals at the parallel runways are considered in the same clusters.

3.3.3 Estimating the centroids of the clusters

Based on the clusters obtained from the trajectory data, we can further obtain the centroid of the flights. In the following Figure 40, we can see the centroid being estimated for departure and arrival clusters for EHRD clusters.

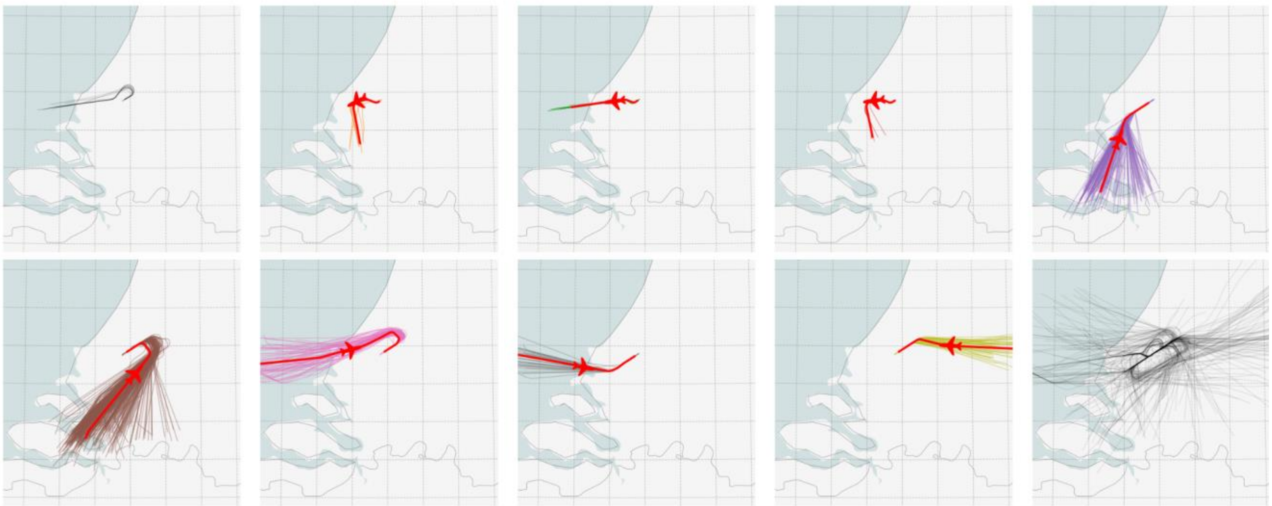

Figure 40 - Estimation of centroid flight path, based on the DBSCAN algorithm.

In this analysis, we can see that the departure strictly follows the standard instrument departure route, while arrivals divert a lot from standard terminal arrival routes due to the vectoring. These path deviations should be considered in the trajectory generation model in a later stage of this task using machine learning based models to generate trajectories.

3.3.4 Analysing the dispersion of the speed and vertical rate at different altitude intervals

Based on the EHRD dataset, we can further analyse the dispersion of key parameters, airspeed and vertical rate for arrival and departure flights in each of the clusters.

In the following Figure 41, we can see the distribution of the vertical rates at different altitudes for all departure flights. The vertical rate is grouped per 1000 ft of altitude. We can see that there are quite some differences between different clusters. For example, cluster=0 and cluster=2 have large differences in vertical rate at different altitudes, while the other two have less variation.

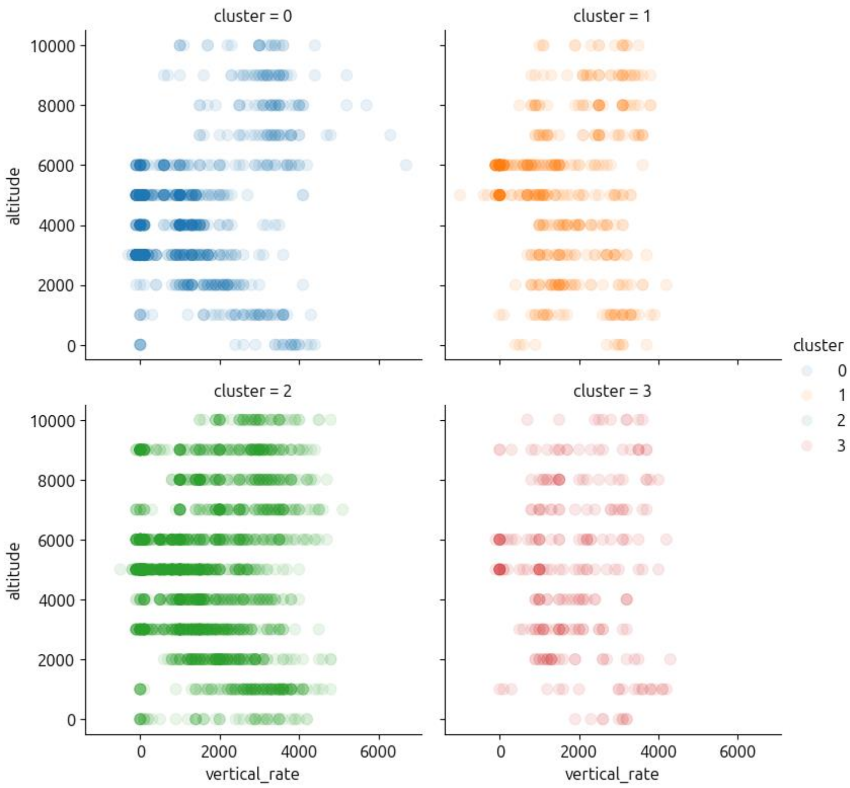

Figure 41 - Dispersion of vertical rate among departure clusters in the EHRD dataset.

In the following Figure 42, we can see the vertical rate profiles and their dispersion for all arrival flights. Unlike large dispersion from the ground tracks, we can see a more consistent pattern for different altitudes among different clusters. We can observe a gradual decrease in the descent rate with the lowering of the altitudes.

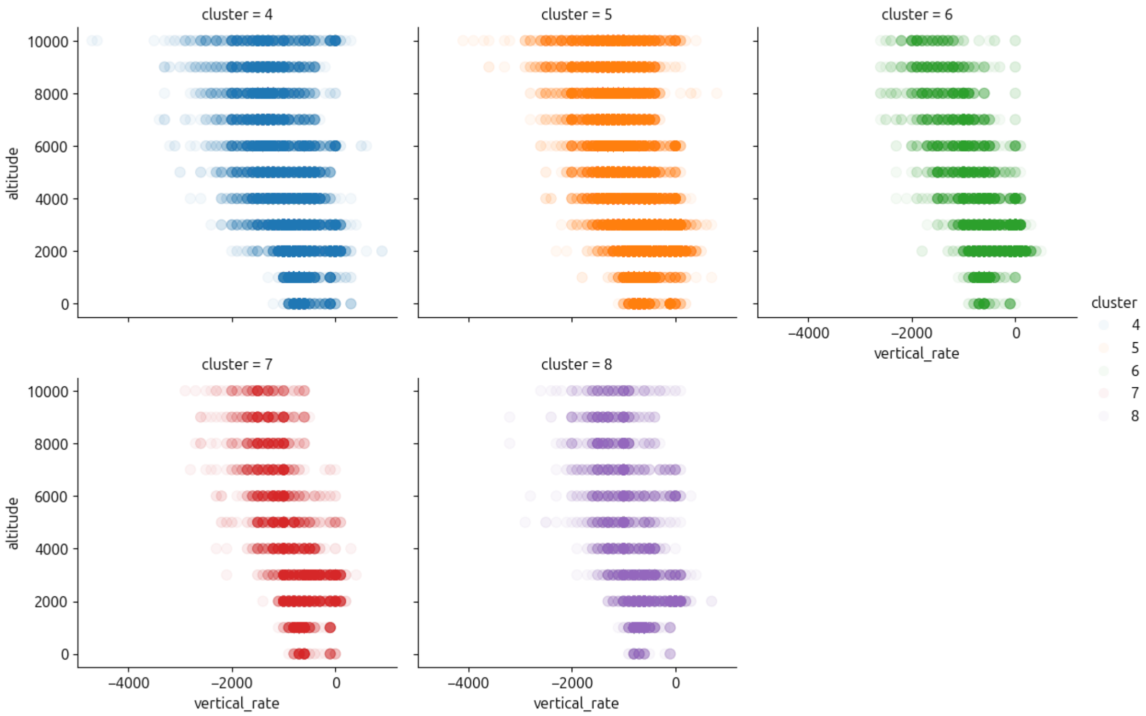

Figure 42 - Dispersion of vertical rate among arrival clusters in the EHRD dataset.

Next, we can take the same approach to analyse the speed profile of the flights from different clusters. In Figure 43, the dispersion of the indicated airspeed at different 1000 ft altitude intervals. Speed from all four departure clusters are shown. Two observations can be made here, 1) the nominal speed and dispersions from different clusters are similar, and 2) there are some data anomalies in some of the trajectories (very low speed at high altitude or very high speed at low altitude). This anomaly is likely caused by the errors contained in the Mode-S downlink data or inference/decoding errors.

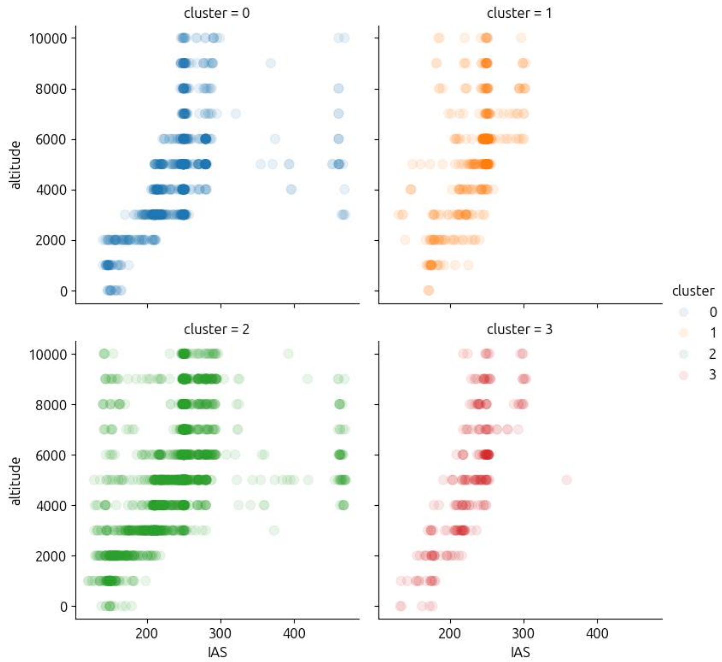

Figure 43 - Dispersion of the indicated airspeed among departure cluster in the EHRD dataset.

Similar analyses are performed regarding the arrival clusters. In Figure 44, we can see the dispersion of the indicated airspeed among all five clusters. For the first two clusters, we observe a larger dispersion, indicating more variability in the airspeed. This is also correlated with the more dispersed ground tracks.

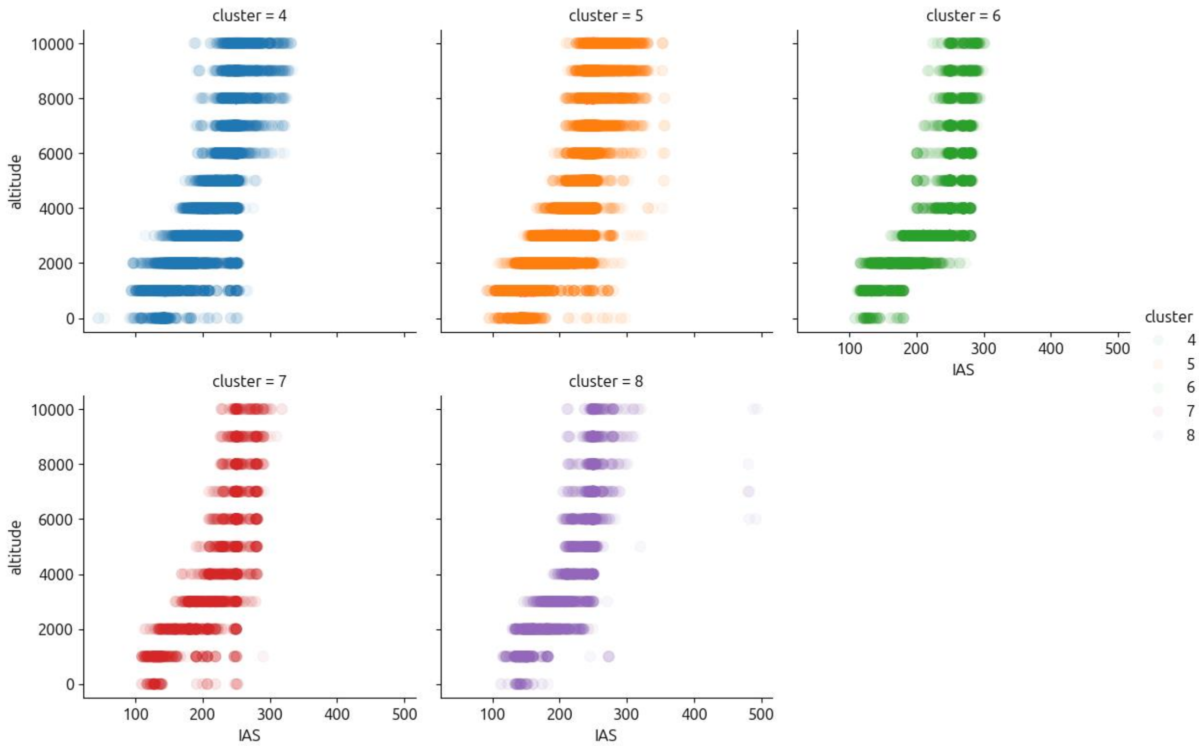

Figure 44 - Dispersion of the indicated airspeed among arrival cluster in the EHRD dataset.

3.4 CONCLUSIONS

The analysis conducted on the flight trajectory data from the airport data from T2.1 has provided critical insights into the operational patterns and variability of departures and arrivals. The study focused on exploratory data analysis, trajectory clustering, estimation of cluster centroids, and dispersion analysis of key flight parameters such as airspeed and vertical rate.

The initial exploratory data analysis showed clear patterns in how departures align with standard instrument departure procedures at EHRD. Visualizations showed both alignments and deviations in flight paths, highlighting areas where operational consistency is strong and where variations occur.

The application of clustering algorithms, DBSCAN, enabled the grouping of flight trajectories into distinct clusters based on their spatial and temporal characteristics. The analysis across multiple airports, including EHRD, EHAM, EIDW, and ESSA, demonstrated varying levels of trajectory dispersion and clustering behaviour, reflecting differences in airport operations, traffic management strategies, and environmental factors. For example, at EHRD, the departure paths closely followed the standard departure routes, whereas arrival paths showed significant deviations due to vectoring procedures. These centroid paths serve as a baseline for modelling and simulating future flight trajectories.

Furthermore, By estimating the centroids of the identified clusters, the study provided a representative flight path for each cluster. These trajectories can be a simple input that can be used as input for later work packages.

The analysis of airspeed and vertical rate dispersion within altitude-based clusters highlighted the variability in flight operations. Significant differences in vertical rate were observed across different departure clusters, particularly at varying altitude levels. For arrivals, while ground tracks were more dispersed, the vertical rate profiles showed a more consistent pattern, indicating a gradual descent with altitude. Airspeed analysis revealed both nominal speeds and anomalies, suggesting potential data issues that need to be addressed in future studies.

The next phase of the project will focus on the generation of flight trajectories in scenarios where real-world flight data is unavailable. This will involve the development of a generative model that can simulate and generate realistic flight paths based on historical data. By leveraging the patterns and clusters identified in the previous analyses, this model will be designed to reproduce typical flight behaviour, while also accounting for the variations and anomalies observed in the existing dataset.

To build this generative model, historical flight data from all airports of interest will be used. Specifically, the clustered trajectories generated earlier. Machine learning techniques, such as auto-encoder neural networks, will be employed to capture the complex dynamics of the flight pattern.

The model will be iteratively refined to ensure that the generated trajectories accurately reflect the operational realities observed in the historical data. This synthetic data will then be used to inform noise and air quality impact assessments, enabling more robust planning and decision-making for future airport operations.

References

[1] EASA - ANP, "Aircraft Noise and Performance (ANP) Data," 2023. [Online]. Available: https://www.easa.europa.eu/en/domains/environment/policy-support-and-research/aircraft-noise-and-performance-anp-data. [Accessed 19 July 2023].

[2] ECAC, "Doc 29 4th Edition: Report on Standard Method of Computing Noise Contours around Civil Airport - Volume 2: Technical Guide," ECAC, Neuilly-sur-Seine, 2016b.

[3] OpenSky Network, "OpenSky aircraft database," [Online]. Available: https://opensky-network.org/aircraft-database. [Accessed July 2023].

[4] "Airlinerlist.com," [Online]. Available: https://plane-list.net/. [Accessed April 2023].

[5] "ICAO Aircraft Engine Emissions Databank," June 2023. [Online]. Available: https://www.easa.europa.eu/en/domains/environment/icao-aircraft-engine-emissions-databank.

[6] EUROCONTROL, "IMPACT: Integrated aircraft noise and emissions modelling platform," 2023b. [Online]. Available: https://www.eurocontrol.int/platform/integrated-aircraft-noise-and-emissions-modelling-platform. [Accessed 11 September 2023].

[7] M. Pretto, P. Giannattasio, M. De Gennaro, A. Zanon e H. K¨uhnelt, «Web data for computing real-world noise from civil aviation,» *Transportation Research Part D: Transport and Environment,* vol. 69, pp. 224-249, 2019.

[8] O. Xavier, "Traffic, a toolbox for processing and analysing air traffic data," *Journal of Open Source,* vol. 4, pp. 1518, 1-3, 2019.

[9] "OurAirports," [Online]. Available: https://ourairports.com/. [Accessed April 2023].

[10] “Rzjets,” [Online]. Available: https://rzjets.net/aircraft/. [Accessed June 2024].

[11] M. Pretto, L. Dorbolò and P. Giannattasio, “[Poster] Exploiting high-resolution ADS-B data for flight operation reconstruction towards environmental impact assessment,” *Journal of Open Aviation Science,* vol. 1, no. 2, 2023.

[12] "scipy.signal.butter — SciPy v1.11.4 Manual," [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.butter.html. [Accessed June 2023].

[13] M. Pretto, P. Giannattasio and M. De Gennaro, “Mixed analysis-synthesis approach for estimating airport noise from civil air traffic,” *Transportation Research Part D: Transport and Environment,* vol. 106, p. 103248, 2022.

[14] M. Pretto, L. Dorbolò, P. Giannattasio and A. Zanon, “Aircraft operation reconstruction and airport noise prediction from high-resolution flight tracking data,” *Transportation Research Part D: Transport and Environment,* 2024 (Under peer review).

[15] G. Koudis, S. Hu, A. Majumdar, R. Jones and M. Stettler, "Airport emissions reductions from reduced thrust takeoff operations," *Transportation Research Part D,* vol. 52, pp. 15-28, 2017.

[16] D. Wales and D. Jonathan, “Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms,” *The Journal of Physical Chemestry A,* vol. 101, no. 28, pp. 5111-5116, 1997.

[17] D. DuBois and G. C. Paynter, “"Fuel Flow Method2" for Estimating Aircraft Emissions,” *Journal of Aerospace,* vol. 115, pp. 1-14, 2006.

[18] ICAO, “Doc 9889, Airport Air Quality Manual, Second Edition,” INTERNATIONAL CIVIL AVIATION ORGANIZATION, 2020.

[19] B. Y. Kim, G. G. Fleming, J. J. Lee, I. A. Waitz, J.-P. Clarke, S. Balasubramanian, A. Malwitz, K. Klima, M. Locke, C. A. Holsclaw, L. Q. Maurice and L. M. Gupta, “System for assessing Aviation’s Global Emissions (SAGE), Part 1: Model description and inventory results,” *Transportation Research Part D: Transport and Environment,* vol. 12, no. 5, pp. 325-316, 2007.

[20] EUROCONTROL, “Data | Aviation Intelligence Unit Portal,” [Online]. Available: https://ansperformance.eu/data/. [Accessed June 2023].

[21] Georgia Institute of Technology, "Project 045 Takeoff/Climb Analysis to Support AEDT Aircraft Performance Model (APM) Development," 2020.

[22] M. Ester, H.-P. Kriegel, J. Sander e X. Xu, «A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,» *Data Mining and Knowledge Discovery,* vol. 2, pp. 226-231, 1998.