Data Driven Approach to Noise Abatement Procedure Classification at Amsterdam Airport Schiphol

Albert Chou

TU Delft

knowledge & development centre
Mainport Schiphol

Data Driven Approach to Noise Abatement Procedure Classification at Amsterdam Airport Schiphol

Thesis report

by

Albert Chou

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on July 16th, 2026 at 12:45

Thesis committee:

Chair:	Dr. M. J. Ribeiro
Supervisors:	Dr.ir. J. Ellerbroek
	ir. M. Zorgdrager
External examiner:	Dr. J. Sun
Place:	Faculty of Aerospace Engineering, Delft
Project Duration:	September, 2025 - July, 2026
Student number:	5010292

Cover Photo: Runway 27 Landing | Javier Crespo (Fellow KDC Student)

An electronic version of this thesis is available at `https://repository.tudelft.nl/`.

Faculty of Aerospace Engineering · Delft University of Technology

Preface

This master's thesis marks the end of my Master's degree in Aerospace Engineering at TU Delft. The research was carried out in collaboration with Royal Schiphol Group under the branch of the Knowledge Development Centre (KDC), with supervision from Joost Ellerbroek and Maarten Zorgdrager.

During my master's program, I was initially uncertain about the type of topic I wanted to pursue. This changed during my internship at Luchtverkeersleiding Nederland (LVNL), where I had the opportunity to interact with multiple master's thesis students at iLabs working on their thesis under the branch of KDC. Through these interactions, I realised that I wanted to work on a topic that involves real-world data and provides the opportunity to apply it to the knowledge that I have obtained throughout my time at TU Delft. I am glad I found this topic on noise abatement procedure classification, as it combines real-world data and theoretical understanding.

At the beginning of the thesis, the differences between NADP procedures and associated regulations felt complex and difficult to grasp. However, I can now confidently say that I can explain them.

First of all, I would like to thank my two supervisors, Joost Ellerbroek and Maarten Zorgdrager, for all their valuable feedback and guidance, and of course for bearing through the time spent during my chaotic progress meetings.

I would like to thank all of the people that I have encountered at iLabs, whether during coffee breaks or by helping each other stay sane throughout the process of the internship and thesis. I would also like to thank Ferdinand Dijkstra for assisting me in finding my thesis topic despite him not being available to supervise me.

Finally, I would like to thank all my friends who supported me throughout this process, in particular my teammates and study mates from Punch at EWI. The coffee breaks and the great time on the field allowed me to reset my mind and not constantly go into stress mode.

Albert Chou
Delft, June 2026

Contents

Nomenclature

List of Abbreviations

ACMS	Aircraft Condition Monitoring System
ADS-B	Automatic Dependent Surveillance–Broadcast
AEDT	Aviation Environmental Design Tool
AGL	Above Ground Level
CAS	Calibrated Airspeed
CDA	Continuous Descent Approach (subset of CDO)
CDO	Continuous Descent Operation
ECAC	European Civil Aviation Conference
EHS	Enhanced Mode-S
FAA	Federal Aviation Administration
FOQA	Flight Operations Quality Assurance
FPP	Fixed-Point Profile
GS	Ground Speed
HMM	Hidden Markov Model
IAS	Indicated Airspeed
ICAO	International Civil Aviation Organization
LSTM	Long Short-Term Memory
MHG	Maximum Amount of Noise
NADP	Noise Abatement Departure Procedure
NAP	Noise Abatement Procedure
NLR	Royal Netherlands Aerospace Centre
NNHS	Nieuwe Normen- en Handhavingsstelsel
RMI	Regeling milieu-informatie luchthaven Schiphol
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RoC	Rate of Climb
SKE	Specific Kinetic Energy
SPE	Specific Potential Energy
TAS	True Airspeed
TEM	Total Energy Model
TSE	Total Specific Energy
VEMMIS	LVNL Surveillance Data System

List of Symbols

Δ	Change in variable
γ	Flight path angle
μ	Mean value
σ	Standard deviation
D	Drag force
ds	Distance increment
dt	Time step
g_0	Gravitational acceleration ($9.81\,\mathrm{m/s^2}$)
h	Altitude
i	Trajectory index
m	Aircraft mass
n	Number of samples
s_k	Normalised cluster score
T	Thrust force
V	Velocity
V_2	Take-off safety speed

Part I

Scientific Article

Data Driven Approach to Noise Abatement Procedure Classification at Amsterdam Airport Schiphol

A. Chou[*]
Supervised by Dr.ir. J. Ellerbroek[*] and ir. M. Zorgdrager[†]
[*]*Control and Operations, Faculty of Aerospace Engineering, Delft University of Technology, the Netherlands*
[†] *Royal Schiphol Group*

Accurate identification of aircraft noise abatement procedures (NAPs) is essential for reliable environmental impact assessment and noise modelling. However, current analyses of the implementation of noise abatement procedures are mainly based on interviews with airlines. Previous studies establish methods to model NADP procedures and estimate flap settings. However, only limited studies have proposed methods for classifying noise abatement procedures. This study presents a data-driven approach for the classification of Noise Abatement Departure Procedure (NADP), Continuous Descent Approaches (CDA) and Reduced Flap Setting Approach using flight trajectory data. The framework integrates energy-based performance, flight performance indicators and machine learning techniques to classify flight trajectories into specific noise abatement procedures. The methodology is applied to operations at Amsterdam Airport Schiphol, evaluated on a dataset comprising over 19,000 inbound and over 19,000 outbound flight data in May 2025, alongside a machine learning training dataset of over 7,000 inbound flights. The results indicate that the developed framework is capable of classifying flight trajectories from enhanced mode-S and radar data to specific noise abatement procedures. However, the developed framework is constrained mainly by data resolution and feature sensitivity. Future work should focus on improving temporal resolution, expanding the dataset and incorporating event-based validation for enhancing model reliability.

I. Introduction

Since the decline of air traffic due to COVID-19, air traffic at Amsterdam Airport Schiphol has been gradually recovering towards pre-pandemic levels. Air traffic has shown a steady annual increase; in 2024, total traffic movements rose by 7.2% [1]. Despite the increase in traffic, Amsterdam Airport Schiphol operates under strict environmental constraints in terms of aircraft noise emissions under the Nieuwe Normen-en Handhanvingsstelsel (NNHS) framework. Within this framework, noise impact is mitigated through the use of a preferential runway system, in which aircraft are directed to specific runways that minimises noise exposure over densely populated areas.

In addition, several noise constraints are imposed. The Maximum Amount of Noise (MHG), representing the upper bound on total annual noise production is limited to 62.8

dB(A) in 2025 [2]. Moreover, daytime exposure is limited to a maximum of 12,000 residential compounds above 58 dB(A) and 186,000 people above 48 dB(A). Whereas during nighttime, stricter limits apply with a maximum of 12,800 residential compounds above 48 dB(A) and 50,000 people above 40 dB(A) [2].

Therefore, the growth in total traffic movement further increases the importance of accurate classification of operational procedures, particularly Noise Abatement Procedures (NAPs), which are designed to minimise noise impact on the surrounding communities.

Annually, the Royal Schiphol Group is required to publish documents outlining Amsterdam Airport Schiphol's usage forecast for the upcoming year ("Gebruiksprognose Schiphol" in Dutch) [2] and an evaluation of the usage forecast. Within these documents, three main NAPs are defined, notably Noise Abatement Departure Procedure (NADP), Continuous Descent Approach (CDA), and Reduced-Flap Setting Approach. The usage forecast mentions the percentages of aircraft operating under each NAP. The percentages are used to simulate the expected impact of Schiphol operations on noise. However, these percentages are based on interviews with airlines and the application of the Pareto principle, which may lead to an over or underestimation of noise impact.

For the noise simulation, a set of fixed-point profiles (FPPs) for various aircraft operations, including NAPs are generated by the Royal Netherlands Aerospace Centre (NLR) [3]. A fixed-point profile is a representation of an aircraft trajectory using a sequence of predefined discrete points along a flight path. Each point specifies parameters such as altitude, distance, true airspeed (TAS) and estimated power setting. These profiles enable noise modelling based on aircraft type and operational procedures using European Civil Aviation Conference's (ECAC) Document 29 noise modelling framework [4], particularly when detailed aerodynamic and flight performance parameters are unavailable.

The origin of these fixed point profiles lies in the ECAC document 29 performance modelling framework, where aircraft trajectories are first constructed using the Aircraft Noise and Performance database. This database provides generic performance data and trajectories are generated by modelling aircraft motion along procedural segments such as climb and acceleration. These model trajectories are subsequently calibrated using radar observations, such that the resulting FPPs match the median behaviour of observed flights at Amsterdam Airport Schiphol [5]. As a result, the generated fixed-point profiles represent typical operational behaviour. However, using these fixed-point profiles limits the ability to model and classify individual flights according to their actual NAP operation.

While the Aviation Environmental Design Tool (AEDT) is widely used by American airports to model and evaluate noise-abatement procedures, it allows for more customisation and variability in modelling various NAPs [6]. However, the AEDT is primarily designed for the Federal Aviation Administration (FAA) regulatory requirements. This limits its applicability to non-FAA regulatory context, including European airports such as Amsterdam Airport Schiphol.

This research aims to develop a data-driven method to estimate Noise Abatement Procedures (NAPs) usage at Amsterdam Airport Schiphol. By moving beyond interview-based assessments that apply the Pareto principle, this approach could provide a more accurate estimation of

Amsterdam Airport Schiphol's NAPs usage.

The structure of this work is as follows. Firstly, section II presents related background information on NAPs and existing modelling and classification methods for air traffic trajectories and NAPs. In section III, the data source and obtainable data are illustrated. Section IV presents the proposed model for this study. Section V, outlines the hypothesis and experimental setup. The results are presented in section VI, and subsequently discussed in relation to the hypothesis in section VII. Finally, section VIII provides the conclusions of the study.

II. Background

This section provides background information on relevant research for the proposed methodology. Specific information on NAPs, Dutch regulations on NAPs, NADP modelling, classification and aircraft configuration estimation.

A. Noise Abatement Procedure (NAPs)

NAPs are operational measures defined by the International Civil Aviation Organisation (ICAO) to reduce aircraft noise exposure in the vicinity of aerodromes. Among these, ICAO distinguishes NADPs and CDOs [7].

NADPs are further categorised into two principal types: NADP1 and NADP2, which differ primarily in their noise-mitigation objectives. NADP1 is designed to minimise noise impact in areas close to the departure runway by prioritising a steeper initial climb before thrust reduction. In contrast, NADP2 aims to reduce noise exposure at greater distances by allowing earlier acceleration and configuration changes [7]. In addition, NADP2 generally results in lower fuel consumption compared to NADP1 [6].

Based on airline interviews and flight-track modelling, ICAO identified a range of operational variations for these procedures, resulting in six NADP1 and eight NADP2 profiles [8]. Each variation differs based on the thrust cutback altitude, acceleration altitude and flap retraction timing. For example, two profiles can have identical thrust cutback and acceleration altitude but the flap retraction timing differs, leading to operationally different NADP profiles. This variability highlights the operational flexibility inherent in NADPs and motivates further analysis of their performance and noise implications. Examples of NADP1 and NADP2 profiles are shown in Figure 1 and Figure 2 respectively.

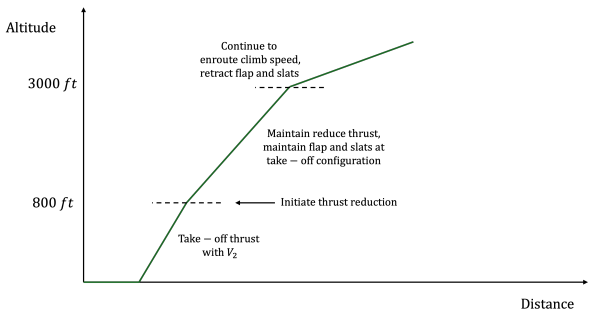


Fig. 1 Example of NADP1 procedures Adapted from ICAO[7]

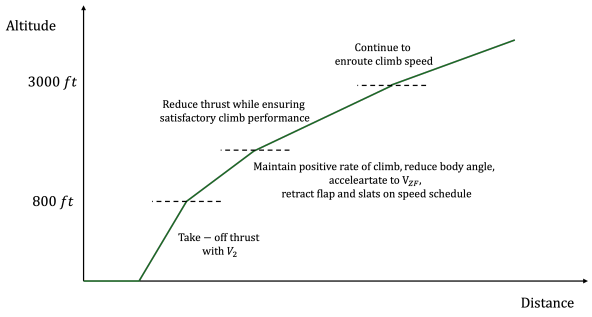


Fig. 2 Example of NADP2 procedures Adapted from ICAO[7]

For noise abatement during descent and approach, ICAO primarily provides the framework of Continuous

Descent Operations (CDO). CDO refers to arrival operations in which aircraft descend continuously with minimal engine thrust and low aerodynamic drag, ideally initiated from the top of descent and maintained until the final approach phase [9]. By avoiding level segments, CDOs reduce thrust variations and associated noise emissions. CDA are subsets of CDO, which focus on maintaining a continuous descent during the approach phase.

In addition to descent profile management, reduced-flap setting approaches are employed as a complementary measure to mitigate noise. Although ICAO does not provide a strict definition for such approaches, it recommends delaying the deployment of the full landing configuration until close to the runway threshold to limit aerodynamic drag and engine thrust during earlier approach segments [7, 10].

B. Dutch Regulations on NAPs

The Dutch government has particular environmental regulations for Amsterdam Airport Schiphol operations, which are depicted in the Schiphol Airport Environmental Information Regulations (in Dutch: Regeling milieu-informatie luchthaven Schiphol (RMI)) [11]. Within the RMI, additional operational classifications are specified for the assessment and reporting of Schiphol Airport operations.

For NADP procedures, NLR defined 4 different NADP procedures that occur at Amsterdam Airport Schiphol, consisting of one NADP1 profile and three NADP2 variants [3, 5]. For the NADP1 profile, the thrust cut-back altitude is assumed to be at 1500 ft, while the acceleration altitude is set at 3000 ft. In contrast, for the three NADP2 profiles, the acceleration altitude and the thrust cut-back altitude assume to occur simultaneously at the altitudes of 800 ft, 1000 ft and 1500 ft.

For descent operations, flights are compliant with a CDA under the RMI definition when all conditions are satisfied. Firstly, the aircraft maintains a continuous descent profile with no level segments below 3500 ft. Secondly, the descent profile is above an arbitrary 2.5 ° glideslope datum line [11].

Flights are compliant with a reduced-flap setting approach when an aircraft performs the final approach without deploying the full landing flap configuration.

C. Noise Abatement Departure Procedure Modelling and Classification

As of the current state of practice, accurate classification of NADP profiles is highly reliant on high-fidelity environmental simulation tools such as AEDT. These tools enable the generation of multiple trajectory variations for a single NADP profile by varying key parameters, including take-off weight, thrust reduction level, and energy share percentage [12]. While this allows for trajectory similarity comparisons, it also highlights one of the core challenges, the same procedural definition can result in a wide range of trajectory shapes.

This variability, combined with the limited availability of reference trajectories in the Schiphol case study, makes it difficult to robustly classify observed flight paths based solely on similarity metrics. Furthermore, such simulation-based approaches are not easily accessible and require significant computational and modelling effort. As a result, trajectory similarity scoring alone becomes insufficient for reliable NADP classification, unless complemented by more interpretable approaches such as event-based

detection and explicit NADP modelling.

To allow the classification of flight trajectories to NADP profiles with relation to the FPP obtained from NLR [3], further modelling and classification techniques would be required for robust NADP detection and classification. Several studies have specifically addressed methods for modelling and classification of NADPs. Lim [6], Behere [12, 13] and Bhanpato [14] each developed approaches to modelling and classification of NADP profiles.

Lim [6] investigated approaches to improve the modelling of NADP profiles to allow better representation to real world operations in the AEDT. Lim proposed an energy-based modelling metric to reduce the variability of aircraft weight, thrust and atmospheric conditions.

Building on this work, Behere et al. applied consensus clustering techniques to cluster flights into similar operational characteristics to allow identification of representative profiles [13]. Their method employs feeding flight trajectories to the AEDT to simulate noise, fuel burn and emissions.

Bhanpato further developed the NADP classification model by combining trajectory similarity classification with a layer of trajectory clustering [14]. The trajectory clustering incorporates distance, altitude, and rate of climb to cluster similar trajectories together. With the combination of both metrics, a consensus NADP classification is obtained by mapping the trajectory to the closest median reference trajectory.

Further extending on this approach, Behere et al. proposed a method to map time-series take-off trajectories to a parametric representation, allowing flight profile classification and inversely map events such as acceleration altitude [12].

Together, these studies highlight the potential of trajectory-based modelling and classification methods to characterise NADP behaviour beyond nominal procedural definitions.

D. Aircraft Configuration Estimation

In addition to the classification and modelling of NADP profiles, another main uncertainty in the classification of NAPs is the configuration of aircraft for the reduced-flap setting approach classification. The estimated configuration of the aircraft allows a binary classification between reduced-flap and full-flap approach. Nichols [15], Jarry [16] and Bhanpato [17] have each investigated different data-driven approaches to estimate aircraft state using machine learning algorithms.

Each algorithm employs the use of Flight Operations Quality Assurance (FOQA) data to train the various supervised learning models. Nichols investigated linear regression, random forest regression and random forest classification for flap setting estimation [15]. Both Jarry and Bhanpato developed models using the Long-Short Term Memory (LSTM) neural network algorithm [16]. However, Bhanpato further compared the use of a stochastic model, the Hidden Markov Model (HMM) to compare the performance of flap configuration estimation [17].

III. Data Sources

This research utilises surveillance data collected from radar tracks and Enhanced Mode-S (EHS) transponders from VEMMIS. VEMMIS is a database that stores all flight information obtained from the Netherlands' air navigation service provider Luchtverkeersleiding Nederland (LVNL). The radar tracks provide from VEMMIS includes positional and ground-based kinematic information about the

aircraft, such as latitude, longitude, ground speed (GS) and barometric altitude derived from secondary surveillance radar. EHS data provides specific aircraft states that are sent from the aircraft transponder. Retrieved aircraft states include True Airspeed (TAS), Indicated Airspeed (IAS), pressure altitude and rate of climb (RoC). Details of the data structure of the coupled radar track and EHS data are shown in Table 1.

Table 1 Data Structure of the Coupled Radar and EHS Data

Aircraft State	Type	Unit
Flight ID	integer	-
Callsign	string	-
Aircraft Type	string	-
Departure	string	-
Arrival	string	-
Time	timestamp	-
Latitude	float	°
Longitude	float	°
Barometric Altitude	float	ft
Ground Speed	float	kts
True Airspeed	float	kts
Indicated Airspeed	float	kts
Rate of Climb	float	ft/min

For the training and validation of flap configuration estimation, Aircraft Condition and Monitoring System (ACMS) data is collected. The ACMS data is obtained from an airline in which only descent operation is provided. ACMS data is representative of FOQA data described in section II, in which the data provides specific aircraft states obtained from onboard aircraft sensors. In addition to aircraft states, flap setting, and landing gear state are provided. Details of the data structure of the ACMS data are shown in Table 2.

Table 2 Data Structure of the ACMS Data

Aircraft State	Type	Unit
Flight ID	string	-
Aircraft Type	string	-
Time	timestamp	-
Latitude	float	°
Longitude	float	°
Barometric Altitude	float	ft
Ground Speed	float	kts
True Airspeed	float	kts
Indicated Airspeed	float	kts
Flap Setting	integer	-

The obtained flap setting values in the provided ACMS differ depending on the aircraft manufacturing company. For Airbus and Embraer aircraft, flap settings are given in discrete flap classes (e.g. 0, 1, 2, etc.). In contrast to that, for Boeing aircraft, flap settings are specified in flap deflection angles such as 0 °, 5 °, 10 °, 15 ° etc.

IV. Methodology

This study develops a trajectory-based classification framework that maps aircraft trajectories to noise abatement procedures. The framework consists of three main components:

1) **Data Processing**: Cleaning, filtering and standardising trajectory data.
2) **Outbound Classifier**: Classification of NADP procedures during outbound flights.
3) **Inbound Classifier**: Classification of CDA compliance and reduced-flap setting approaches during inbound flights.

The outbound and inbound classifiers are designed as independent components within the framework. While they rely on different modelling approaches, both components follow

a similar structure by extracting trajectory features, detecting operational events and combining multiple indicators to perform a classification.

The outputs of these classifiers are integrated at the final stage, where the two components jointly serve as inputs to the annual noise load estimation and realisation at Amsterdam Airport Schiphol.

The overall framework and the interaction between the components and the sub-components is illustrated in Figure 3. Each component is described in detail in the following subsections.

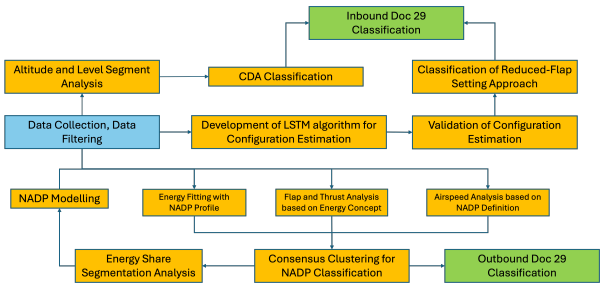


Fig. 3 Model Methodology Block Diagram

A. Data Processing

The data processing stage prepares the trajectory data for subsequent classification ensuring consistency and comparability across all data sources. This stage is applied uniformly to both outbound and inbound trajectories.

Radar track, enhanced Mode-S data from VEMMIS, and ACMS data from an airline are first filtered to include only altitude points below FL100, then pre-filtering of the radar and enhanced Mode-S data is performed. This pre-filtering is required to resample all measurements to uniform timestamps of 4.8 seconds, accounting for the radar sampling time.

Additionally, anomalous flights are filtered out, such as flights that diverted back to Amsterdam Airport Schiphol and aircraft that declared emergency events during flight. For ACMS data, flight trajectories with erroneous flap settings and erroneous latitude and longitude are filtered out.

Furthermore, all data types and variables are smoothed using the SciPy Savitzky-Golay function, with the use of a filter window of five and fitting to a quadratic polynomial [18]. This is commonly used to filter and smooth ADS-B and radar track data [19].

B. Outbound Classifier: NADP Classification

The outbound classifier identifies noise abatement departure procedures by analysing aircraft climb trajectories. Since different NADP procedures are characterised by thrust cut-back altitudes, flap retraction altitude and acceleration timing, the classification relies on capturing these operational differences from trajectory data.

Due to variability in operational execution, NADP procedures cannot be uniquely identified using a single indicator. Instead, the proposed method combines multiple metrics to form a robust classification.

The NADP classification consists of three main steps:

1) Evaluation of performance metrics.
2) Identification of similar operational behaviour through clustering
3) Fusion of multiple indicators into a final probabilistic classification

The overall framework and the interaction between the steps is illustrated in Figure 4.

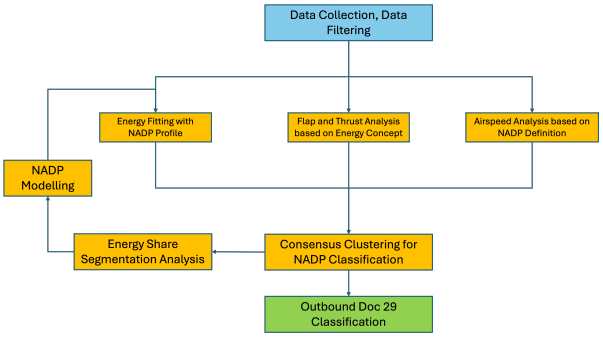


Fig. 4 Outbound Classifier Methodology Block Diagram

1. Clustering and Classification

Since individual performance indicators capture different aspects of departure behaviour, they may produce inconsistent or conflicting detections. To address this, trajectories are grouped based on overall similarity in their operational characteristics.

Each flight trajectory is represented by a set of confidence scores associated with different NADP profiles. For each performance metric, confidence scores are computed for the different NADP profiles, capturing both global trajectory behaviour and discrete operational events.

To identify similar patterns across flights and combine information from multiple performance metrics, a similarity matrix is constructed using cosine similarity between the confidence score vectors. This representation captures the relative similarity between trajectories independent of absolute magnitude differences.

Clustering A spectral clustering algorithm is applied to partition the data into 4 distinct clusters, each representing a group of flights with similar operational behaviour. The cluster labels are assigned using the k-means algorithm [20]. This clustering step enables grouping of trajectories based on overall behavioural similarity, independent of any single performance metric.

Cluster-based Probability Estimation For each cluster label, a probability distribution over NADP profiles is estimated. This is achieved by analysing statistical properties of the metric-derived confidence scores within the cluster.

For each performance metric, the mean (μ) and standard deviation (σ) of each corresponding metric's values are computed within the cluster. A normalised score (s_k) is derived by Equation 1, providing a measure of the signal strength relative to variability.

$$s_k = \frac{\mu}{\sigma} \tag{1}$$

These scores are normalised to form a probabilistic distribution for each NADP class. The contributions from all performance metrics are aggregated using equal weighting.

Final Classification The final NADP classification is obtained by combining two sources of information, the metric-derived confidence scores and cluster-derived probability distribution. All derived confidences are assigned equal weight, and the final score is computed as a weighted sum across all metrics. The resulting vector is then normalised to ensure that it forms a valid probability distribution. The final NADP profile classification is obtained by selecting the NADP profile that corresponds to the highest probability distribution.

2. Performance Metrics

To generate the confidence scores used in the clustering and classification process, the model employs three separate performance metrics:

- Energy-Based Trajectory Similarity Metric

- Energy-Based Flight Performance Metric
- Airspeed-Based Metric

Each metric captures different aspects of the departure behaviour and contributes to distinguishing between NADP profiles. Together, these metrics provide complementary information enabling a more robust distinction between NADP profiles.This corresponds to the first layer of the outbound classifier block diagram as shown in Figure 4.

Energy-Based Trajectory Similarity Metric This metric evaluates how closely the flight trajectory follows the predefined NADP reference profiles. By comparing the total specific energy between the trajectory and the reference NADP profile, the metric captures overall similarity in climb behaviour.

The total specific energy is defined as the sum of the specific kinetic energy (SKE) and specific potential energy (SPE), given by Equation 2 and Equation 3, respectively.

$$SKE = \frac{1}{2}V^2 \tag{2}$$

$$SPE = gh \tag{3}$$

As shown in Equation 2 and Equation 3, mass is independent of energy. Additionally, for all energy-based performance metrics the velocity variable is related to true airspeed (TAS).

Based on the definition of total specific energy (TSE), the trajectory similarity metric compares the flight trajectory to the NADP reference profiles from the FPP. The similarity scores are computed using the root mean square error (RMSE) between the total specific energy (TSE) of the trajectory and that of the fixed-point NADP profile per 100 ft, as shown in Equation 4.

$$\text{Score} = \sqrt{\frac{\sum_n (TSE_{Trajectory} - TSE_{NADP})^2}{n}} \tag{4}$$

Within the fuzzy logic framework [21], the subsequent similarity score of each NADP profile is expressed as a confidence score using the inverse logistic function, as shown in Equation 5.

$$confidence_i = \frac{1}{1 + \frac{score_i}{score_{median}}^2} \tag{5}$$

This metric therefore provides a global measure of trajectory similarity for each NADP profile, which is used as an input to the subsequent clustering and classification framework.

Energy-Based Flight Performance Metric This metric identifies operational events during the departure phase by analysing changes in the aircraft's energy state. Based on the total energy model (TEM), it captures variations in climb dynamics, enabling the detection of thrust cutback and flap/slat retraction. These events provide important information about the aircraft state and configuration, which describes the differences between different NADP profiles.

The energy-based flight performance metric employs the concept of energy equilibrium and the distribution of energy change, as described by the total energy model in Equation 6 [22].

$$TEM = \frac{T - D}{m} = \frac{d(TAS)}{dt} + g_0 \frac{RoC}{TAS} \tag{6}$$

For each flight, only data above 650 ft AGL are included, since thrust reduction and flap/slat retraction are typically

completed at a prescribed minimum altitude of 800 ft [7]. The changes in RoC, acceleration and TEM are used to identify the altitudes at which thrust cutback and flap/slat retraction occur.

For thrust cutback, the incremental change in TEM relative to the previous measurement becomes negative, indicating a reduction in net energy rate. For the aircraft to maintain velocity and to compensate for the reduction in energy rate, the change in RoC relative to the previous measurement also becomes negative.

To represent this behaviour within a fuzzy logic framework, linear left-hand-side (LHS) membership functions are used, modelling conditions where lower input values corresponds to higher confidence. Here, "left-hand-side" refers to the lower range of input domain, where the membership value is maximal and decreases linearly as the input increases. The corresponding left-hand-side membership function is defined in Equation 7.

$$\mu_{\text{LHS}}(x) = \begin{cases} 1, & x \leq b \\ \dfrac{a-x}{a-b}, & b < x < a \\ 0, & x \geq a \end{cases} \tag{7}$$

Linear left-hand-side membership function is used for analysing both incremental changes in TEM and RoC. The corresponding bounds for each state variable are provided in Table 3.

Whereas during flap/slat retraction, thrust remains unchanged, and the increase in TEM is primarily driven by a reduction of drag. Consequently, a positive incremental change in TEM occurs, in combination with a decreasing RoC due to a decrease in the maximum lift coefficient and an increase in acceleration.

To capture this behaviour, linear right-hand-side (RHS) membership functions are employed. In contrast, the "right-hand-side" membership function refers to the upper range of the input domain, where the membership value increases linearly with the input variable. The RHS membership function is defined in Equation 8.

$$\mu_{\text{RHS}}(x) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases} \tag{8}$$

For flap/slat retraction behaviour, a combination of linear right-hand-side membership functions for incremental change in TEM and acceleration, together with a left-hand-side membership function for incremental change in RoC is employed. The corresponding bounds for each state variable are provided in Table 3.

Table 3 Linear membership function State Variable Bounds for Event Detection

Event	Variable	Type	Bounds
Thrust cutback	$dTEM$	LHS	$(-0.05, -0.00)$
	$dRoC$	LHS	$(-100, -0)$
Flap/slat retraction	$dTEM$	RHS	$(0, 0.05)$
	$dRoC$	LHS	$(-100, 0)$
	$acceleration$	RHS	$(0.2, 2)$
Acceleration Altitude	IAS_i	RHS	$(IAS_{i-1}, IAS_{i-1} + 2)$
	IAS_{i+1}	RHS	$(IAS_{i-1}, IAS_{i-1} + 5)$
	IAS_i	RHS	$(V_2, V_2 + 5)$

The membership function values are combined using the product t-norm fuzzy logic for each event detection. An event is considered detected when the resulting product t-norm fuzzy logic confidence exceeds 30%. This relatively low threshold is selected to compensate for the conservative

nature of the product t-norm aggregation and the strict bounds imposed on the event detection criteria.

The first instance of the event detection is used as the corresponding thrust cutback altitude and flap/slat retraction altitude. To aggregate the two event detections and express them as a confidence score for each NADP profile, the thrust cutback altitude is evaluated using the triangular membership function, whereas the flap/ slats retraction altitude is evaluated using the linear membership function. The resulting confidence score is obtained as the product t-norm of the two membership products. The corresponding bounds for each NADP profile are shown in Table 4.

Table 4 Fuzzy Logic Framework for TEM NADP Profiles

Profile	Variable	Function	Altitude Bounds
NADP1	h_{thrust}	Triangular	(650, 1500, 3000)
	h_{flap}	RHS	(2500, 3500)
NADP2 800 ft	h_{thrust}	Triangular	(650, 800, 1250)
	h_{flap}	LHS	(3000, 1500)
NADP2 1000 ft	h_{thrust}	Triangular	(800, 1000, 1750)
	h_{flap}	LHS	(3000, 1500)
NADP2 1500 ft	h_{thrust}	Triangular	(1000, 1500, 3000)
	h_{flap}	LHS	(3000, 1500)

This metric therefore provides event-based indicators of aircraft performance, specifically the altitudes of thrust cutback and flap/slat retraction. The derived confidences are used as inputs to the subsequent clustering and classification framework, enabling the differentiation between global measure and event-based operational behaviour.

Airspeed-Based Metric This metric identifies operational events during the departure phase by analysing changes in the aircraft's indicated airspeed. Based on the incremental changes in indicated airspeed, the transition between take-off speed and initial acceleration climb regime could be captured. This transition behaviour is a key distinguishing feature between different NADP procedures, particularly sub-variants of NADP2 profiles.

The airspeed-based metric analyses the altitude range over which the initial acceleration occurs. For this metric, indicated airspeed (IAS) and altitude are the only input variables considered. This ensures that the analysis focuses on the climb segment where acceleration reflects the transition from take-off speed to climb speed.

Acceleration altitude is identified based on changes in indicated airspeed (IAS) for altitudes above 650 ft AGL. During the acceleration phase, the aircraft transitions from a quasi-constant IAS climb to a regime where IAS increases more rapidly. Three separate incremental IAS conditions characterise this behaviour. First, a positive incremental change relative to the current IAS. Secondly, a continuous positive incremental change relative to the subsequent IAS. Lastly, the positive incremental change relative to the V_2 speed. The V_2 speed is computed as the minimum speed between 100 ft and 650 ft AGL. The exact bounds for each condition are provided in Table 3. The membership functions are combined using the product t-norm fuzzy logic, and the event is considered detected when the confidence exceeds 30%. The first instance of the event detection is used as the corresponding acceleration altitude.

To evaluate the detected acceleration altitude with respect to different NADP profiles, triangular membership functions are used. This function assigns the highest confi-

dence to altitudes near the nominal NADP reference and gradually reduce confidence. The triangular membership function is defined in Equation 9.

$$\mu_{\text{triangular}}(x) = \begin{cases} 0, & x \leq a \\ \dfrac{x-a}{b-a}, & a < x \leq b \\ \dfrac{c-x}{c-b}, & b < x < c \\ 0, & x \geq c \end{cases} \tag{9}$$

The bounds (a, b, c) define the altitude range over which each NADP profile is most representative. The corresponding bounds for each profile are provided in Table 5.

Table 5 Fuzzy Logic Framework for Acceleration Altitude NADP Profiles

Profile	Variable	Function	Altitude Bounds
NADP1	$h_{\text{acceleration}}$	Triangular	$(2500, 3000, 6000)$
NADP2 800 ft	$h_{\text{acceleration}}$	Triangular	$(650, 800, 1250)$
NADP2 1000 ft	$h_{\text{acceleration}}$	Triangular	$(850, 1000, 1750)$
NADP2 1500 ft	$h_{\text{acceleration}}$	Triangular	$(1000, 1500, 3000)$

This metric therefore provides an estimate of the acceleration altitude, which is used as input to the subsequent clustering and classification framework. In combination with the other two performance metrics, it contributes to a comprehensive representation of NADP profile characteristics.

3. NADP Modelling and Refinement

To improve classification robustness, the NADP reference profiles are extended to account for operational variability. Instead of relying on a single reference trajectory per NADP profile, multiple representative trajectories are generated based on energy share distributions. These distributions are modelled using gamma functions across altitude segments.

By sampling multiple percentiles of the energy share distribution, a set of reference trajectories is constructed for each NADP profile. This allows the energy-based similarity metric to better capture variability in observed flight behaviour. This corresponds to the feedback loop of energy share segmentation analysis and NADP modelling shown in Figure 4. This refinement step reduces sensitivity to variation in aircraft performance and improves robustness of the classification.

The refined reference set is then used to update the energy-based similarity metric, improving the consistency of both clustering and classification.

To achieve this, the gamma distribution of the energy share associated with each NADP profile is analysed at multiple altitude segments for each aircraft type. The altitude segments are 0 – 800ft, 800 – 1000ft, 1000 – 1500ft and 1500 – 3000ft and 3000 – 5000ft. These altitude segments are defined based on the NADP profile definitions discussed in section II.

For NADP modelling, the prescribed FPP from NLR is used as a baseline. Based on the FPP, the aircraft is assumed to encounter a constant 8 kts headwind. Hence, based on this assumption, the ground speed, mean ground speed, and time between each segment are derived. To follow finer segmentations, all segments are resampled to be 4.8 seconds for consistency with the radar sampling time.

Furthermore, to construct NADP profile envelopes, a set of percentiles [0.2, 0.3, 0.4, 0.5] is used to generate multiple reference trajectories. The corresponding adjusted

energy share value is obtained by evaluating the inverse cumulative distribution function of the gamma distribution at the selected percentile.

The new reference trajectories are then computed by keeping the altitude of each node constant while adjusting the travelled distance and TAS according to the energy share formula shown in Equation 10.

$$Energy\ Share\ \% = \frac{\Delta Kinetic\ Energy}{\Delta Kinetic\ Energy + \Delta Potential\ Energy} \tag{10}$$

Lastly, the travelled distance for each segment is adjusted using kinematic relationships based on the derived true airspeed, as shown in Equation 11

$$ds = \frac{GS_i + GS_{i+1}}{2} dt \tag{11}$$

With the generation of multiple reference trajectories, the energy-based trajectory similarity metrics are recomputed. This leads to an updated similarity matrix, which is subsequently used to refine both the NADP consensus clustering and the NADP profile classification.

C. Inbound Classifier: Continuous Descent Approach and Reduced-Flap Setting Approach

The inbound classifier identifies continuous descent approaches and reduced-flap setting approaches by analysing aircraft approach trajectories. These two noise abatement procedures capture different operational characteristics, therefore modelled independently. However, the combination of the two classifiers provides a complementary description of the overall arrival behaviour.

The overall framework and the interaction between the two aspects of inbound classifier is illustrated in Figure 5.

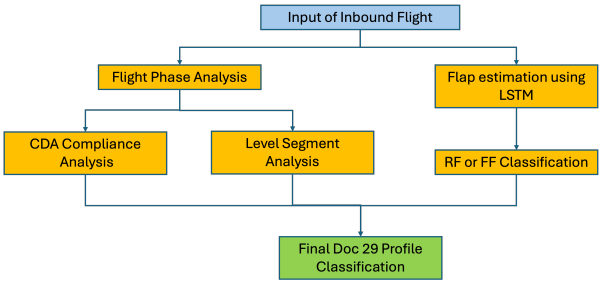


Fig. 5 Inbound Classifier Methodology Block Diagram

1. Continuous Descent Approach Classification

The CDA classifier determines whether an aircraft follows a continuous descent profile or exhibits level-flight segments during approach. This is achieved by analysing flight phases through changes in altitude and flight path angles. The CDA classification process is further analysed in three main steps:

1) Flight phase detection
2) Level flight analysis based on ICAO and RMI criteria
3) CDA flight analysis based on ICAO and RMI criteria

These three main steps are also depicted on the left-hand-side of the inbound classifier block diagram as shown in Figure 5.

Flight Phase Detection For the classification of CDA, flight trajectories are segmented into flight phases of climb, level, and descent, based on changes in altitude and the calculated flight path angle (γ), as shown in Equation 12.

$$\gamma = arcsin\left(\frac{RoC}{TAS}\right) \tag{12}$$

For flight phase detection, the climb phase is identified when the current altitude exceeds the previous altitude by over 25 ft, and the corresponding flight path angle γ is higher than 0.75° and sustained over two consecutive

trajectory points.

The level flight phase is detected when the absolute change in altitude between consecutive points (both previous–current and current–next) is below 25 ft, and the absolute flight path angle γ is less than 0.75°. Additionally, the identified level segment must have a minimum ground travel distance exceeding 500 m. All remaining trajectory segments are classified as the descent phase until final touchdown.

For noise modelling purposes and Dutch regulation compliance, two complementary definitions are used to evaluate CDA compliance, reflecting both operational behaviour and regulatory compliance: ICAO framework and RMI compliance.

For ICAO framework, flights that exhibit level segments or climb segments are classified as non-CDA flights. Furthermore, the detected level altitudes are mapped to the closest standard level altitudes defined in the FPP, which are 4000 ft, 3000 ft and 2000 ft. Similarly, the corresponding level segment lengths are mapped to the closest standard distances of 5 km, 10 km, 20 km and 40 km.

For ICAO compliance CDA flights, the average flight path angle γ is evaluated below 6000 ft and before a distance of 12 km from the runway threshold [5]. Furthermore, the calculated average γ is mapped to the closest standard CDA γ defined in the FPP, which are 2°, 2.25°, 2.5°, 3° and 3.5°. These analyses subsequently lead to the formulation of the FPP profile identifiers.

For RMI compliance, a flight trajectory is classified as a CDA if no climb or level segments occur below 3500 ft, and the flight trajectory remains above a 2.5° datum-line [11] as shown in Figure 6.

Flights that do not satisfy the above conditions are classified as level segment flights. Based on the first instance of RMI CDA violation, these flights are mapped to either a 3000 ft or 2000 ft level flight [11].

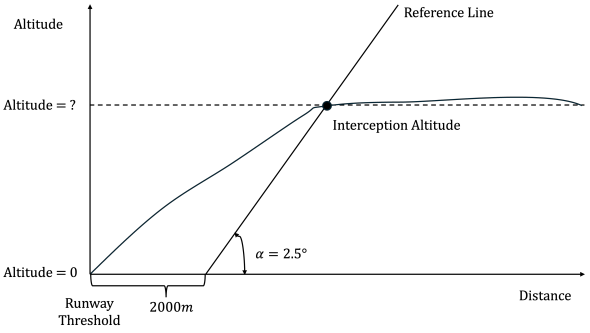


Fig. 6 RMI Continuous Descent Approach Classification Definition Adapted from the RMI [11]

2. Reduced-Flaps Setting Approach Classification

This component classifies whether an aircraft performs a reduced-flap or full-flap landing configuration during approach. This distinction between the two configurations influences the noise generation. Since flap configuration cannot be directly observed from surveillance data, it must be inferred from trajectory behaviour using a data-driven model.

For the reduced-flaps setting approach classification, an LSTM neural network is employed. The model uses kinematic, positional, and energy-based inputs, namely: ground speed (GS), calibrated airspeed (CAS), true airspeed (TAS), barometric altitude, latitude, longitude, runway identifier, distance to landing, specific kinetic energy (SKE), and specific potential energy (SPE).

Data Preparation The flight trajectories are split into windows with a sequence length of 30, in which the stride between each window is tuned to 10. Furthermore, the ACMS dataset per aircraft type is split into training, vali-

dation and testing data with the ratio of 70/15/15. For the training and validation dataset, the batch size is chosen to be 32. Whereas for the testing dataset, all flight trajectories are considered in one batch, which ensures the whole flight trajectory is not split into separate batches.

To prevent the model from learning a particular order, the training set is shuffled by the dataloader. Whereas for the validation and testing dataset, shuffle is disabled from the dataloader to ensure consistency between flight trajectory sequences [17].

Model Architecture and Training For the LSTM training model, the model undergoes batch normalisation, and then the input variables are fed to the LSTM layer. The LSTM layer is tuned with 16, 32 and 64 hidden neurons, also tuned with 1, 2 and 3 LSTM layers.

To prevent data over-fitting, an output dropout rate of 0.3 is used. This means that after each training step, 30% are randomly removed before passing to the next training step [23]. As this LSTM model outputs a classifier, cross-entropy loss is used for training and validation, and the Adam optimiser is used for training.

To further prevent data over-fitting, weight decay is employed in the Adam optimiser [23], with a value of 0.001. Learning rate decay is also incorporated with a decay rate of 0.9, patience of 40 epochs. These parameters are tuned to avoid premature depreciation of the learning rate.

The training process also incorporates an early stopping mechanism that balances training and validation loss by assigning weights of 35% and 65% respectively. If after 80 epochs, the weighted score loss does not further decrease, the observed best model would be retained.

Operation-Informed Training Constraints To enforce operational realistic behaviour, additional penalty terms are introduced during training.

To prevent flaps from retraction during descent operations, a penalty function is included in the LSTM training model. In the event of such behaviour, the penalty function is defined by Equation 13.

$$\frac{1}{N}\sum_{i=1}^{N} max(-\text{diff}_i, 0)^2 \tag{13}$$

Furthermore, to prevent the extension or retraction of flaps below the minimum required stabilised approach altitude of 500 ft [24], an additional penalty function is introduced in the LSTM training model. This penalty is applied when such behaviour occurs below 500 ft and is defined by Equation 14.

$$\frac{0.05}{N}\sum_{i=1}^{N} \text{diff}^2_{\text{Below500ft}} \tag{14}$$

Classification Output Lastly, the classification of reduced-flap or full-flap approach is formulated as a binary classification. Since flap configurations are not directly comparable across aircraft types due to number of available flap settings and their definition. The analysis is performed per aircraft type and interpreted relative to its configuration envelope. The final classification is obtained by comparing the predicted final flap setting with the maximum available flap setting for each aircraft type. This allows a consistent distinction between reduced-flap and full-flap operations without explicit normalisation. The available flap settings per aircraft type is shown in Table 6.

Table 6 Available Flap Settings per Aircraft Type

ICAO Aircraft Code	Available Flap Settings
A21N	0, 1, 2, 3, 4
B737	0, 1, 2, 5, 10, 15, 25, 30, 40
B738	0, 1, 2, 5, 10, 15, 25, 30, 40
B739	0, 1, 2, 5, 10, 15, 25, 30, 40
B772	0, 1, 5, 15, 20, 25, 30
B77W	0, 1, 5, 15, 20, 25, 30
B789	0, 1, 5, 10, 15, 20, 25, 30
B78X	0, 1, 5, 10, 15, 20, 25, 30
E190	0, 1, 2, 3, 4, 5, 6
E295	0, 1, 2, 3, 5, 6
E75L	0, 1, 2, 3, 5, 6

V. Hypothesis

Based on the existing literature and theoretical frameworks on noise abatement procedures. The following hypotheses are proposed:

- **H1:** Accounting for variation in the energy share factor using default fixed-point profiles improves the accuracy of classifying aircraft trajectories into NADP profiles.
- **H2:** The energy-based trajectory similarity metric is the strongest predictor for classifying trajectories into NADP profiles.
- **H3:** Combining the continuous descent approach CDA definitions from the RMI and ICAO improves the robustness of CDA classification.
- **H4:** Including landing runway length in approach trajectory data improves the classification performance of reduced-flap setting approaches when using a sequential neural network algorithm.

The independent and dependent variables corresponding to each hypothesis is summarised in Table 7 and Table 8 respectively.

Table 7 Summary of Hypothesis Independent Variables

Hypothesis	Independent Variable
H1	Energy share variation
H2	Performance metric type
H3	CDA definition (ICAO + RMI)
H4	Runway length input

Table 8 Summary of Hypothesis Dependent Variables

Hypothesis	Dependent Variable
H1	NADP classification accuracy
H2	NADP classification accuracy
H3	CDA classification robustness
H4	Reduced-Flap classification accuracy

VI. Results

The results of the three different noise abatement procedures will be discussed separately. However, for precise FPP profile identification in noise modelling, the CDA classifier and the reduced-flap setting approach classifier are combined.

A. NADP Classifier

The results of the NADP classifier indicate that departure operations are predominantly classified as NADP2-type procedures, with NADP2 1000 ft emerging as the dominant profile.

Final Classification The final classification results show that the majority of flights are assigned to the NADP2 1000 ft profile, followed by the NADP2 1500 ft and NADP2 800 ft, with NADP1 being the minority operation.

To obtain the final classification, a fusion approach is applied that combines the cluster-level probability distributions with the metric-level confidence scores. This weighted fusion integrates both the similarity-based structure identified by clustering and the metric-level information, providing a unified classification for each trajectory.

The resulting NADP classification percentages are shown in Table 9 and compared to realisation results. The realisation results are based on interviews with airlines operating at Amsterdam Airport Schiphol. In cases where no specific NADP procedure is defined, flights are default classified as following an NADP1 profile.

Table 9 Final Fusion NADP Classification Percentages

NADP Profile	Percentage	Realisation	Percentage Difference
NADP1	5.6%	15.8%	-10.2%
NADP2 800 ft	13.0%	65.9%	-52.9%
NADP2 1000 ft	61.2%	16.2%	+45.0%
NADP2 1500 ft	20.2%	0.3%	+19.9%
Unknown	-	1.8%	-

As presented in Table 9, the largest discrepancy between classification results and realisation data occurs for the NADP2 800 ft and NADP2 1000 ft procedures. However, when combining the classification percentages of NADP2 800 ft and NADP2 1000 ft, the discrepancy compared to the realisation reduces to 7.2%. This discrepancy suggests that NADP2 800 ft and NADP2 1000 ft are difficult to distinguish.

Clustering Results To understand the structure underlying the final classification, the clustering results are analysed. Spectral clustering groups trajectories based on similarity in their combined metric representation.

Each cluster (1-4) represents a group of trajectories with similar operational behaviour, rather than predefined NADP categories. The cluster confidence values indicate the relative likelihood of each NADP profile within a given cluster, derived from the aggregated performance metric scores. The resulting cluster-level confidence distribution are shown in Table 10.

Table 10 Spectral Cluster NADP Profile Confidence and Profile Percentage

NADP Profile	Cluster 1 Confidence	Cluster 2 Confidence	Cluster 3 Confidence	Cluster 4 Confidence	Profile Percentage
NADP1	0.084	0.081	0.078	0.357	10.6%
NADP2 800 ft	0.139	0.281	0.372	0.154	25.2%
NADP2 1000 ft	0.331	0.400	0.351	0.255	44.1%
NADP2 1500 ft	0.446	0.236	0.198	0.234	20.1%

In a perfect scenario, Table 10 should look like an identity matrix, with each NADP procedure corresponding to a cluster with 100% confidence. However, this is not the case as shown on Table 10. Each cluster is associated with a specific NADP profile by identifying the profile with the highest confidence value. In particular, for clusters 1, 2, and 4, the maximum confidence exceeds the second-highest value by at least 10%. This suggests that these clusters exhibit a clear separation in the confidence values.

In contrast, cluster 3 exhibits a less pronounced distinction between the confidence values, with a relatively small difference of only 2% between the highest and second highest confidence. Specifically, the clusters show comparable confidence levels for NADP2 800 ft and NADP2 1000 ft, indicating that the underlying characteristics captured by the metrics are not sufficiently distinct to fully separate these profiles within this cluster. Hence, the combined use of cluster-level confidence and metric-level confidence

provides a more reliable basis for distinguishing between closely related NADP profiles.

Performance Metric Results The individual performance metrics produce different classification distributions, indicating that each metric captures a distinct aspect of NADP procedures. The resulting distribution obtained from individual performance metrics are shown in Table 11.

Table 11 Trajectory Similarity and Flight Performance Metrics per NADP Profile

NADP Profile	Trajectory Based Similarity Metric	Energy-Based Flight Performance Metric	Airspeed Based Metric
NADP1	7.8%	5.7%	1.5%
NADP2 800 ft	54.0%	14.4%	11.8%
NADP2 1000 ft	8.4%	66.5%	49.7%
NADP2 1500 ft	29.8%	29.8%	37.0%
Unknown NADP	0.0%	1.3%	0.0%

As shown in Table 11, the trajectory similarity metric indicates a strong operational dominance for NADP2 800 ft profile, with 54.0% of trajectories classified under this category.

In contrast, the energy-based flight performance and airspeed-based metrics predominantly classify trajectories as NADP2 1000 ft (66.5% and 49.7%, respectively). These metrics rely on event-based and transition-based characteristics, such as thrust cutback, flap/slats retraction and acceleration altitude.

The trajectory similarity results presented above correspond to the refined version of the metric, in which NADP modelling and variation in energy share are incorporated to construct profile envelopes. To assess the impact of this refinement, a comparison with the original trajectory similarity metric—based solely on the fixed-point profile (FPP) reference trajectory is presented in Table 12.

Table 12 Comparison of Preliminary and Refined Trajectory Similarity Metrics for NADP Profile Classification

NADP Profile	Only FPP Reference Trajectory	Incorporation of Profile Envelope
NADP1	9.9%	7.8%
NADP2 800 ft	23.7%	54.0%
NADP2 1000 ft	18.5%	8.4%
NADP2 1500 ft	47.9%	29.8%
Unknown NADP	0.0%	0.0%

As shown from Table 12, incorporating profile envelope refinement leads to substantial redistribution of classifications. The proportion of NADP1 decreased from 9.9% to 7.8%. Furthermore, the dominant NADP profile shifted from NADP2 1500 ft to NADP2 800 ft, moving closer to the operational benchmark based on interviews with airlines.

Additionally, when selecting energy shares using percentiles above 0.5 of the inverse cumulative distribution function, the proportion of trajectories classified as NADP1 increased to approximately 20% and higher. This highlights the sensitivity of the method to energy-share assumptions.

The discrepancy between the trajectory similarity metric and the event-based metrics highlights that different aspects of departure behaviour are being captured. While the trajectory similarity metric reflects the overall energy distribution along the trajectory, the flight performance and airspeed metrics focus on discrete operational transitions. This divergence motivates the need for a clustering-based integration of the metrics.

Sensitivity Analysis To evaluate the relative contribution of each performance metric to the final classification, the agreement between individual metrics and the final NADP

classification is computed.

Agreement represents how often an individual performance metric assigns the same NADP profile as the final classification across all trajectories. This percentage provides insight into which performance metric is the strongest predictor. The agreement between individual metrics and the final NADP classification is summarised in Table 13.

Table 13 Agreement of Individual Metrics with Final NADP Classification

Metrics	Agreement
Energy-Based Trajectory Similarity	29.3%
Energy-Based Flight Performance	59.2%
Airspeed Based	72.2%

The result from Table 13 indicates that the airspeed-based performance metric exhibits the highest agreement (72.2%), while the energy-based trajectory similarity metric has the lowest agreement (29.3%).

This indicates that event-based features have greater influence on the final classification, while the energy-based trajectory similarity plays a complementary role.

To assess the robustness of the obtained classification results, a sensitivity analysis is performed. The analysis considers variation in the weighting of each performance metric and cluster contribution in the final fusion process.

The weights of the trajectory similarity, energy-based flight performance, and airspeed metrics, as well as the clustering contribution are varied over the range [0, 1] in increments of 0.05. All weight combinations are constrained to be strictly positive and normalised such that their sum equals one. This results in a total of 969 valid weight combinations.

For each combination, the final NADP classification is recomputed. The resulting minimum and maximum classification percentages are summarised in Table 14 and the distribution of outcome is shown in Figure 7.

Table 14 Sensitivity Analysis of NADP Classification

NADP Profile	Minimum	Maximum	Range
NADP1	0.7%	10.6%	9.9%
NADP2 800 ft	7.3%	34.8%	27.5%
NADP2 1000 ft	33.5%	62.0%	28.5%
NADP2 1500 ft	12.5%	33.8%	21.2%

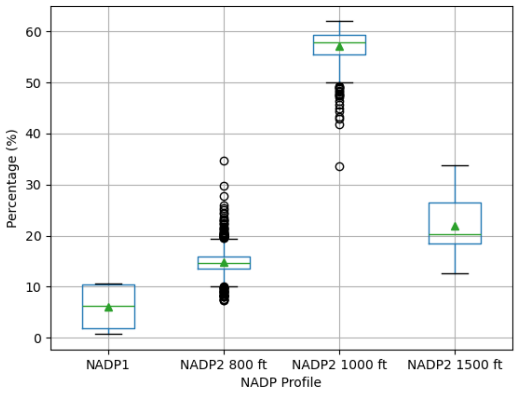


Fig. 7 NADP Classification Sensitivity Distribution

The result indicates that the largest variability is observed for NADP2 800 ft and NADP2 1000 ft profiles. This variability can be attributed to the small differences in the operational altitudes, which makes the two procedures difficult to distinguish. In contrast, the NADP1 profile exhibits a smaller range, indicating relatively stable classification across all weight combinations.

The distribution of classification outcomes in Figure 7 further illustrates these trends. The NADP2 800 ft profile exhibits a relatively narrow interquartile range but includes a large number of outliers, indicating that while typical classifications remain concentrated, certain weight combinations lead to substantial deviations.

In contrast, the NADP2 1000 ft profile shows a slightly

wider interquartile range, but its outliers are all located on the lower end. Furthermore, approximately 99% of all evaluated weight configurations identified NADP2 1000 ft as the dominant operation, indicating a strong robustness of this classification.

Finally, the NADP1 and NADP2 1500 ft profiles exhibit relatively stable behaviour, with limited dispersion and no significant outliers observed.

Overall, the results demonstrate that while individual performance metrics capture different aspects of departure behaviour, their integration through clustering, fusion, and sensitivity analysis enables a robust identification of dominant NADP procedures, with remaining uncertainty primarily occurring between closely related NADP2 sub-profiles.

B. Continuous Descent Approach Classifier

For the classification of CDA flights, a comparison is made between compliance with the ICAO framework and RMI definition. The resulting classifications for CDA and level flight operations are presented in Table 15.

Table 15 Comparison of CDA Classification Methods

Flight Operations	RMI Classification	ICAO Classification	Realisation
CDA Flights	38.5%	32.4%	38.8%
2000 ft Level Flights	46.7%	47.5%	45.2%
3000 ft Level Flights	14.8%	9.9%	16.0%
4000 ft Level Flights	-	10.2%	-

The classification of CDA flights is relatively consistent across both classification methods and aligns closely with the realisation data.

In contrast, the classification of level flight operations shows greater deviation between the two methods and realisation, particularly for 3000 ft level flight operations.

When both classification methods are combined, only flights that satisfy both definitions are identified as CDA. This results in a reduction of classified CDA flights to 28%, with underlying inconsistencies between the two classifications shown in Figure 8. The confusion matrix shows that 78.3% of the flights are classified consistently between the two methods. However, disagreements between the two methods consist of up to 21.7% of the flights, where approximately 60% of the disagreements correspond to flights classified as CDA under the RMI definition but not under the ICAO framework.

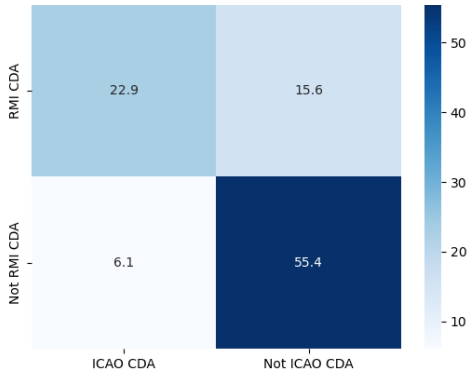


Fig. 8 Confusion Matrix of Different CDA Classification

C. Reduced-Flap Setting Approach Classifier

The classification performance of the reduced-flap setting approach classifier is evaluated using accuracy, defined as the percentage of trajectories for which the predicted final flap configuration either reduced-flap or full-flap matches the reference configuration obtained from the ACMS dataset. The corresponding LSTM architecture for each evaluated aircraft type is presented in Table 16.

The majority of the LSTM models achieve a classification accuracy above 80% for the prediction of the maximum

Table 16 LSTM Model Performance per Aircraft Type

Aircraft Type	Layers Layers	Hidden Neurons	Accuracy	Over Prediction	Number of Samples	Number of Available Flaps
A21N	2	16	92.9%	100.0%	278	5
B737	2	32	92.9%	0.0%	463	9
B738	2	32	86.5%	0.0%	1885	9
B739	2	64	89.5%	0.0%	255	9
B772	2	32	69.6%	95.6%	304	7
B77W	1	64	73.0%	90.0%	247	7
B789	2	32	76.7%	85.7%	201	8
B78X	2	64	63.2%	92.8%	257	8
E190	1	32	99.2%	0.0%	1677	7
E295	1	32	100.0%	0.0%	945	6
E75L	1	16	100.0%	0.0%	1265	6

deployed flap setting. In contrast, lower accuracy levels are observed for wide-body jet types, namely the B772, B77W, B789 and B78X, with the lowest accuracy recorded at 63.2%.

In addition to that, the over-prediction rates are generally higher for lower accuracy LSTM models. This indicates the model predicts the flight to be flown under full-flap instead of reduced-flap, allowing a more conservative estimation for noise modelling.

To assess the contribution of runway length to LSTM classifier performance, an additional LSTM model was trained without including the runway identifier as an input feature. In addition to that, the number of hidden neurons per layer is kept constant. The resulting performance and model configurations are summarised in Table 17.

The results show that the exclusion of runway information leads to only marginal changes in classification accuracy across all aircraft types. In most cases, comparable model performance is achieved with a simpler LSTM architectural structure.

However, when enforcing the same architecture as the runway-informed models, the performance of the models without runway information often degrades significantly, in some cases resulting in extremely low classification accuracy, such as 0%.

Table 17 Impact of Runway Information on Model Performance and Architecture

Aircraft	With Runway		Without Runway	
	Layers	Accuracy	Layers	Accuracy
A21N	2	92.9%	2	85.7%
B737	2	92.9%	1	92.9%
B738	2	86.5%	2	86.5%
B739	2	89.5%	1	89.5%
B772	2	69.6%	1	69.6%
B77W	1	73.0%	1	73.0%
B789	2	76.7%	1	70%
B78X	2	63.2%	1	57.9%
E190	1	99.2%	1	99.2%
E295	1	100%	1	100%
E75L	1	100%	1	98.9%

VII. Discussion

The results of the three noise abatement procedure classifiers highlight both the potential and limitations of data-driven classification of operational procedures.

A. NADP Classification

The NADP classification results demonstrate a clear divergence between energy-based trajectory similarity metrics, energy-based performance metric, the airspeed-based metric and the final fusion classification.

Refining the trajectory similarity metrics improves the classification of NADP1 operations, bringing the estimated proportion closer to the expected operational benchmark. Furthermore, the refined results suggest that the majority of flights operate under the NADP2 800 ft procedure, consistent with insights from airline interviews. These findings support hypothesis **H1**, demonstrating that incorporating variation in the energy share factor to the FPP enhances the accuracy in using the trajectory similarity metric to classify NADP profiles.

However, a key limitation emerges from the sensitivity

of the method, particularly in the use of energy share values. When higher energy share values are selected, specifically using percentiles above 0.5 of the inverse cumulative distribution function, the portion of trajectories classified as NADP1 increases significantly.

This behaviour can be explained by the effect of applying a higher energy share on the resulting trajectory characteristics. Increasing the energy share implies that a larger portion of the available energy is allocated to acceleration performance, which alters the trajectory shape. As a result, the constant energy share feature of NADP1 profiles diminishes and overlaps with the distinguishing thrust cut-back features of NADP2 profiles. This leads to increased overlaps between the operational regimes of NADP1 and NADP2 procedures, reducing their separability and causing a systematic shift of classifications toward NADP1.

The discrepancy between trajectory similarity and flight performance-related metrics highlights the difficulty in distinguishing between NADP2 procedures. While the trajectory similarity metrics analyse the geometric and energy-based characteristics of the flight path, the energy-based performance metric and airspeed metrics aim to capture operational events such as thrust cutback, flap/slats retraction and acceleration altitude. However, these event-based detection metrics are sensitive to temporal resolution from enhanced mode-S data and sampling frequency from the radar. In addition to sampling frequency and temporal resolution, system delay could also affect the event-based detection by shifting the detection to a later trajectory node than the actual event instance.

The use of spectral clustering serves as a unifying framework to integrate the information obtained from the various metrics by grouping flights based on similarity. Confidence values are subsequently assigned to each cluster as shown in Table 10, enabling the interpretation of the dominant NADP profile. While most clusters exhibit a clear dominance of a single NADP profile, cluster 3 still shows ambiguity between the NADP2 800 ft and NADP2 1000 ft.

By comparing the highest and the second-highest confidence values, the clusters also reveal transitional behaviour between closely related NADP procedures. For example, cluster 1 is predominantly associated with the NADP2 1500 ft profile. However, the relatively high secondary confidence for NADP2 1000 ft indicates a gradual transition between these two operational regimes. This behaviour is also observed in clusters 2 and 3, which suggests that the clustering process captures not only distinct procedural categories but also intermediate behaviours that lie between distinct procedural categories.

The final fusion classification results in the majority of the operations classified to NADP2 1000 ft, which shows a large discrepancy between the classification and the realisation from interviews with airlines. However, when combining the percentages of classified NADP2 800 ft and NADP2 1000 ft, the discrepancy between the realisations is minimised. This further supports the observed phenomenon that operational NADP behaviour exists on a continuum rather than in strictly discrete categories.

Furthermore, the shift from NADP2 800 ft to NADP2 1000 ft may also be attributed to the reliance of several metrics on event detection. Given the previously mentioned sensitivities in sampling frequency, temporal resolution and system delay, the identification of the precise timing of the event detection poses difficulty. Since the distinction

between NADP2 800 ft and NADP2 1000 ft is primarily defined by the event transition, such uncertainties could potentially lead to classification bias towards the next discrete category in the continuum spectrum.

The sensitivity analysis in Table 14 demonstrates that while the dominant classification of NADP2 1000 ft remains robust across 99% of weighting configurations, significant variability exists in the classification of secondary profiles. This indicates that the classification framework remains sensitive to the parameter selection and weighting assumption.

However, across the majority of weighting configurations, NADP1 consistently remains the least represented profile. This observation aligns with the expected operational benchmark, suggesting that despite the classification framework remaining sensitive to parameter selection and weighting assumptions, it retains a reasonable level of identifying relatively low frequency procedures.

Finally, the agreement analysis shown in Table 13 indicates that airspeed-based metrics contribute most strongly to the final classification, thereby rejecting hypothesis **H2**

B. CDA Classification

The CDA classification results exhibit a relatively strong agreement between the RMI regulations and the ICAO framework, particularly in CDA classification. However, discrepancies arise during the level flight operation classification, which is due to the underlying differences between the two frameworks.

When combining the two frameworks, as shown in Figure 8, 78.3% of the CDA classifications are consistent. However, 15.6% of the flights are classified as CDA-compliant under the RMI definition but not under the ICAO framework. This discrepancy aligns with the observed 10.2% 4000ft level flight operations presented in Table 15. The primary cause of this discrepancy lies in the RMI regulations, which restrict level flight classifications to only 2000 ft and 3000 ft, while other level flight segments above these thresholds, such as level flight above 3500ft are still considered CDA-compliant. In contrast, the ICAO framework explicitly accounts for such higher altitude level segments, leading to their classifications as non-CDA operations.

Conversely, 6.1% of the flights are identified as CDA-compliant under the ICAO framework but not under the RMI definition. This misalignment can be attributed to the geometric constraint imposed by the RMI definition, which requires the descent trajectory to be above a 2.5° datum reference line. As a result, flights classified as CDA-compliant under the ICAO framework with shallower flight path angle, such as γ with 2° or 2.25°, fail to meet RMI CDA criteria.

These differences in geometric constraints further contribute to the discrepancy observed in the level flight classification between the two frameworks. RMI regulations define level flight based on the altitude at which it is no longer compliant with the CDA criteria. This classification essentially assigns the transition point from continuous descent to level flight. In contrast, the ICAO's framework, with references also to the FPP[5], allows for multiple-level segments within a descent profile.

Consequently, the methodological difference leads to systematic differences in the classification outcomes. In particular, it explains the higher portion of 2000ft level flight operations identified under the ICAO framework compared to the RMI classification.

All in all, these findings support hypothesis **H3**, as the combined use of RMI and ICAO definition provides a more robust framework for CDA classification. The combination of frameworks captures both geometric constraints and operational descent characteristics. This enables the classification to reflect both regulatory compliance and variability observed in real-world operations.

C. Reduced-Flap Setting Approach Classification

The reduced-flap setting approach classification result exhibits that LSTM-based sequence modelling is generally effective in identifying flap deployment behaviour from trajectory data. For most narrow-body aircraft types, classification accuracy above 85% are achieved. However, generically, the LSTM models tend to under-predict the flap deployment behaviour.

As a result, when coupled with noise modelling, this may lead to underestimation of the noise emissions, as reduced-flap configurations are associated with lower aerodynamic drag and subsequently lower noise levels compared to full-flap operations.

In contrast, for wide-body aircraft, classification accuracy is averaged around 70%. This reduction could be attributed to having a limited sample size available for these aircraft types, which restricts the LSTM model's ability to learn a representative pattern.

Furthermore, in contrast to narrow-body jets, the LSTM models for wide-body jets tend to over-predict flap deployment behaviour. Therefore, the classification for wide-body jets results in a conservative manner.

Lastly, when comparing the inclusion of the runway identifier as an input to the LSTM model Table 17, only minimal changes in model performance are observed. Although most models require a reduction in architectural complexity from two layers to one layer. This reduction of model complexity can be attributed to the increase in input dimensionality introduced by the runway identifier, where each runway is represented as a binary-encoded feature.

Finally, the limited impact of the runway identifier may be explained by the structure of the dataset, in which cumulative distance implicitly captures aspects of the runway length through the inclusion of landing distance. As a result, this finding does not support hypothesis **H4**.

VIII. Conclusion

This study developed and evaluated a data-driven approach for the classification of noise abatement procedures, including NADP, CDA and reduced-flap setting approaches.

For NADP classification, the results demonstrate that incorporating variability in the energy share factor within the FPP improves the accuracy of the energy-based trajectory similarity classification, supporting hypothesis **H1**.

However, substantial discrepancies remain between classification outputs and airline realisations, particularly for distinguishing between different NADP2 sub-variants. These differences highlight that operational behaviour does not follow strictly discrete categories, but rather exists along a continuum. While the fusion approach provides robust dominant classifications, the strong influence of airspeed-based metrics leads to the rejection of hypothesis **H2**.

For CDA classification, relatively strong agreements are observed between RMI regulation and the ICAO framework in identifying CDA operations, while discrepancies between operational reality and regulatory definitions are observed. The combined use of both frameworks enhances

the classification robustness of CDA operations, supporting hypothesis **H3**.

For the reduced-flap setting approach classification, the LSTM models demonstrate high performance for narrow-body jets, indicating the method is suitable for classifying reduced-flap setting approach operations. However, a decrease in accuracy is observed for wide-body jet due to the limited availability of training data. Additionally, the minimal impact of runway identifier as an input suggests that trajectory derived features are sufficient, provided that trajectory data includes ground segments during landing, consequently leading to rejection of hypothesis **H4**.

Overall, this research demonstrates that through data-driven approaches it is possible to capture key patterns and operational events to classify flight trajectory to the corresponding noise abatement procedure flown. Future work should focus on improving temporal resolution, expanding the dataset and developing event-based validation set to enable more direct comparison with airline realisations.

References

[1] Royal Schiphol Group, "Annual Traffic Reviews," Available: https://www.schiphol.nl/en/schiphol-group/traffic-review/, 2026. Accessed: 2026/04/25.

[2] Royal Schiphol Group, "Usage forecast," Available: https://www.schiphol.nl/en/you-and-schiphol/gebruiksprognose-alles-wat-je-maar-wilt-weten/, 2024. Accessed: 2026/04/25.

[3] Heppe, G., "Appendices van de voorschriften voor de berekening van de geluidsbelasting in Lden en Lnight voor Schiphol," Tech. Rep. NLR-CR-96650 L, Nederlands Lucht- en Ruimtevaartcentrum (NLR), 2014.

[4] European Civil Aviation Conference, *ECAC Doc 29: Report on Standard Method of Computing Noise Contours around Civil Airports*, 4th ed., Vol. 2, 2016.

[5] Heblij, S. J., and Derei, J., "Methodenrapport Doc29," Technical Report NLR-CR-2019-076, Netherlands Aerospace Centre (NLR), Amsterdam, The Netherlands, 2019.

[6] Lim, D., Behere, A., Jin, Y. D., Li, Y., Kirby, M., Gao, Z., and Mavris, D. N., "Improved Noise Abatement Departure Procedure Modeling for Aviation Environmental Impact Assessment," *AIAA SciTech 2020 Forum*, 2020. https://doi.org/10.2514/6.2020-1730.

[7] International Civil Aviation Organization, *Procedures for Air Navigation Services — Aircraft Operations (PANS-OPS), Volume I: Flight Procedures*, 5th ed., ICAO Doc 8168, International Civil Aviation Organization (ICAO), Montreal, Canada, 2006.

[8] International Civil Aviation Organization, "Review of Noise Abatement Procedure Research & Development and Implementation Results," Preliminary edition Doc 9888, International Civil Aviation Organization (ICAO), Montreal, Canada, 2007.

[9] International Civil Aviation Organization, *Continuous Descent Operations (CDO) Manual*, 1st ed., ICAO Doc 9931, International Civil Aviation Organization (ICAO), Montreal, Canada, 2010.

[10] International Civil Aviation Organization, *Report on Standard Method of Computing Noise Contours around Civil Airports, Volume 2: Technical Guide*, 4th ed., ICAO Doc 29, International Civil Aviation Organization (ICAO), Montreal, Canada, 2016.

[11] Nederlandse Wetgevende Macht, "Wet luchtvaartwet," , 07 2023. URL https://wetten.overheid.nl/BWBR0014722/2023-07-01/0, accessed: 2025/10/25.

[12] Behere, A., and Mavris, D. N., "A Method for the Parametric Representation of Take-off Time-Series Trajectory Data for Environmental Impact Assessment," *AIAA AVIATION 2023 Forum*, 2023.

[13] Behere, A., Isakson, L., Puranik, T. G., Li, Y., Kirby, M., and Mavris, D. N., "Aircraft Landing and Takeoff Operations Clustering for Efficient Environmental Impact Assessment," *AIAA AVIATION 2020 Forum*, 2020. https://doi.org/10.2514/6.2020-2583.

[14] Bhanpato, J., Puranik, T. G., and Mavris, D. N., "Data-Driven Analysis of Departure Procedures for Aviation Noise Mitigation," *Engineering Proceedings*, Vol. 13, No. 1, 2021, p. 2. https://doi.org/10.3390/engproc2021013002.

[15] Nichols, C., and Cook, T., "Estimating Aircraft State from Surveillance Data Using Statistical Learning," , 2022. https://doi.org/10.2514/6.2022-3424.

[16] Jarry, G., Delahaye, D., and Feron, E., "Approach and landing aircraft on-board parameters estimation with LSTM networks," *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT 2020)*, IEEE, 2020. https://doi.org/10.1109/AIDA-AT48540.2020.9049199.

[17] Bhanpato, J., Behere, A., Moharir, A., and Mavris, D. N., "Aircraft Configuration Prediction for Arrival Operations from Trajectory Data," , 2024. https://doi.org/10.2514/6.2024-2618.

[18] Savitzky, A., and Golay, M. J. E., "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, Vol. 36, No. 8, 1964.

[19] Sun, J., Ellerbroek, J., and Hoekstra, J., "Large-Scale Flight Phase Identification from ADS-B Data Using Machine Learning Methods," *7th International Conference on Research in Air Transportation (ICRAT)*, Delft University of Technology, 2016.

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.

[21] Zadeh, L. A., "Fuzzy Sets," *Information and Control*, Vol. 8, No. 3, 1965, pp. 338–353.

[22] Eurocontrol Experimental Centre, "User Manual for the Base of Aircraft Data (BADA) Revision 3.6," Eec note no. 10/04, European Organisation for the Safety of Air Navigation (EUROCONTROL), Bretigny-sur-Orge, France, 2004.

[23] Smith, L. N., "A Disciplined Approach to Neural Network Hyper-Parameters: Part 1 – Learning Rate, Batch Size, Momentum, and Weight Decay," *arXiv preprint arXiv:1803.09820*, 2018.

[24] EUROCONTROL, "Human Factors Module," https://skybrary.aero/sites/default/files/bookshelf/864.pdf, 2008. Accessed: 14-January-2026.

Part II

Literature Review and Research Definition

2

Literature Review

2.1. Noise Abatement Procedures

Noise Abatement Procedures (NAPs) are operational procedures established by the International Civil Aviation Organisation (ICAO), aiming to provide a global framework to minimise aircraft noise exposure around the aerodrome. The detailed operational guidance for different NAPs is provided in [1]. ICAO defines two aeroplane operating procedures, namely the Noise Abatement Departure Procedures (NADPs) and Continuous Descent Operations (CDO).

2.1.1. Noise Abatement Departure Procedures (NADP)

Within the category of Noise Abatement Departure Procedure, NADPs are further refined into two main types: NADP1 and NADP2. NADP1 aims to reduce the noise impact in close proximity to the end of the departure runway. In this procedure, the initial climb speed before reaching 800 ft above aerodrome elevation should not be less than 10 kts above V_2. Furthermore, a power or thrust cutback is performed between 800 ft and 3000 ft above aerodrome elevation. Within the 800 to 3000 ft range, the climb speed of the aircraft maintains between 10 and 20 kts above V_2. Lastly, the flaps and or slats retraction is delayed until the aircraft reaches 3000 ft above aerodrome elevation and the aircraft climb profile transitions to normal en-route climb speed [1].

NADP2 is designed to reduce noise impact in areas further away from the end of the departure runway. In this procedure, the initial climb speed before reaching 800 ft above aerodrome elevation should be between 10 to 20 kts above V_2. The flaps and or slats retraction is initiated between 800 ft and 3000 ft above aerodrome elevation. Within this altitude range, the power or thrust cutback is also performed, while ensuring satisfactory acceleration performance. Lastly, once the aircraft reaches 3000 ft above aerodrome elevation, the aircraft's climb profile transitions to normal en-route climb speed [1]. An example of NADP1 and NADP2 is shown in Figure 2.1 and Figure 2.2 respectively [1].

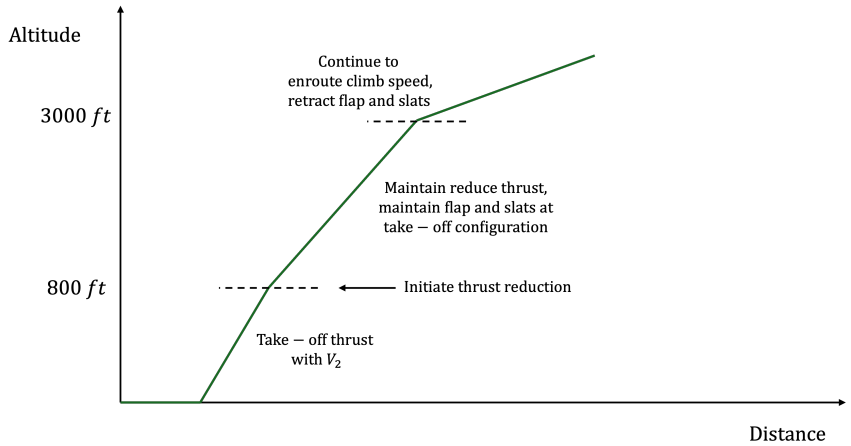


Figure 2.1: Example of NADP1 Procedure Adapted from ICAO [1]

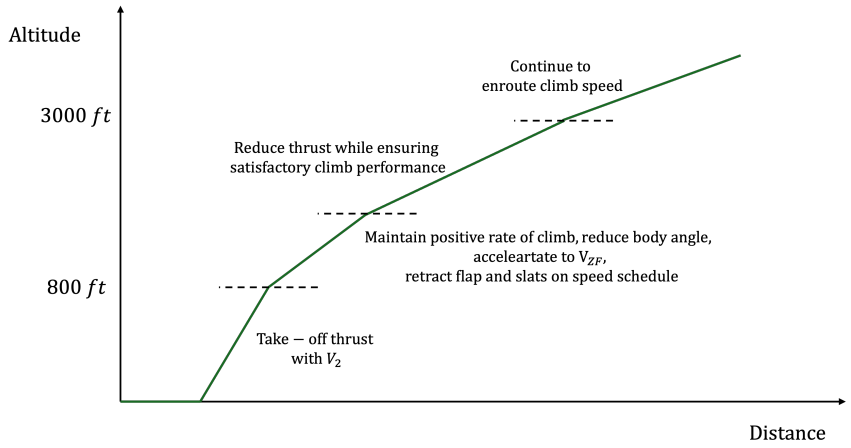


Figure 2.2: Example of NADP2 Procedures Adapted from ICAO [1]

The major difference between NADP1 and NADP2 is that flap/ slats are retracted at a later stage of the climb, resulting in a later initial acceleration altitude. Additionally, in 2007, ICAO established 14 different NADP profiles based on interviews with airlines and modelling of flight trajectories, resulting in six NADP1 profiles and eight NADP2 profiles [2]. These 14 different NADP profiles are based on different initial acceleration altitudes, thrust cut back altitudes and flap/ slats retraction altitudes.

The Dutch government has particular environmental regulations for Amsterdam Airport Schiphol operations, which are depicted in the Schiphol Airport Environmental Information Regulations (in Dutch: Regeling milieu-informatie luchthaven Schiphol (RMI)) [3]. The RMI defines the environmental and operational data that the airport operator and air navigation service provider must record and report to regulatory authorities, including noise levels, aircraft movements, and emissions. To ensure standardised and consistent reporting, the RMI references the technical data provided by Royal Netherlands Aerospace Centre (NLR) in [4]. The technical data in [4] includes a set of fixed-point profiles (FPPs) for various aircraft operations including NAPs. Within each fixed-point profile, the aircraft type, operation type, distance from the runway, stage length, altitude, true airspeed (TAS) and thrust setting are provided. The generated FPPs are derived from the median flight trajectories observed from flight trajectory data [5]. However, using these representative profiles limits the ability to model and classify individual flights according to their actual NAP operations.

For Amsterdam Airport Schiphol operations, further distinct subcategories of the NADP2 type are defined by the NLR, which are the sub-variants of 800 ft, 1000 ft and 1500 ft. These sub-variants describe the initial acceleration altitude to be at 800 ft, 1000 ft and 1500 ft. However, in the FPPs, these 800 ft, 1000 ft, and 1500 ft sub-variants assume the thrust cutback and flaps/ slats retraction occur simultaneously at the prescribed altitude. Furthermore, comparing the FPP to the ICAO definitions, both NADP1 and NADP2 profiles do not exhibit an increase in speed from V_2 to 800 ft, the speed is assumed to be constant. Based on the Aeronautical Information Publication (AIP) from Luchtverkeersleiding Nederland (LVNL), the default departure operation is to perform an NADP2 operation, unless specified otherwise [6]. However, as of the status quo, there are no accurate measures to monitor compliance and distinguish which flight trajectory is operating under NADP1, NADP2 800 ft, NADP2 1000 ft and NADP2 1500 ft.

2.1.2. Noise Abatement Descent Procedures

For NAPs relating to descent, ICAO mainly specifies and provides a framework for the concept of CDO. ICAO describes CDO as an operation that enables arriving aircraft to descend continuously by employing idle thrust in a low drag configuration before the final approach fix (FAF) [7]. ICAO also describes the optimum scenario as initiating a CDO at the top of descent. In the RMI, a particular subset of CDO is considered, which is the Continuous Descent Approach (CDA). Based on the RMI, a flight is compliant with CDA operations as long as, from below 3500 ft and above 1500 ft, the aircraft has no level segments. In addition to that, from the runway threshold, the aircraft should always be above an arbitrarily drawn 2.5° glideslope line.

Despite the lack of an exact definition of the reduced-flap setting approach from ICAO, ICAO [1] stated that the full landing configuration should not be deployed until less than 5 nm from the threshold of the intended landing runway [1]. This approach aims to minimise aerodynamic drag and, consequently, reduce engine thrust and associated noise during the earlier segments of the approach. Based on the RMI and the supporting technical documentation provided by NLR [3, 4], a more specific definition is provided for Amsterdam Airport Schiphol operations. As stated in the RMI, the reduced-flap setting approach is defined as when an aircraft performs the final approach without deploying the full landing flap configuration. When comparing the full flap setting approach to the reduced-flap setting approach in the FPPs, the only difference is in the minimum TAS. However, purely based on the modelled minimum TAS does not allow accurate classification, as flight conditions vary leading to varying minimum TAS.

2.2. NADP Profile Modelling and Clustering

Within the FPP generated by the NLR, for each aircraft type, and depending on the stage length, the distance from departure to arrival, there is one modelled NADP1, NADP2 800 ft, NADP2 1000 ft and NADP2 1500 ft flight trajectory. Hence, to further improve the potential to model various trajectories for each sub-variant of NADP operation. NADP modelling and classification techniques are researched.

Several studies have specifically focused on the modelling and classification of NADPs. Lim [8] investigated methods to improve the modelling of NADP profiles for environmental impact assessment. Building on this work, Behere et al. and Bhanpato et al. extended the research by developing and applying clustering and classification algorithms for the analysis of NADP [9, 10, 11].

Building upon the established NADP framework defined by ICAO, Lim [8] focused specifically on enhancing the definition and representation of NADP profiles. In the study, Lim noted that all airlines were using procedures that were combinations of a set of discrete NADP profiles, and that all airline departure procedures could be classified as either NADP1 or NADP2. To further model new departure profiles in addition to the 20 NADP profiles described in [2], the proposed method models the profiles based on the final speed schedule at FL100, the flap schedule, and the energy share percentage. This approach minimises variations due to aircraft weight, atmospheric conditions, and thrust settings. In the paper, Lim described the energy share percentage as shown in Equation 2.1.

$$Energy\ Share\ \% = \frac{\Delta Kinetic\ Energy}{\Delta Kinetic\ Energy + \Delta Potential\ Energy} \tag{2.1}$$

With the insight from Lim [8], Behere developed a consensus clustering model to cluster NADP profiles [9]. The consensus clustering model allows multiple individual metric-based clustering algorithms to combine into a single clustering solution. This enables multiple perspectives of the data to be incorporated into the algorithm, providing greater dimensionality and insight into the NADP profiles. The model employs three main categories of similarity metrics, namely performance-based, emission-based, and noise-based. The different categories allow the clustering process to consider not only the flight trajectory characteristics but also their environmental and operational implications. The schematic of the methodology developed by Behere for the consensus clustering model is shown in Figure 2.3.

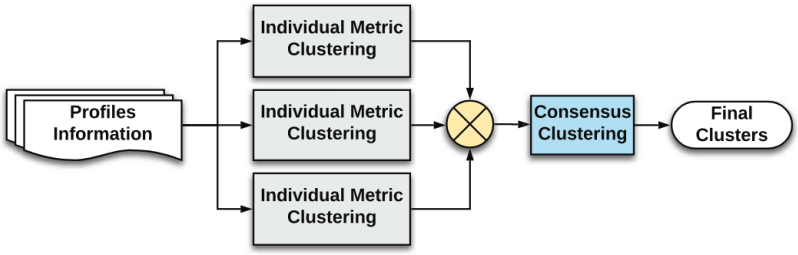


Figure 2.3: Consensus Clustering Schematic developed by Behere [9]

Bhanpato further developed the NADP classification model. Bhanpato combined trajectory similarity classification with a layer of trajectory clustering [10]. The trajectory clustering incorporates distance, altitude and rate of climb to cluster similar trajectories together. With the combination of both metrics, a consensus NADP classification is obtained by mapping the trajectory to the closest median reference trajectory.

Behere further developed another model using take-off time-series trajectory data to classify the NADP profiles flown [11]. Firstly, Behere resamples time-series trajectory data into a common basis based on cumulative ground track distance in nautical miles. Furthermore, based on the pre-computed NADP trajectories, each flight trajectory is assigned a similarity score based on the altitude root mean square error (RMSE) between the NADP profile and the flight trajectory per cumulative ground track for every NADP profile. Based on the minimum RMSE, the flight trajectory will be classified as an NADP profile. The schematic of the classification model developed by Behere is shown in Figure 2.4.

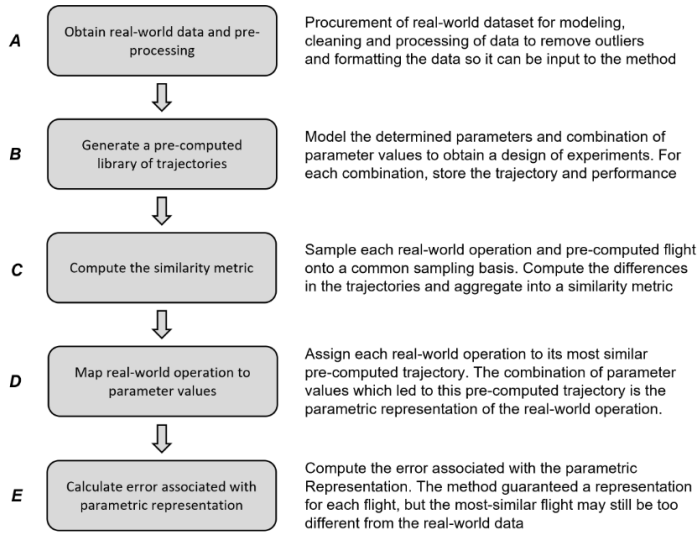


Figure 2.4: Trajectory Classification Schematic developed by Behere [11]

While all four studies [8, 9, 11, 10] present insights into the modelling and classification of NADP profiles, the modelling of NADP profiles remains dependent on the use of the Federal Aviation Administration (FAA) developed support tool, the Aviation Environmental Design Tool (AEDT). In addition, the proposed classification methods [11, 10] employ trajectory similarity and do not investigate the detection of key NADP events such as flap retraction, thrust reduction and acceleration.

2.3. Configuration Estimation

In the previous sections, different classification and modelling methods are aimed towards NADP profiles. To identify reduced-flap setting approaches, the detection of flap setting is a necessity. Nichols, Jarry, and Bhanpato each established a data-driven model to estimate aircraft configuration using various machine learning algorithms [12, 13, 14].

Nichols proposed a different supervised learning method to estimate aircraft state using surveillance data [12]. The supervised learning models were trained from data obtained from the FAA level-D qualified full-motion flight simulator, with inputs that are obtainable from ADS-B data. Within the research, supervised learning models such as linear regression, random forest regression and random forest classification are explored. In addition to the different supervised learning models, Nichols also investigated the incorporation of the Dynamic Principal Component Analysis (DPCA) technique on the dataset to allow inclusion of the time feature. In conclusion, from the results, the incorporation of DPCA does not lead to higher accuracy in flap estimation, and the best supervised model learning is the random forest classification.

Jarry further developed the concept of using flight data records to estimate aircraft state [13]. Jarry

developed the model based on the Long Short-Term Memory (LSTM) neural network algorithm. The advantage of using the LSTM algorithm compared to the Recurrent Neural Network (RNN) is that the LSTM neural network algorithm performs better for time-series datasets. This is due to the algorithm allowing the neural network model to learn long-term dependencies and not be limited to short-term dependencies. The research of Jarry relies on the model input of radar available data, such as ground speed, vertical speed and altitude. In addition to that, derivable inputs such as specific kinetic energy, specific potential energy and specific total energy. The LSTM model is run on two different types of aircraft for two different airports; both accuracies are relatively similar, hence confirming the model is suitable for different aircraft and airport approaches.

Bhanpato further developed the research on comparing the LSTM neural network algorithm to the Hidden Markov Model (HMM) model [14] for configuration estimation. Hidden Markov model uses the Markov process to infer hidden states from observations. Bhanpato also compared how the difference between input features affects the model configuration estimation. From the research, it is concluded that the LSTM algorithm outperforms the HMM model.

While all three research [12, 13, 14] present different algorithms for aircraft configuration estimation. However, all three research focuses on the arrival procedure, and does not discuss aircraft configuration estimation for departure procedures.

3

Research Activity

In this chapter, the research objective of the project is presented and research questions are formulated to bridge the knowledge gap between the research objective and the literature review shown in Chapter 2.

3.1. Research Question

Research Objective

The primary objective of this research is to develop a model for classifying flight track data into specific noise abatement procedures at Amsterdam Airport Schiphol, to support noise impact assessment and operational compliance monitoring.

This research objective aims to provide an analytical approach towards NAP compliance analysis in comparison to the current interview-based analysis.

Research Question 1

To what extent can Noise Abatement Departure Procedures (NADPs) be accurately classified using fixed-point profiles (FPPs) derived by NLR?

Research Question 1 addresses the applicability of classifying flight trajectories to particular NADP profiles based on the limited trajectories provided in the FPP. In addition, the research question also aims to explore the current dependency on the FAA's AEDT tool for modelling NADP profiles and explore alternative modelling approaches, as motivated by the limitations and gaps identified in [8, 9, 11]. Stemming from research question 1, this leads to multiple sub-questions as shown below:

1. How can the use of NLR's FPP be incorporated to model NADP profile envelopes?
2. How well do the NADP profiles specified in the RMI align with the ICAO NADP profile definitions?
3. How does the analysis of historical flight paths enable more realistic modelling of NADP profile envelopes?
4. To what extent does the incorporation of enhanced Mode-S data improve the NADP profile classification in comparison to radar track data?
5. To what extent does analysing specific energy allow more robust NADP profile modelling?
6. Which performance metrics will aid in improving the NADP profile classification when combined in the consensus clustering?

Research Question 2

To what extent can the combination of RMI regulations and ICAO framework increase robustness in classifying flight trajectories as Continuous Descent Approaches?

Research Question 2 compares the classification fidelity between using ICAO and RMI definition and further compares the combination of the two frameworks together. Stemming from research question 2, this leads to multiple sub-questions as shown below:

1. To what extent is applying only RMI regulations sufficient for CDA analysis?
2. Is purely using the RMI regulation sufficient for noise modelling purposes?

Research Question 3

How does estimating aircraft configuration allow for reduced-flap setting approach classification?

As of the status quo, there are limited literature studies on the reduced-flap setting approach. Research Question 3, aims to combine the knowledge of configuration estimation and further expand on this knowledge to bridge the gap of reduced-flap setting approach detection. Stemming from research question 3, this leads to sub-questions as shown below:

1. How does including a runway identifier improve accuracy of configuration estimation?
2. How can the use of a single airline dataset be generalised for all airline operations?
3. To what extent is training a machine learning model to purely predict final flap configuration sufficient for reduced-flap setting approach classification?

3.2. Evaluation of Research Question

Research Question 1 is addressed by demonstrating that flight trajectories can be effectively classified into NADP profiles despite the limitations associated with the minimal reference trajectories provided in the FPP. This is achieved through the incorporation of an energy-based trajectory similarity metric as one of the primary performance indicators. Furthermore, NADP profile envelopes can be integrated into the FPP-based representations using an initial fusion classification derived from historical flight trajectory data, thereby enhancing the representativeness of the reference profiles.

The developed methodology relies on the inclusion of enhanced Mode-S data across all performance metrics. While the proposed methods were also evaluated using datasets containing only radar track data, the resulting classification performance was found to be insufficient. Finally, the combination of multiple performance metrics within a consensus clustering and weighted fusion framework improves overall classification, while also revealing sensitivity to weight selection between metrics.

Research Question 2 is addressed by evaluating the underlying differences between the RMI regulations and ICAO framework. This is achieved through the development of two separate CDA classifications approaches, followed by their combined application. This result demonstrates that the combined framework yields a more robust classification, where shallow descent paths below 2.5° are no longer classified as CDA, and level segments below 6000 ft are explicitly excluded from CDA classification.

The findings indicate that the application of the RMI regulation alone is sufficient for regulatory compliance analysis. However, for noise modelling purposes, the combined framework or the ICAO framework independently provides a better representation of operational behaviour, hence improving the accuracy of noise modelling outcomes.

Research Question 3 is addressed by demonstrating flap deployment estimation models can be effectively utilised for reduced-flap setting approach classification. This is achieved through the development of an LSTM neural network model for estimating flap deployment, followed by a comparison between the predicted final flap setting and the maximum allowed flap setting. The result demonstrates high classification accuracy for narrow-body jets, indicating trajectory information is sufficient to capture flap deployment behaviour. However, for wide-body jets the accuracy decreases significantly, which is likely caused by the limited data size. The findings also indicate that runway identifier does not improve accuracy of configuration estimation, provided that trajectory data includes ground segments during landing.

Lastly, training machine learning model directly from trajectory data to binary reduced flap classification was found to be insufficient. The model tends to learn highly specific flight behaviours, leading to over-fitting and poor generalisation. Furthermore, such an approach requires the full trajectory to be processed in

a single sequence rather than in segmented batches, resulting in heavy computational complexity and training time.

References

[1] International Civil Aviation Organization. *Procedures for Air Navigation Services — Aircraft Operations (PANS-OPS), Volume I: Flight Procedures*. 5th. ICAO Doc 8168. Volume I — Flight Procedures. Montreal, Canada: International Civil Aviation Organization (ICAO), 2006.

[2] International Civil Aviation Organization (ICAO). *Review of Noise Abatement Procedure Research & Development and Implementation Results*. Preliminary edition Doc 9888. Approved by the Secretary General; discussion of survey results. Montreal, Canada: International Civil Aviation Organization, 2007.

[3] Nederlandse Wetgevende Macht. *Wet luchtvaartwet*. Accessed: 2025/10/25. July 2023. URL: https://wetten.overheid.nl/BWBR0014722/2023-07-01/0.

[4] G. Heppe. *Appendices van de voorschriften voor de berekening van de geluidsbelasting in Lden en Lnight voor Schiphol*. Tech. rep. NLR-CR-96650 L. Version 12.3. Nederlands Lucht- en Ruimtevaartcentrum (NLR), 2014.

[5] S. J. Heblij et al. *Methodenrapport Doc29*. Technical Report NLR-CR-2019-076. Amsterdam, The Netherlands: Netherlands Aerospace Centre (NLR), 2019.

[6] Luchtverkeersleiding Nederland. *EHAM - AMSTERDAM/SCHIPHOL Aeronautical Information Publication*. Available:https://eaip.lvnl.nl/web/eaip/AIRAC%20AMDT%2001-2026_2026_01_22/index.html. 2026.

[7] International Civil Aviation Organization. *Continuous Descent Operations (CDO) Manual*. 1st. ICAO Doc 9931, AN/476. First Edition – 2010. Montreal, Canada: International Civil Aviation Organization (ICAO), 2010.

[8] D. Lim et al. "Improved Noise Abatement Departure Procedure Modeling for Aviation Environmental Impact Assessment". In: *AIAA SciTech 2020 Forum*. Jan. 2020. DOI: 10.2514/6.2020-1730.

[9] A. Behere et al. "Aircraft Landing and Takeoff Operations Clustering for Efficient Environmental Impact Assessment". In: *AIAA AVIATION 2020 Forum*. June 2020. DOI: 10.2514/6.2020-2583.

[10] J. Bhanpato et al. "Data-Driven Analysis of Departure Procedures for Aviation Noise Mitigation". In: *Engineering Proceedings* 13.1 (2021), p. 2. DOI: 10.3390/engproc2021013002.

[11] A. Behere et al. "A Method for the Parametric Representation of Take-off Time-Series Trajectory Data for Environmental Impact Assessment". In: *AIAA AVIATION 2023 Forum*. June 2023.

[12] C. Nichols et al. "Estimating Aircraft State from Surveillance Data Using Statistical Learning". In: *AIAA AVIATION 2022 Forum* (2022). DOI: 10.2514/6.2022-3424.

[13] G. Jarry et al. "Approach and landing aircraft on-board parameters estimation with LSTM networks". In: *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT 2020)*. IEEE, 2020. DOI: 10.1109/AIDA-AT48540.2020.9049199.

[14] J. Bhanpato et al. "Aircraft Configuration Prediction for Arrival Operations from Trajectory Data". In: *AIAA SCITECH 2024 Forum* (2024). DOI: 10.2514/6.2024-2618.

A

Sensitivity Analysis of NADP Classification

This appendix contains more information on the sensitivity analysis of the NADP classification, in particularly the correlation between each metrics weight and influence on profile classification.

Further sensitivity analysis was conducted to evaluate the influence of weight assignment on the final fusion classifier used for NADP profile classification. A total of 969 different weight combinations were explored, where each weight was varied in increments and decrements of 0.05. The correlation between each weight and NADP profile is shown through a heat map in Figure A.1.

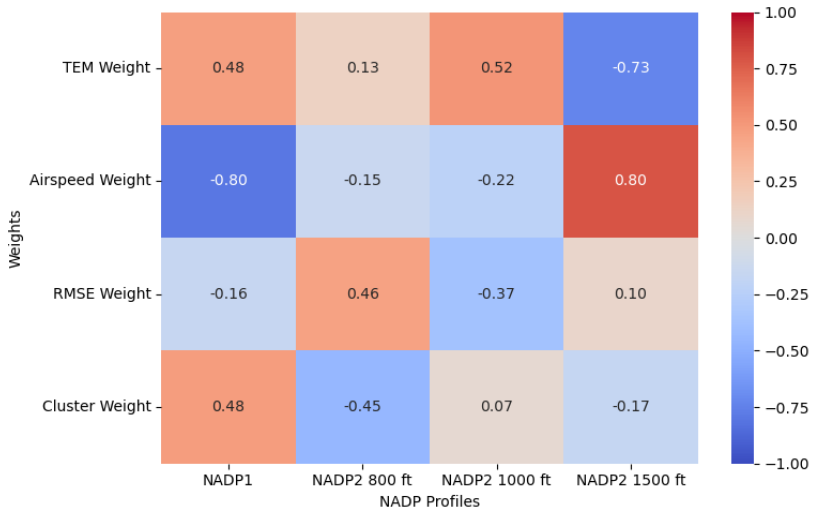


Figure A.1: Weight Influence on NADP Classification

The resulting heat map indicates that the classification outcome is distinctly sensitive to both the airspeed-based metric and the energy-based performance metric. In particular, the airspeed metric exhibits a strong positive influence (0.80) on the NADP2 1500 ft classification, while demonstrating a strong negative influence (-0.80) on the NADP1 classification. This behavior is expected, as the airspeed metric achieves the highest classification rate for the NADP2 1500 ft profile and the lowest classification rate for NADP1.

Similarly, the energy-based performance metrics shows a relatively strong negative influence (-0.73) on NADP2 1500ft classification. While NADP1 and NADP2 1000ft classification depicts moderate positive influence.

Overall, these results highlights that the classification framework can be governed by a trade-off between airspeed based metric and energy-based performance metric, where increasing emphasis or confidence in one metric systematically shifts the classifier between the high-altitude NADP2 classifcation and NADP1.

B

NADP Trajectory Classification

This appendix presents trajectory visualisations for outbound flights. For each NADP profile, a representative trajectory plot is provided based on flights classified under the corresponding profile. Each trajectory is shown as a function of distance and includes key operational markers: the acceleration altitude, thrust cutback altitude, and flap retraction altitude.

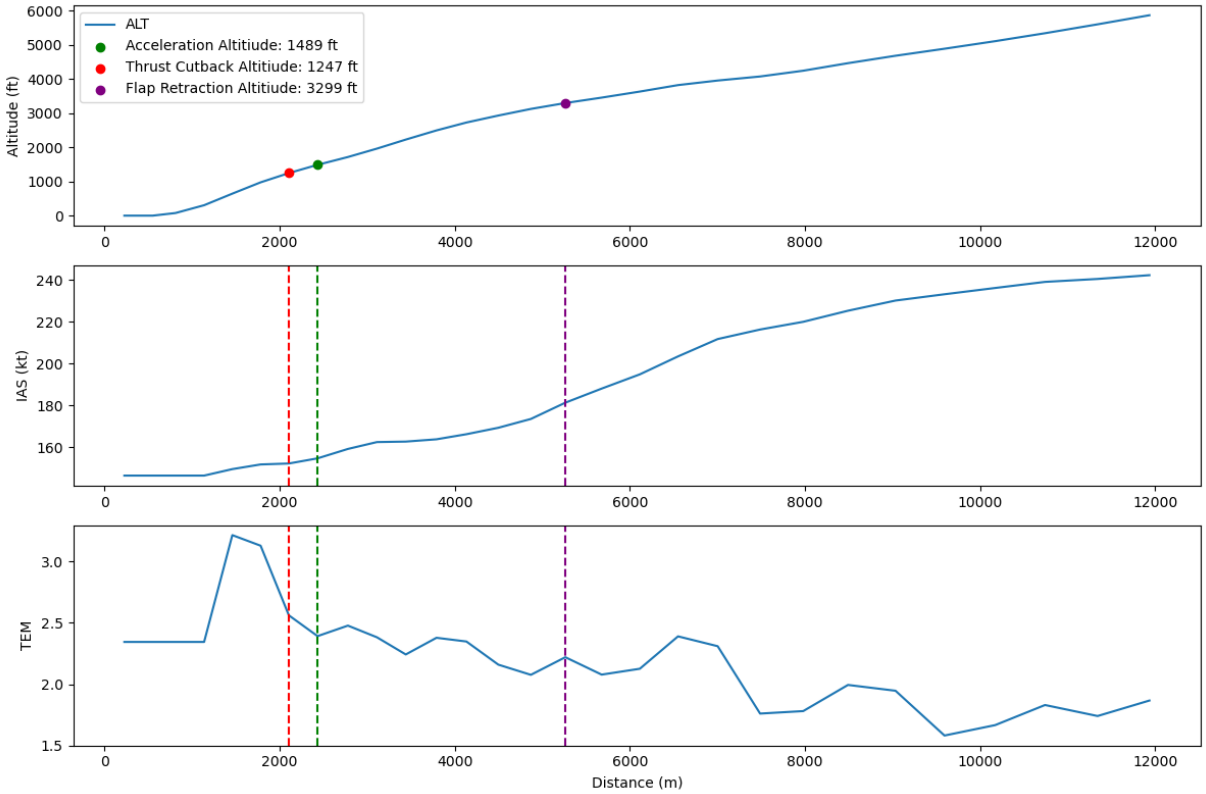


Figure B.1: Sample NADP 1 Flight

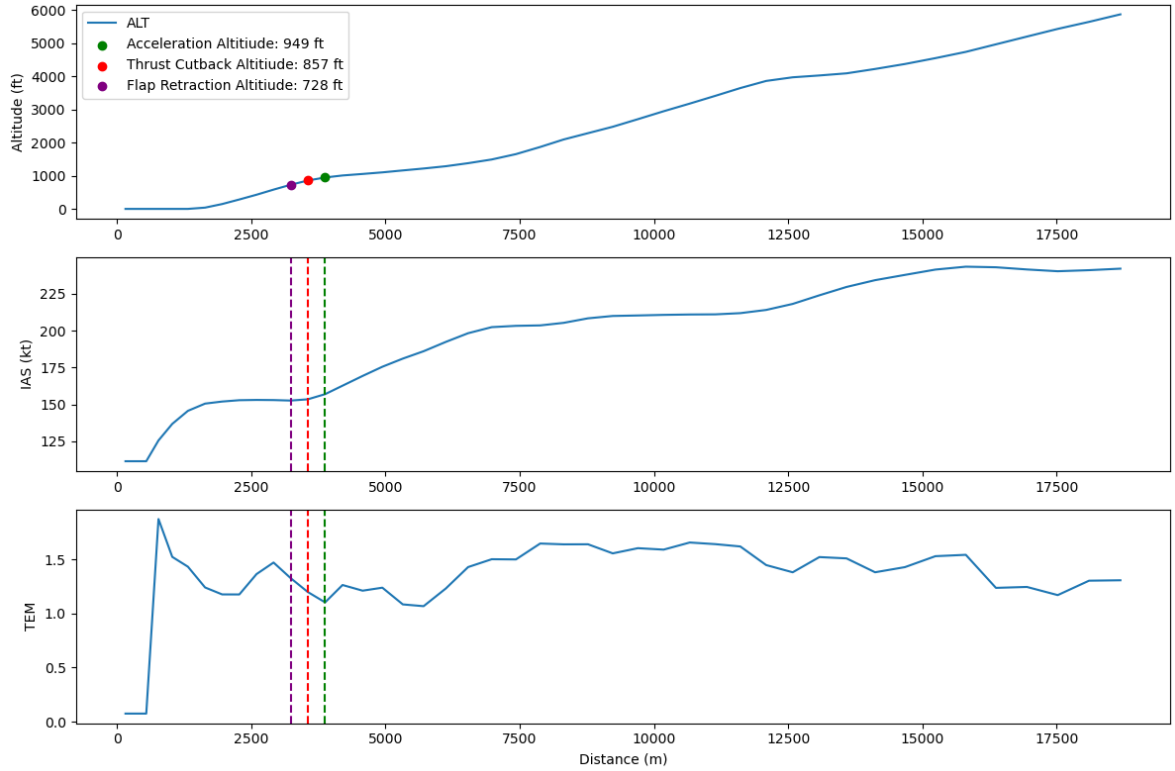


Figure B.2: Sample NADP 2 800 ft Flight

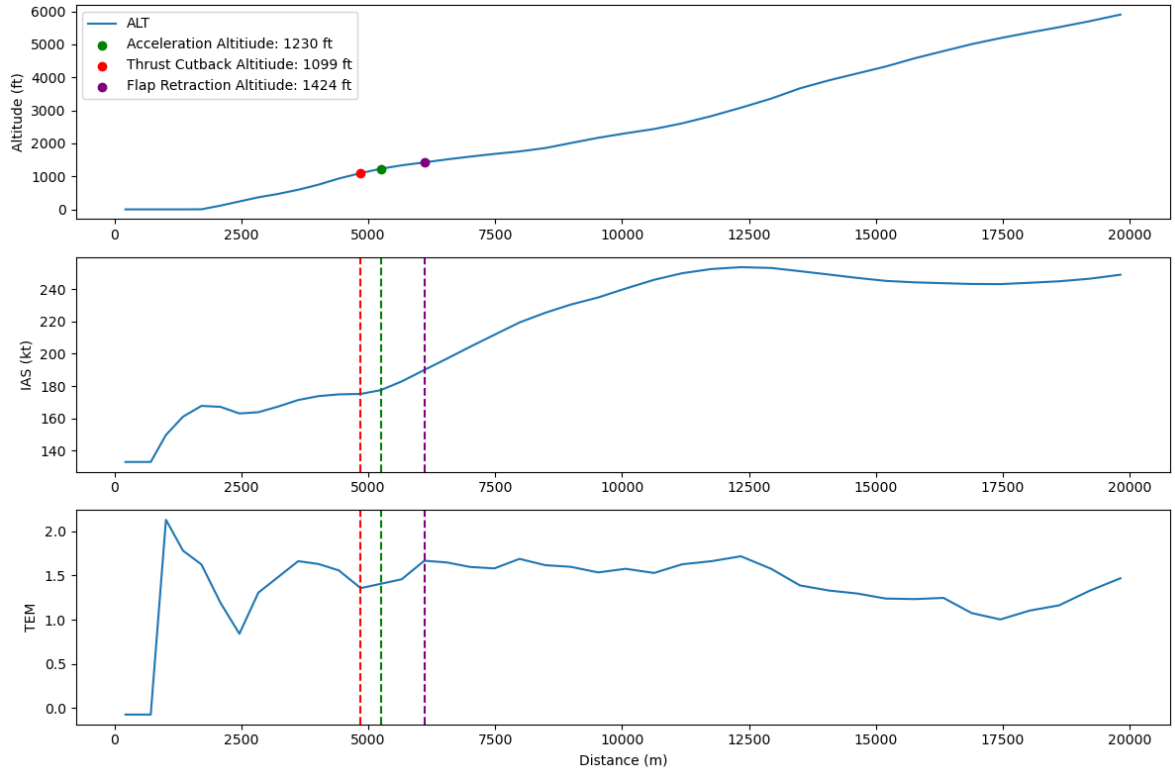


Figure B.3: Sample NADP 2 1000 ft Flight

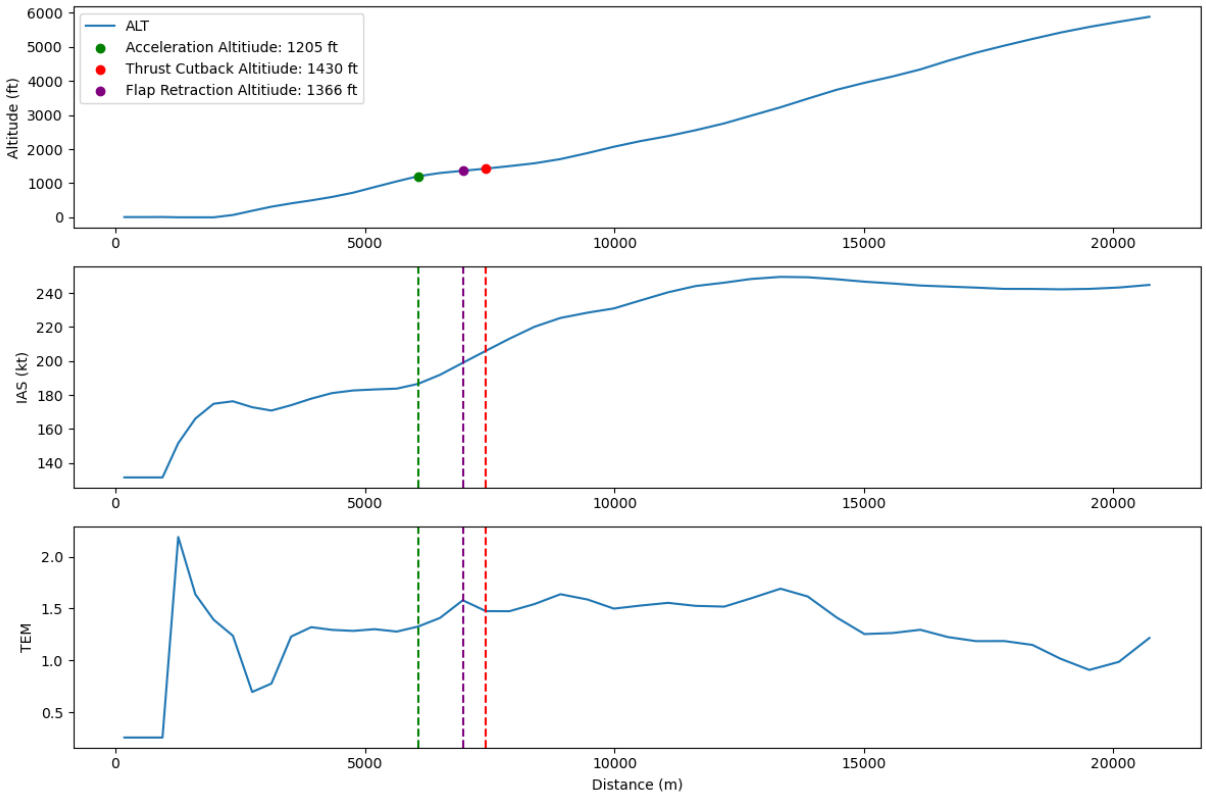
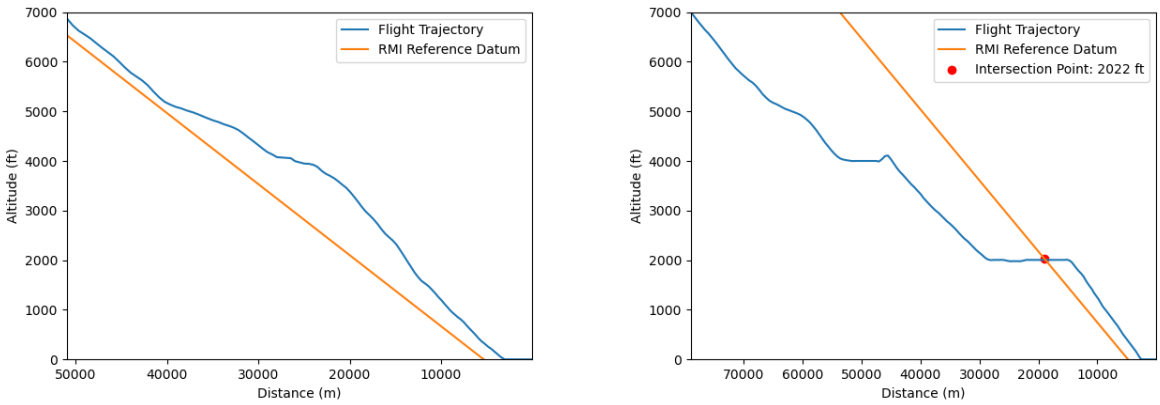


Figure B.4: Sample NADP 2 1500 ft Flight

C

CDA Classification Comparison

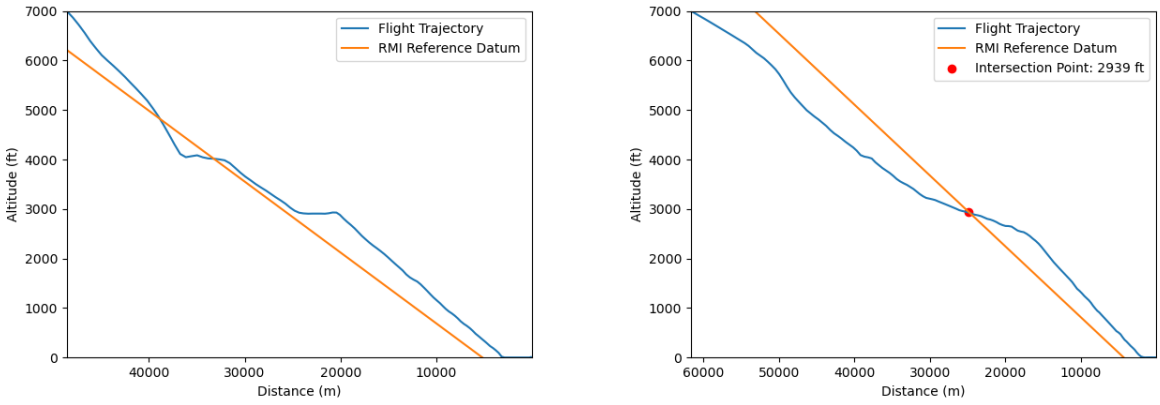
This appendix provides visualisation of the differences between the RMI regulation and the ICAO framework for CDA classification through trajectory plots, as shown in Figure C.1 and Figure C.2



(a) CDA Compliant

(b) Level Flight Compliant

Figure C.1: Cases for Both ICAO and RMI Agreement



(a) CDA Compliant with RMI not ICAO

(b) CDA Compliant with ICAO not RMI

Figure C.2: Cases for ICAO and RMI Disagreement

As shown in figure C.2a, the flight trajectory is compliant with the RMI regulation, as all trajectory points

below 3500ft is above a 2.5° datum line. However, it is not ICAO framework compliant due to the presence of level segment at 4000ft and 3000ft. In contrast, in figure C.2b, the trajectory is compliant with the ICAO framework. Nevertheless, around 3000ft the trajectory falls below the constant 2.5° datum line and therefore it is classified as level flight under RMI regulations.

D

LSTM Model Training

This appendix provides an overview of the training of LSTM models of each aircraft type. The accompanying plots serves as a verification function to confirm that the model performance does not deteriorate as a result of over-fitting.

Over-fitting behaviour is evaluated based on the trend of validation loss over the course of training epochs. A consistent increase in validation loss indicates a presence of over-fitting. In contrast, if the validation loss stabilises or further decreases after an increase, the model is not considered to exhibit sustained over-fitting behaviour.

The training loss, validation loss and combined training score per aircraft type model is shown below. The presented training curves shows no consistent increase in validation loss behaviour, indicating that no sustained over-fitting behaviour occurs during training.

Furthermore, for most aircraft types, training loss continues to decrease even when validation loss stabilises. This behaviour highlights the need to balance both training and validation loss when applying the early stopping mechanism.

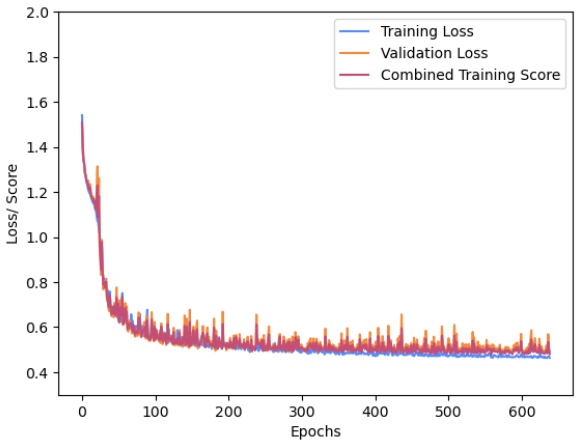


Figure D.1: A21N Training Loss, Validation Loss and Combined Training Score

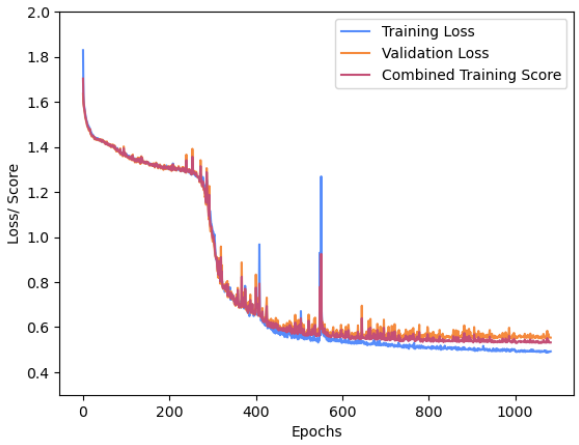


Figure D.2: B737 Training Loss, Validation Loss and Combined Training Score

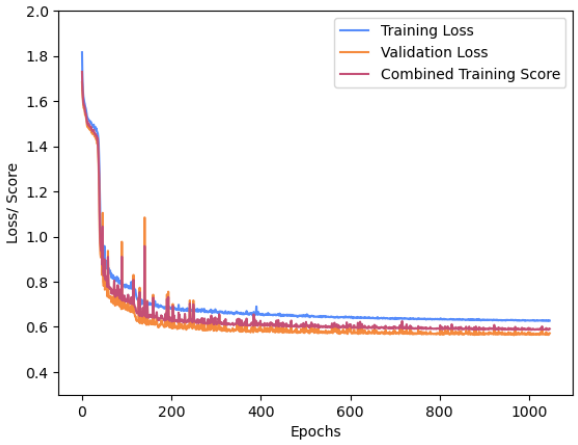


Figure D.3: B738 Training Loss, Validation Loss and Combined Training Score

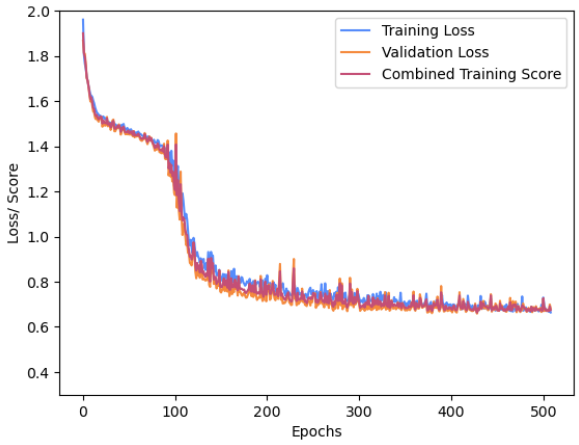


Figure D.4: B739 Training Loss, Validation Loss and Combined Training Score

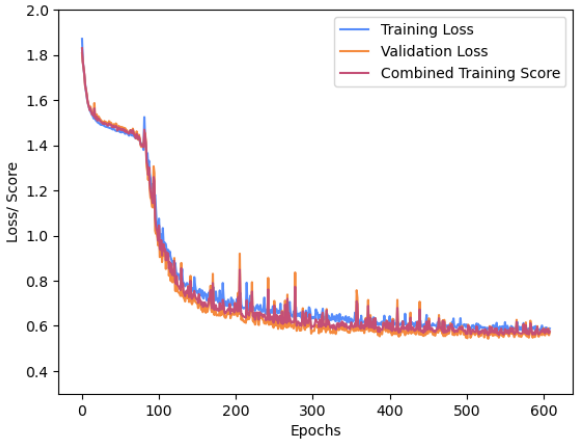


Figure D.5: B772 Training Loss, Validation Loss and Combined Training Score

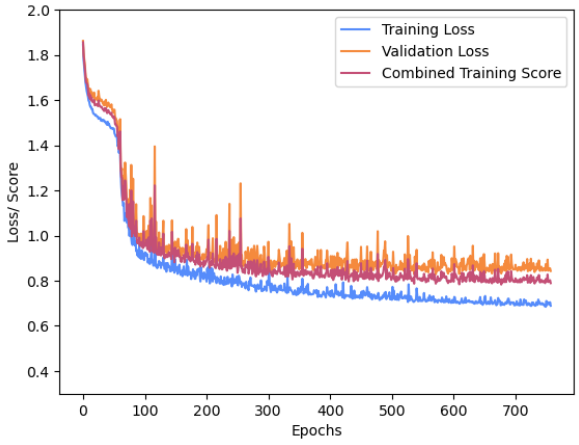


Figure D.6: B77W Training Loss, Validation Loss and Combined Training Score

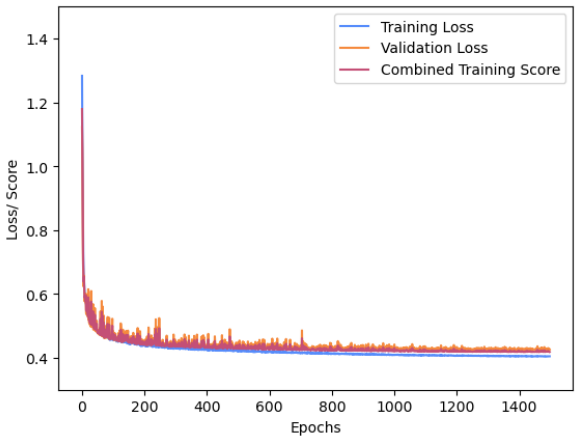


Figure D.7: E190 Training Loss, Validation Loss and Combined Training Score

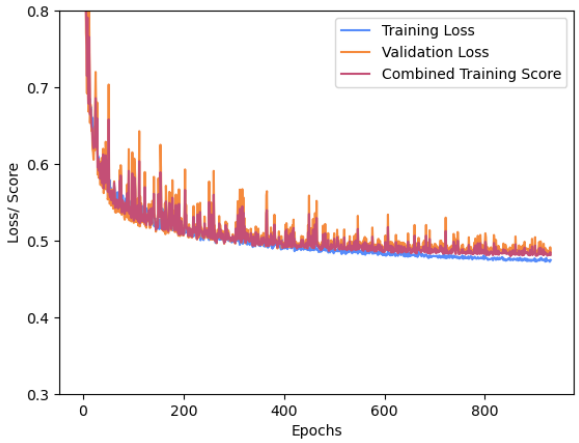


Figure D.8: E295 Training Loss, Validation Loss and Combined Training Score

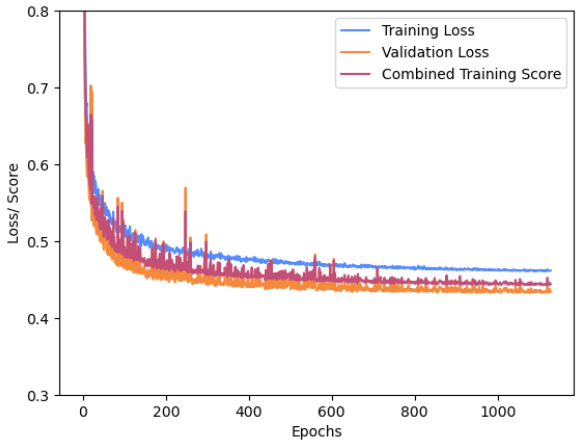


Figure D.9: E75L Training Loss, Validation Loss and Combined Training Score