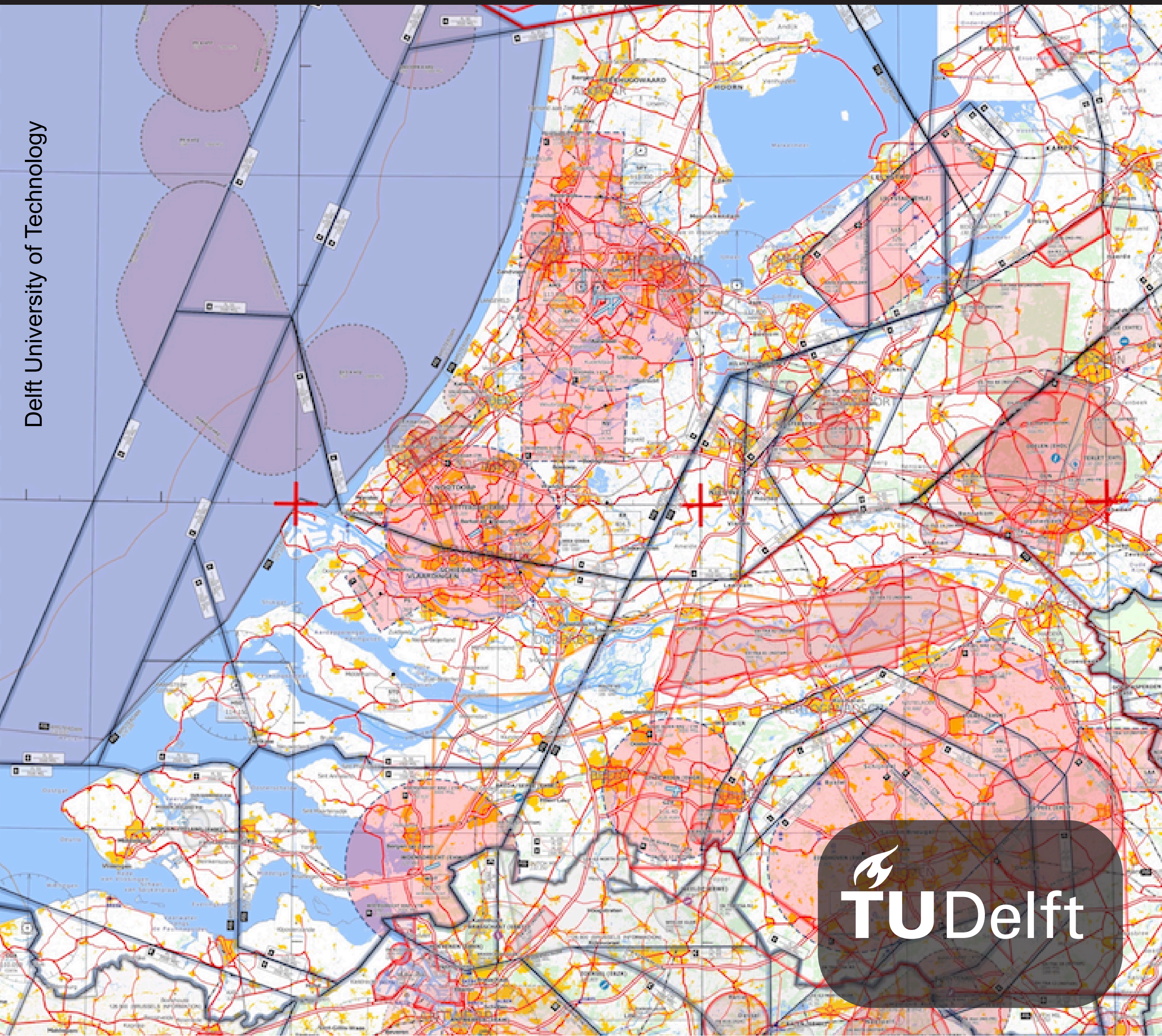
Modeling and Detecting Anomalous Safety Events in Approach Flights Using ADS-B Data

Master thesis

Alberto Bonifazi



Delft University of Technology

TU Delft

This page is intentionally left blank.
Picture credits: openflightmaps

Modeling and Detecting Anomalous Safety Events in Approach Flights Using ADS-B Data

Master thesis

by

Alberto Bonifazi

in partial fulfillment of the requirements for the degree of

Master of Science
in Aerospace Engineering,
Control & Operations

at Delft University of Technology

Delivery date: Thursday 3rd December, 2020

Readers:	Dr. J. Sun	Delft University of Technology
	MSc. G. van Baren	Inspectie Leefomgeving en Transport
	Prof. J. Hoekstra	Delft University of Technology
	Dr. O. A. Sharpanskykh	Delft University of Technology
Student number:	4431057	

TU Delft

This page is intentionally left blank.

Acknowledgements

This thesis concludes five intense years of studies in Delft. At the end of this journey, I can look back satisfied with everything I have achieved. I have to thank Delft University of Technology for the endless opportunities it has to offer and for the great people that it attracts from all over the world. I would like to thank all my friends with whom I shared a part of this experience.

This milestone has been the toughest challenge of my study to overcome. I could not have done it without the help of my supervisors Junzi Sun, Gerben van Baren and Jacco Hoekstra. I appreciated their openness which made asking for feedback and talking about problems very natural. What struck me the most has been their resilience and flexibility in these strange times. This attitude helped a lot in keeping going with the work.

Last, but not least I would like to thank my family for welcoming me back in Bari and supporting me in any way they could throughout my life.

A. Bonifazi
Bari, December 2020

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations related to data mining methods

AE Autoencoder
ARIMA Auto Regressive Integrated Moving Averages
CNN Convolutional Neural Network
DAE Deep Autoencoder
DBN Deep Belief Network
DBSCAN Density-Based Spatial Clustering of Applications with Noise
DT-MIL Deep Temporal Multiple-Instance Learning
ELM Extreme Learning Machines
GAN Generative Adversarial Network
GLOSH Global-Local Outlier Score from Hierarchies
GMM Gaussian Mixture Model
GRU Gated Recurrent Unit
ICA Independent Component Analysis
IF Isolation Forest
KDE Kernel Density Estimation
kNN K-Nearest Neighbours
IMF Inductive Monitoring System
LOF Local Outlier Factor
LoOP Local Outlier Probability
LSTM Long Short-Term Memory
MKAD Multiple Kernel Anomaly Detection
NN Neural Network
OC-SVM One-Class Support Vector Machine
OPTICS Ordering Points To Identify the Clustering Structure
PCA Principal Component Analysis
RBM Riemann Boltzmann Machine
RL Reinforcement Learning
RNN Recurrent Neural Network
SMS-VAR Semi Markov Switching Vector Autoregressive model
STORN Stochastic Recurrent Network
SVM Support Vector Machine
VAE Variational Autoencoder

VAR Vector Auto-Regressive

Abbreviations related to aviation

1090ES 1090 MHz Extended Squitter
ACAS Airborne Collision Avoidance System
ACARS Aircraft Communication Addressing and Reporting System
ADS-B Automatic Dependent Surveillance–Broadcast
AIS Aeronautical Information Service
ANSP Air Navigation Service Provider
APP APProach Control
ASAS Airborne Separation Assurance/Assistance System
ASDE Airport Surface Detection Equipment
ASIAS Aviation Safety Information Analysis and Sharing
ASR Airport Surveillance Radar
ATC Air Traffic Control
ATIS Automatic Terminal Information Service
ATM Air Traffic Management
CTA ConTrol Area
CTR Controlled Traffic Region
DDR Demand Data Repository (Eurocontrol)
EHAM Schiphol (Amsterdam, ICAO code)
EHS Enhanced Surveillance
ES Extended Squitter
FAAFederal Aviation Agency
FAF Final Approach Fix
FDM Flight Data Monitoring
FDR Flight Data Recording
FIR Flight information Region
FIS Flight information Service
FL Flight Level (= each 100 ft above 1013.25 Pa)
FMS Flight Management System
FOQA Flight Operations Quality Assurance
GNSS Global Navigation Satellite System
GPS Global Positioning System
IAF Initial Approach Fix
ICAO International Civil Aviation Organisation
IFR Instrument Flight Rules
ILS Instrument Landing System
METAR METeorological Aerodrome Report

MLAT MultiLATeration
MLS Microwave Landing System
MSL Mean Sea Level
NAS National Airspace System
NASA National Aeronautics and Space Administration
NextGen Next Generation Air Transport System (USA)
PAR Precision Approach Radar
RADAR Radio Detection And Ranging
SES Single European Sky
SESAR Single European Sky ATM Research
SJU SESAR Joint Undertaking (=the SESAR organisation)
SID Standard Instrument Departure route
SMR Surface Movement Radar
SSR Secondary Surveillance Radar
STAR STandard Arrival Route
SUA Special Use Airspace
SWIM System Wide Iformation Managemen
TCAS Traffic alerting and Collision Avoidance Systems
TMA TerMinal Control Area
TWR Tower, Aerodrome Control
UTA Upper conTrol Area
VFR Visual Flight Rules
WGS World Geodetic System (e.g. WGS 84)

1

Scientific Paper

The scientific paper provides a succinct overview of the work performed and the results obtained. The purpose of the appendices is to expand some of the concepts presented. The paper begins on the next page.

Modeling and Detecting Anomalous Safety Events in Approach Flights Using ADS-B Data

Alberto Bonifazi[1], Junzi Sun[1], Gerben van Baren[2], Jacco Hoekstra[1]
[1]*Control & Operations, Delft University of Technology, Delft, the Netherlands*
[2]*Inspectie Leefomgeving en Transport, Minister van Infrastructuur en Waterstaat, Den Haag, the Netherlands*

***Abstract*—This paper shows that it is possible to produce safety knowledge by mining Automatic Dependent Surveillance-Broadcast (ADS-B) data. The methodology combines exceedance detection and anomaly detection techniques to identify anomalous safety events in approach flights. One of these events is unstable approaches, which are identified with a rule-based algorithm and a Gaussian Mixture Model (GMM). The first model relies on the idea that an aircraft during the final approach needs to be flying within a certain horizontal area. The second one extracts the energy characteristics of the aircraft using ADS-B data, and later trains the GMM which is used for anomaly detection. Also, go-arounds are detected in the data using fuzzy logic with four S-functions to model the dynamics of a go-around. After identifying these events, indicators are constructed by aggregating the results to monitor safety performance. These models are applied to the ADS-B data from 2018 of the Schiphol Airport area in Amsterdam. Thus, it is possible to derive insights for runways and months, and it is possible to combine these indicators with extra variables such as meteorological data.**

***Index Terms*—anomalous safety event, safety monitoring, ADS-B, Schiphol Airport, data mining and anomaly detection**

I. Introduction

Air traffic management is one of the most complex systems that humans have ever created. In this system, ensuring safe aircraft operations has the utmost importance. For this reason, operators perform Flight Data Monitoring (FDM) programs, called Flight Operations Quality Assurance (FOQA) in the United States. To enhance safety, aviation authorities around the world have promoted safety information sharing reporting mechanisms. The most notable ones are the European Co-ordination Centre for Accident and Incident Reporting Systems (ECCAIRS), and the Aviation Safety Information Analysis and Sharing (ASIAS) from the FAA. Although the majority of stakeholders join these initiatives, much of the safety knowledge they generate remains within the boundaries of their organization. The reason is that only serious occurrences must be communicated. Furthermore, it is challenging for researchers to work with because it is confidential, and thus, not-easily and not-widely accessible.

In this context, authorities are responsible to ensure the safety of operations in civil aviation. They monitor its performance through the reporting system, and by building statistics of relevant occurrences. This provides a high-level overview because it is based on safety occurrences, rather than specific aircraft data. Automatic Dependent Surveillance-Broadcast (ADS-B) technology can complement this information, and it is compulsory for all aircraft from 2020. The system consists of a transmitter and a receiver. ADS-B Out, the transmitter, broadcasts continuously, on average every second, the aircraft position, velocity and track angle. [1]

Furthermore, ADS-B data are not encrypted, there are no restrictions on its use, and the receiver is cheap. These characteristics make these data easily retrievable and widely accessible. Combining these data with data-mining techniques allows producing safety knowledge. For safety monitoring, it means that it can be used to gain an independent point of view on aircraft operations.

This paper aims to contribute to the field of safety monitoring with a concrete use case centered on the final approach operations at Schiphol Airport in Amsterdam, The Netherlands. The goal is to produce insights based on the safety knowledge extracted from ADS-B data.

The final approach phase is selected because it is where 65% of accidents occurred between 2011 and 2015 [2]. Furthermore, within this flight phase, the research focuses on detecting two types of anomalous safety events: unstable approaches and go-arounds. The first event is a factor in 14% of the accidents occurring during approach [2]. The second one is a standard procedure that can be initiated by the pilot or ATC for different reasons, such as an unstable approach, conflicting traffic, or adverse weather. For this reason, there is a strong link between go-arounds and anomalous safety events.

Extracting safety knowledge from aircraft data is an active field of research. The industry follows it closely. Currently, they extract safety knowledge from aircraft data using exceedance detection algorithms, which are dependent on some thresholds. The main pitfall of this method is that it fails in detecting unknown events. Instead, researchers focus on anomaly detection techniques, which have the key advantage of detecting those. [3] However, their disadvantage is a high false-positive rate that is generally around 70%.

These two strategies are combined in this paper. Wang et al. [4] focus on exceedance detection and develop a methodology to detect unstable approaches using radar data. The rule-based algorithm is designed using knowledge from safety regulations. For the detection of go-arounds, Proud [5] offers a strategy based on a set of rules defined on the attributes of ADS-B data.

The most well-recognized technique in anomaly detection is the Once Class Support Vector Machine (OC-SVM) variant called the Multiple Kernel Anomaly Detection (MKAD) algorithm, which is developed at NASA by DAS et al. [6].

This algorithm has been tested extensively for approximately 10 years for various phases of flight, and it is able to detect unstable approaches and go-arounds. [7] [8]

However, its focus is on discovering flight level anomalies, and it has shown the best results when used with a single type of aircraft and FOQA data. In this paper, the use case is different since the strategy is applied to ADS-B data and on multiple types of aircraft at the same time. To overcome these limitations, Puranik et al. [9] preprocess the data using energy metrics [10], and instantaneous anomalies in general aviation operations are detected using a Gaussian Mixture Model (GMM).

This paper offers the basis for the development of an independent safety monitoring mechanism. It shows how a novel methodology can detect unstable approaches and go-arounds in ADS-B data. A general overview is shown in Fig. 1. In the following sections, it becomes clearer how all the pieces interconnect. Section II covers aspects regarding the data in use, the preprocessing strategy, and the postprocessing one. Section III and section IV discuss the detection of anomalous safety events using a rule-based algorithm, GMM, and fuzzy logic. Section III is specific to the detection of unstable approaches and section IV to the detection of go-arounds. The findings are presented in section V and section VI. The result consists of a set of safety indicators that can be aggregated in different ways to produce insights.

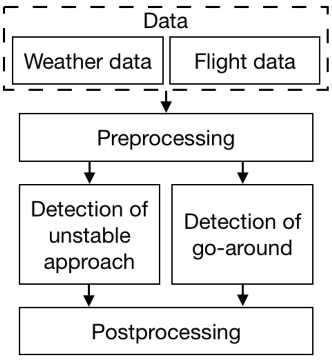


Fig. 1. The work flow of this paper

II. Data

This section provides an overview of the data and a description of the manipulation strategies before and after using the anomalous safety events detection methods.

A. Data Overview

The data consist of flight data and weather data. Flight data include ADS-B data, an aircraft database from OpenSky[1], and Schiphol's Aeronautical Information Services (AIS) publucations[2]. While the weather data comprise of METeorological Aerodrome Report (METAR) reports and Global Forecast System (GFS) data from NOAA[3]. These data are all publicly available and free to download.

ADS-B data constitute the backbone of this research and the most valuable source of information. In this paper, the ADS-B data are collected through the antenna positioned on top of the Aerospace Faculty of Delft University of Technology. ADS-B data provide the timestamp, the unique ICAO identifier of the aircraft, its ground-speed, its rate of climb, its position, and its track angle. [11] The source of most of this data is a satellite radio-navigation system on-board of the aircraft. For this reason, there is a direct relation between the accuracy of this instrument and the accuracy of the ADS-B data.

The aircraft database contains more information on the aircraft under consideration: the manufacturer, the model, the national registration ID, the operator, the owner, and the ICAO aircraft type. The database is particularly useful when preprocessing the data. Also, Schiphol's data from the AIS publication are used during the preprocessing step, and in the rule-based algorithm. In particular, the Instrument Approach Charts and the specific information of the runways are of interest.

METAR reports indicate the weather perceived at the airport on any particular day. In this case, METAR reports from Schiphol are downloaded from IOWA ASOS network[4]. These reports are generated every 30 minutes, and they include many weather variables. The ones that are used in this research are the timestamp, the temperature, the dew point temperature, the wind direction, the wind speed, the pressure, the visibility, the wind gust, the cloud coverage, and the weather codes. However, this information is provided only for the airport. This means that wind speed and direction are accurate only at low altitudes. The GFS data from NOAA is used to complement this information for higher altitudes. This dataset offers wind data at intervals of 700 ft with updates every 6 hours.

B. Preprocessing

In this step, the raw data are cleaned and combined to obtain something ready for further analysis. It will become clear how this research resolves some of the limitations of ADS-B data. A preliminary procedure consists of removing general aviation aircraft, helicopters, and ground vehicles using OpenSky aircraft database.

It is possible to know exactly which aircraft communicated the ADS-B data point because the message includes the ICAO's identifier. However, the same aircraft might land and take-off multiple times on the same day. For this reason, the data points of a particular aircraft are further divided into trajectories. This is a crucial step as we will be using trajectories later in the analysis. Given the fact that we only use the area around Schiphol for the analysis, it is easy to detect inbound and outbound traffic, which gets labeled respectively as approaching, and taking-off traffic. In this way, we can further refine the trajectories used for the analysis by

[1]opensky-network.org/aircraft-database
[2]en.lvnl.nl/information-for-airmen/publications-for-airmen
[3]ncdc.noaa.gov/data-access/model-data/model-datasets
[4]https://mesonet.agron.iastate.edu/request/download.phtml

selecting only approaching traffic. Throughout this process, incomplete trajectories are removed. Furthermore, using data from Schiphol's AIS publication, it is possible to determine the landing runway for each approaching trajectory. This is useful information when building statistics of safety indicators.

ADS-B's track angle data are a valuable source of information, but these are not always accurate. Identifying the causes of this behavior is beyond the scope of this work, but a track angle fix is proposed. It is observed that sometimes when an aircraft performs a go-around, or after landing when it moves along the taxiways, its track angle indicator doesn't follow the aircraft movements. The track angle communicated via ADS-B data doesn't change whereas it is clear it should have. To solve this issue, the track angle information communicated from the ADS-B data is compared to a bearing estimated using its position. If the difference between the two is higher than 60 deg, the estimate is used as track angle data. To limit the influence of poor measurements in the track angle estimation, a window of 40 s is considered and a minimum amount of 5 points.

Furthermore, ADS-B data provide ground speed and barometric altitude. These are dependent on the weather, and thus limit the comparison of aircraft in different meteorological conditions. Nonetheless, it is possible to remove their dependency using METAR reports and GFS NOAA data. The barometric altitude assumes standard temperature and pressure. These two values can be adjusted using the METAR reports, as follows:

$$
\begin{aligned}
T_A &= T + a \cdot h_A \\
P_A &= P \cdot \left(\frac{T_A}{T}\right)^{\frac{-g}{a \cdot R}} \\
P_{A,M} &= P_A \\
T_{A,M} &= T_M \cdot \left(\frac{P_{A,M}}{P_M}\right)^{\frac{a \cdot R}{-g}} \\
h_{A,M} &= \frac{T_{A,M} - T_M}{a}
\end{aligned}
\tag{1}
$$

In (1), P_A, T_A and h_A refer respectively to pressure, temperature and altitude of the aircraft assuming standard atmospheric conditions. The variables P, T, g, a and R are International Standard Atmosphere constants. While, $P_{A,M}$, $T_{A,M}$ and $h_{A,M}$ are the airplane's pressure, temperature and altitude calculated using P_M and T_M from the METAR report.

Ground speed is dependent on the wind speed, and it can be corrected to obtain true airspeed by subtracting ADS-B ground speed from the wind speed. METAR reports are used to correct speed up to an altitude of 100 m, where the difference in wind speed is estimated to be approximately 1.5 m/s. [5] After this point, wind information is extracted from the GFS NOAA data.

Furthermore, it is interesting to understand the relationship between safety anomalous events and the weather. For this reason, it is important to establish a way to assess the severity of the weather with a meaningful score. Thus, EUROCONTROL's ATMAP weather algorithm [12] is implemented in Python. In this metric, a high score corresponds to poor weather with 4 being the threshold for a weather condition that disrupts airport operations. In addition, it determines also if an aircraft is flying in VMC or IMC. [13]

C. Postprocessing

After identifying anomalous safety events, it is possible to construct safety indicators. These indicators combine the knowledge obtained from the data to drive new insights.

In this way, it is possible to monitor how the number of identified events by every strategy changes depending on an extra feature or multiple ones. These features are the weather score, the month of the year, and the landing runway.

In addition, for the horizontal compliance strategy and go-around detection, there are extra specific features. For the first one, there is also the stabilization altitude. While for the second one, there are the distance to the closest aircraft and if the aircraft is reported to be unstable. In the case of go-around detection, the possibility to vary all these features is meant to provide insights on the possible causes of go-arounds.

III. Detection of Unstable Approach

There are many definitions of an unstable approach depending on the operator and the entity. All these definitions have in common an approach that is not aligned with the correct flight path, that is too fast, or too slow. Generally, a definition describes the requirements for a stable approach, rather than prescribing when one is unstable. [14]

Furthermore, there is a link with the meteorological condition because these conditions need to be satisfied at 1000 feet above airport elevation in IMC, or 500 feet above airport elevation in VMC. Otherwise, a go-around should be executed. This paper proposes two strategies to detect unstable approaches: one checks if an airplane operates outside the horizontal boundaries of stable approach operations, and the other analyses if an aircraft is within normal energy bounds.

A. Horizontal Compliance

This method assumes that all approaches in Schiphol are ILS approaches, and for this reason must be flown within 1-dot of the localizer. [14] 1-dot represents a dot on the course deviation indicator (CDI), the definition of dot depends on the instrument. For the ILS-intercept, it corresponds to 1 deg. Instead, for a VOR, a dot corresponds to 2 deg.

It is important to define an area of horizontal stable approach operations for each runway. In this case, information is collected from the Instrument Approach Chart. The ILS-intercept is defined as the line connecting the Final Approach Fix (FAF), Runway Threshold (THR), and Runway Localizer (LOC). A horizontal compliance region is constructed as the area comprised within 1 deg of the ILS-intercept. It results in areas with the same dimension and shape, but positioned differently depending on the runway.

In Fig. 2, it is possible to visualize the horizontal compliance area for each runway in gray. It is possible to see also the

[5] http://euanmearns.com/high-altitude-wind-power-reviewed/

runway in blue, the decision gate at 1000ft in sky blue, and the one at 500ft in green. As in the regulations, an aircraft is unstable if it stabilizes after 1000ft in Instrument Meteorological Condition (IMC), and after 500ft in Visual Meteorological Condition (VMC).

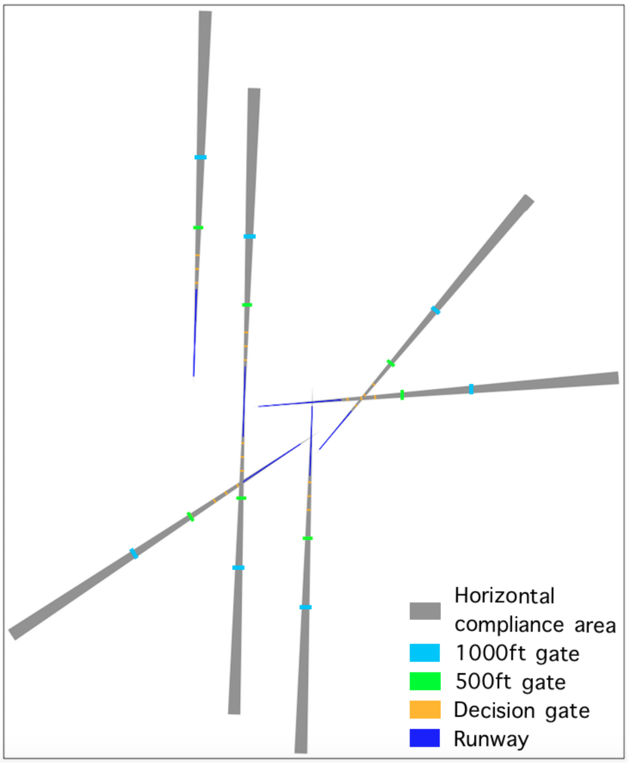


Fig. 2. The horizontal compliance area in gray corresponds to the area in which an aircraft needs to be flying once it intercepts the ILS. An aircraft performs a stable approach if it intercepts the ILS before the 1000ft light-blue gate in IMC, and before the 500ft green gate in VMC.

B. Energy Compliance

This strategy starts from the assumption that an unstable approach has abnormal energy levels, which is quite common in literature [15] [16] [17]. Thus, a set of energy features is derived from ADS-B data. Subsequently, anomaly detection is performed using a GMM. This is an unsupervised learning model, as it does not require any a priori knowledge of what is anomalous.

The key advantages of the GMM over other unsupervised learning models are that it can be trained directly on multivariate data and it outputs its probability of being normal. What a GMM does is clustering normal operations together using a weighted sum of Gaussian component densities. The strategy is composed of two steps: creating energy features and anomaly detection.

1) Energy Features Generation: The energy features are the following: specific total energy (e), specific kinetic energy (e_k), specific potential energy (e_u), energy angle (γ_p), specific rate of total energy ($\dot{e}$), specific rate of kinetic energy ($\dot{e}_k$) and specific rate of potential energy ($\dot{e}_u$).

These features are similar to the ones used by Puranik and Marvis in their work. [15] Since the mass of the aircraft is not available from the ADS-B data, these features represent the aircraft's specific energy and its change. In addition, the energy angle is a measure of how the flight-path angle can change given the current energy level. [18] These metrics are computed as follows:

$$
\begin{aligned}
e &= e_k + e_u \\
e_k &= 0.5 \cdot v^2 \\
e_u &= h \cdot g \\
\gamma_p &= \arcsin\left(\frac{roc}{v}\right) + \frac{\dot{v}}{g} \\
\dot{e} &= \dot{e}_k + \dot{e}_u \\
\dot{e}_k &= v \cdot \dot{v} \\
\dot{e}_u &= roc \cdot g
\end{aligned}
\tag{2}
$$

In (2), v corresponds to the true airspeed, h to the altitude above the ground. These are both computed as explained in the preprocessing section. Furthermore, g is the gravitational acceleration constant near Earth's surface and roc is the rate of climb.

Once the energy features are generated, the data are re-sampled, which consists of obtaining a new representation suited for the use with GMM. In this case, data are sampled in space using a median window comprising 0.5 NM, which corresponds roughly to 12 sec. The time interval varies greatly depending on the flight stage. The analysis considers aircraft flying between 0.5 to 10NM from the runway threshold. This is the area interested by final approach procedures, as it can be seen in the instrument approach charts. The last portion between 0 and 0.5 is not used because it is highly inaccurate. Fig. 3 shows how the specific potential energy, the specific kinetic energy and the specific total energy vary during the approach phase.

2) Anomaly Detection: The anomaly detection step is performed with a model based on a GMM. Because aircraft's behavior changes while performing a final approach, the model comprises three separate GMM and it will be referred to as *3-GMM model*. In this way, each GMM learns how the aircraft behaves in a particular phase. Otherwise, if a single GMM is trained on the full spectrum of operations, the model of aircraft behavior would be too general, and it would fail at recognizing anomalies.

The first GMM learns what means being stable during the final approach. In practice, this GMM is trained on aircraft flying at a distance from the runway threshold between 0.5 NM and 4 NM. This covers the most interesting aircraft operations because it corresponds with the decision area described by the Flight Safety Foundation as it includes the gates of 500ft and 1000ft. The second GMM learns how aircraft descend and intercept the ILS. It comprises the final approach area that goes at a distance from the runway threshold between 4 NM to 7 NM. The last GMM learns of a broader spectrum of aircraft behaviors as the area expands 3NM starting at a

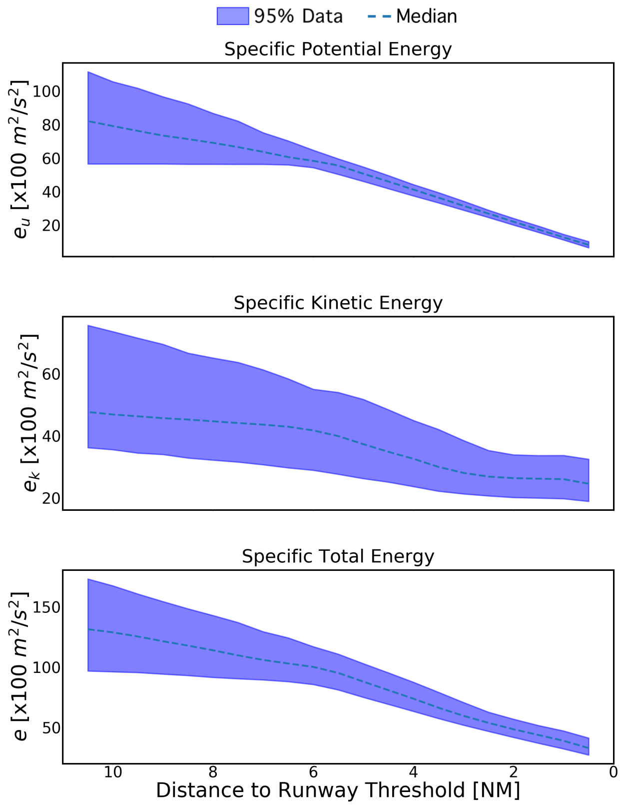


Fig. 3. Variation of the specific potential energy, the specific kinetic energy and the specific total energy depending on the distance to runway threshold

distance of 7NM from the runway threshold. At this stage, an aircraft could be descending, turning, or flying level.

The features selected to train the 3-GMM model include the energy features and two more features: time to runway threshold and distance to threshold. These are particularly important because each energy level is closely linked to a particular position and time before landing.

A GMM is fully defined by the following equation:

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{k} w_i \mathrm{g}(\mathbf{x}|\mu_i, \Sigma_i) \qquad (3)$$

Equation (3) represents a parametric probability density function that is a weighted (w_i) sum of Gaussian components. $g(\mathbf{x}|\mu_i, \Sigma_i)$ indicates a single component, where Σ is the covariance matrix that captures the relation between the different features. k is the number of components in the mixture. The majority of these parameters can be obtained from the data using the expectation-maximization (EM) algorithm. Only two need to be specified beforehand: the type of covariance matrix and the number of components. This approach is very common in literature [9] [19].

The type of covariance matrix has direct consequences on the shape of the Gaussian components. [9] and [19] use a diagonal covariance matrix because of the lower computational cost, which restricts the shape of the Gaussian components forcing it to be oriented along the coordinate axes. During the experimentation process, the computational time of using any kind of covariance matrix is reasonable. Thus, this thesis uses a full covariance matrix for the 3-GMM model such that the components are free of assuming any shape.

The number of components is chosen using the Calinski–Harabascz (C-H) index, as suggested by Puranik in [9]. This is an internal evaluation criterion that measures how compact components are and how well separated they are.

Based on this information, the 3-GMM model can be trained and it can be used for anomaly detection. Every time the 3-GMM model is trained, the data are normalized and only 95% is used. This is an attempt to remove anomalies from the training data. In this way, the 3-GMM model should be more sensible to anomalies once it is tested.

After the 3-GMM model is trained, every data point has a certain probability of belonging to the distribution. An anomaly threshold is selected such that contains points with the least probability of being normal. The percentage of points included can be changed depending on the need. Puranik [9] suggests to use 0.05%, or 0.1%. The anomaly detection goes as follows:

- Each trajectory is fed to the 3-GMM model, which will return the probability of being normal for each of its points.
- If the probability of a point is above the anomaly threshold, the point is labeled as anomalous.
- A trajectory is considered anomalous if at least two points are anomalous in the last 7 NM.

IV. Detection of Go-Around

A go-around is a normal operation that occurs when landing is aborted. However, pilots don't perform it very often, which leads to two risk scenarios. *A*: If the go-around is executed, it is a deviation from the intended operation; and *B*: Pilots may refrain from performing this maneuver. The risk in scenario *A* arises from the fact that performing this maneuver increase the workload for the pilots. Furthermore, accommodating a go-around in busy airspace also increases the workload for air traffic controllers.[6] In scenario *B*, although guidelines recommend executing a go-around, pilots decide not to. This means that the approach is continued and may remain unstable. Thus, it can lead to loss of control, runway excursion, or controlled flight into terrain.

On top of the inherent risks of go-arounds, it is important to understand what are the circumstances of a go-around because they are potentially anomalous safety events themselves. Generally, these are an unstable approach, conflicting traffic, or adverse weather.

The detection of go-arounds from ADS-B data follows a two-steps approach: identification of a possible go-around and evaluation of the go-around score.

[6] https://www.skybrary.aero/index.php/Go-around_Execution

1) Identification of a Possible Go-Around: A go-around consists of climbing to a predetermined altitude prescribed in the instrument landing procedures, and once at the correct altitude turning 360 degrees around the runway. Thus, a possible go-around is identified when an airplane changes its course and starts climbing.

This step employs the technique developed by Sun et al. in [20]. Applying this algorithm allows identifying rapid changes in aircraft behavior. Sun's phase detector can distinguish between 5 different phases: climb (CL), ground(GND), descent (DE), level (LVL), and cruise (CR). As shown by Proud [5], this is particularly useful in case of a go-around as the algorithm detects changes from DE to LVL/CL.

The change in phase is the indication for the beginning of a possible go-around, this will be referred to as the *starting position* in the text. Four ADS-B variable are analyzed: rate of climb, altitude, ground speed, and track angle. During a go-around, these variables are expected to change in a very specific way.

The rate of climb and the altitude are an indication for the aircraft climbing and gaining altitude. It is expected that these two variables will be changing immediately after the *starting position*. In particular, the model analyzes the *delta* in the altitude from the *starting position*.

Instead, the ground speed and the track angle are expected to change some time after the *starting position*. Indeed, they will start varying after the initial climb. The track angle keeps changing until the aircraft is aligned with a runway for landing. In this case what is considered by the algorithm is the *delta* in the ground speed and the track angle from the *starting position*.

For these reasons, each variable is monitored for a specific interval after the *starting position*. The rate of climb and the altitude are analyzed from the *starting position* to the next 2 minutes. Whereas the ground speed and the track angle are monitored from the *starting position* to the next 10 minutes. This value is chosen because a go-around typically adds a flight delay of this amount.[7] A moving average with a window of 15 seconds is used to reduce the susceptibility of the model to outliers.

2) Evaluation of Go-around Score: The idea is to evaluate how much the data resemble a go-around. We describe the expected behavior of these variables using 4 S-functions, which output 1 for maximum similarity and 0 for none. Every variable has its specific S-shape function with the following general definition:

$$S(x; a, b) = \begin{cases} 0, & \text{if } x \leq a \\ 2\left(\frac{x-a}{b-a}\right)^2, & \text{if } a \leq x \leq \frac{a+b}{2} \\ 1 - 2\left(\frac{x-b}{b-a}\right)^2, & \text{if } \frac{a+b}{2} \leq x \leq b \\ 1, & \text{if } x \geq b \end{cases} \tag{4}$$

[7]https://www.casa.gov.au/safety-management/advice-air-travellers/go-arounds

In (4), a is the lower bound of the sloped part of the curve, while b is the upper bound. Fig. 4 offers a visual interpretation of the S-function. In this case, the functions for each variable are defined as follows, where delta represents the difference between the current position and the *starting position*:

$$\begin{aligned} \Delta X(m) &= \mathrm{S}(m; 30, 300) && [deg] \\ \Delta H(n) &= \mathrm{S}(n; 100, 1000) && [ft] \\ \Delta V(p) &= \mathrm{S}(p; 5, 80) && [kn] \\ ROC(q) &= \mathrm{S}(q; 10, 1000) && [fpm] \end{aligned} \tag{5}$$

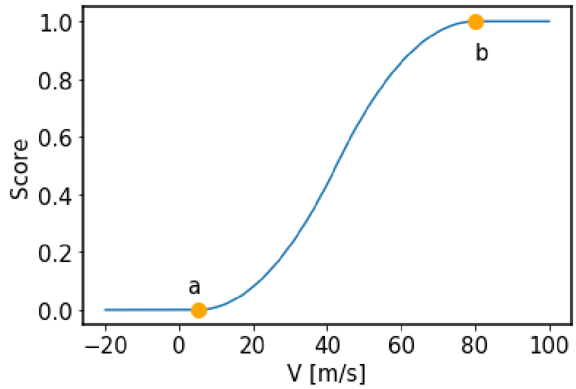


Fig. 4. S-function of the ground speed

These 4 S-functions in (5) represent a simplified version of the behavior of an aircraft during a go-around, where X corresponds to the track angle, H to the altitude, V to the ground speed, and ROC to the rate of climb. By plugging the actual data in these functions, we obtain scores. Mathematically, there should be points in the trajectory considered where these functions score 1. This is not always the case as ADS-B data are prone to errors. For this reason, the model flags a trajectory as a go-around in case the average of the maximum score obtained from the 4 S-functions is higher than 0.5.

Fig. 5 provides direct insights into the working mechanism of this strategy because it shows how the ADS-B variables evolve during a go-around procedure with the red horizontal line corresponding to the 0.5 detection threshold. In particular, time zero corresponds to the starting position that is the point in which the phase changes from descent (DE) to climb (CL)/ level (LVL). Overall it is possible to see that Sun's phase detector reveals that during a go-around the phase changes from climb to level, and then it returns to descent. In this case, a go-around is clearly detected because all variables have at least a point above the 0.5 detection threshold. As expected, the altitude and the rate of climb vary rapidly.In the rate of climb plot, there is a sudden jump. This is not surprising as this variable represents the instantaneous change in altitude. Furthermore, once the pilot has reached the go-around altitude, this variable drops as the pilot will not increase the altitude anymore. Whereas, the altitude score remains high for all the go-around procedure, and then starts decreasing gradually

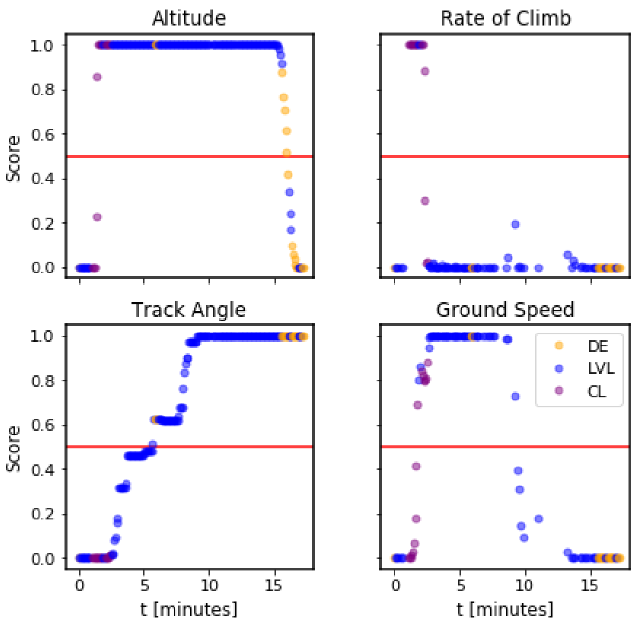


Fig. 5. Evolution of the scores of the various ADS-B variables during a go-around with the red line corresponding to the detection threshold

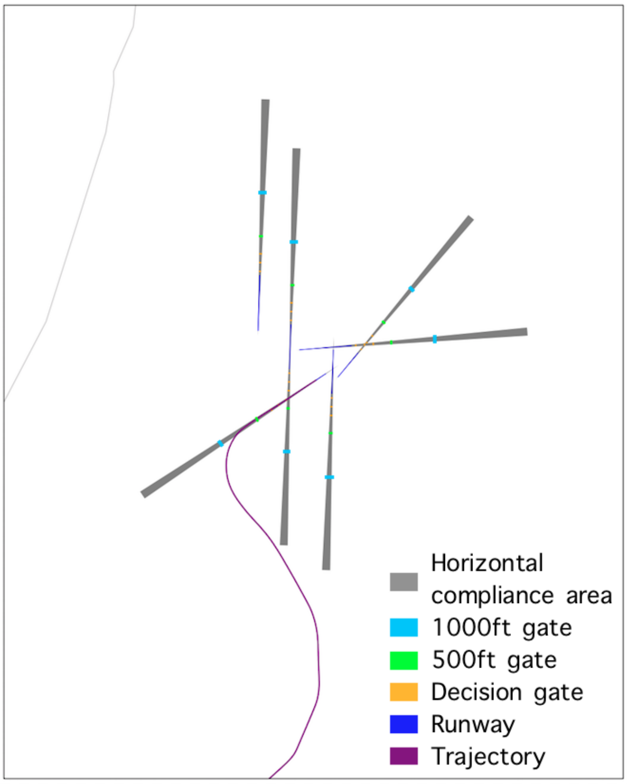


Fig. 6. A flight that stabilizes after the 1000ft light blue gate and before the 500ft green gate.

during the next descent. It is interesting to see how rapidly the altitude score changes in the climbing part compared to the descending one. The track angle and ground speed take longer to change. In particular, the track angle stops changing once the airplane has terminated looping around the airport. It is clearly visible that this happens at around 10 minutes, and then the aircraft maintains its bearing until the end of the flight.

V. Results

The results are produced using the data collected by the antenna on top of the Aerospace Faculty of Delft University of Technology during the year 2018. This section shows examples of the detected events, the validation for the three detection models developed for the identification of anomalous safety events, and insights drawn from the monitoring in the year 2018.

A. Examples of the Detected Events

This subsection offers a graphic overview of the detected events while providing insights into the working strategy of the algorithms.

1) Horizontal Compliance: Fig. 6 shows an example of the type of trajectory that the horizontal compliance strategy is able to detect. In this case, it shows a trajectory, purple line, that stabilizes at 500 ft, this gate is represented in green.

2) Energy Compliance: The flight shown in Fig. 7 represents an unstable approach. The red dots are points belonging to this particular aircraft, while the purple area comprises 95% of data, and the dotted line its median. It is possible to see that it is considered anomalous although some of the parameters vary within normal energy bounds. This is where the multivariate nature of GMM anomaly detection comes into play because it discovers anomalies based on a combination of features.

In Fig. 7, it is possible to understand why the 3-GMM model classifies the trajectory as unstable. Looking at the specific potential energy, it is clear that the aircraft is approaching higher than usual. The reason might be that a high speed is maintained until the beginning of the final approach phase, as can be seen in the specific kinetic energy plot. This situation leads to an overall higher than usual specific total energy level. The pilot is trying to dissipate all this excess energy, which is visible in the rate plots. To better understand the situation from a time-frame perspective, there is the time to runway threshold plot. Indeed, the aircraft is advancing faster than usual towards the threshold.

This flight is also present in the validation list, which is described in detail in subsection V-B. This occurrence is accompanied by the following explanation ”after accepting a short line-up the approach is unstable because the aircraft is high during the approach”. This situation is, indeed, visible in the energy metrics plot.

3) Detection of Go-Around: An example of a go-around detected on runway 06 by the algorithm is presented in Fig. 8.

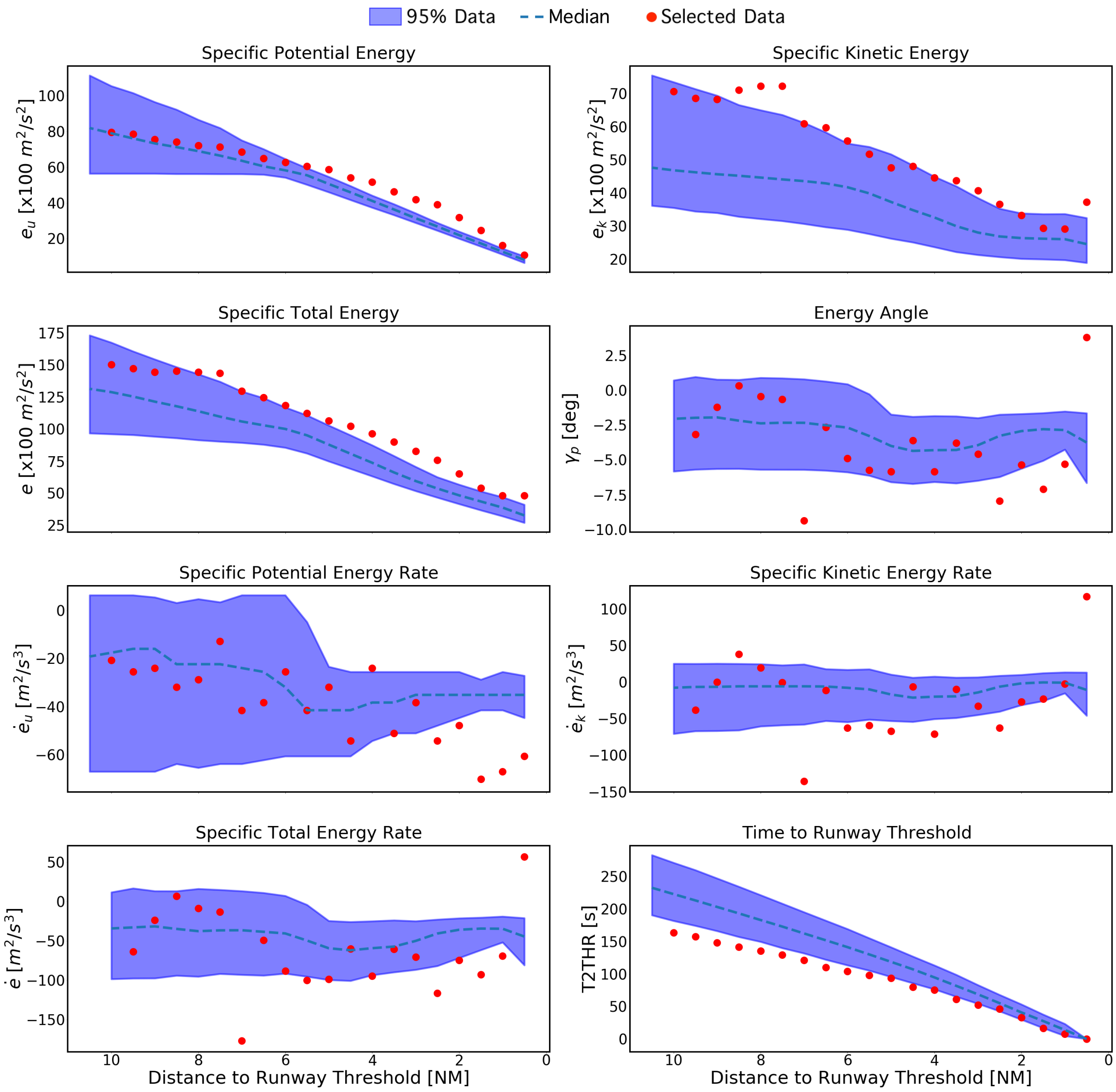


Fig. 7. The energy metrics of an aircraft performing an unstable approach

B. Validation

Validating the results of the detection algorithms is challenging because data is unlabeled. However, we could rely on a list of known go-arounds and unstable approaches provided by ILT that is the part of the Ministry of Infrastructure in charge of the airports' oversight in the Netherlands. The list, which will be called the validation list, comprises 65 go-arounds and 48 unstable approaches.

There is an important methodological difference that needs to be made. There is the possibility that a flight present in the validation list is not present in the ADS-B data. However, whereas for go-around detection, it is possible to determine without doubts if this happens by visualizing the trajectories graphically. For unstable approaches, this is not the case because plotting the trajectory doesn't provide sufficient information on the energy of that aircraft. Also, by inspecting the energy metrics, we can get more insights, but it is hard to say objectively that a flight is unstable. Indeed, this is shown by LI et al. in [19] where 4 experts could not agree on which situation poses safety concerns.

Thus in the following paragraphs, for go-arounds, the indication of "not present" will refer to a flight not being present. While for unstable approaches, it will refer to an aircraft not present in the available ADS-B data.

1) Horizontal Compliance: There are two unstable approaches detected for 2018 using this strategy, none of which is present in the validation list. Furthermore, by inspecting the detected cases, one detection is a false-positive as it is caused by poor data quality. The analysis shows that landing aircraft are most times within horizontal stability limits.

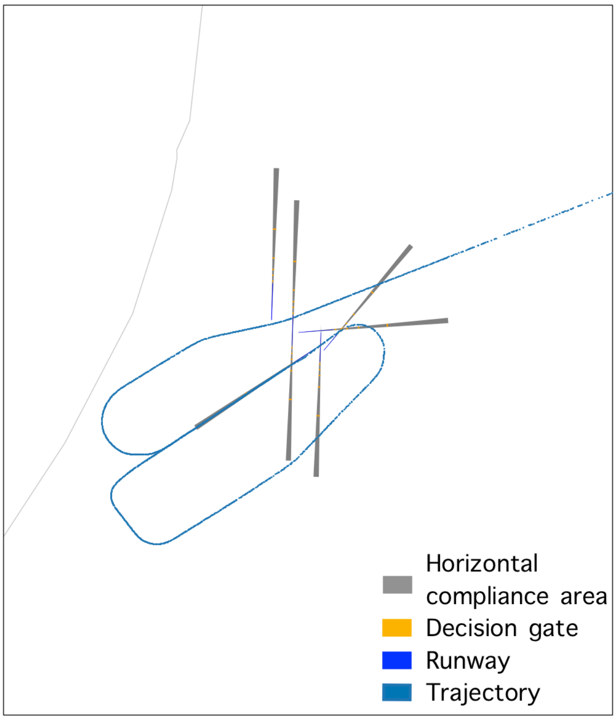


Fig. 8. An airplane performs a go-around and then lands on runway 06.

2) Energy compliance: Comparing the validation list with the ADS-B data reveals that only 31 airplanes are present in it. This detection model relies on the anomaly detection step executed using the 3-GMM. One of the parameters that is possible to vary in a GMM is the threshold. The goal is to select a threshold that includes all anomalous safety occurrences while limiting as much as possible the number of false-positives.

Table I shows the relation between a particular GMM threshold, the number of detected events, the detection accuracy, and the ratio of positives. The number of detected events refers to the occurrences that are also present in the validation list. The detection accuracy is the ratio of the number of detected events over the total number of events found in the validation list. Finally, the ratio of positives provides an overview of the amount of approaches that it is considered unstable in an year relative to the total number of approaches of that year.

As expected, increasing the threshold increases the number of detected events with the last row showing that when the threshold equals 10 % the detection accuracy rises to roughly 94%. However, 43% of landing trajectories are labeled as unstable. It is likely that the higher percentage of positives results underlies a high percentage of false positives.

According to Boeing, approximately 3% of the approaches are unstable. [21] For this reason, we would choose a threshold that is 0.1% or 0.5%. In this case, 0.1 is selected to limit the number of false positives. This corresponds to a detection accuracy of 26%, and it reveals an overall number of unstable

TABLE I
COMPARING THE NUMBER OF UNSTABLE APPROACHES DETECTED FROM THE VALIDATION LIST DEPENDING ON A GMM THRESHOLD

GMM threshold [%]	Detected[a]	Detection accuracy[b] [%]	Ratio of positives[c] [%]
0.01	3	9.68	0.50
0.05	6	19.36	1.08
0.1	8	25.81	1.60
0.5	9	29.03	4.78
1	11	35.48	7.55
2	15	48.39	12.70
3	17	54.84	17.55
5	21	67.74	26.23
10	29	93.55	42.80

[a]Number of detected unstable approaches, which are present in the validation list.
[b]Ratio between detected unstable approaches also present in the validation list and the total number of elements in the validation list
[c]Ratio between number of detected unstable approaches and overall number of detected landing aircraft

approaches equal to 3000 for the year 2018. A higher threshold could be selected after an accurate analysis of a safety expert that balances the number of false-positives and new anomalies.

3) Detection of Go-Around: In this paragraph, two validation tests are performed: comparing the results with the validation list and manual inspection for the detection of false-positives.

Table II shows the results of applying the go-around detection model to the aircraft from the validation list. "Not present" refers to go-around trajectories not present in the data, while undetected to go-arounds present in the data that are not detected by the algorithm.

TABLE II
OVERVIEW OF THE NUMBER OF GO-AROUNDS DETECTED COMPARED TO THE GO-AROUNDS PRESENT IN THE VALIDATION LIST

Detected[a]	Undetected[b]	Not present[c]
46	1	18

[a]Number of detected go-arounds, which are present in the validation list.
[b]Number of undetected go-arounds, which are present in the validation list.
[c]Number of go-arounds not present in the data, which are present in the validation list.

This test shows a detection accuracy of approximately 98%. However, it also shows that 28% of the trajectories are not present in the data. Thus, the first step to improve the detection of go-arounds is better data availability.

Another way to evaluate this algorithm is by manually inspecting its output. This means logging when the result is a go-around. The results are shown in Table III.

TABLE III
OVERVIEW OF THE OVERALL NUMBER OF GO-AROUNDS DETECTED, THE TRUE POSITIVES AND THE FALSE POSITIVES

Detected	True positives	False positives
292	285	7

False positives occur only 2% of the time, and the large majority of the detected trajectories 98% are go-arounds.

C. Monitoring the Safety Indicators

In this section, safety indicators are constructed by aggregating the results of the energy compliance model and the go-around detection model. With the knowledge extracted from the ADS-B data, it is possible to gain insights into operations by analyzing the relationship between different variables.

1) Energy Compliance: Table IV shows the relation between unstable approaches, months of the year, and weather conditions. The weather column in the table represents unstable approaches happening in the degraded weather situation.

TABLE IV
COMPARING THE NUMBER OF DETECTED UNSTABLE APPROACHES DEPENDING ON THE MONTH AND THE WEATHER CONDITION

Month	Unstable[a]	Weather[b]	Total[c]
Jan.	297	77	15269
Feb.	182	33	14370
Mar.	304	30	16794
Apr.	215	8	17645
May	256	36	17212
Jun.	259	8	18276
Jul.	298	7	19409
Aug.	265	48	19554
Sep.	213	48	18948
Oct.	226	26	19292
Nov.	214	14	16227
Dec.	270	72	16249

[a]Number of unstable approaches
[b]Number of unstable approaches with a weather score higher than 4
[c]Number of approaches detected in the data

This table shows that the months with the highest number of unstable approaches are March, January, and July with approximately 300 unstable approaches per month. Instead, February is the month with the least unstable approaches around 180. It is interesting to analyze the relationship between weather conditions and unstable approaches. As expected, the months with the largest portion of unstable approaches happening with poor weather are January and December. In these months, approximately 25% of the unstable approaches are linked with the weather. Whereas, only 3% of the time in July and June.

Furthermore, it is possible to visualize how the unstable approaches vary depending on the runway and the weather condition. This is shown in Table V. The runway with the most unstable approaches is 36R with 809, almost 30% of all. The runway where there seems to exist a strong link between degraded weather and unstable approaches is runway 27. Almost 40% throughout the year, with peaks in January and December, where poor weather is concurrent to unstable approaches 60% and 50% of the time respectively.

2) Detection of Go-Around: It is of interest to determine what are the circumstances of go-arounds. For this reason, Table VI and Table VII are used to investigate the relationship between the number of go-arounds, the weather condition, the unstable approach, and the separation to closest aircraft.

TABLE V
COMPARING THE NUMBER OF DETECTED UNSTABLE APPROACHES DEPENDING ON THE RUNWAY AND THE WEATHER CONDITION

Runway	Unstable[a]	Weather[b]	Total[c]
06	528	57	48394
18C	409	20	35393
18R	662	86	66637
22	107	21	4018
27	351	131	17304
36C	133	27	9742
36R	809	65	27757

[a]Number of unstable approaches
[b]Number of unstable approaches with a weather score higher than 4
[c]Number of approaches detected in the data

Compared to the analysis of unstable approaches, there are two extra columns. The unstable column contains information on whether the go-around is preceded by an unstable approach. The separation column indicates if the closest aircraft to the one performing the go-around is at a distance between 1.5NM and 3NM. These thresholds are chosen because there is no aircraft closer than 1.5 NM, and 3NM is the minima for terminal airspace operations.

TABLE VI
COMPARING THE NUMBER OF DETECTED GO-AROUNDS DEPENDING ON THE MONTH, THE WEATHER CONDITION, THE UNSTABLE APPROACH AND THE SEPARATION TO CLOSEST AIRCRAFT

Month	Total[a]	Weather[b]	Unstable[c]	Separation[d]
Jan.	44	29	2	1
Feb.	12	1	3	1
Mar.	29	6	4	3
Apr.	22	0	9	1
May	32	5	9	6
Jun.	26	2	8	1
Jul.	28	2	7	1
Aug.	22	4	3	1
Sep.	23	13	3	2
Oct.	11	1	4	1
Nov.	10	1	1	0
Dec.	26	12	7	1

[a]Number of go-arounds
[b]Number of go-arounds with a weather score higher than 4
[c]Number of go-arounds linked with an unstable approach
[d]Number of go-arounds where the separation to the closest aircraft is between 1.5NM and 3NM.

From Table VI, 25% of go-arounds happen in December and January. Probably, the reason is that weather conditions are worse. Indeed, 60% of go-arounds in this period is linked with the degraded weather condition. Also in September, approximately 60% of go-arounds are linked with poor weather. April is the month where go-arounds are preceded by an unstable approach most often, 40% of the time. Followed by October, 36%. Separation seems to influence very little the number of go-arounds.

Table VII shows how the number of go-arounds varies depending on the runway. The majority happens on runway 18R (68 go-arounds) and 27 (65 go-arounds). On this last runway, 55% of go-arounds are linked with a poor weather condition.

TABLE VII
COMPARING THE NUMBER OF DETECTED GO-AROUNDS DEPENDING ON THE RUNWAY, THE WEATHER CONDITION, THE UNSTABLE APPROACH AND THE SEPARATION TO CLOSEST AIRCRAFT

Runway	**Total**[a]	**Weather**[b]	**Unstable**[c]	**Separation**[d]
06	49	2	14	2
18C	33	3	5	5
18R	68	19	14	2
22	11	6	2	2
27	65	36	5	1
36C	12	3	4	1
36R	47	7	16	6

[a]Number of go-arounds
[b]Number of go-arounds with a weather score higher than 4
[c]Number of go-arounds linked with an unstable approach
[d]Number of go-arounds where the separation to the closest aircraft is between 1.5NM and 3NM.

In particular, on runway 27 out of 25 go-arounds happening in January, 22 are concurrent with a difficult weather situation. Also, runway 18 has 28% of go-arounds associated with degraded weather. The peaks happen in January and September where go-arounds are associated with poor weather 73% and 87% of the time respectively. Unstable approaches precede 34% of go-arounds on runway 36R and 29% on runway 06. In particular, in April 80% of go-arounds on runway 06 are preceded by an unstable approach.

VI. DISCUSSION

This paper shows a new data-driven approach to monitor unstable approaches and go-arounds using ADS-B data. It proposes three models: two for detecting unstable approaches and one for go-arounds. Overall all methods would benefit by having better data availability. This would allow reiterating the algorithms to improve detection accuracy. A limitation common to all methods is the absence of a proper validation dataset. This makes it difficult to produce metrics such as false-negative and false-positives.

The horizontal compliance algorithm is the one that shows the most limited applicability. This is an exceedance detection method inspired by the Flight Safety Foundation stable approach guidelines and by the work of Wang et al. [4]. Airplanes have shown to always comply with the lateral stabilization criteria, thus the method has shown low efficacy in detecting unstable approaches.

The energy compliance algorithm, instead, shows promising results with a detection accuracy of 26% on the validation list. After improving the availability of data especially at lower altitudes, the threshold used for anomaly detection could be reiterated by a safety expert to improve the detection accuracy. A GMM is chosen as the anomaly detection strategy because it provides interpretability and is directly usable with multivariate data. Instead, many anomaly detection strategies reduce the feature space and identify trajectory level anomalies rather than point level anomalies. The ability of detecting point level anomalies makes the model more resilient in case some points are missing because in this method the analysis is performed point by point. While if the analysis was performed on a trajectory level, incomplete trajectories would have to be removed completely. The effect would be a further limitation on the analysis.

One of the limitations of the energy compliance method is linked with the preprocessing of the true airspeed used in the energy metrics. The GFS NOAA wind data is available at intervals of 700 ft and updated every 6 hours. Furthermore, there is no information about the wind vertical velocity.

Go-around detection is the best performing model with a false positive rate of only 2%, and over 98% detection accuracy on the validation list. Minor variations of this model can potentially detect other aircraft operations, such as holding patterns, which provide a measure of the ATC workload. This strategy detects go-arounds based on the similarity between ADS-B data and a predefined model. The underlying assumptions are well-reviewed by Proud in [5].

The thresholds used for the S-functions of the go-around detection strategy are chosen based on experiments. Further studies can be performed to improve them. However, the optimum values are likely close to the ones selected. Indeed, this is visible in the extremely low false-positive rate and high detection accuracy.

A go-around might be caused by an anomalous safety event. For this reason, this paper attempts to understand the most likely circumstances of go-arounds. Another aspect that can be analyzed is the wake-category of the preceding aircraft. Nevertheless, some factors remain difficult to investigate such as an ATC instruction because of an occupied runway.

Better data availability can directly improve the result of this analysis, as 28% of the go-arounds in the validation list is not present in the data, and 35% for unstable approaches. Furthermore, this overview does not take into account incomplete flights, thus these numbers are likely higher. There are fewer data available at a lower altitude because this study relies on an antenna 40km away. This impacts the detection of unstable approaches more than the detection of go-arounds because this last maneuver is performed at higher altitudes.

Yet, it is not possible to provide comprehensive statistics on false-negatives and false-positives because it does not exist a proper validation dataset. In the case of go-arounds, false-positive statistics are provided by examining the output of the method, which provides a clear idea of false-positives. The same method does not apply to false-negatives since it would require an inspection of thousands of trajectories one by one.

VII. CONCLUSIONS

Different models are tested to construct the basis of a safety monitoring system. These methods are able to extract safety knowledge from aviation data. Two of them focus on detecting unstable approaches. The first one relies on the idea that a stable approach is constrained within certain horizontal bounds. The second one assumes that an unstable approach is characterized by an anomalous energy level. Thus, energy features are derived from ADS-B data, and anomalies are revealed using a GMM. The third model detects go-arounds using fuzzy logic and S-functions. These functions model the dynamics of go-arounds and provide a similarity score.

It is shown that these models can detect anomalous safety events on ADS-B data from 2018 around Schiphol Airport. These results are aggregated to derive useful insights. For example, December and January are the months in which a degraded weather condition has the highest impact on the presence of unstable approaches and go-arounds. Also, on runway 27 unstable approaches and go-arounds are often linked with an adverse weather situation.

From the results, it can be concluded that the data-driven methodology proposed in this paper has the potential to enable independent monitoring of aircraft operations using aviation and meteorological open data.

In the future, the focus will be on detecting other events that may impact the safety of operations. Although some of the strategies discussed in this paper can be modified and used to detect other anomalous safety events, more strategies will be required to efficiently detect them. Flight phases of particular interest are take-off and ground operations. This analysis would benefit greatly from better data availability especially at lower altitudes, which becomes fundamental for the analysis of ground operations.

References

[1] C. Pradera and H. Teper, "ADS-B and other means of surveillance implementation status," SESAR, Tech. Rep., may 2018.

[2] IATA, "Unstable Approaches Risk Mitigation Policies, Procedures and Best Practices 2nd Edition," Tech. Rep., 2016.

[3] A. Dahleni Kraemer, E. Villani, A. D. Kraemer, and E. Villani, "On the gap between aircraft FDI methods in industry and academy: challenges and directions," in *AIAA Scitech 2019 Forum*. American Institute of Aeronautics and Astronautics Inc, AIAA, 2019, p. 506.

[4] Z. Wang, L. Sherry, and J. Shortle, "Airspace risk management using surveillance track data: Stabilized approaches," in *ICNS 2015 - Innovation in Operations, Implementation Benefits and Integration of the CNS Infrastructure, Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc., jun 2015, pp. W31–W314.

[5] S. R. Proud, "Go-Around Detection Using Crowd-Sourced ADS-B Position Data," *Aerospace*, vol. 7(2), pp. 16–, feb 2020.

[6] S. Das, B. L. Matthews, A. N. Srivastava, and N. C. Oza, "Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 47–56.

[7] S. Das, B. L. Matthews, and R. Lawrence, "Fleet level anomaly detection of aviation safety data," in *2011 IEEE International Conference on Prognostics and Health Management, PHM 2011 - Conference Proceedings*, 2011, pp. 1–10.

[8] B. Matthews, D. Nielsen, J. Schade, K. Chan, and M. Kiniry, "Comparative study of metroplex airspace and procedures using machine learning to discover flight track anomalies," in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, IEEE. Institute of Electrical and Electronics Engineers Inc., oct 2015, pp. 2G41–2G415.

[9] T. G. Puranik and D. N. Mavris, "Identification of Instantaneous Anomalies in General Aviation Operations Using Energy Metrics," *Journal of Aerospace Information Systems*, vol. 17, no. 1, pp. 51–65, jan 2020. [Online]. Available: http://dx.doi.org/10.2514/1.I010772

[10] T. Puranik, H. Jimenez, and D. Mavris, "Energy-based metrics for safety analysis of general aviation operations," *Journal of Aircraft*, vol. 54, no. 6, pp. 2285–2297, 2017.

[11] J. Sun, "Open Aircraft Performance Modeling: Based on an Analysis of Aircraft Surveillance Data," Ph.D. dissertation, TU Delft, 2019.

[12] Performance Review Unit and ATMAP MET working group, "Algorithm to describe weather conditions at European airports," Tech. Rep., 2011.

[13] ICAO, "ICAO Annex 2: Rules of the Air, Chapter 4," ICAO, Edition 42.0, nov 2009.

[14] Flight Safety Foundation, "ALAR BRIEFING NOTE 7.1: Stabilized Approach," Tech. Rep., 2009.

[15] T. G. Puranik and D. N. Mavris, "Identification of Instantaneous Anomalies in General Aviation Operations Using Energy Metrics," 2019.

[16] G. Jarry, D. Delahaye, F. Nicol, and E. Feron, "Aircraft atypical approach detection using functional principal component analysis," *Journal of Air Transport Management*, vol. 84, p. 101787, may 2020.

[17] R. Deshmukh and I. Hwang, "Anomaly Detection Using Temporal Logic Based Learning for Terminal Airspace Operations," in *AIAA Scitech 2019 Forum*. American Institute of Aeronautics and Astronautics Inc, AIAA, 2019, p. 682.

[18] M. C. L. Van Den Hoven, P. M. A. De Jong, C. Borst, M. Mulder, and M. M. Van Paassen, "Investigation of Energy Management during Approach-Evaluating the Total Energy-Based Perspective Flight-Path Display," 2010.

[19] L. Li, R. J. Hansman, R. Palacios, and R. Welsch, "Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 45–57, mar 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X16000188

[20] J. Sun, J. Ellerbroek, and J. Hoekstra, "Flight extraction and phase identification for large automatic dependent surveillance-broadcast datasets," *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 566–571, aug 2017.

[21] The Boeing Edge, "Boeing Aeromagazine Issue 54 Quarter 02 2014." [Online]. Available: http://www.boeing.com/commercial/aeromagazine/articles/2014_q2/pdf/AERO_2014q2.pdf

A

Research Framework and Methodology

This chapter provides the general research framework of which the presented paper in Chapter 1 is a realization. Section A.1 introduces the problem, which is the starting point of this thesis. Section A.2 formalizes the objective and research questions. Furthermore, Section A.3 clarifies the contribution of the thesis to the field of research. Section A.4 explains how the questions are going to be answered and how the objective is going to be reached. After reading this section, the choices made when selecting the relevant literature will become clear.

A.1. Problem Definition

In the Netherlands, Schiphol Airport is the main port connecting the country with the rest of the world and it is the busiest airport in Europe by aircraft movements[1]. The different actors operating at Schiphol Airport aims to capture as much air transport demand as possible. For this purpose, Royal Schiphol Group has built a secondary airport in the Flevopolder, 70 km away. Furthermore, it is building a new terminal at Schiphol to accommodate 14 million passengers with an expected opening in 2023.[2]

However, an increase in aircraft movements rises multiple questions, especially for a complex infrastructure such as Schiphol Airport. It is not only a matter of the economic benefit it would have on the Dutch economy, but it also raises concerns for the safety of operations and the people living around the airport. These aspects have been part of the political agenda for more than 20 years.

The safety of operations at Schiphol has been high on the political agenda in the 90s, after the Bijlmer incident of 1992. The key measure taken at that time was the establishment of the Integral Safety Management System (ISMS) between actors at Schiphol: Royal Schiphol Group, Air Traffic Control the Netherlands, airlines, and airground services.[1] The work on the ISMS has accelerated in 2017 when the Roadmap for Safety Improvement at Schiphol was launched[3]. This renewed effort was a direct

[1] data from `www.aci.aero`

[2] data from `www.schiphol.nl/en/schiphol-as-a-neighbour/page/lelystad-airport/` and `news.schiphol.com/amsterdam-airport-schiphol-presents-new-terminal/`

[3] data from `https://integralsafetyschiphol.com/`

consequence of a report published by the Dutch Safety Board the same year. The report highlighted how ”Schiphol is reaching its limits”, and since 2014, the growth in movements was accompanied by an increase of incidents. According to their analysis, actors at Schiphol Airport had to rethink their safety and risk mitigation strategy before being able to increase air traffic after November 2020.[2]

Until November 2020, aircraft movements at Schipol Airport are limited to 500,000 due to an agreement signed in 2008 which aimed at the reduction of noise pollution of the surrounding residential areas[4]. Furthermore, the secondary airport in the Flevopolder is ready, but it is closed for concerns about the level of nitrogen emissions that would be produced by the operating aircraft.

In this complex scenario, the Dutch government is responsible to ensure the safety of operations in civil aviation. Within the government, this delicate responsibility falls to the Ministry of Infrastructure and Water Management. Inside the Ministry, the Human Environment and Transport Inspectorate (ILT), in Dutch Inspectie Leefomgeving en Transport, monitors and oversees safety at Schiphol Airport. This research is conducted in close collaboration with ILT. Currently, they assess the safety performance analyzing occurrence reports and building statistics of relevant occurrences. These reports are produced by the different parties operating at the airport and include any event that could hinder safety. With the higher availability of aviation data, it is possible to gain greater insights from normal operations.

The idea is to obtain anomalous aviation events from the data and to analyze the occurrences to gain insights in the form of safety indicators. Some examples of indicators are the number of unstabilized approaches, identification of traffic density, the average minimum distance between landing/take-off aircraft, and the number of go-arounds. It is important to clearly define the area in which these operations take place. It includes all areas that regulate the inflow and outflow of aircraft traffic. In this analysis, the term Schiphol Airspace (SA) is used to indicate the area that includes:

- Schiphol’s Holding stacks: this is the area where aircraft wait before landing in case of runway unavailability
- Schiphol’s Terminal control areas (TMA): the area that extends up to FL095 surrounding the airport
- Schiphol’s Control zones (CTR): the area that extends from the up to 3000 ft surrounding the airport
- Schiphol’s Runways: used for take-off and landing
- Schiphol’s taxiways: used for moving between parking areas and runways
- Schiphol’s parking areas: used for the parking of aircraft

The primary data used is ADS-B data collected using the antenna on top of the Aerospace Faculty of TU Delft. A preliminary analysis has shown that because of the distance to the airport there is little availability of ground movements data. For this reason,

[4]data from https://www.schiphol.nl/en/schiphol-as-a-neighbour/page/schiphol-and-the-future/

other databases are used to improve the situation. This comes at the cost of extra pre-processing steps.

In this thesis, finding anomalous aviation events means finding incidents and their precursors. From ICAO, an incident is an occurrence, other than an accident, associated with the operation of an aircraft which affects or could affect the safety of operation.[5] In more detail, using Statler's definition[3], an incident is a finite sequence of states. As shown in Figure A.1, the first and last are safe states. While all other states are compromised or anomalous states, with at least one which is anomalous. If the final state is anomalous, instead of safe, the sequence of states belongs to an accident. A precursor is a compromised state which, if not corrected, leads to an incident or an accident.

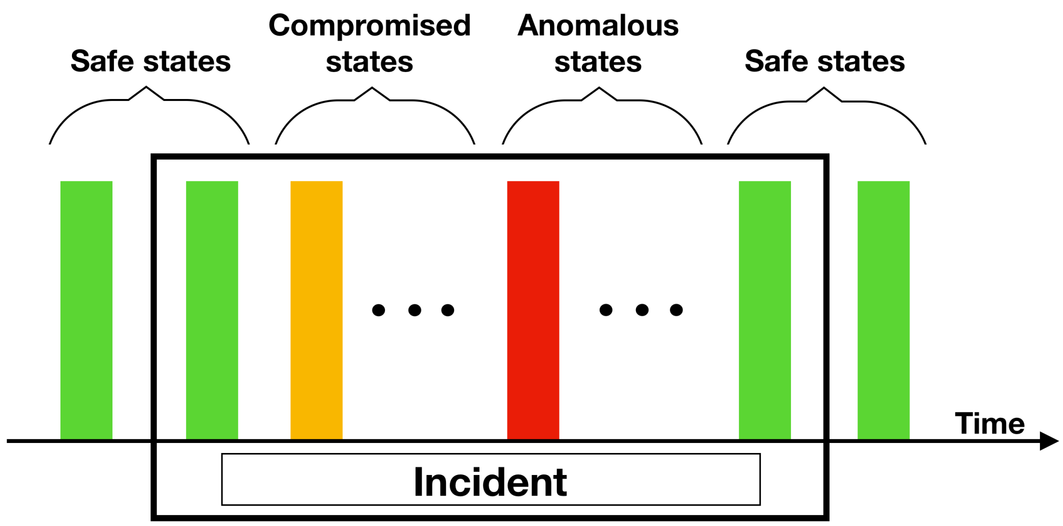


Figure A.1: The figure shows the representation of an incident. (Adapted from [3])

What makes this task particularly challenging is the imbalance of the dataset, namely the fact that anomalous aviation safety events are a handful against millions of regular flights. On top of this complexity, once anomalous aviation events are found, it is not guaranteed that these are operationally significant. It is important to be aware of the difference between operationally anomalous and statistically anomalous.

Algorithms recognize statistically abnormal events, which include safe events and operationally anomalous ones. Figure A.2 shows the output of a detection algorithm in relation to these different events. In the figure, the rectangle represents the output of the algorithm that differentiates events for being statistically normal, green, or anomalous, yellow. Within the statistically anomalous ones, the algorithm recognizes false alarms, which are safe events, known and potentially unknown problems. This study is interested in operationally anomalous events, which include known and unknown problems. For this reason, different strategies are used to minimize the number of false positives, which are described in Section A.4.

[5]ICAO Annex 13

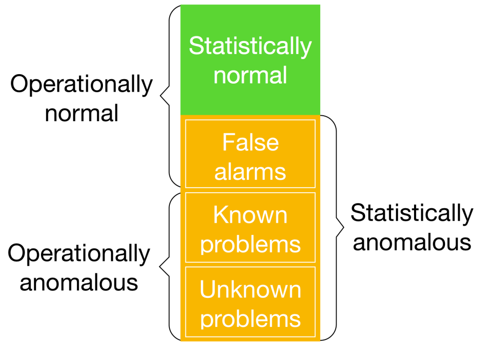


Figure A.2: The figure shows a generalized output for a detection algorithm.

A.2. Research Objective and Questions

This section explains what is the objective of this thesis and which research questions are answered to achieve it. The definition of the research objective and questions follows from the identification of the key challenges in Section A.1.

A.2.1. Research Objective

The objective of this research is:

> **"To propose ILT a set of safety indicators that provide insights into the safety performance of aircraft operations around Schiphol Airport by mining ADS-B data."**

The objective is very practical and it contributes to bridge the gap between research and real-world applications of ADS-B technology.

A.2.2. Research Questions

It follows the main research question:

> **"How can ADS-B data be used to gain insights into the safety performance of Schiphol Airport?"**

To answer this question, several research questions are identified. These questions together with sub-questions and sub-goals achieve the research objective, as shown in Table A.1. Answering these questions aim to deliver the most valuable operational tool. After obtaining a result from the algorithm, time is spent in determining how it can be operationalized.

Table A.1: This table shows research questions, sub-questions and sub-goals.

Research Questions (RQ)	Sub-questions	Sub-goals (all "build" tasks refer to programming in Python)
1 What areas can be analyzed given the data available?	1.1 Which data should be discarded? 1.2 Which smoothing methods are applicable? 1.3 Which outlier removal technique can be used?	- Merging database - Manual cleaning of data - Removal of outliers from data - Map with data density over Schiphol - Table with data density at different altitudes - Build pre-processing tool
2 What data mining strategies can be developed to identify incidents and precursors for the final approach phase?	2.1 Which threshold can be used to construct an incident detection model? 2.2 Which set of feature(ex. energy) is more appropriate to detect anomalies? 2.3 Which algorithms are more indicated for an anomaly detection model? 2.4 Which algorithms are more indicated for a precursor detection model?	- Build incident detection model and tool - Detect anomalous events - Build precursor detection model and tool - Detect precursor
3 What are the limitations of the outcome of these models? What is the operational significance of the result?	3.1 What tests can be performed to identify limits in anomalous events? 3.2 What tests can be performed to identify limits in precursors?	- Validate anomalous events - Validate precursors
4 How can the results of the models be used to design safety indicators, which reveal the underlying safety performance of Schiphol Airport?	4.1 How often does a certain anomalous event or precursor happen? 4.2 Is there a condition (ex. weather) that affects the appearance of anomalous events or precursors? 4.3 What are the similarities between anomalous events and precursors (ex. a trajectory characteristic of these events, a particular speed profile)?	- Finalize tool that takes ADS-B data as input and computes safety indicators statistics

This research takes place within a limited amount of time, 9 months. Given the time constraint, it is unlikely that all operational phases are analyzed. For this reason, the thesis is set-up such that answering all questions entails analyzing a single operational phase. Once the pre-processing step is overcome, one operational phase is selected. At this point, the anomalous event detection and the precursor detection are performed only for the selected operational phase. The results can be directly used to produce safety indicators for this particular phase.

In coordination with ILT, anomalous safety events of particular interest are identified to achieve the most useful operational result given the limited amount of time. After a preliminary analysis the following operations are identified: unstabilised approach, go-around, rejected take-off, and runway crossings. Since 50% of the accidents happen in the landing phase, this will be the first focus of the analysis. The same methodology can be used to continue the research for other operational phases.

A.3. Impact and Contribution

The work of this thesis has an impact in three different ways that are summarized in this section. It contributes to the body of knowledge, to ILT, and in general to the Air Traffic Management safety research segment.

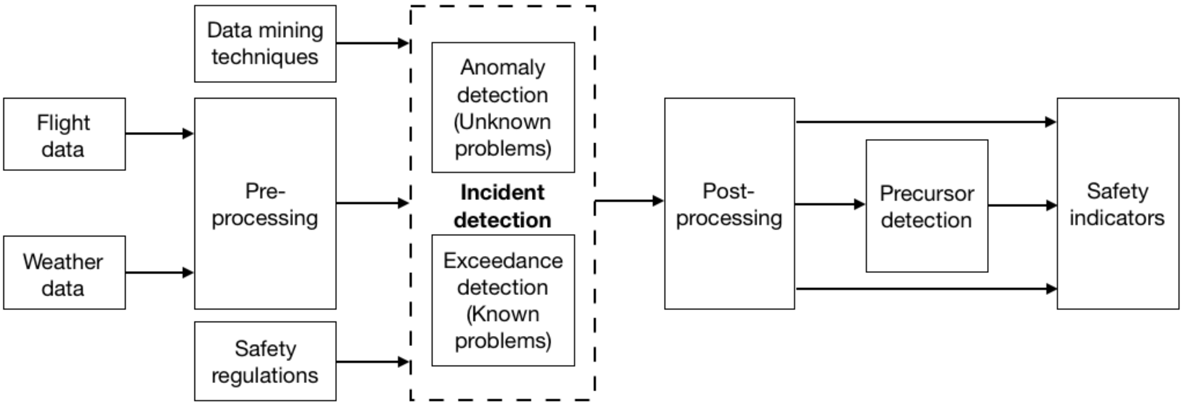


Figure A.3: The picture shows the work flow of this thesis.

The contribution of this thesis to the body of knowledge consists of an innovative methodology that uses data-driven indicators to produce safety insights and a concrete use case of it. This contributes in bridging the gap between research and real-world applications of ADS-B technology. In this way, it advances the research regarding proactive risk assessment using ADS-B data.

ADS-B data is independent. This characteristic makes its applications very valuable to air traffic safety body. With this methodology is possible to independently monitor aircraft operation. Given that the case study for this analysis is Schiphol Airport, ILT can directly use the results of this thesis to gain safety insights.

Furthermore, the research has far-reaching applications because the same methodology can be used anywhere in the world. ADS-B data is not encrypted and anyone with the correct receiver set-up can record it without any restriction in much of the world. This represents a cost-effective way to monitor safety and operational performance. It might constitute a piece of an efficient monitoring tool that could be useful in developing countries, where accident rates are much higher.

A.4. Methodology

In this section, the methodology used to achieve the project objective is presented. Flight data is used to perform the analysis. In particular, ADS-B is the main resource throughout the work, which may be integrated with weather and flight procedures data. ADS-B data is collected using the antenna on top of the Aerospace Faculty of TU Delft. A preliminary analysis has shown that because of the distance to the airport there is little availability of ground movements data. For this reason, the OpenSky database[6] is used to improve the coverage. These data constitute the starting point of the work, as shown in Figure A.3.

The next step consists of pre-processing the data. In particular, ADS-B data is collected from different sources. Thus, databases are merged and data undergo manual cleaning in which data not useful for the analysis is removed, for instance, negative

[6]https://opensky-network.org/

altitudes, helicopters, and ground vehicles. More pre-processing steps might be required depending on the algorithm.

For example, these can be selecting only one type of aircraft, the same phase of flight, and adding extra features to the ADS-B basic ones (ex. energy). Pre-processing answers the first research question. Once data is cleaned, it will be clear in which area it will be possible to perform the analysis.

After pre-processing, the data is ready to be analyzed to find anomalous aviation events. The idea is to use a combination of exceedance detection and anomaly detection techniques to obtain the most effective and operationalizable result. For this reason, it is privileged a solution with low complexity and a low false-positive rate compared to complex models.

As anticipated in the previous section, the first focus is on the final approach phase to detect unstable approaches and go-arounds. This is the most critical flight phase since it is where 50% of all accidents happen. This step partially answers the second research question.

When statistically anomalous events are discovered, post-processing is applied to validate the operational significance of the algorithm's results. The validation consists of performing sensitivity analysis on the tuning parameters of the algorithms to understand how the result varies, comparing if the trajectory is anomalous for multiple algorithms, and visual inspection possibly with an expert. Furthermore, anomalous safety events can be checked against data from occurrence reports. In this way, the third research question is fully addressed.

Once anomalous aviation safety events are detected, this labeled data can be used to identify precursors. For instance, it would be interesting to understand whether go-arounds happen because of conflicting traffic, poor weather, or an unstable approach. At this step, the second research question is fully answered.

Finally, safety indicators are designed by combining statistics of incidents and precursors. Completing this step answers the fourth research question and reaches the objective of the thesis presented in Section A.2.

B

Literature Study

The literature study is presented in this chapter. The goal is to highlight relevant work that could be useful while performing the thesis. It starts with an overview of the research areas in Section B.1. The data used for the analysis is described in Section B.2 and it includes an overview of the current data availability. In Section B.3, it continues with an analysis of pre-processing techniques that can be used for these data. Afterward techniques for finding incidents in data are presented in Section B.4 and Section B.5. After finding incidents, precursors are discovered with methods from Section B.6. There will be an overview of techniques highlighting their relevance in the field and above all their applicability to the present thesis. Section B.7 summarizes the key findings.

B.1. Research Areas

This research is based on multiple research areas. It finds its foundations and motivation in the field of aviation safety. This field produces much of the regulations that govern air traffic operations. In this thesis, these regulations are used to derive a set of rules that is accountable in the ADS-B data. The core driver of this research overlaps with this research field. The goal is minimizing the occurrence and probability of occurrence of accidents, incidents, and precursors.

In this thesis, this goal is achieved via proactive risk assessment. Safety risk assessment in aviation is experiencing profound changes. The technological innovations, introduced with the modernization of the ATM, Air Traffic Management, architecture, are causing a shift from reactive to proactive risk management. With the reactive approach, risks are identified after a serious event occurs. In this case, mitigation strategies are identified after performing in-depth analysis to reconstruct the causes using safety models such as the Swiss Cheese barrier and the Bowtie. Instead, proactive risk management wants to avoid the occurrence of safety events. To achieve this result, it studies what might pose a risk to safe operations before any accident or incident occurs.

Proactive risk assessment is possible because of the availability of large amounts of data and data mining techniques. Data mining is another fundamental research area for this thesis. In fact, preprocessing, anomaly and precursor detection are largely resulting from it. Data mining uses techniques from statistics, signal processing, machine learning, and process control. Within data mining, the most interesting field is anomaly detection. The research focuses on identifying data that deviates from normality. Because ADS-B data is unlabeled, the focus of this research is on unsupervised learning algorithms. The key assumption of these is that only a few samples are anomalous, and this is also the case for anomalous aviation safety events.[4]

These data-driven algorithms can be used because of the advancements made in the usage of ADS-B data.[5] There is a growing field of experts that wants to apply this data to different areas of aviation research. The reason is that aviation data has generally been expensive and difficult to access. Instead, ADS-B has opposite characteristics being easily retrievable and cheap. In fact, anyone with the correct receiver set-up can record this data. These characteristics facilitate the independent monitoring of air traffic.

B.2. ADS-B Data

ADS-B stands for Automatic Dependent Surveillance-Broadcast: ”automatic” because it transmits data without any input of the pilot and ”dependent” because it depends on the aircraft’s navigation system. Furthermore, it is a ”surveillance” technology composed of two distinct instruments: a transmitter and a receiver. ADS-B out, the transmitter, allows any equipped aircraft to ”broadcast” continuously its position enabling anyone with ADS-B in, the receiver, to know where it is located.[6]

ADS-B data is the main data source for this master thesis, as such, it has a critical position for the outcome of the project. For this reason, it is important to understand the limitation arising from this technology. Background information are presented in Section B.2.1. While Section B.2.2 describes possible inaccuracy in the data. Finally, Section B.2.3 addresses why the limited availability of data at particular altitudes is a problem and how it can be overcome.

B.2.1. Background

ADS-B represents one of the key technological enablers to modernize the ATM architecture all over the world, and it is a pillar in the European and American ATM Master Plan. Since the Eleventh ICAO Air Navigation Conference of 2003, regulators recommended the implementation of the ADS-B technology for the cost-effective improvement of safety and capacity.[7] For instance, it has been employed in areas which were previously not covered by any surveillance technology because it was not economically or physically feasible, such as the Australian inland or over Hudson Bay in Canada. From 2020, ADS-B Out equipment became compulsory for all air traffic in designated air-space in Europe and US.

In Europe, ADS-B out uses the 1090-ES ICAO data-link standard. Among the 3 existing standards, this is the most common one and its name stands for ”1090” is the MHz frequency, which is typically used by aviation surveillance system; ”E” is extended because its message consists of 112 bits, whereas the ACAS [1] has 56 bits; ”S” is squitter for the autonomous broadcasting capability.

Every ADS-B message contains the unique ICAO identifier of the aircraft together with a piece of specific information. There are 9 different types. In particular, the type of information is raveled through a code encoded at bits 33 - 37 of the ADS-B message. This code is shown in Table B.1 together with the respective encoded content. [5]

Table B.1: The table, adapted from [5], shows the type code used and data contained in the ADS-B message.

Type Code	Content
1 - 4	Aircraft identification
5 - 8	Surface position
9 - 18	Airborne position (w/ Baro Altitude)
19	Airborne velocities
20 - 22	Airborne position (w/ GNSS Height)
23 - 27	Reserved
28	Aircraft status
29	Target state and status information
31	Aircraft operation status

The source of most of this data is a satellite radio-navigation system. It might also use other on-board instruments such as the inertial platform or the altimeter. For this reason, there is a direct relation between the accuracy of these instruments and the accuracy of the ADS-B data

B.2.2. Data Quality

For this thesis, it is important to understand if the quality of the data is sufficient to perform the analysis. Verbraak et al. [8] performed an analysis to understand if ADS-B can be used as the primary mean of surveillance, thus focusing on data quality and signal quality. His research is extremely relevant because he used ADS-B data collected using the set-up located on top of the aerospace building, which is the main source of data also for this thesis. There are two main components of data quality: latency and accuracy. Latency is the delay between measuring and receiving the position caused by the processing of data. Accuracy means how accurate is the position measured compared to the real one. Verbraak found an average latency of 20 ms and an average horizontal offset of 21 m with cross-track accuracy better than 51 m 95% of the time. [8]

Another study on the quality of ADS-B data has been conducted by Ali et al. [9], the analysis was performed on a limited dataset of 9 aircraft and resulted in an average latency of 1 sec and a root mean square error (RMS) horizontal offset of 188 meters, whereas Verbraak found a RMS of 250 meters. It is important to note that Ali cleaned its dataset removing data coming from the faulty ADS-B out set-up.

[1] ACAS: Airborne Collision Avoidance System

Ali et al. performed also a detailed analysis of ADS-B system errors on 57 British Airways aircraft. [10] He revealed that some of these errors are random while others are systematic and specific to certain aircraft plus avionics equipment. The most common types of error are position jumps, ADS-B update interval, and GPS clock. These errors risk to jeopardize the implementation of ADS-B surveillance technology. Thus, it is important to put in place mitigation strategies. ICAO has released a list with known ADS-B avionics problems. [7]

B.2.3. Availability

Another important aspect is the availability of ADS-B data, which depends on the position of the ADS-B in the receiver. The primary source of data for this thesis is the database collected using the set up on top of the Aerospace Faculty of TU Delft. Table B.2 shows the average availability of data at different altitudes in a day for 2018. For all altitudes, there is a reasonable amount of data except at ground level where there are only 50 points per flight on average. Considering the average latency of 1 sec this means 1-2 minute of trajectory, while aircraft spend much more time in ground operations. In Schiphol, it can take more than 15 minutes to reach the Polderbaan runway.

This is not enough for an analysis of ground movement and integration with more databases is required. A possibility would be integrating it using OpenSky database because it is easily accessible for researchers and an amateur ADS-B in receiver is registered at 52°12'N 4°53'E, which is only 15 Km from Schiphol. However, this is based on a community receiver network and it presents unique challenges because of the integrity of the data. [11]

Table B.2: Average availability of TU Delft ADS-B data at different altitudes in a day for 2018.

Altitude range [ft]	# fid	# of points	ratio
10000 - 8000	1271.43	89937.57	70.74
8000 - 6000	1269.42	124469.35	98.05
6000 - 4000	1268.03	134823.59	106.33
4000 - 2000	1266.02	200105.11	158.06
2000 - 1000	1243.75	121920.53	98.03
1000 - 500	1223.85	50215.02	41.03
500 - 100	1119.17	28434.37	25.41
100 - 0	673.03	33801.15	50.22

B.3. Pre-processing

Pre-processing the data is a fundamental step especially when dealing with data from different sources. ADS-B data comes in the form of a continuous-time series. The goal is to obtain data that is regular and easy to interact with. In this section, general pre-processing steps will be described. It is likely that when applying the different methods to detect anomalies more pre-processing steps will be required, like re-sampling in time or space.

It is required to merge the two databases: units will need to be converted, duplicates removed, and possible outliers. As presented by Sun et al. [12], common pre-processing steps are data scaling, interpolation, and smoothing. When applying machine learning algorithms, the range of data influences its performance. For instance, Euclidean distance varies depending on if meters, degrees, or a mix is used. For this reason data is scaled, Equation B.1 scales x_i, an instance of the distribution X, between [0, s_{max}].

$$x_i^{'} = \frac{x_i - min(X)}{max(X) - min(X)} \cdot s_{max} \tag{B.1}$$

Some machine learning algorithms also require the data to have 0 mean and unit variance. The standardization procedure is shown in Equation B.2.

$$x_i^{'} = \frac{x_i - mean(X)}{var(X)} \tag{B.2}$$

Interpolating and smoothing is often required when dealing with real-world data. For instance, the data collected for an aircraft presents some gaps or irregularities. Interpolation is used when given a set of measurements we want to estimate what would be the value of a point between known ones. Smoothing is used when the data present small measurement errors and/or fluctuations. Common methods include low-pass filters, moving average, and splines.

Another important pre-processing step is dividing aircraft operations in different flights. In fact, ADS-B data provide information about which aircraft is transmitting the message, and not about the specific flight. The same aircraft lands and take-offs multiple times on the same day. In this thesis, all operations are happening around Schiphol Airport. For this reason, a rule-based method is designed to detect when an aircraft lands and take-offs.

B.4. Exceedance Detection

Exceedance detection is currently the most common methodology used to track anomalous aviation safety events. Classic exceedance detection consists in analyzing Flight Operation Quality Assurance (FOQA) data using threshold based algorithms. These thresholds are set by experts based on knowledge of previous incidents and risks. The general process is as follows. During a flight, an aircraft records information about its position, speed, pilot inputs, engine setting and many more. There are more than 100 features recorded. After the flight, this data is collected and analyzed by an exceedance detection algorithm. The events highlighted by the algorithms are analyzed by a safety expert that decides if the event requires further investigation.

FOQA data and exceedance detection algorithms are proprietary information. This thesis uses ADS-B data as the main source of information, which means that airline graded exceedance detection algorithms can not be used. For these reasons, exceedance detection in this thesis is performed using research papers, policies from regulators, and operating manuals from manufacturers. Among the regulators, key information is retrieved from LVNL, ICAO, CANSO, EASA, Eurocontrol, and IATA. It follows an example of how data from regulations and research papers are used in the exceedance detection method for unstabilised approach.

Unstabilised approach
Detecting unstabilised approaches is crucial as 50% of the accidents happen in this phase of flight.[2] According to the Flight Safety Foundation a stabilised approach happens when the following is satisfied:

- The aircraft is on the correct flight path
- Only small changes in heading/pitch are necessary to maintain the correct flight path
- The airspeed is not more than VREF + 20kts indicated speed and not less than VREF
- The aircraft is in the correct landing configuration
- Sink rate is no greater than 1000 feet/minute; if an approach requires a sink rate greater than 1000 feet/minute a special briefing should be conducted
- Power setting is appropriate for the aircraft configuration and is not below the minimum power for the approach as defined by the operating manual
- All briefings and checklists have been conducted
- Specific types of approach are stabilized if they also fulfill the following:
  - ILS approaches must be flown within one dot of the glide-slope and localizer
  - a Category II or III approach must be flown within the expanded localizer band
  - during a circling approach, wings should be level on final when the aircraft reaches 300 feet above airport elevation;
- Unique approach conditions or abnormal conditions requiring a deviation from the above elements of a stabilized approach require a special briefing.

An approach that becomes unstabilised below 1000 feet above airport elevation in IMC or 500 feet above airport elevation in VMC requires an immediate go-around. If it fails to do so, it may result in airborne loss of control, runway excursion, or controlled flight into terrain.[3]

Many of the items provided in the official guidelines can be directly analyzed using ADS-B data. Wang et al. [13] propose a method to identify stabilised approaches using radar data, which provides information similar to ADS-B. They define four performance factors: changes in speed, rate of descent, alignment with the runway centerline, alignment with the glidepath. If an aircraft is out of performance then it is performing an unstable approach. Their results for Newark Liberty International Airport shows that 65 % of the approaches are stabilised. In the thesis, a similar approach will be followed when incidents are identified with the exceedance detection methodology.

[2] data from: `https://flightsafety.org/asw-article/commercial-accident-final-approach/`
[3] Page 44 of `https://www.skybrary.aero/index.php/Hindsight_17`

B.5. Anomaly Detection

In this section, a review of different anomaly detection methods is provided. To evaluate their applicability for detecting anomalies at Schiphol Airport, certain aspects are carefully evaluated. These are the data used, the phase of flight, the number of flights, the type of A/C, and the amount of false positive. Following the division made by Basora et al. [4], the different methods analyzed are:

- Section B.5.1 **Distance based methods:** These methods determine anomalous operations based on distance/similarity metrics.
- Section B.5.2 **Boundary based methods:** These methods determine a boundary that separates normal operations from the anomalous ones
- Section B.5.3 **Statistical based methods:** These methods identify anomalies by assigning a certain probability to each data-point.
- Section B.5.4 **Neural Network based methods:** These methods use a neural network to identify anomalies.

These provide an overview going from classical approaches to recent developments.

B.5.1. Distance Based Methods

Distance based methods use a measure of similarity or distance between different points. The techniques that find wider usage in the identification of aviation safety anomalies are:

- k-Nearest Neighbours: This method is based on finding anomalies by computing the distance of all points to its k nearest points. If the distance is higher than a certain threshold, this point is considered to be an outlier.
- Clustering: The idea is to cluster data based on similarity using an unsupervised or semi-supervised technique. Anomalies are points that are not included in any cluster.

Oehling et al. [14] use Flight Operational Quality Assurance (FOQA) data recorded by airborne flight sensors with the Local Outlier Probability (LoOP) algorithm to detect anomalies in the approach phase, more specifically, the last 10NM. Generally, FOQA data is analyzed using threshold models that compare the operational data of flights to fixed thresholds.

LoOP, developed by Kriegel et al. [15], is based on the assumption that the density of k-Nearest Neighbour around an outlier is lower than around a normal data point. Its advantage over other density based approaches is that it provides an outlier in the range [0,1] which can be directly interpreted as the probability that the point is an outlier.

Oehling et al. [14] focus on the final landing phase because it is where 30% of accident occurs, and the most common type of accidents are: loss of control in flight (LOC-I), controlled flight into terrain (CFIT) and runway excursions (RE). The causes of these accidents are an aircraft entering stall for LOC-I, an aircraft being in the wrong position for CFIT, and an aircraft having too much energy for RE. To highlight these factors, these are the features used in the model:

- Lateral distance from runway centreline (NM)
- Height above runway threshold (feet)
- Average airspeed (knots)
- Angle difference between aircraft track and runway track (°)
- Average flight path angle (°)
- Average flight path acceleration (g)

This approach could be used with some adaptation in the thesis by deriving the quantities above from ADS-B data. It is extremely important to take into account that the study is performed on 1 million A320 and A321 flights. The algorithm highlighted 134 events of which 112 were false positive (84%). On the same dataset, an exceedance detection method found 2465 true anomalies and 6171 false positives (71 %).

In [16] and [17], Li et al. use an anomaly detection method based on the clustering algorithm DBSCAN. This method is called ClusterAD and it uses FOQA data to detect abnormal take-offs and landings. The analysis is performed separately for the 2 phases. The take-off phase is the period when a large amount of power is applied. When preprocessing the data, each flight is sampled every second for this phase. Instead, the landing phase is sampled in space from the touchdown point. ClusterAD consists of generating very high-dimensional vectors using 67 flight parameters from the FOQA dataset, which is then reduced in dimension using PCA before applying the clustering algorithm. This approach can identify high/low energy approaches, unusual pitch excursions, abnormal flap settings, and high wind conditions.

DBSCAN is a density based spatial clustering developed by Ester et al. [18] with the idea of grouping together points in dense areas leaving the rest as outliers. Figure B.1 shows the working mechanism of the algorithm. MINPT and ϵ are the main parameters of the algorithm. In this case, MINPT is set to 4 and ϵ is represented by the circle around each point. A point is classified as a core point if it is surrounded by at least MINPT within a distance ϵ, reachable point if it is not a core point and it has a core point within ϵ, and outliers if there are no core points within ϵ. Clusters are formed by joining connected core and reachable points. [19]

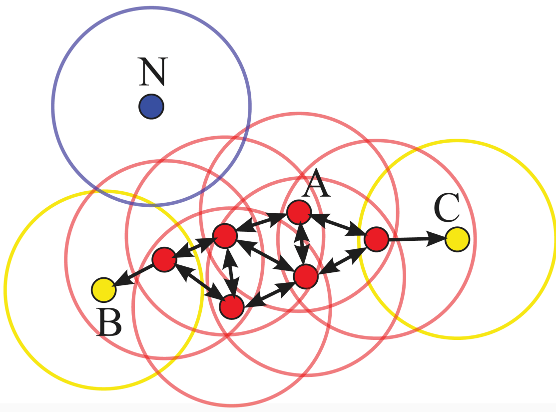


Figure B.1: Overview of DBSCAN algorithm.[19]

In Li's application of DBSCAN the value MINPT is selected based on a sensitivity analysis. The value of MINPT is fixed while ϵ varies between its maximum and minimum value. After a few iterations, MINPT is picked in the area in which the number of detected anomalies is independent of the parameters. Later, ϵ is selected depending on an acceptable percentage of outliers detected.

In [17], Li performs a detail comparison of exceedence detection, ClusterAD and MKAD, which will be discussed in the next subsection. 25,519 A320 flights are analyzed and it shows that just a few percentage are detected by all algorithms. Exceedance detection has 3 levels of severity: level 3 (severe exceedance), level 2 (moderate exceedance), and level 1 (mild exceedance). To give an idea: level 3 includes 3 % of flights, level 2 14 %, and level 1 74%. Generally level1 flights are not analysed because most of them are not operationally significant and it is too expensive to analyze them. By varying the detection threshold of MKAD and ClusterAD between 1% and 10%, as it can be seen in Table B.3. The table also shows the flights present at the same time in Exceedance detection (level 3 only) and ClusterAD, or MKAD. The comparison is performed on level 3 exceedance detection because the number of outlier is expected to be lower since its the most severe case.

Table B.3: This table summarizes the differences between MKAD and ClusterAD for different thresholds and exceedance detection (level3). [17]

Technique / Threshold [%]	1	3	5	10
ClusterAD only [flights]	244	606	919	1584
MKAD only [flights]	170	557	851	1528
Both [flights]	33	147	355	955
Exceedance detection in ClusterAD [flights]	39	93	143	220
Exceedance detection in MKAD [flights]	12	31	53	86

ClusterAD shows approximately 3 times more common flight with the exceedance detection method. This is expected because ClusterAD detects anomalies based on the deviation from nominal values. This nominal values are close to the target value used in exceedance detection. Based on this consideration and the fact that level 3 anomalies are approximately 3% of the dataset, the false positive of ClusterAD can be approximated to be 80%. The author reveals that MKAD works best with discrete data, while ClusterAD works best with continuous data. It is noteworthy that ClusterAD does not provide a score for the anomaly and that all anomalies have the same "score", while MKAD differentiate between anomalies using a scoring criteria. This is important because with MKAD it is possible to limit the search to anomalies with an high score.

Churchill et al. [20] present a DBSCAN clustering strategy based on a two steps approach to detect anomalies in aircraft ground movements. First, they cluster trajectory in space and then they cluster in time to obtain different kinds of insights. Their idea is to find standard paths using the DBSCAN algorithm while identifying anomalies as a by-product. This approach could be directly adapted for the use case of Schiphol airport to extrapolate useful safety indicators in ground movements, as they use ADS-B data. This method is tested for 10,000 flights at Charlotte Douglas International Airport.

A different kind of clustering is used by Iverson et al.[21], this time they build an Inductive Monitoring System (IMS) based on clustering, which is used for safety monitoring for the International Space Station (ISS). The key difference with the earlier examples is that they train a model on normal safe operations of the ISS. While in the previous cases, anomalies are included in the training dataset. Afterward, the trained model detects anomalies based on their distance from the cluster centroid. The downside of this approach is that 2 steps are required and to generate the training dataset another technique needs to be used to clear it from anomalies.

The main downside of these methods is that they suffer from the curse of dimensionality that is they do not scale well with high dimensional data-set. For applications with flight data which is generally high dimensional, it means that fewer dimensions are used, i.e. only a subset of the available ones or the dimensionality needs to be reduced using techniques such as PCA[22], t-SNE[23] or autoencoders (discussed later). [4]

B.5.2. Boundary Based Methods

These methods use a training dataset that represents the normal data from which boundaries of regular operations are learned. The famous python package Scikit-learn offers a way to easily implement these algorithms, which are often used in anomaly detection. Figure B.2 offers an intuitive explanation of the different algorithms applied to a 2D datasets.[24]

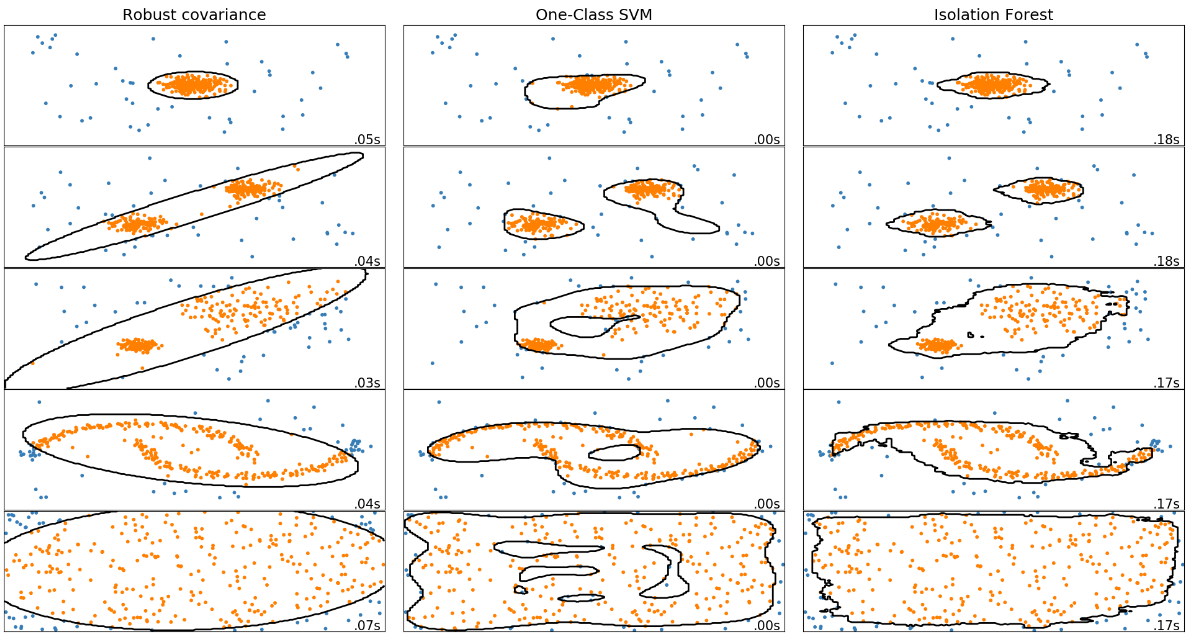


Figure B.2: Overview of classical anomaly detection algorithms.[24]

Among these algorithms, the one that finds wider usage in identifying safety risks in aviation is one-class support vector machine (OC-SVM), which is the variant of SVM designed to detect outliers. This technique finds a hyperball in high dimensional space that comprises only good observation leaving the outliers outside this border, which is described by the following Equation B.3.[25]

$$\begin{aligned} \min_{\mathbf{R},\xi} \quad & \mathbf{R^2} + \frac{\mathbf{1}}{\mathbf{vn}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \xi_i \\ \text{s.t.} \quad & ||\phi(x_i) - b||^2 \leq R^2 + \xi_i \quad i = 1, \dots, n \\ & \xi_i \geq 0 \qquad\qquad\qquad\qquad i = 1, \dots, n \end{aligned} \tag{B.3}$$

What Equation B.3 expresses concretely is finding the smallest possible hyperball by minimizing **R**, the radius of the hyperball, and ξ, the slack variable. Such that, the square Euclidean distance of the point **x**, which in kernel space is **b**, to the center of the hyperball, is smaller or equal to $\mathbf{R^2} + \xi_i$. Where $\xi_i \geq 0$is necessary to allow for outliers detection. In $\mathbf{1}/\mathbf{vn}$, the weight factor **n** represents the number of samples and it is used to normalize and make the result independent of the training size. **v** is the user design parameter used to determine the amount of slack.[25]

Based on OC-SVM is the Multiple Kernel Anomaly Detection (MKAD) algorithm developed by Das et al. at NASA.[26] In this technique, multiple kernels are employed to analyze discrete and continuous data at the same time. It finds application in the analysis of FOQA, surveillance data, and even occurrence reports. In this paper, out of the 500 parameters in the Flight Operations Quality Assurance (FOQA), the 39 most meaningful features are selected. It can detect approach anomaly, high airspeed, gusty wind anomaly, and go-around.

Das at al. in [27] and Matthews et al. in [28] test MKAD on a larger dataset made of multiple aircraft. It can detect high energy approaches, due to high speed or high altitude, and turbulent approaches, where a loss of lift causes roll or pitch changes. The basic idea is that aviation data is highly multivariate, which is impossible to compare directly. For this reason, Kernel functions are used to map data to a different dimension where similarity can be assessed. At this point, a flight is anomalous if it is different from most other flights. MKAD is the most studied and used anomaly detection algorithm in aviation. It is very often used as a benchmark for new algorithms.

Matthews et al. [29] use MKAD on surveillance data which includes the following features latitude, longitude, altitude, and time. When training MKAD also the distance to the closest neighboring aircraft is added as a feature. The training dataset is made of the 30 days preceding the test day. The 90 most statistically anomalous flights are analyzed by an expert, which considered 33 to be operationally significant.

A key problem of MKAD is the number of false positives produced by the algorithm, another experiment with a domain expert shows that out of 98 statistically anomalous flights only 20 are operationally significant. For this reason, M. Sharma et al. [30] at NASA focused on the way to overcome this limitation using Active Learning (AL). This approach is based on an expert that reviews the results of the algorithm and asses if the anomaly is operationally significant. The approach is based on the idea that there is a cost associated with labeling the dataset, the expert who reviews the result of MKAD. The process is as follows. The expert reviews an initial set of flights marked as the most anomalous by MKAD. In this way, it is possible to train an initial classifier that determines which anomaly is operationally significant and which not. Based on this classifier, the expert is interrogated again with a new set of flights and its labels are used to improve the classifier itself. The expert not only provides a label, but he can also provide an explanation, which the researchers try to add in MKAD using a new kernel. The result shows a slight improvement over MKAD. In particular, comparing the top 46 anomalous flights of AL and MKAD reveals that AL identifies 15 operational anomalies and MKAD 11.

Another application of OC-SVM is present in the paper of Puranik et al. [31]. A two-step strategy is used to detect anomalies in the approach phase in general aviation using FOQA data. Before performing the anomaly detection, they have a feature generation step in which they transform the data using some energy metrics from [32]. Using energy metrics is particularly important because many incidents are due to poor energy management of the aircraft. The model is shown in Figure B.3, DBSCAN is used to obtain the number of clusters in the data set and OC-SVM is trained to detect anomalies.

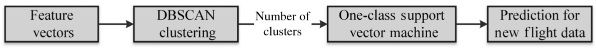


Figure B.3: Strategy of Puranik to identify anomalies in general aviation flights.[31]

Testing is performed on different energy metrics on a dataset of 1000 flights, only the top 5% anomalous flights are considered. These energy metrics use different kinds of features with the most limited one using only information about speed and altitude. Interestingly 80% of the flights detected with each method are the same. This means that using ADS-B data should give similar results. Additionally, the results are insensitive of runway, airport, and aircraft model. This means that the same model can be used with different aircraft landing at different runways, or even different airports. For future work, this paper raises the point that it would be interesting to see how operations change based on the weather. Comparing these results with the ones obtained from ClusterAD reveals that 60% of the anomalies are the same, while the differences arise from anomalies that are very feature dependent.

Temporal Logic-Based Learning is a novel anomaly detection approach that is capable of to infer signal temporal logic (STL) formulae resembling natural language from data. STL is a language used to specify system parameters, a possible output is ”if x is greater than y, then within T_1 seconds, it will drop below y and remain below y for at least T_2”.[33]

Deshmukh at al. [34] applied this technique to trajectories obtained from ADS-B DATA in terminal airspace. They call their methodology ”TempAD” and it is composed of two steps: first they apply DBSCAN to cluster together similar trajectories, and later they use temporal logic learning to define boundaries of normal operations with an optimization problem similar to Equation B.3. Operations are bounded in terms of horizontal (latitude and longitude), altitude, speed, and energy. It is important to note that bounds are not combined when detecting outliers in the sense that the algorithm is trained on each bound and tested separately. For instance, this method reveals if the aircraft is out of limits in altitude, but it does not recognize if the combination of speed and altitude is anomalous.

In a subsequent publication, Deshmukh at al. [35] transformed TempAD in an incremental learning algorithm. The authors realize that TempAD is not able to account for large amounts of data because when trained with data collected over multiple days it becomes too conservative. The cause is that air-traffic operations have a periodic pattern and they change depending on the time of the day and the season. For this reason, the incremental learning algorithm updates the operation boundaries every day. This approach can recognize go-around, S-turn (path stretch), overspeed/underspeed, late interception of glideslope, energy excess/deficit, and change of runway anomalies.

B.5.3. Statistical Based Methods

Statistical methods rely on the concept of assigning a certain probability to each datapoint. Normal data will fall in high probability region. While anomalies will fall in low probability regions.

Gaussian Mixture Model

Gaussian Mixture Model (GMM) is a probabilistic model and it assumes that all datapoints are generated by several weighted Gaussian distributions. Their wide usage in aviation derives from the fact they handle multivariate data explicitly. This model detects anomaly by calculating the probability that each data point belongs to a certain Gaussian distribution, and if this probability is lower than a certain threshold it will consider it as an outlier.

There are two common ways of estimating Gaussian distributions. The first one relies on an iterative expectation-maximization (EM) algorithm what it does is maximizing the likelihood that each data-point belongs to a certain number of Gaussians. In this approach, the number of Gaussian is assumed to be known beforehand, and the algorithm can find a local optimum. A way to check that the algorithm was applied correctly is by investigating that the resulting Gaussians do not contain singularities, i.e. a component that corresponds to a single point, with mean equals to the point and variance equals 0.

The second approach is called Bayesian Gaussian Mixture, the key difference is that this algorithm can estimate the number of Gaussians, but it needs a higher number of hyper-parameters. The most important of which is the weight-concentration parameter. If a low value is chosen, a few components will carry most of the weight. If instead, a high value is chosen, many components will play a role in the model. Table B.4 summarizes the differences between the two estimation algorithms. A shortcoming of this model is that if the training dataset has many outliers, it will fail in recognizing it as an outlier. For this reason, the dataset should be cleaned of this data, which might be very challenging.

Table B.4: The table summarizes the key difference between the two estimation algorithms, adapted from [24].

	EM	Bayesian
Pro	Speed	Automatic selection of components
	No bias	Less sensible to change in parameters
Cons	Very sensible to change in parameters	Slower than EM
	Singularities	Implicit bias

The main application of GMM in aviation is detecting an instantaneous anomaly. For instance, Puranik and Li, who are presented in the previous sections, use GMM to add instantaneous anomaly detection to their flight anomaly detection algorithms. Puranik et al. [36] use GMM because it deals with multivariate data explicitly with EM estimator. They find the number of components by using the Calinski-Harabascz clustering criteria because it provides the clearest definition of clusters. In this paper, the authors not only identify instantaneous anomaly but also monitor teach aircraft operation with anomalous probability score.

Furthermore, there is a comparison with exceedance detection. It is shown that there is a negative correlation between the two methods. A point has a high exceedance level, but a low anomalous probability score. However, if the points considered are only the top 1% anomalous points resulting from the GMM method, they are also anomalies for the exceedance detection algorithm. In this application, GMM can detect anomalies in take-off and landing.

Similarly, after developing ClusterAD using DBSCAN, Li et al. [37] investigate the use of GMM. On a high level, the key difference is that the new approach can detect an abnormal data point instead of an abnormal flight phase. This is advantageous when an expert has to review the flight because he can analyze directly the anomalous part. Normal operations are captured by the Gaussian components of the model as ILS approach, different kinds of visual approach and touchdown. Li notes that full covariance matrices are unnecessary even when flight parameters are not independent. Therefore diagonal covariance matrices can be used which have the advantage of being more computationally efficient. As Puranik, Li uses EM estimator, but this time the number of components is chosen based on the Bayesian Information Criterion(BIC), which finds the optimal balance between model accuracy and complexity.

The authors analyze a dataset of 10528 A320 flights with this algorithm and they analyzed the 53 most anomalous flight with a group of experts that revealed a 80% false positive rate. Furthermore, they compare this new algorithm with the previous ClusterAD, MKAD and exceedance detection. The result is that this new algorithm produces results more in line with level 3 exceedance detection. 26% of the anomalies detected by GMM with a threshold of 3% are also in level 3% exceedance detection. This means that the percentage of false positive can be considered to be around 70%.

The main challenges in applying the work of Li and Puranik to the thesis are the difference in data and the different aircraft. The data available for the thesis is ADS-B data, while in the work of Li and Puranik FOQA data is used. The work of Puranik and Li focuses on a specific type of aircraft, while the thesis will need to manage most aircraft landing at Schiphol. The second problem can be overcome by categorizing the aircraft in different categories and performing different analyses per category.

Principal Component Analysis
Principal Component Analysis(PCA) is a linear transformation algorithm that transforms the feature space such that variance is maximized on the components. The principal component has a maximum variance, and the variance decreases for the successive components. In particular, it is guaranteed that if reconstruction is performed using only the principal component, the obtained 1-D result minimizes L2 error, the sum of the square differences between the original value and the reconstructed value. Flight data is often non-linear, and researchers use modified PCA able to deal with non-linearities. For instance, Zhang et al. [38] developed an efficient Kernel PCA (KPCA) algorithm optimized specifically for flight data.

Jarry et al. [39] use Functional PCA (FPCA) and unsupervised learning with track radar data to detect Non-Compliant Approach (NCA) that is a precursor of Non-Stabilised Approach(NSA), which may lead to catastrophic accident Control Flight Into Terrain (CFIT). The 2-step approach consists of applying FPCA on total energy trajectories, and later applying hierarchical clustering. The total energy is used because an excess of energy is the key reason for non-compliance. The radar data consists of longitude, latitude, altitude, ground speed, time, vertical speed, heading, and aircraft type. These are the same parameters provided by ADS-B, which make this method adaptable for the thesis. FPCA allows reducing the dimension by estimating a truncated Karhunen-Loève decomposition. It is based on the fact that aircraft radar positions are functional data, and it can be considered as realizations of an underlying stochastic trajectory function.

Furthermore, to allow for a comparison between different flights that have different operational speeds, the reference system is based on the curvilinear distance from the runway threshold. The author analyzes a small interval of trajectory at the time to provide segment non-compliance score by using a sliding window and applying recursively the algorithm. To compute the non-compliance score, Hierarchical DBSCAN (HDBSCAN) is used after which the Global-Local Outlier Score from Hierarchies (GLOSH) algorithm is applied to obtain a score between 0 and 1.

HDBSCAN constructs the minimum spanning tree among the data and computes a density dendrogram based on the size of the edges between the data. Clusters are selected from the dendrogram based on the number of points per each sub-element. According to the authors, this approach is very sensitive to a change in sliding window size. Jarry et al. [40] extensively tested this approach at multiple airports with good results. The reason for the development of this method arises from the fact that an exceedance detection methodology has a high false-positive rate. In this paper, the author mention a system from the French government called ELVIRA which has a false positive rate of 98 %. While the FPCA method has a false positive rate of 70 %.

Independent Component Analysis
Independent Component Analysis (ICA) is a statistical technique that consists of a linear transformation of some observed data to a new feature space. ICA assumes that the observed data originates from a mix of non-Gaussian and mutually independent latent variables. ICA reveals the latent variables, or independent components, which maximize feature reconstruction such that it is possible to predict the new feature from the original one, and vice-versa. Jiang et al. [41] applied ICA in the aviation sector. In particular, they use a complex network approach based on ICA to monitor air traffic congestion outside the Terminal Maneuvering area.

Regression model based
This approach consists of training a regression model and highlighting anomalies based on the distance between the expected value produced by the regression model and the actual value. In the case of aircraft data, the vector autoregressive (VAR) model is used to take into account multiple time series variables using a single model.

Melnyk et al. [42] [43] models the flights using a dynamic Bayesian network combining discrete and continuous variables taken from the FOQA database. The model is a semi Markov switching vector autoregressive model (SMS-VAR) and it works as follows. A single VAR model built using continuous data does not represent the entire flight. For this reason, different VAR models are used depending on the flight phase, which is revealed by a combination of the flight switches (change in autopilot). The flight phase is not observed in data and it is modeled as a latent parameter (hidden Markov process). An example of flight phases identified for go-around: descending, leveling-off at a certain altitude, final approach, sudden take-off, and circling.

After training the model on the full dataset, it finds anomalies by predicting the next state given the current one. Then, it computes the dissimilarity using KL divergence between the two states. Testing the algorithm on simulated data shows that the model works well with a percentage of anomalies up to 10%, afterward the model starts to learn the anomalies and the model fails. Testing a VAR, SMS-VAR and MKAD on 20,000 flights and taking the 100 most statistically anomalous flight reveals that VAR detects 17 operational anomalies, SMS-VAR 49, and MKAD 38. Although this approach seems to perform better than MKAD, it will not be used in the present thesis because in order to achieve the same level of performance flight switches information should be used as well. Otherwise, this would be a simple VAR model that performs worse than MKAD.

B.5.4. Neural Network Based Methods

The principle of the neural network is to resemble the working mechanism of the human brain. It is a simulation involving interconnected neurons. In general, these are organized in layers, and each neuron of a layer is connected to all neurons of 2 neighboring layers. The first layer receives the input data, and the last layer contains the output of the network. In between, there are the hidden layers, which allow gaining a deeper 'understanding' of the data. This structure is shown in Figure B.4. "Understanding" means finding a relationship between data. Thus a neural net represents a hidden mathematical function that reproduces the relation of a training set.

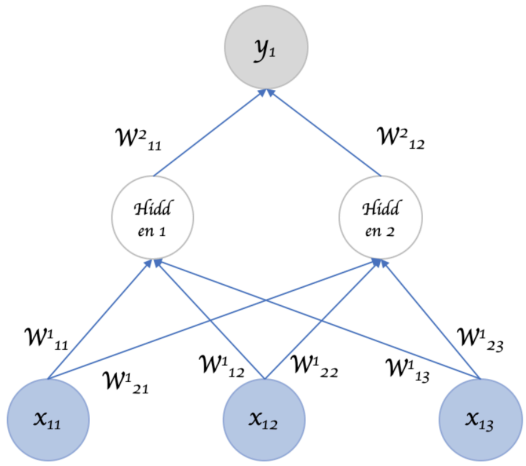


Figure B.4: Structure of common neural network.[44]

There exist different types of neural networks used in anomaly detection in the aviation section, which will be analyzed in the following sections.

Recurrent Neural Network

Recurrent Neural Networks (RNN) have a different structure compared to the one showed in Figure B.4. This type of network is designed to take series type input with no predetermined size. As it is possible to see in Figure B.5, the key difference with Figure B.4 is that the weights are applied "recurrently" for all inputs and that there is a connection between the different input sequences. This connection is implemented via a hidden state, which is calculated from the current input and the hidden state at the previous step. This feature is also called memory capability as if RNN remembers what happened at the previous timestamp. The idea behind this is that in this way it can understand the sequential nature of data, and in this case of flight data.

Nanduri et al.[45] use RNN based models, Long Short Term Memory (LSTM) and Gated Recurrent Units (GRU), with FOQA data. LSTM introduces extra elements to the basic RNN, most importantly the cell state. In this way, LSTM networks are able to understand if the information is important in a given context. GRU is a variation of LSTM that merges non-relevant information and the input.

In an experiment, both RNN and MKAD are trained on trajectories generated by X-Plane. In particular, there are 478 training examples and 22 test samples (11 of which contain an anomaly). This test set-up raises concerns as the amount of data used is very limited and it is possible that if this model is used with real data results will differ from the one of this author. The output of the RNN model is the predicted state of the aircraft. An anomaly is detected when the error between real state and predicted

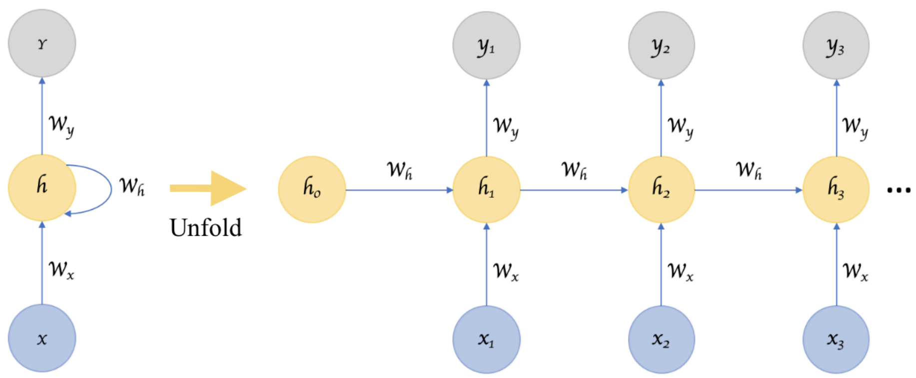


Figure B.5: Structure of recurrent neural network.[44]

state is high. RNN models did not have false positives, as well as MKAD. Furthermore, RNN models show a much higher recall than MKAD classifying correctly 80% of anomalies compared with 50% of MKAD. Among the two RNN models, GRU shows the best performance as the MSE is higher, which means that there is less ambiguity in detecting anomalies. Recurrent Neural Networks are able to detect 9 out of 11 canonical anomalies in the test dataset, while MKAD 6 out 11.

Autoencoder

Autoencoders are feed-forward neural networks with an equal number of input and output nodes. It is composed of two parts: an encoder and a decoder. Figure B.6 shows the general structure of an autoencoder. As it is possible to see in Figure B.6, an autoencoder reduces the feature space of the input in the encoding part and tries to recreate it in the decoding part. The objective of the autoencoder is to minimize the error of the reconstructed result. Anomalies are detected when this error is high.

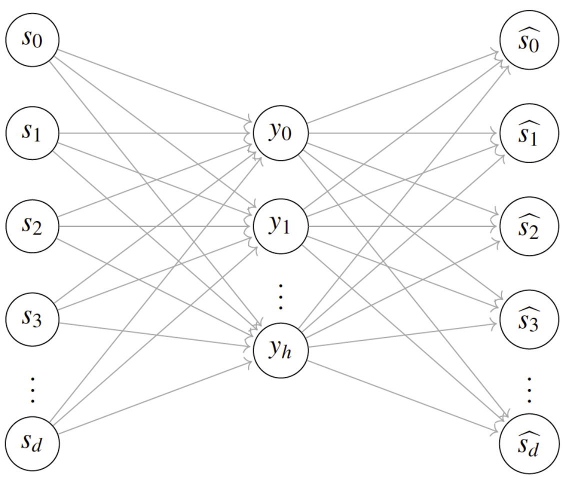


Figure B.6: Structure of an autoencoder.[46]

Olive et al. [46] analyze one year of ADS-B data between different cities to detect anomalies in-flight trajectory and identify controllers command. Their analysis excludes terminal maneuvering area operations. They use an autoencoder with 150 input and output neurons, 64 neurons in the hidden layer, sigmoid activation function, and mean squared error as the loss function. They show that the reconstruction error is higher for anomalies caused by weather conditions and traffic regulations. While ATC deconfliction causes anomalies with a lower error. Interestingly they use data from the OpenSky network.

Expanding on the previous work, Olive and Basora [47] present a framework to identify anomalies in trajectories based on a two step approach: clustering and anomaly detection. First, they apply DSBSCAN to identify trajectory clusters, then in each cluster, they use an autoencoder with a regularization term to identify anomalies. They analyze 7 months of data from the OpenSky database for the LFBBPT sector controlled by Bordeaux ACC.

Janakiraman and Nielsen [48] implement a special type of autoencoder called Extreme Learning Machine (ELM) to detect anomalies in the last 60 miles before landing. ELM maps the input data to a random hyperspace, like autoencoders, then spectral embedding is applied. Anomalies are detected based on their distance from the embedded space origin. They use real surveillance data with these parameters: latitude, longitude, altitude, and distance from the closest neighbor. The last parameter is used to consider the loss of separation anomalies. They compare the ELM algorithm with MKAD on a dataset of 43000 flights and a test dataset of 115 flights producing similar results with the difference that ELM 2X faster. ELM has a very high false alarm rate (more than 70%), while for MKAD is 50 %. It is important to note that MKAD is trained using the same parameters employed for ELM. The anomalies detected by ELM are high-speed landing, overshooting their final turn before landing and overtaking other landing aircraft.

B.6. Precursors Detection

This section describes different methods to identify precursors to anomalies. After incidents are identified, different algorithms exist to identify precursors. These methods are similar to the anomaly detection ones with the key difference that this time the algorithms are modified to take most advantage from labeled data. In fact, the data was labeled as anomalous or normal, after applying exceedance detection and anomaly detection. Precursor detection not only reveals which compromised states lead to an incident, but also correcting action taken to avoid it.

Reinforcement learning

Reinforcement Learning (RL) is based on the idea that an agent learns to perform a particular task by interacting autonomously with the environment. In order to learn the task, at every iteration, the active agent performs an action in response to the observation from the environment. Depending on the action, the agent receives a reward and strives to maximize the cumulative reward in the long-term by considering the expected future rewards.

Janakiraman et al. [49] at NASA uses a mix of reinforcement learning and inverse reinforcement learning to identify precursors. The idea is that the algorithm searches for suboptimal actions in the adverse time series which increase the risk of an incident.

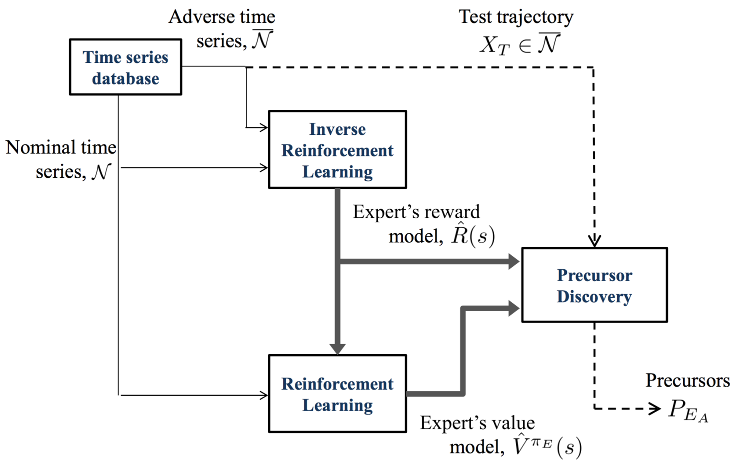


Figure B.7: Structure of precursor detection algorithm.[49]

The algorithm is shown in Figure B.7. Before looking for precursors, it is necessary to label the time series (a flight) to identify flights with an adverse event. Flights with an incident are considered the result of an non-expert agent, while nominal flights the result of an expert one. First, inverse reinforcement learning is performed to deduce the expert's reward model, which represents the instantaneous consequence of a decision. Second, the long-term consequence of a decision is derived using reinforcement learning on the expert reward model and the nominal time series to obtain the expert's value model. Precursor discovery uses the expert's value model and reward model on the adverse time series to identify a state which should have been different and has an high probability to lead to an anomaly.

Multiple-instance learning

After developing the precursor model based on inverse reinforcement learning, Janakiraman [50] develops a novel approach based on multiple-instance learning (MIL) and deep recurrent neural networks (DRNN). MIL is a supervised learning algorithm particularly indicated for weakly-supervised data. The idea of MIL assumes that data is grouped in bags and that a bag is positive if one element is positive and negative if all elements are negative. Once trained, MIL learns to predict element's label. Before applying this precursor algorithm, flights needs to be labeled using an anomaly detection method. In this model, a single flight represents a bag and an element of the bag is the sequence of measurements up to the current time. The standard MIL is not able to understand the temporal relation between elements, and DRNN is introduced to add this capability. The results present precursors to high-speed exceedance (HSE) during landing. The method is compared to the inverse reinforcement learning approach and it shows a lower false alarm rate and higher accuracy.

Temporal Logic Learning

Deshmukh at al.[51] design a precursor algorithm based on TempAD. Fundamentally, once anomalies are recognized using TempAD, the idea is to re-run the anomaly detection algorithm with labels and using only the part of the trajectory that precedes the occurrence of the anomaly. In looks for an occurrence in the states preceding an anomaly such that if the occurrence occurs it is grantee that the anomaly will happen.

The algorithm analyzes not only where the feature where the anomaly occurred, but also other features to identify a possible cause from another plausible feature space. It is important to note that the feature space is expanded with the horizontal and vertical distance of the closest neighboring aircraft.

FPCA

A very different approach is used by Olive et al. [52]. The goal this time is very specific. The idea is to quantify the risk of runway excursion by evaluating precursors using Mode-S data. Instead of having an algorithm looking for precursors autonomously, the idea is to use an existing fault tree models developed by Future Sky Safety and estimate the probabilities of contributing factors using ADS-B data. Using FPCA on particular features of the data reveals that the secondary components resulting from the decomposition carry important information about anomalous behavior.

Regression Trees

Herrema et al. [53] deal with the problem of analyzing abnormal Arrival Runway Occupancy Time (AROT). First, abnormal AROTs are found modeling AROT using a normal distribution and identifying them based on a deviation of more than 2σ. Afterwards a regression tree is built from the anomalous AROTs by iteratively finding predictors. At each iteration, the variables and the split point are chosen such that MSE between predictions and AROT is minimized.

B.7. Summary

The previous sections have described in detail many anomalous safety event detection methods and precursor detection methods applied in aviation. Table B.5 shows an overview of different strategies focusing on the technique, the type of data, the flight phase, the aircraft type, and the false positive rate.

Typical exceedance detection algorithms used in the industry are proprietary and they are not publicly available. Furthermore, they wouldn't be usable in the specific case of the thesis because they use many features from the FOQA data, which are not present in the ADS-B data. In this thesis, exceedance detection is performed designing a set of rule-based algorithms. The knowledge for these rules is taken from safety regulations. This is inspired by the work of Wang et al.[13] that use radar data and a set of predefined threshold to identify anomalous aviation events in the approach phase.

It is noteworthy that the most well-recognized technique in anomaly detection is the OC-SVM variant called MKAD and developed at NASA by DAS et al.[26]. This algorithm is the benchmark used in research when developing novel approaches and it has been tested extensively for approximately 10 years, Das et al.[27] and Matthews et al.[28] [29].

As previously anticipated, the false positive rate is an important characteristic when evaluating detection algorithms. The performance of these algorithms is far from ideal with an average false positive rate of approximately 70%. The best performing technique is SMS-VAR developed by Melnyk et al.[42] at NASA. This is not usable with

Table B.5: This table shows a summary of the strategy identified to achieve the objective. In the table, "/" means that different models are developed based on the same strategy depending on the flight phase. When talking about A/C type, "multiple" refers to a single model that can be used at the same time with multiple A/C models. False positive data is not specified when absent in the paper.

Strategy	Technique: citation (data, flight phase, A/C type, false positive[%])
Exceedance detection	- Rule-based: [13] (radar data, approach, multiple)
Distance based methods	- k-NN: [14](FOQA, approach, A320, 84) - DBSCAN: [17](FOQA, approach, B777/A320, 80) [20](ADS-B, ground movements, multiple)
Boundary based methods	- OC-SVM: [26](FOQA/radar, take off/approach, multiple, 65-75) [31](FOQA, approach, general aviation) [34](ADS-B, approach, multiple)
Statistical based methods	- GMM: [36](FOQA-radar, take off/approach, general aviation) [37](FOQA, take off/approach, A320, 70) - PCA: [39](radar, approach, multiple) - SMS-VAR: [42](FOQA, approach, multiple, 50)
Neural Network based methods	- RNN: [45](simulated FOQA, approach, multiple) - Autoencoders: [46](ADS-B, en-route, multiple) - ELM: [48](radar, approach, multiple, 70)
Precursor detection	- RL: [49](labeled data, approach/take-off, multiple) - DT-MIL: [50](labeled data, approach, multiple) - OC-SVM: [51](labeled data, approach, multiple) - PCA: [52](existing fault tree + ADS-B, ground movements, multiple)

ADS-B data because it is highly dependent on the pilot button switches. The algorithm works in such a way that when it detects a switch it changes its dynamics.

There is an important consideration to make about the false positive rate. The values provided are not exact because it heavily relies on expert judgment. The concept of what is operational significant varies depending on the expert. This is shown by LI et al. in [37] where 4 experts could not agree on which situation poses safety concerns.

Table B.5 shows that researchers have mainly focused on detecting anomalous safety events in the approach phase. This is crucial as 50% of the accidents happen in this phase of flight including airborne loss of control, runway excursion, or controlled flight into terrain.[4]

Interestingly, distance based methods are generally less applicable to multiple aircraft at the same time. The reason is that these models, such as Oehling et al.[14] and Li et al.[17], use features directly and identify anomalies based on the distance from cluster. On the other hand, all other anomaly detection models implicitly project the feature in other spaces where anomalies are computed taking into account also the inter-relations of the features. It is important to underline that these researchers used FOQA data, which provides features that are more aircraft dependent.

Furthermore, the applicability of a single model to multiple aircraft types grows when features are preprocessed in a meaningful way. Puranik et al.[31] [36], Deshmukh

[4] data from: `https://flightsafety.org/asw-article/commercial-accident-final-approach/`

et al.[34] and Jarry et al.[39] augment the set of features available with some energy metrics[32]. This is more important when ADS-B and radar data are used, as it enriches the limited amount of features available.

Techniques that have been recently developed to identify anomalies include Neural Networks, different types of networks have been implemented by different researchers: Naduri et al.[45] works with Recurrent Neural Networks, Olive et al.[46] with Autoencoders, and Janakiraman et al.[48] with Extreme Learning Machines. Results from these methods are promising, and many researchers are working with Autoencoders. Although, as it is possible to see from the table, the false positive rate remains in line with more classical approaches for the moment. Goel et al. [54] test VAR, RNN-LSTM, and autoencoder-LSTM in reconstructing time series of aircraft data. They perform some experiments on real and synthetic data. It shows that VAR performs better than both LSTMs. When a value can be predicted by looking at a few previous steps, a linear model (VAR) is better.

Precursor detection techniques are mainly semi-supervised approaches that use the identified anomalous aviation events as labeled data, such as Janakiraman et al.[49] [50] and Deshmukh at al.[51]. A very different approach is used by Olive et al.[52]. Instead of having an algorithm looking for precursors autonomously, the idea is to use an existing fault tree model developed by Future Sky Safety and estimate the probabilities of contributing factors using PCA with ADS-B data.

C Preprocessing

This chapter provides an overview of the data used and the preprocessing strategy experimented with. Preprocessing is a required activity when dealing with large amounts of data. Often data contains outliers and noise that can jeopardize the result of the analysis. Outliers are erroneous data points, and noise is an oscillation in the recorded data. The reasons for these phenomena are errors in the raw data captured with the antenna and possible errors in processing the raw data.

C.1. Manual Cleaning

In this phase, the intuition of possible useless data and possible outliers is used to delete and modify data. The steps are removing duplicates, change in measurement units, and merging the Open Sky and the Delft databases Furthermore, operations out of the limits of the Schiphol Airspace are removed. Trajectories with few points are removed because they can't be analyzed. Also, trajectories belonging to a particular type of vehicle are removed such as ground vehicles and helicopters. Finally, the trajectory is divided into different pieces depending on the aircraft operation: landing, take-off, and ground movements.

C.2. Divide Flights

This section describes the divide flights function. This function assigns to each aircraft, identified through the ICAO code, its flight phases. In fact, the same aircraft might land and take-off multiple times on the same day. There are 4 possible phases: land (**L**), take-off(**T**), unidentified(**C**), and ground operations(**G**). The unidentified phase is a special case that includes aircraft cruising and aircraft taking off and landing without exiting Schiphol Airspace. These are identified based on 2 steps:

- Identify changes in phase
- Identify ground movements

Identify changes in phase assumes that if data is not received within 20 minutes, it must have changed. Once these breaking points are collected in a list, phases are assigned as follows:

```
list_break_points = [ex. t5, t22, t44]
for every element in list_break _point:
  # compute overall tendency of aircraft, is it going up or down?
  up_down = flight_df.loc[between break points, 'alt'].diff().sum()
  if up_down > 100:  #goes up
  # phase_attribute is a list long as the number of points recorded for a
   given aircraft
    phase_attribute[between break points] = 'T'  # all points between
   break points are representitive of a take-off
  elif up_down < 100:  #goes down
    phase_attribute[between break points] = 'L'
  else:
    phase_attribute[between break points] = 'C'
```

Identify ground movements assumes that ground movements are all the ones taking place below 100ft, and it assigns **G** to all these points in the phase_attribute list.

C.3. Normalization

Normalization is used to prepare the data for the following steps. Normalizing the data ensures that all data is uniformly scaled between 0 and 1. In this way, it is possible to use the same algorithms for data that have different statistical characteristics. For instance, altitude has values ranging from 10,000 ft to 0, while speed has values between 500 kts and 0 kts. When a threshold-based algorithm is used, a different threshold value would be required. While if the data is standardized the same value can be used. The standardization procedure is shown in Equation C.1.

$$x_i^{'} = \frac{x_i - mean(X)}{var(X)} \tag{C.1}$$

Every trajectory is analyzed separately, and every feature of the trajectory is normalized. Now the data is ready for the remotion of outliers.

C.4. Deletion of Outliers

Removing outliers improves the quality of the analysis. First of all, it improves the accuracy of the smoothing procedure. It will also improve the result of the overall result of the research when performing the analysis. Batista et al. [55] use a Hampel filter to remove outliers from radar data. Hampel filter is based on the 3σ rule, which assumes the data to be normally distributed and the outlier to be a point further than 3 standard deviations.

$$MAD = median\left(|X - median(X)|\right) \tag{C.2}$$

The Hampel filter runs a moving window through the data and replaces outliers with the median of the sample based on a robust 3 σ rule. Instead of using the mean of the window, it uses the median. While instead of the standard deviation, it used the median absolute deviation (MAD) calculates as shown in Equation C.2. The algorithm has two tuning parameters: the size of the window size, and the threshold for outlier removal, expressed in the number of median absolute deviations. Limitations of this technique are the masking effect means, outliers are undetected because there are multiple ones next to each other, and the swamping effect, a good data point is considered an outlier because of the presence of multiple outliers in its proximity.

C.5. Smoothing

The purpose of smoothing the data is to remove noise. Different techniques are experimented with and researched. It follows an overview of the discarded techniques:

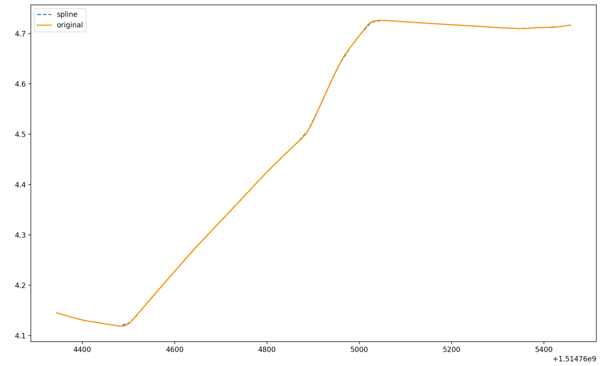
- Forward-backward filter is excluded because it provides sub-optimal results that vary considerably the behavior of data.
- Savitzky-Golay filter is a low-pass filter that runs a moving average over the data. It is found to change the behavior of data.
- Kalman smoother relies on a Kalman filter to determine the most probable position of the aircraft given the sensor data available. A shortcoming of this technique is that an acceleration needs to be assumed to construct the aircraft motion matrix. This research analyzes operations in the Terminal Schiphol Airspace that are all operations with variable acceleration. For this reason, this method is excluded as it is required to define a possible function of the acceleration.
- Generalized additive models (GAM) are linear combinations of base functions used to fit a set of data. There is an optimization routine that runs to produce a result with little least square error. This approach works quite well, but it uses a complex model that requires some time to optimize.

After some tuning, splines are found to produce the most efficient result. A spline is a piecewise polynomial and knots are the points where the different parts join. In this case, 3rd order B-splines are used. The main challenge is tuning the knots parameter. A basic choice would be choosing equally spaced knots, this offers sub-optimal solutions. Then Dung et al. [56] propose a method that places the knots at variable positions while minimizing the least square error of the current segment. The implementation works as follows. Serial bisection is applied to the data to find the largest segments for which the least square error is lower than a certain threshold.

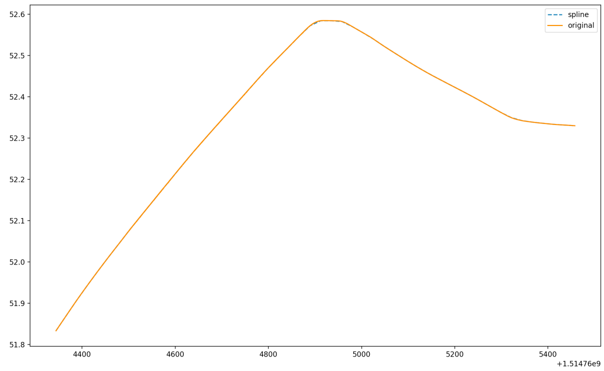
The result of the smoothing procedure is a set of spline functions where noise is removed, as can be seen in the next section. Having a set of functions describing every feature is particularly advantageous because it allows resampling at any frequency, and it can be stored extremely efficiently. Instead of storing thousands of rows per trajectory, now a single row containing the spline coefficients is enough. **This method showed to be unreliable. Smoothing has the side-effect of removing too much information from the data, as can be seen in Figure C.1.** The result is for a trajectory selected randomly on 1st January 2018. It shows that the preprocessing strategy can effectively remove noise and compute smooth polynomials.

C.6. Track Angle

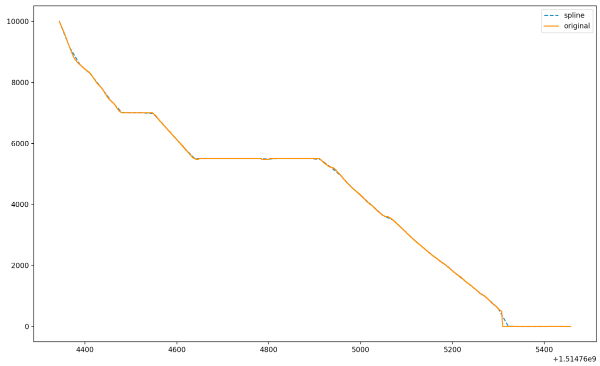
Track angle data has shown to be a valuable source of information, but its accuracy is highly susceptible to rapid aircraft movements. It is observed that sometimes when an aircraft performs a go-around, or after landing when it moves along the taxiways, its



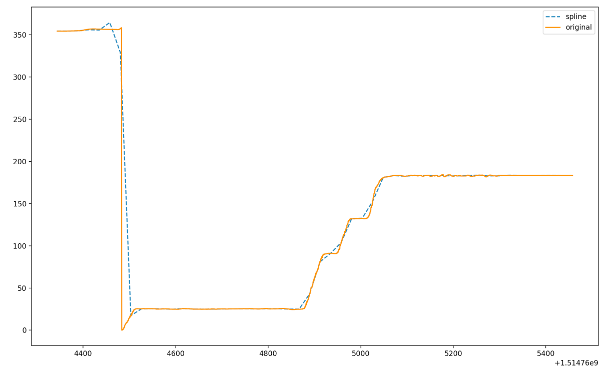
(a) Longitude



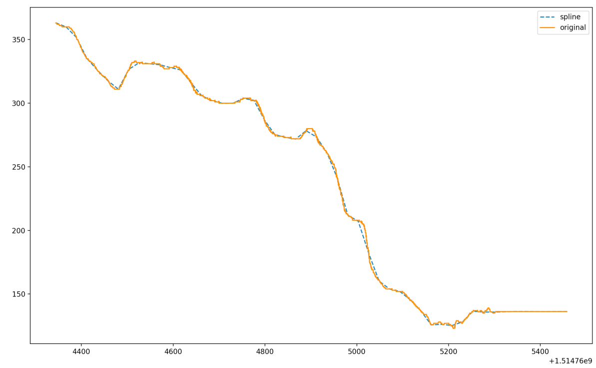
(b) Latitude



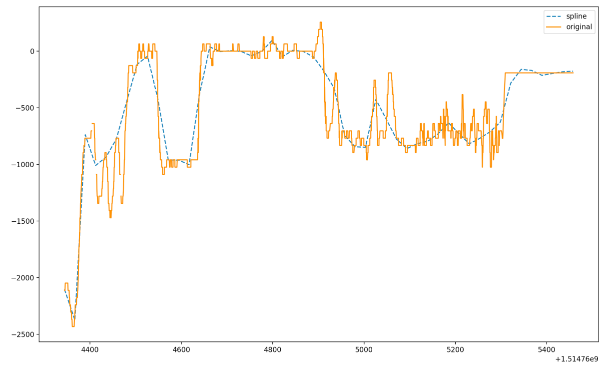
(c) Altitude



(d) Heading



(c) Ground Speed



(d) Rate of Climb

Figure C.1: The figure shows the output of the smoothing strategy for a single trajectory. Data fit with a spline is in blue and the original one is in orange.

track angle indicator doesn't follow the aircraft movements. The track angle communicated via ADS-B data doesn't change whereas it is clear it should have. To solve this issue, an algorithm has been developed such that the track angle information communicated from the ADS-B data is compared to an estimate of it computed using latitude and longitude information. If the difference between the two is higher than 60 deg, the estimate is used as track angle data. To limit the influence of poor measurements in the track angle estimation, a window of 40 s is considered and a minimum amount of 5 points.

C.7. AIP the Netherlands

Information from AIP Netherlands produced by LVNL is used to define the ILS intercept on a map, useful for detecting unstable approaches. In particular, information is collected from the instrument landing procedures charts. The intercept is defined as the line connecting the Final Approach Fix (FAF), Runway Threshold (THR), and Runway Localizer (LOC). Table C.1 has an overview of these values for runways with an ILS procedure in place.

This information is used to define the allowed horizontal bound for all runways with an ILS intercept procedure in place: 06, 18C, 18R, 22, 27, 36C, 36R.

C.8. METAR Data

METAR reports can be used to have an idea of how the weather was like on a particular day. In this case, METAR reports from Schiphol are downloaded from IOWA ASOS network[1]. In particular, information about wind speed and direction is collected for each trajectory. Furthermore, using METAR reports is possible to determine if an aircraft is flying in VMC or IMC. This is performed using Table C.2.[57]

Table C.2: This table shows the definition of VMC and IMC condition.

Altitude [ft]	VMC	IMC
<3000	- Visibility $\geq$ 5km - In sight of the surface and clear of clouds	- Visibility < 5km - Not clear of clouds
$\geq$3000	- Visibility $\geq$ 5km - Distance from clounds: 1500m horizontally and 1000ft vertically	- Visibility < 5km - Not clear of clouds
>10000	- Visibility $\geq$ 8km - Distance from clounds: 1500m horizontally and 1000ft vertically	- Visibility < 8km - Not clear of clouds

[1]https://mesonet.agron.iastate .edu/request/download.phtml

Table C.1: This table shows an overview of the waypoints used to define the ideal intercept.

Runway	Wpt_type	Lat	Lon	AMSL [ft]	Hdg True [deg]
6	EH609 - FAF	52°14’04”N	004°35’45”E	2000	057.92
6	EH616 - INT	52°15’13”N	004°38’43”E	1310	057.92
6	THR06	52°17’20.78”N	004°44’14.01”E	50	057.92
6	LOC06	52°18’25.71”N	004°47’03.10”E	-11,1	057.92
22	EH661 - FAF	52°23’28”N	004°54’49”E	2000	221.27
22	EH651 - INT	52°21’51”N	004°52’29”E	1310	221.27
22	THR22	52°18’50.51”N	004°48’10.89”E	46	221.27
22	LOC22	52°17’55.25”N	004°46’51.82”E	-13,7	221.27
18C	EH630 - FAF	52°26’01.7”N	004°44’58”E	2000	183.22
18C	EH626 - INT	52°23’52.6”N	004°44’45.3”E	1310	183.22
18C	THR18C	52°19’53.03”N	004°44’24.11”E	50	183.22
18C	LOC18C	52°18’00.50”N	004°44’13.78”E	-12	183.22
18R	EH621 - FAF	52°27’45.7”N	004°43’16”E	2000	183.20
18R	EH622 - INT	52°25’36.2”N	004°43’04.1”E	1310	183.20
18R	THR18R	52°21’36.93”N	004°42’42.21”E	50	183.20
18R	LOC18R	52°19’32.22”N	004°42’30.93”E	-13	183.20
27	EH639 - FAF	52°19’26”N	004°57’50”E	2000	266.82
27	EH640 - INT	52°19’19”N	004°54’19”E	1310	266.82
27	THR27	52°19’06.15”N	004°47’48.81”E	50	266.82
27	LOC27	52°18’59.67”N	004°44’39.71”E	-12,1	266.82
36C	EH632 - FAF	52°12’12”N	004°43’42”E	2000	003.22
36C	EH633 - INT	52°14’22”N	004°43’54”E	1310	003.22
36C	THR36C	52°18’20.99”N	004°44’15.66”E	50	003.22
36C	LOC36C	52°20’02.33”N	004°44’24.97”E	-12	003.22
36R	EH636 - FAF	52°11’18”N	004°46’04”E	2000	003.25
36R	EH635 - INT	52°13’28”N	004°46’16”E	1310	003.25
36R	THR36R	52°17’26.97”N	004°46’38.45”E	50	003.25
36R	LOC36R	52°19’24.61”N	004°46’49.34”E	-11,1	003.25

C.9. Obtaining Altitude Above Ground and Airspeed

ADS-B data provide respectively ground speed and barometric altitude. Ground speed is dependent on the windspeed. Barometric altitude assumes standard temperature and pressure. These two values can be adjusted using the METAR reports. In Equation C.3, it is possible to compute the altitude of the aircraft knowing p_{ISA}, T_{ISA}, p_{METAR}, T_{METAR} and the constants.

$$\begin{aligned}
T_{AC_{ISA}} &= T_{ISA} + a \cdot h_{AC_{ISA}} \\
p_{AC_{ISA}} &= p_{ISA} \cdot \left(\frac{T_{AC_{ISA}}}{T_{ISA}}\right)^{\frac{-g}{a \cdot R}} \\
p_{AC_{ISA}} &= p_{AC_{METAR}} \\
T_{AC_{METAR}} &= T_{METAR} \cdot \left(\frac{p_{AC_{METAR}}}{p_{METAR}}\right)^{\frac{a \cdot R}{-g}} \\
h_{AC_{METAR}} &= \frac{T_{AC_{METAR}} - T_{METAR}}{a}
\end{aligned} \tag{C.3}$$

However, the data from the METAR report is obtained at ground level, and at higher altitudes, there might be large differences in wind speed and direction. For this reason, METAR reports are used to correct speed up to an altitude of 100 m, where the difference in wind speed is estimated to be approximately 1.5 m/s. [2] After this point, data is extracted from the National Climatic Data Center (NCDC), a department of NOAA.[3] This dataset offers wind data at intervals of 700 ft with updates every 6 hours.

For both the METAR and NOAA datasets, the direction is expressed, in tens of degrees, from which the wind is blowing with respect to the true north.[4] While aircraft direction is expressed in terms of where it is going to. For this reason, the correction is, as follows:

$$TrueAirSpeed = GroundSpeed - WindSpeed \tag{C.4}$$

C.10. Weather Algorithm

It is important to establish a way to assess the severity of the weather to determine its contribution to a go-around. This section describes an algorithm that attaches a numerical score to the quality of the weather. The METAR reports are the input data for the algorithm. These reports are generated every 30 minutes at Schiphol airport, and they contain many weather features (i.e. precipitation type, visibility, cloud base, wind speed). Much of the content of this chapter is based on the technical report ”Algorithm to describe weather conditions at European airports” published in 2011 by EUROCONTROL.[58]

The objective of this technical report is to describe the ATMAP weather algorithm, which ”assesses the weather conditions independently from traffic congestion”.[58] The algorithm described by EUROCONTROL works as follows:

[2] http://euanmearns .com/high-altitude-wind-power-reviewed/
[3] https://www.ncdc.noaa .gov/data-access/model-data/model-datasets/global-forcast-system-gfs
[4] https://www.wmo.int/pages/prog/www/WDM/Guides/Guide-binary-2.html

1. Each METAR report is evaluated considering weather class
2. The scores of each class are summed, and the overall score is obtained for a single METAR report
3. The quality of a given day is evaluated averaging the values of all METAR scores

The end goal of EUROCONTROL's algorithm is to determine the overall quality of the weather in a single day. They assume that if the score is higher than 1.5 the day is bad. Shultz et al. [59] work with METAR reports from Frankfurt airport and they recommend to increase the threshold to 2.8. From this description, it is clear that for this thesis we are interested specifically in points number 1 and 2. After a description of how the score for each METAR is calculated, a new threshold is defined to assess the overall severity of a single report.

There are 5 weather classes to assess the severity score for a METAR report: ceiling & visibility, wind, precipitations, freezing conditions, and dangerous phenomena. Each class analyses certain elements of the METAR reports (i.e. wind speed), and it associates a coefficient indicating the level of severity to a set of thresholds. These thresholds are decided by experts at EUROCONTROL. The coefficient constitutes the score obtained by a METAR report for that weather class. The tables below show how the thresholds are linked to a given coefficient. Each weather class has a single table, except for the dangerous class. For this class, coefficients are calculated for all the tables, and the final score for this class is the highest of them. Table C.3 explains the symbol found in the following tables, which are also the ones found in METAR reports.

Table C.3: This table explains the symbols present in the weather class tables.

- = slight	+ = Heavy	BC = Patches	BL = Blowing	FEW = Few
BR = Mist	DR = Low Drifting	DS = Dust Storm	DU = Widespread Dust	SCT = Scattered
DZ = Drizzle	FG = Fog	FC = Funnel Cloud	FU = Smoke	BKN = Broken
FZ = Freezing	GR = Hail	GS = Small Hail	HZ = Haze	OVC = Overcast
IC = Ice Crystals	MI = Shallow	PL = Ice Pellets	PO = Dust Devils	CB = Cumulonimbus
RA = Rain	SA = Sand	SG = Snow Grains	SH = Shower	TCU = Tower Cumulus
SN = Snow	UP = Unid. Precip.	SS = Sandstorm	TS = Thunderstorm	VA = Volcanic Ash

Table C.4: Ceiling & visibility class ATMAP algorithm.

coefficient	visibility [m] OR	(vertical visibility [ft]	AND octas [-])
0	>800	>300	OVC, BKS
2	>550	>200	OVC, BKS
4	>350	>100	OVC, BKS
5	<= 350	<= 100	OVC, BKS

Table C.5: Wind class ATMAP algorithm.

coefficient (with gust)	wind speed [kt]
0 (+1)	<= 15
1 (+1)	>15
2 (+1)	>20
4 (+1)	>30

Table C.6: Precipitations class ATMAP algorithm.

coefficient	type of precipitations
0	no precipitation
1	RA, UP, DZ, IC
2	-SN, SG, +RA
3	FZ, SN, + SN

Table C.7: Freezing class ATMAP algorithm.

coefficient	(temperature [C] AND	type of precipitations (Notes: TT means True Temperature [C], and DP Dew Point [C]))
0	>3	any
1	<= 3	DZ, IC, RA, UP, FG, GR, GS, PL, TT - DP <3
3	<= 3	-SN, SG, +RA, RASN, BR
4	<= 3	SN, +SN, SHSN, FZ
4	<= -15	any

Table C.8: Dangerous phenomena class 1: CB and TCU condition without precipitation ATMAP algorithm.

octas	coefficient CB	coefficient TCU
FEW	4	3
SCT	6	5
BKN	10	8
OVC	12	10

Table C.9: Dangerous phenomena class 2: CB and TCU condition with shower precipitation ATMAP.

(octas TCU OR	octas CB)	coefficient -SH	coefficient SH
FEW	-	4	6
SCT	FEW	8	12
BKN	SCT	10	15
OVC	BKN	12	20
-	OVC	18	24

Table C.10: Dangerous phenomena class 3

coefficient	type of precipitations
18	GS
24	FC, DS, SS, VA, SA, GR, PL, TS
30	+TS

Once the METAR report score for each class is obtained, the overall score is computed. The weather of a METAR report is considered to be bad if its overall score is equal or higher than 4. The reasons for this value rely on the explanation of each class present in EUROCONTROL's report[58]. For instance, coefficient 4 is associated with:

- **Ceiling & visibility class:** With these conditions operations become more complex, CAT II approaches should be conducted, Low Visibility Operations are activated and the landing intervals between aircraft will increase.
- **Wind class:** With wind speeds higher than 30kts, it starts having a severe impact on airport operations (higher impact on ground speed and more and more aircraft reach the crosswind airworthiness limits).
- **Freezing class:** Freezing conditions become very harsh and difficult to mitigate, even for Scandinavian airports.

Performing an analysis on METAR reports from January 2018 reveals that 207 reports have a value higher than 4, which represents 14% of the total number of reports. In terms of time, this means approximately that 4.5 days had severe weather conditions. Table C.11 shows an overview of how the number of severe METAR reports changes with different thresholds.

Table C.11: This table shows how the number of severe reports depends on the threshold for January 2018.

Threshold	# severe reports	% total
3	221	15
4	207	14
5	122	8
6	99	6.6
10	81	5.5
12	54	3.6
14	29	2.6
20	10	0.7

Note: Values in Table C.11 are computed considering METAR reports received at any time of the day (incl.nights).

C.11. Aircraft Database

An aircraft database is a collection of ICAO24 codes, registration numbers, vehicle type, type of aircraft, manufacturer, operator, and owner. This database is very useful to distinctly identify an aircraft given an identification parameter. Basically the ICAO provided by the ADS-B data is enriched with more information that can be useful in the analysis. For this thesis, the aircraft database from OpenSky is used because during some experiments it seemed to be the most complete one.[5] Table C.12 shows an overview of the available data.

Table C.12: This table shows the features offered by OpenSky aircraft database.

Icao24	**National registration**	**Manufacturer**	**Model**
Operator	**Owner**	**Status**	**Year built**
Engines	**First flight**	**Seat configuration**	**Icao aircraft type**

C.12. Results

This section presents the result of the preprocessing performed on the data. Table C.13 shows a comparison of the data available between the Delft database and the combined Delft-Open Sky database. The tables are produced after the same manual cleaning procedures are applied to these two databases. This is the average data of a day taking a sample of four days. There is a clear improvement in the data available when comparing the two tables: more than 1.5X data at all altitudes, and 4 times more for ground movements.

Examining the number of usable trajectories, approximately 10% more trajectories at all altitudes and approximately 1.5X more ground trajectory. A trajectory is defined as a series of continuous data points with a distance of less than 5 minutes between each other. The ratio column represents the average amount of points per trajectory.

[5]https://opensky-network.org/aircraft-database

Table C.13: These tables show a comparison of the average data available per day between the 9th and 12th of January 2018. The same preprocessing techniques are applied to obtain these two tables, namely only manual cleaning. The table on the left shows the data available from Delft. The one on the right shows the data available combining the two databases.

Altitude [ft]	# traj	# of points	ratio	Altitude [ft]	# traj	# of points	ratio
10000 - 8000	1138.0	72058.0	63.32	10000 - 8000	1282.67	126080.67	98.3
8000 - 6000	1133.0	98771.25	87.18	8000 - 6000	1277.67	164209	128.72
6000 - 4000	1124.0	111976.5	99.62	6000 - 4000	1260	190615.67	151.28
4000 - 2000	1114.75	176721.0	158.53	4000 - 2000	1238.33	306796	247.75
2000 - 1000	1082.5	76064.75	70.27	2000 - 1000	1206	152605	126.54
1000 - 500	1052.25	34341.25	32.64	1000 - 500	1164	51704	44.42
500 - 100	1008.0	16817.75	16.68	500 - 100	1148.67	30520.16	26.57
100 - 0	403.0	11162.0	27.7	100 - 0	641.33	44232	68.97

Looking at the ratio column, it is clear that by merging the two databases the number of points per trajectory increases considerably at all altitudes, and as for the other parameters, the highest increase happens at the ground level. This is likely a direct consequence of the fact that OpenSky adds crowdsource antennas spread across the Netherlands. Considering the trajectories present both above 9000ft and below 500ft reveals that on average operations at Schiphol Airport from 100FL to 100 ft take 580 s, approximately 10 minutes. The average latency is 0.85s. It is possible to estimate an average traveling time per altitude range knowing the latency. For Delft only data is 450 s, while for combined data, it is 700 s. It is possible to conclude that for operations between 100FL and 100ft even using only Delft data offers a reasonable amount of data and combining the two datasets offers a clearer picture. For ground operations, data seems to be scarce even when combing the two datasets. More knowledge regarding operations at Schiphol is required to confirm these findings.

Note that a trajectory might have more points than the one shown in the table. The reason is that the analysis is per altitude level and the same trajectory might appear in multiple altitudes because it has points at different levels.

Verification for tables

To test that duplicates between the two databases are removed the preprocessing process applied to produce these tables is performed merging the same database. The Delft database is used to perform this test. The result is exactly the one presented with the table on the left meaning that all duplicates are removed successfully.

Density maps

Figure C.2 shows the point density over Schiphol operational area. As expected, with growing altitude operations get further from the airport. Looking at the subfigures a and b, it shows that there were certain runways particularly used on the day of the analysis.

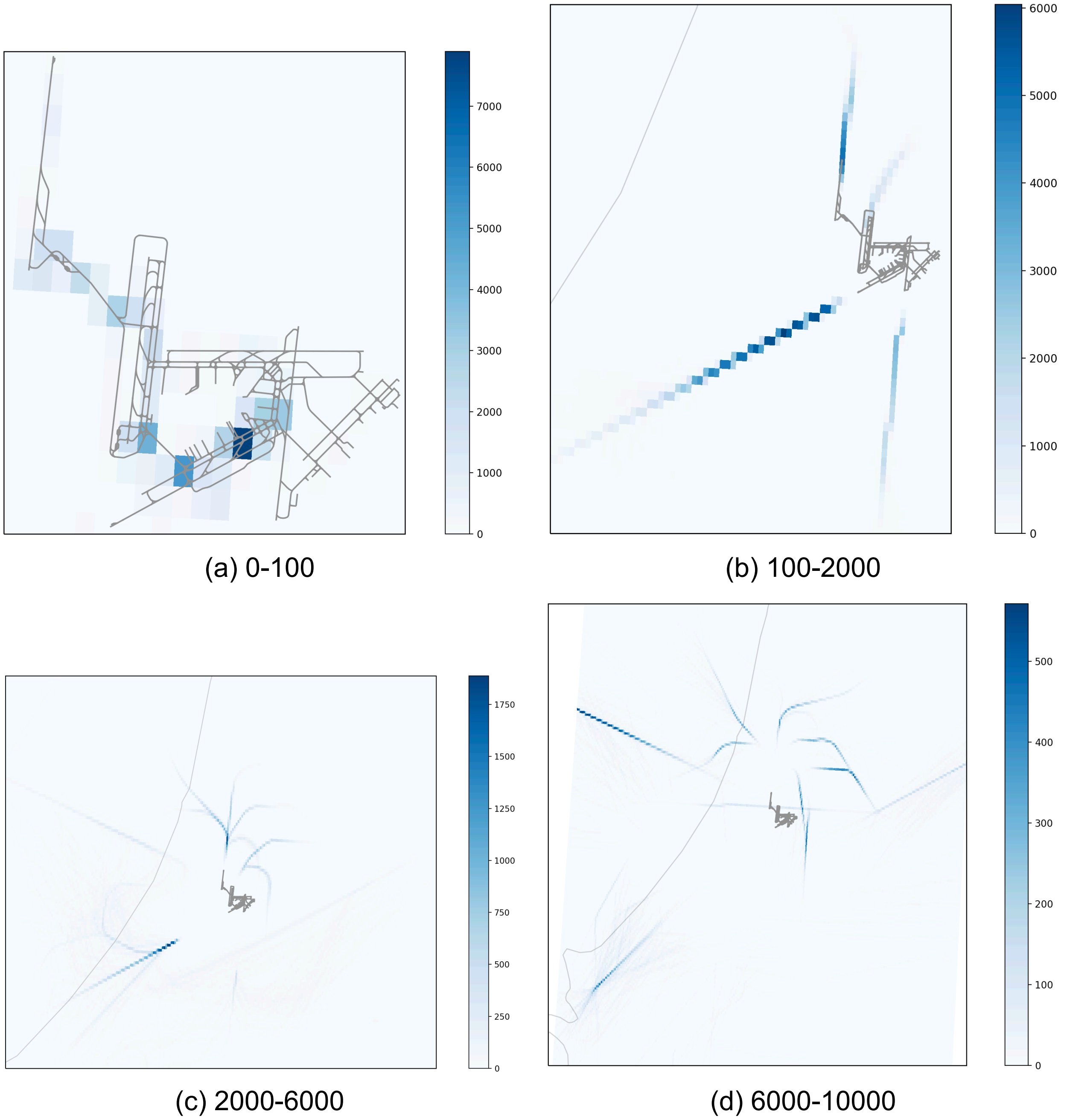


Figure C.2: The figure shows a data density map of Schiphol Airspace, Schiphol operational area. The data is collected in a day on the 8th January 2018. There are 4 maps for four different altitude levels specified below the figure in feet.

D

Data Anomaly Analysis

After preprocessing the data, the next step is understanding the limitation of this data. This chapter elaborates on this topic and offers insights discovered during the work. First, it offers an overview of the methodology, and subsequently, it describes the results obtained. The content of this chapter is a collection of experiments performed while working on the exceedance detection of unstable approaches. For this reason, it is an error analysis specific for this phase of flight. The decision of dedicating an entire chapter to this topic arises from the fact that some of the insights and results are applicable to multiple phases of flight.

D.1. Methodology

The purpose of this chapter is to identify potential sources of error in the analysis of safety events. The focus is on understanding the limitations of the main positional data used (ADS-B). While working in the identification of potential safety events during the final approach phase, visualizing many trajectories has revealed that a significant number of trajectories contain a constant error in the data. Furthermore, the analysis has shown that the data for some trajectories are very imprecise with many unpredictable jumps. The following figures offer an example of these undesirable situations.

In Figure D.2 and Figure D.1, the gray areas (one for each runway) indicate the ILS intercept with allowed tolerance. The perpendicular yellow lines indicate where the aircraft should be at 1000 ft, 500 ft, 300 ft, 200 ft, and 100 ft respectively, if it was following the instrument approach chart from LVNL.

The idea is to select the portion of trajectory relevant to the final approach and to compute the error of each point from its smoothed trajectory and the ideal ILS intercept (in the figures above, it would be the line running in the middle of the gray area). The error of each point from these lines are then used to compute statistical error metrics.

The relevant portion of the trajectory is selected based on 3 constraints: the aircraft is below 2500 ft, the aircraft is aligned with the runway with a tolerance of +/- 5 deg, and the aircraft is at no more than 6.2 nm from the runway threshold.

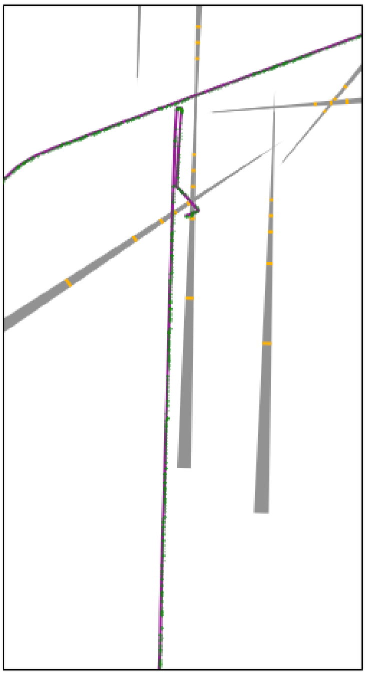


Figure D.1: The figure shows an example of trajectory with a constant shift, raw data in green, and smoothing spline in purple.

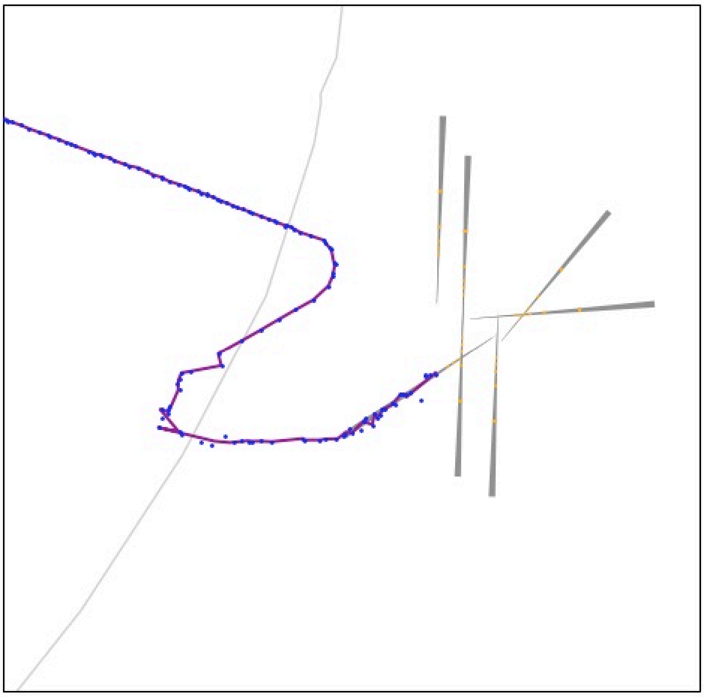


Figure D.2: The figure shows an example of a trajectory with jumpy data points (blue) fitted with a smoothing spline (purple).

The experiments are performed using different metrics to measure a representative error for each trajectory and a measure of the dispersion of the error. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Median Absolute Error (MedAE) have been initially taken into consideration for measuring the representative error. As it is often the case when working with large amounts of data, MedAE is the best fit for this purpose because it is inherently more resistant to outliers. On the other hand, RMSE and MAE show inconsistent results, especially RMSE which emphasizes the weight of outliers. Following similar reasoning, as a measure for dispersion of the error the Median Absolute Deviation is used, instead of the variance.

D.2. Results

This section presents the result of the experiments described above in Section D.2.1 and Section D.2.2.

D.2.1. Error from Smoothed Trajectory

These two box plots (Figure D.4 and Figure D.3) show that trajectories have on average a low error with a low dispersion. Figure D.5 shows an example of the outliers highlighted using this strategy. However, empirical analysis of the trajectories revealed that the smoothing spline approximation used as the truth value for the error has different limitations. As also shown from Figure D.2, the spline offers a good approximation in case there is a low amount of error while it tends to follow the outliers. This means that if outliers are present in the data, they will be present in the smoothing

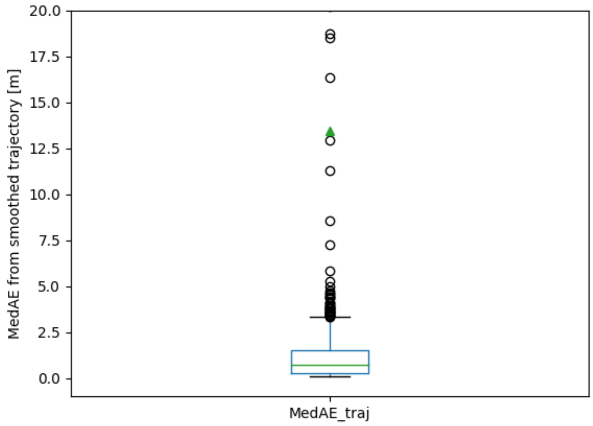


Figure D.3: The figure shows the box plot of the Median Absolute error between the data points and the spline smoothed trajectory.

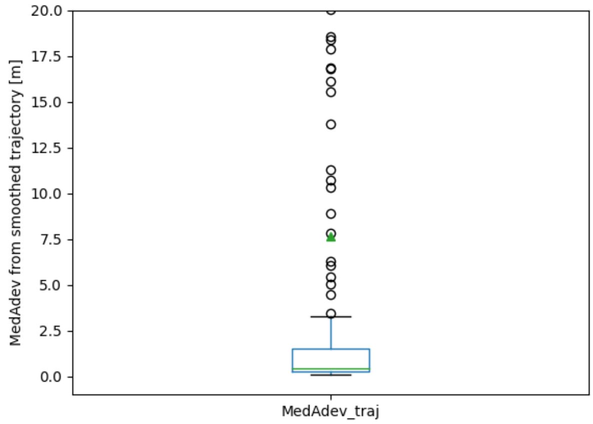


Figure D.4: The figure shows the box plot of the Median Absolute deviation of the error between the data points and the spline smoothed trajectory.

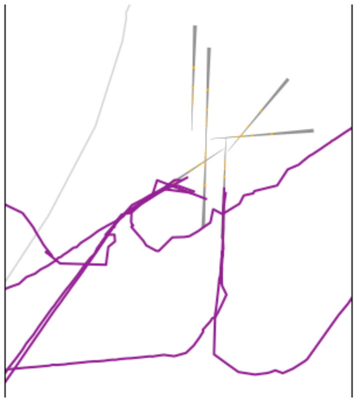


Figure D.5: The figure shows some high scoring outliers highlighted by the Median Absolute Error, and Median Absolute Deviation.

spline representation as well. Appendix C offers an overview of the working mechanism of the spline. A solution would be to increase the amount of error allowed, but the side effect is that the spline would follow less closely the trajectory and it would lose important information contained in the data.

Ideally, the goal was to define a criterion based on these parameters to remove jumpy trajectories. Further investigating the problem has revealed that the open sky database was the cause of much of this error. In Appendix C, it is shown that the OpenSky database enriches the data available form the Delft one, and it is useful especially at low altitudes. As an example, for the 8-1-2018 using the OpenSky data adds approximately 100 extra landings compared to the 500 present in the Delft database. For the same day, approximately 400,000 data points are lost when using only Delft data (approx. 500,000) for this particular phase of flight. This might cause some occurrences to go undetected because not present in the data. For instance, the go-around performed by aircraft with registration PH-EXD and ICAO 485086 is not present in the Delft data. It is clear that the OpenSky data is a crucial source of information, but all limitations must be considered. In addition, the Hampel filter discussed in Appendix C seems to miss many of these outliers. The reason is probably the fact that in case there are multiple outliers in the same time window it loses its efficacy.

D.2.2. Error from ILS Intercept

As mentioned earlier, some trajectories seem to have a constant offset. For this reason, the Median Absolute Error is computed and analyzed using box plots and histograms shown in Figure D.6 and Figure D.7.

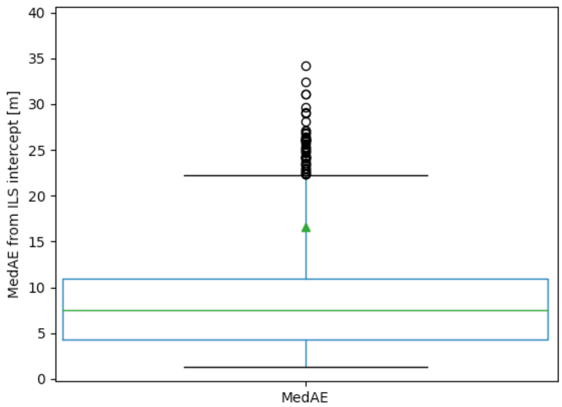


Figure D.6: The figure shows the box plot of the Median Absolute error between the data points and the ILS intercept.

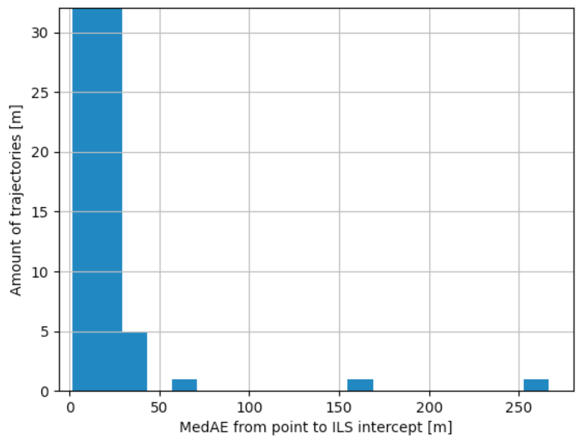


Figure D.7: The figure shows the histogram of the Median Absolute error between the data points and the ILS intercept.

Figure D.6 shows that 75% of the landings have a median distance from the ILS intercept within 11 m, this is in line with the fact that the runways have a width of 45 m (aircraft can be maximum 22.5 m to the left or right of intercept). Considering that the mean error is approximately 15 m, it means that generally once an aircraft intercepts the ILS it descends following it with very small deviations. In fact, considering the tolerance cone of each ILS as shown in the figures plotting trajectories, when the aircraft is at 6.2 nm at the beginning of the final approach phase it can be in a horizontal range of 200 m from the ILS.

Figure D.7 shows that at 40-50 m trajectories can be considered outliers. There are few trajectories after this threshold at irregular steps. An example of these trajectories is shown in Figure D.8 with a threshold of 40 m.

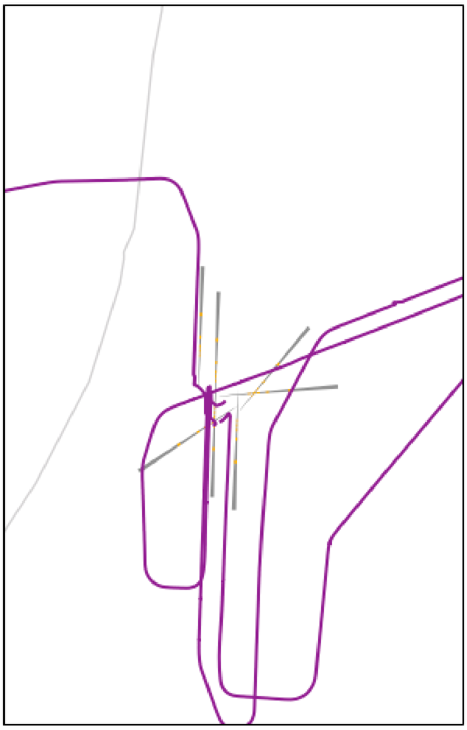


Figure D.8: The figure shows outliers highlighted by the Median Absolute Error with a threshold of 40 m.

E

Unstable Approach

This chapter provides extra information on the unstable approach methods, including some reference material. Unstable approaches are considered one of the leading causes of accidents. It is estimated that 50% of the accidents happen in this phase of flight.[1] It is estimated that 3% of landings are unstable. Although unstable landings should always be followed by a go-around, a study performed by Boeing reveals that this happens 3% of the time.[60] When comparing these numbers with the ones provided by the Staat van Schiphol, it reveals that in Schiphol approximately 0.03% of landings are unstable.[61]

E.1. Definition

There is no single definition of an unstable approach, every operator and entity can have a different one. All these definitions have in common an approach that is not aligned with the correct flight path, too fast, or too slow. For this thesis, the definition of the Flight Safety Foundation is used as a base for the evaluation of unstable approaches. According to them, a stable approach happens when the following is satisfied:

- The aircraft is on the correct flight path
- Only small changes in heading/pitch are necessary to maintain the correct flight path
- The airspeed is not more than VREF + 20kts indicated speed and not less than VREF
- The aircraft is in the correct landing configuration
- Sink rate is no greater than 1000 feet/minute; if an approach requires a sink rate greater than 1000 feet/minute a special briefing should be conducted
- Power setting is appropriate for the aircraft configuration and is not below the minimum power for the approach as defined by the operating manual
- All briefings and checklists have been conducted
- Specific types of approach are stabilized if they also fulfill the following:

[1] data from: https://flightsafety .org/asw-article/commercial-accident-final-approach/

  - ILS approaches must be flown within one dot of the glide-slope and localizer
  - a Category II or III approach must be flown within the expanded localizer band
  - during a circling approach, wings should be level on final when the aircraft reaches 300 feet above airport elevation;

- Unique approach conditions or abnormal conditions requiring a deviation from the above elements of a stabilized approach require a special briefing.

An approach that becomes unstable below 1000 feet above airport elevation in IMC or 500 feet above airport elevation in VMC requires an immediate go-around.

ADS-B data doesn't contain all the information required to verify that the requirements above are full-filled.

E.2. Extra Details on Energy Compliance

E.2.1. Preprocessing

Data contains various errors, whose source is difficult to identify. Outlier removal means removing data points with unreasonable value. Limits are defined for the speed, height, and rate of climb as shown in Table E.1.

Table E.1: This table shows the limits for possible approach operations.

variable	min	max
Airspeed [m/s]	0	140
Height [ft]	0	6000
Rate of climb [fpm]	-3000	2000

Considering the table above, airspeed has a maximum value of 140 m/s. This value is based on the fact that large category D jets have a maximum speed of 95 m/s for the final approach, and 135 m/s in the case of a missed approach.[2] Analyzing the instrument approach charts lead to the conclusion that aircraft are normally at a maximum height of 4000 ft for runway 36C at 10 NM. For this reason, the height is assigned an upper limit of 6000ft taking into account some contingency. Finally, for the rate of climb, -3000 fpm is the minimum allowed value considering the general rules of the Flight Safety Foundation, which can be found in the section above. It mandates that a go-around should be performed in case the rate of climb is below -1000 fpm. The limit is the theoretical decision value multiplied by 3 to take into account some contingency. The upper limit of 2000 fpm takes into account the case in which an aircraft performs a go-around.

During the outlier removal process, it is found that 7% of Open Sky data exceeds the limits defined in Table E.1, while the percentage is close to 1% for the Delft data. For this reason, this analysis uses only Delft data and this preprocessing step doesn't need to be performed.

[2]data from: www.skybrary.aero/index.php/Approach_Speed_Categorisation

E.2.2. GMM Components

The number of components is chosen using the Calinski–Harabascz (C-H) index, as suggested by Puranik in [36]. This is an internal evaluation criterion that measures how compact are components and how well separated they are. Several GMMs are trained with a different number of components using the training data. Figure E.1 shows the result of the C-H index for the different number of components for the 3-GMM model. A higher value of this index is better, and thus we select 2 components for the first GMM, 3 for the second one, and 2 for the third one.

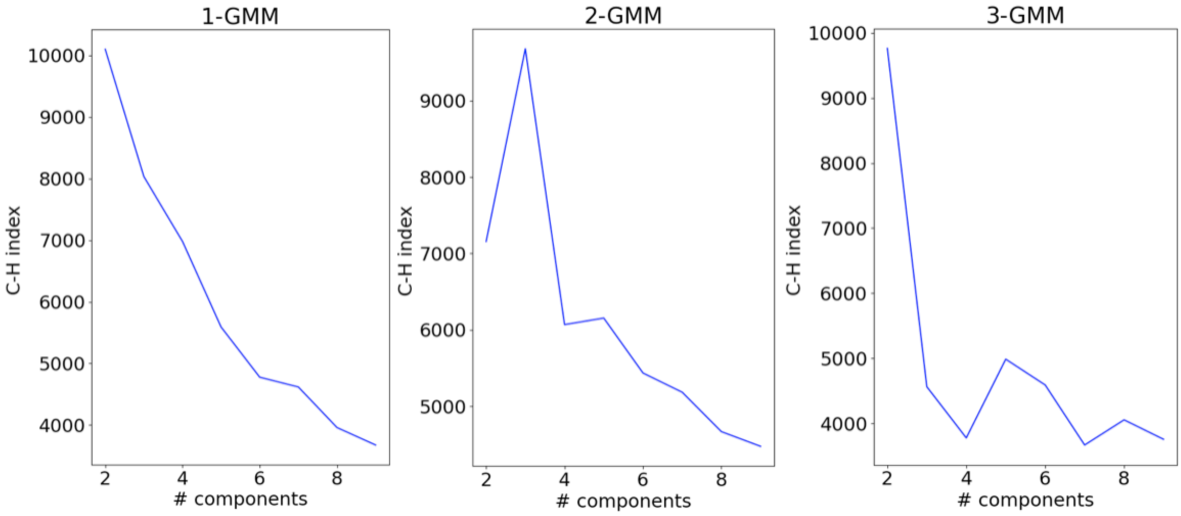


Figure E.1: The figure shows the results of the C-H inex for the 3 components of the GMM.

E.2.3. Results

In this example, the 3-GMM model is trained on 5 days of data from the 5th to 10th of January 2018. In this way, there are approximately 3000 trajectories available for training. This number is chosen based on the previous study by Puranik [36] that also uses 3000 flights. After training, the number of anomalous trajectories resulting from this methodology with a threshold of 0.1% is shown in Table E.2.

Table E.2: This table shows the amount of anomalous trajectory resulting from the 3-GMM model (threshold of 0.1%) trained with 3000 trajectory collected from 5th to 10th of January.

<table>
<tr><th>Distance to runway [NM]</th><th># unique [% of</th><th>anomalous total]</th><th>trajectories</th></tr>
<tr><td>0.5 - 4</td><td>34 [1.2]</td><td rowspan="2">52 [1.8]</td><td rowspan="3">91 [3.1]</td></tr>
<tr><td>4 - 7</td><td>24 [0.8]</td></tr>
<tr><td>7 - 10</td><td>49 [1.7]</td><td>\\</td></tr>
</table>

The most interesting part is the one closest to the runway threshold because it is where the aircraft needs to comply with the stabilized approach criteria. In this area, 1.2 % of the trajectories result to be anomalous.

The subsequent area from 4 to 7 NM can be of interest when looking for unstable approach precursors, or as an extension of the previous one. An aircraft unstable between 4 and 7 NM may manage to return to the stable condition before reaching the decision gate. This assumption is confirmed by looking at the second column. There are 52 anomalous unique trajectories flying between 0.5 and 7 NM, and only 34 between 0.5 and 4 NM. It means that 18 trajectories stabilize before reaching the decision gate. Instead, 6 of them remain unstable after entering the last portion of the final approach path.

The last part, between 7 and 10 NM, has the highest number of anomalous trajectories. This is expected because this area includes a wider spectrum of operations.Here, an aircraft could be flying at 2000, 3000, 4000 ft depending on ATC commands. The last column of the table indicates the total number of unique anomalous trajectory from 0.5 to 10 NM.

Table E.3: This table shows the validation of the energy compliance algorithm with the validation list depending on the threshold and the dataset.

		0.01	0.05	0.1	0.5	1	2	3	5	10
Detected	Open+Delft	4	6	8	11	16	17	19	24	28
	Delft	3	6	8	9	11	15	17	21	29
Detection accuracy [%]	Open+Delft	12.90	19.36	25.81	35.48	51.61	54.84	61.29	77.42	90.32
	Delft	9.68	19.36	25.81	29.03	35.48	48.39	54.84	67.74	93.55
Positives [%]	Open+Delft	2.74	3.74	4.42	7.58	10.55	15.40	19.96	27.35	40.66
	Delft	0.50	1.08	1.60	4.78	7.55	12.70	17.55	26.23	42.80

In Chapter 1, a simplified version of Table E.3 is presented. In this section, there is an extra column containing the results from the OpenSky data.Table E.3 shows the result of validating the energy compliance algorithm with the validation list depending on the threshold and the dataset used. It shows how the overall number of detected trajectories varies depending on the threshold and dataset. Furthermore, there is the detection accuracy, and how the overall number of positive detection varies. It is important to note that to perform a fair comparison and highlight the differences only aircraft present in both datasets are considered. In the Delft data, there are 31 trajectories and in OpenSky, there are 44 trajectories.

It is interesting to see that the OpenSky dataset performs overall better than using only the Delft data. This is probably the consequence of having more data points in the first compared to the second database. This directly leads to an increase in detection accuracy. However, by analyzing the third row showing the overall percentage of autonomous trajectories returned by the algorithm, it is clear that using Opensky data more often return an unstable approach. This is caused by the lower quality of data present in this dataset. A higher percentage of positives results likely underlies a high percentage of false positives. It is possible to see that as the threshold is set higher and higher the detection of unstable approaches equalizes between the two datasets reaching for a threshold equal to 10% a 40% portion of unstable trajectories.

F Go-around

A go-around is a maneuver that happens when landing is aborted. It can be initiated by the pilot or ATC for different reasons, such as unstable landing and obstacles on the runway. It is a procedure that consists in climbing to a predetermined altitude prescribed in the instrument landing procedures, and once at the correct altitude turning 360 degrees around the runway.

F.1. Extra Examples

This section shows some examples of trajectories identified by the model.

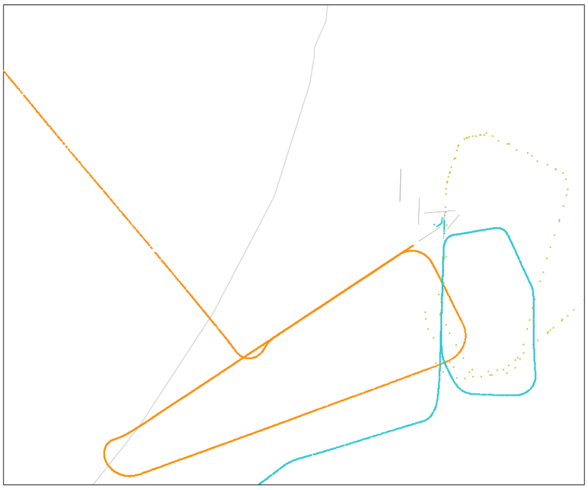


Figure F.1: Go-arounds detected on the 8th January 2018.

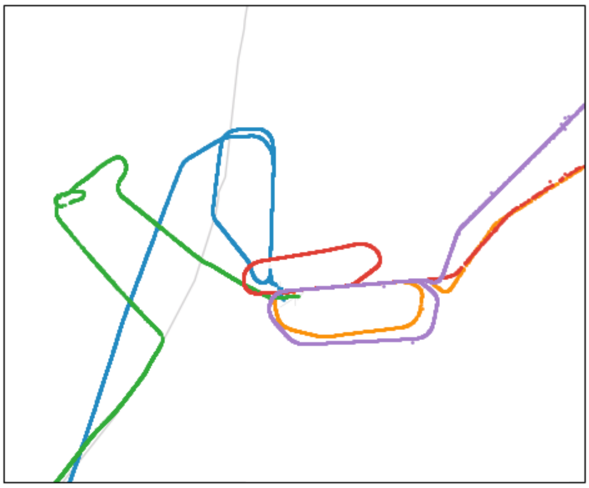


Figure F.2: Go-arounds detected on the 29th January 2018.

Preliminary tests of the model show that go-arounds are detected for trajectories with a high and low density of data, as shown in Figure F.1. Figure F.2 shows that aircrafts performing holding loops at low altitudes might be flagged as go-around as well. This behavior needs to be examined further.

F.2. Conflicting Traffic

A go-around might be an indicator for conflicting traffic. An aircraft might perform a go-around because separation can not be guaranteed. For this reason, ATC might command a go-around. In Schiphol TMA, the separation minima are:

- 1000ft vertically or
- 3nm horizontally

There is an exception in case parallel runways are in use. For this airport independent parallel approaches are possible since the distance between the parallel runways is larger than 3400ft. This type of parallel approach will be discussed further. Aircraft are separated vertically by flying at different altitudes until they intercept the ILS. Table F.1 shows the standard procedures altitudes when flying at different parallel runways and Figure F.3 offers an horizontal overview of the operation.

Table F.1: This table shows an overview of the different altitudes when flying parallel approaches.

Parallel runways	Altitudes [ft]	Center-runway distance [ft]	Buffer area [ft] ([NM])
36C & 36R	36C: 4000 and 36R: 3000	8200	3000 (0.5)
18C & 18R	18C: 3000 and 18R: 2000	6200	2000 (0.33)

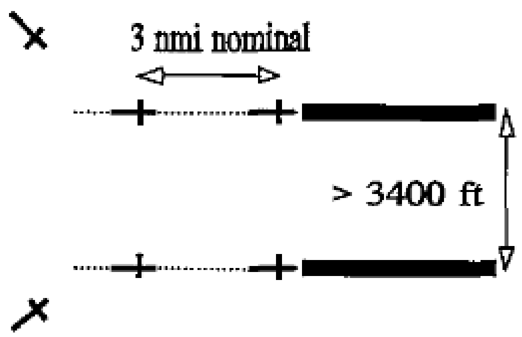


Figure F.3: This figure offers an overview of the horizontal requirements for independent parallel runway approach.[62]

There is an approach controller for each runway. He monitors that its aircraft don't enter the No Transgression Zone (NTZ). NTZ is a corridor established between the parallel runways to ensure separation. If an aircraft enters this zone, immediate correcting action from the controller is required.[63] The no-transgression zone (NTZ) is at least 2000 feet (610 meters) in width and established equidistant between the extended runway centerlines.

It is important to note that in the CTR, when airplanes are in line of sight, the controller can deviate from the described minima.[1]

[1] data from: https://en.lvnl.nl/safety/achieving-safety/separation-of-aircraft

F.3. Comparison with OpenSky

In Chapter 1, a simplified version of Table F.2 is presented. In this section, there is an extra column containing the results from the OpenSky data. This table shows the validation of the go-around detection method using the validation list.

Table F.2: This table shows a comparison of validating the go-around detection method with different datasets.

	Open+Delft	Delft
Detected	50	46
Undetected	6	1
Not present	9	18

As expected, the combination of the two datasets improves the number of detected go-arounds, and the data availability from 72% to 86%. Interestingly, also the amount of undetected go-arounds is much higher, for open+Delft is 11% of the available go-arounds, while for Delft it is only 2%. This behavior is probably caused by the lower data quality of the OpenSky and the fact that these trajectories have fewer points compared to the average detected trajectory. The difference is of an order of magnitude, with these trajectories having approximately 100 points compared with the average detected trajectory possessing 1500 points.

G
Monitoring Indicators

This chapter is dedicated to the dashboard built as a basis for a monitoring tool.

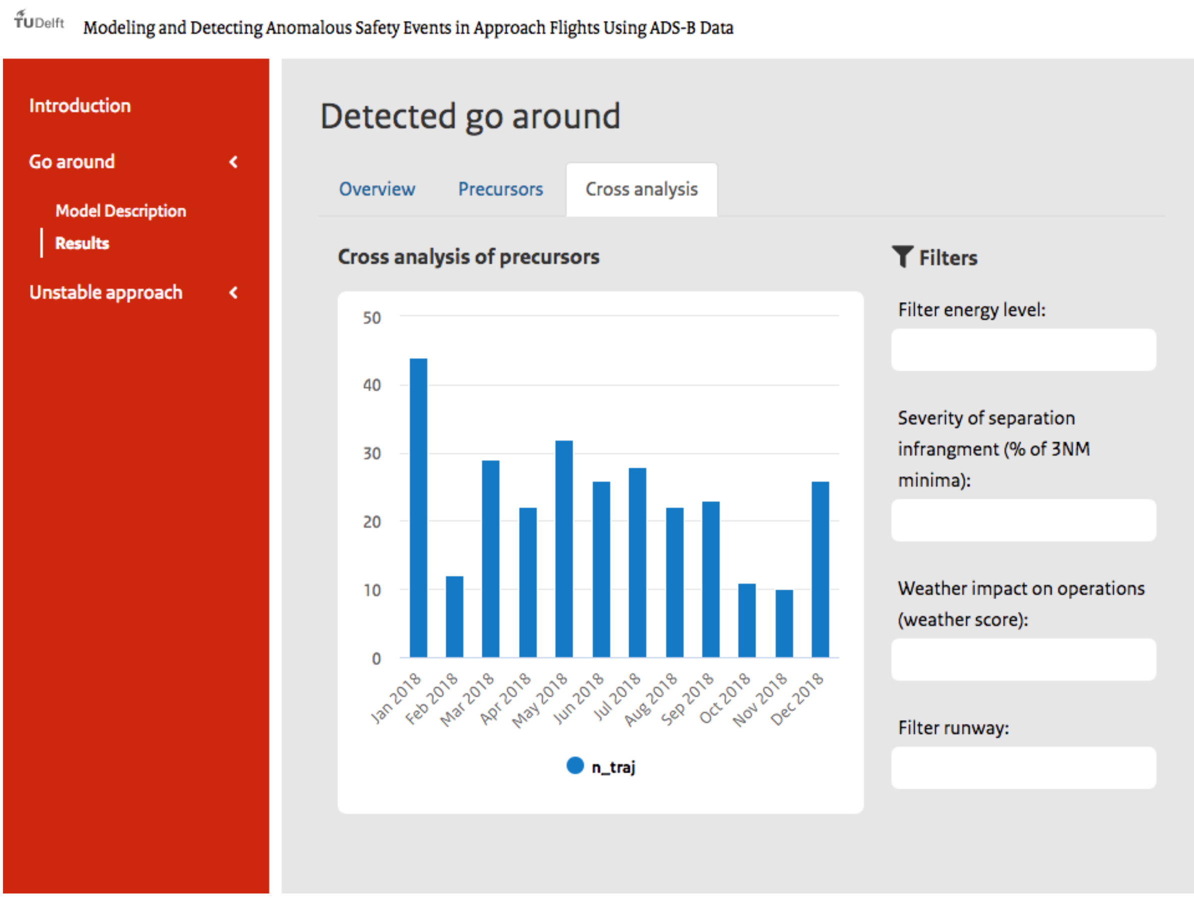


Figure G.1: This figure shows the monitoring dashboard.

G.1. Overview

The purpose of this dashboard is to develop an interactive interface to analyze the results. It represents the basis of a safety monitoring system. This dashboard is developed in R using the Shiny package. Figure G.1 shows one of the interactive plot areas. In this plot, it is possible to visualize the number of go-arounds per month and filter it depending on several conditions simultaneously. For instance, it is possible to select go-arounds happening with adverse weather on a particular runway. This allows drawing unique insights into aircraft operations.

The dashboard is organized as follows:

- Introduction contains info about the project.
- Go-around, it has two tabs:
  - Model Description with info on how the detection of events works.
  - Results contains three interactive plot areas to analyze the relationship between go-arounds, months of the year, runway, weather, unstable approach, and separation to closest aircraft.
- Unstable approaches, it has two tabs:
  - Model Description with info on how the detection of events works.
  - Results contains one interactive plot to analyze the relationship between unstable approaches, months of the year, runway, and weather.

Bibliography

[1] A. Hale, “Regulating airport safety: The case of Schiphol,” *Safety Science*, vol. 37, no. 2-3, pp. 127–149, mar 2001.

[2] Dutch Safety Board, “Schiphol air traffic safety,” Dutch Safety Board, Tech. Rep., 2017.

[3] I. C. Statler, “The Aviation System Monitoring and Modeling (ASMM) Project: A Documentation of its History and Accomplishments: 1999-2005,” NASA Ames Research Center, Tech. Rep. June, 2007. [Online]. Available: http://ntrs.nasa.gov/search.jsp?R=20080015665

[4] L. Basora, X. Olive, and T. Dubot, “Recent Advances in Anomaly Detection Methods Applied to Aviation,” *Aerospace*, vol. 6, no. 11, p. 117, oct 2019. [Online]. Available: https://www.mdpi.com/2226-4310/6/11/117http://dx.doi.org/10.3390/aerospace6110117

[5] J. Sun, “Open Aircraft Performance Modeling: Based on an Analysis of Aircraft Surveillance Data,” Ph.D. dissertation, TU Delft, 2019. [Online]. Available: https://mode-s.org/

[6] C. Pradera and H. Teper, “ADS-B and other means of surveillance implementation status,” SESAR, Tech. Rep., may 2018.

[7] ICAO Asia and Pacific Office, “ADS-B implementation and operations guidance document,” ICAO, Edition 11.0, jul 2018.

[8] T. L. Verbraak, J. Ellerbroek, J. Sun, and J. M. Hoekstra, “Large-Scale ADS-B Data and Signal Quality Analysis,” *Twelfth USA/Europe Air Traffic Management Research and Development Seminar*, 2017.

[9] B. Syd Ali, A. Majumdar, W. Y. Ochieng, W. Schuster, B. S. Ali, A. Majumdar, W. Y. Ochieng, and W. Schuster, “ADS-B: The Case for London Terminal Manoeuvring Area (LTMA),” *Tenth USA/Europe Air Traffic Management Research and Development Seminar*, 2013.

[10] B. Syd Ali, W. Schuster, W. Ochieng, and A. Majumdar, “Analysis of anomalies in ADS-B and its GPS data,” *GPS Solutions*, vol. 20, no. 3, pp. 429–438, mar 2015. [Online]. Available: http://dx.doi.org/10.1007/s10291-015-0453-5

[11] M. Schäfer, M. Strohmeier, M. Smith, M. Fuchs, V. Lenders, and I. Martinovic, “OpenSky Report 2018: Assessing the Integrity of Crowdsourced Mode S and ADS-B Data,” *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, vol. 2018-Septe, dec 2018.

[12] J. Sun, J. Ellerbroek, and J. Hoekstra, “Large-Scale Flight Phase Identification from ADS-B Data Using Machine Learning Methods,” Tech. Rep., 2016.

[13] Z. Wang, L. Sherry, and J. Shortle, “Airspace risk management using surveillance track data: Stabilized approaches,” in *ICNS 2015 - Innovation in Operations, Implementation Benefits and Integration of the CNS Infrastructure, Conference Proceedings*. Institute of Electrical and Electronics Engineers Inc., jun 2015, pp. W31–W314.

[14] J. Oehling and D. J. Barry, “Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data,” *Safety Science*, vol. 114, pp. 89–104, apr 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925753518308269

[15] H. P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, “LoOP: Local Outlier Probabilities,” in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, ser. CIKM ’09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 1649–1652. [Online]. Available: https://doi.org/10.1145/1645953.1646195

[16] L. Li, M. Gariel, R. J. Hansman, and R. H. Palacios, “Anomaly detection in onboard-recorded flight data using cluster analysis,” *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*, pp. 4A4–1–4A4–11, 2011.

[17] L. Li, S. Das, R. J. Hansman, R. Palacios, A. N. Srivastava, R. John Hansman, R. Palacios, and A. N. Srivastava, “Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations,” *Journal of Aerospace Information Systems*, vol. 12, no. 9, pp. 587–598, 2015. [Online]. Available: https://doi.org/10.2514/1.I010329

[18] M. Ester, H. P. Kriegel, J. Sander, X. Xu, and Others, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Kdd*, vol. 96, no. 34, Portland, OR, USA, 1996, pp. 226–231.

[19] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, “DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN,” *ACM Transactions on Database Systems*, vol. 42, no. 3, pp. 1–21, jul 2017.

[20] A. M. Churchill and M. Bloem, “Clustering Aircraft Trajectories on the Airport Surface,” in *Proceedings of the 13th USA/Europe Air Traffic Management Research and Development Seminar*. EUROCONTROL, 2019.

[21] D. L. Iverson, R. Martin, M. Schwabacher, L. Spirkovska, W. Taylor, R. MacKey, J. P. Castle, and V. Baskaran, “General Purpose Data-Driven System Monitoring for Space Operations,” *Journal of Aerospace Computing, Information, and Communication*, vol. 9, no. 2, pp. 26–44, sep 2012.

[22] I. T. Jolliffe, *Principal components in regression analysis*. Springer, 1986.

[23] L. van der Maaten, G. Hinton, L. Van Der Maaten, and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, nov 2008.

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

[25] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, J. C. Platt, and J. Piatt, "Support vector method for novelty detection," in *Advances in neural information processing systems*. Neural information processing systems foundation, 2000, pp. 582–588.

[26] S. Das, B. L. Matthews, A. N. Srivastava, and N. C. Oza, "Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 47–56.

[27] S. Das, B. L. Matthews, and R. Lawrence, "Fleet level anomaly detection of aviation safety data," in *2011 IEEE International Conference on Prognostics and Health Management, PHM 2011 - Conference Proceedings*, 2011, pp. 1–10.

[28] B. Matthews, N. Oza, S. Das, K. Bhaduri, K. Das, and R. Martin, "Discovering anomalous aviation safety events using scalable data mining algorithms," *Journal of Aerospace Information Systems*, vol. 10, no. 10, pp. 467–475, oct 2013.

[29] B. Matthews, D. Nielsen, J. Schade, K. Chan, and M. Kiniry, "Comparative study of metroplex airspace and procedures using machine learning to discover flight track anomalies," in *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*, IEEE. Institute of Electrical and Electronics Engineers Inc., oct 2015, pp. 2G41–2G415.

[30] M. Sharma, K. Das, M. Bilgic, B. Matthews, D. Nielsen, and N. Oza, "Active Learning with Rationales for Identifying Operationally Significant Anomalies in Aviation," in *Machine Learning and Knowledge Discovery in Databases*, B. Berendt, B. Bringmann, É. Fromont, G. Garriga, P. Miettinen, N. Tatti, and V. Tresp, Eds. Cham: Springer International Publishing, 2016, pp. 209–225.

[31] T. G. Puranik and D. N. Mavris, "Anomaly detection in general-aviation operations using energy metrics and flight-data records," *Journal of Aerospace Information Systems*, vol. 15, no. 1, pp. 22–36, 2018.

[32] T. Puranik, H. Jimenez, and D. Mavris, "Energy-based metrics for safety analysis of general aviation operations," *Journal of Aircraft*, vol. 54, no. 6, pp. 2285–2297, 2017.

[33] Z. Kong, A. Jones, A. M. Ayala, E. A. Gol, C. Belta, A. Medina Ayala, E. Aydin Gol, and C. Belta, "Temporal logic inference for classification and

prediction from data," in *Proceedings of the 17th international conference on Hybrid systems: computation and control*. New York, New York, USA: Association for Computing Machinery, 2014, pp. 273–282. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2562059.2562146

[34] R. Deshmukh and I. Hwang, "Anomaly Detection Using Temporal Logic Based Learning for Terminal Airspace Operations," in *AIAA Scitech 2019 Forum*. American Institute of Aeronautics and Astronautics Inc, AIAA, 2019, p. 682.

[35] ——, "Incremental-Learning-Based Unsupervised Anomaly Detection Algorithm for Terminal Airspace Operations," *Journal of Aerospace Information Systems*, vol. 16, no. 9, pp. 362–384, 2019.

[36] T. G. Puranik and D. N. Mavris, "Identification of Instantaneous Anomalies in General Aviation Operations Using Energy Metrics," *Journal of Aerospace Information Systems*, vol. 17, no. 1, pp. 51–65, jan 2020. [Online]. Available: http://dx.doi.org/10.2514/1.I010772

[37] L. Li, R. J. Hansman, R. Palacios, and R. Welsch, "Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 45–57, mar 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X16000188

[38] X. Zhang, J. Chen, and Q. Gan, "Anomaly Detection for Aviation Safety Based on an Improved KPCA Algorithm," *Journal of Electrical and Computer Engineering*, vol. 2017, mar 2017.

[39] G. Jarry, D. Delahaye, F. Nicol, and E. Feron, "Aircraft atypical approach detection using functional principal component analysis," *Journal of Air Transport Management*, vol. 84, p. 101787, may 2020.

[40] G. Jarry, D. Delahaye, E. Feron, and E. Féron, "Trajectory APproach AnalysiS: A Post-operational Aircraft Approach Analysis Tool," in *SID 2019, 9th SESAR Innovation Days*. Athenes, Greece: SESAR Joint Undertaking, dec 2019. [Online]. Available: https://hal-enac.archives-ouvertes.fr/hal-02385671

[41] X. Jiang, X. Wen, M. Wu, M. Song, and C. Tu, "A complex network analysis approach for identifying air traffic congestion based on independent component analysis," *Physica A Statistical Mechanics and its Applications*, vol. 523, pp. 364–381, jun 2019.

[42] I. Melnyk, A. Banerjee, B. Matthews, N. Oza, H. Valizadegan, A. Banerjee, N. Oza, B. Matthews, N. Oza, H. Valizadegan, A. Banerjee, and N. Oza, "Semi-Markov switching vector autoregressive model-based anomaly detection in aviation systems," *Journal of Aerospace Information Systems*, vol. 13, no. 4, pp. 161–173, apr 2016. [Online]. Available: http://dx.doi.org/10.2514/1.I010394

[43] I. Melnyk, B. Matthews, H. Valizadegan, A. Banerjee, and N. Oza, "Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems," *Journal*

of Aerospace Information Systems, vol. 13, no. 4, pp. 161–173, apr 2016. [Online]. Available: http://dx.doi.org/10.2514/1.I010394

[44] M. Venkatachalam. Recurrent Neural Networks. [Online]. Available: https://towardsdatascience.com/recurrent-neural-networks-d4642c9bc7ce

[45] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using Recurrent Neural Networks (RNN)," in *ICNS 2016: Securing an Integrated CNS System to Meet Future Challenges*. Institute of Electrical and Electronics Engineers Inc., jun 2016, pp. 5C2–1.

[46] X. Olive, J. Grignard, T. Dubot, and J. Saint-Lot, "Detecting Controllers' Actions in Past Mode S Data by Autoencoder-Based Anomaly Detection," *Eighth SESAR Innovation Days*, 2018. [Online]. Available: www.liveatc.net

[47] X. Olive and L. Basora, "Identifying Anomalies in past en-route Trajectories with Clustering and Anomaly Detection Methods," in *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar*. EUROCONTROL, 2019.

[48] V. M. Janakiraman and D. Nielsen, "Anomaly detection in aviation data using extreme learning machines," in *2016 International Joint Conference on Neural Networks (IJCNN)*, vol. 2016-Octob. Institute of Electrical and Electronics Engineers Inc., jul 2016, pp. 1993–2000.

[49] V. M. Janakiraman, B. Matthews, and N. Oza, "Discovery of precursors to adverse events using time series data," in *16th SIAM International Conference on Data Mining 2016, SDM 2016*. Society for Industrial and Applied Mathematics Publications, 2016, pp. 639–647.

[50] V. M. Janakiraman, "Explaining aviation safety incidents using deep temporal multiple instance learning," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery, jul 2018, pp. 406–415.

[51] R. Deshmukh, D. Sun, and I. Hwang, "Data-Driven Precursor Detection Algorithm for Terminal Airspace Operations," in *13th USA/Europe Air Traffic Management Research and Development Seminar*. EUROCONTROL, 2019.

[52] X. Olive and P. Bieber, "Quantitative assessments of runway excursion precursors using Mode S data," *arXiv preprint arXiv:1903.11964*, 2019.

[53] F. Herrema, V. Treve, B. Desart, R. Curran, and D. Visser, "A novel machine learning model to predict abnormal Runway Occupancy Times and observe related precursors," in *12th USA/Europe Air Traffic Management Research and Development Seminar*. EUROCONTROL, 2017.

[54] H. Goel, I. Melnyk, N. Oza, B. Matthews, and A. Banerjee, "Multivariate Aviation Time Series Modeling: VARs vs. LSTMs."

[55] A. B. B. Júnior and P. S. M. da Pires, “An approach to outlier detection and smoothing applied to a trajectography radar data,” *Journal of Aerospace Technology and Management*, vol. 6, no. 3, pp. 237–248, 2014.

[56] V. T. Dung and T. Tjahjowidodo, “A direct method to solve optimal knots of Bspline curves: An application for non-uniform B-spline curves fitting,” *PLoS ONE*, vol. 12, no. 3, mar 2017.

[57] ICAO, “ICAO Annex 2: Rules of the Air, Chapter 4,” ICAO, Edition 42.0, nov 2009.

[58] Performance Review Unit and ATMAP MET working group, “Algorithm to describe weather conditions at European airports,” Tech. Rep., 2011.

[59] M. Schultz, S. Lorenz, R. Schmitz, and L. Delgado, “Weather Impact on Airport Performance,” 2018.

[60] The Boeing Edge, “Boeing Aeromagazine Issue 54 Quarter 02 2014.” [Online]. Available: http://www.boeing.com/commercial/aeromagazine/articles/2014_q2/pdf/AERO_2014q2.pdf

[61] Inspectie Leefomgeving en Transport (Ministerie van Infrastructuur en Waterstaat), “Staat van Schiphol 2019.” [Online]. Available: https://www.ilent.nl/documenten/publicaties/2020/02/07/staat-van-schiphol-2019

[62] G. Wong, “Development of precision runway monitor system for increasing capacity of parallel runway operations,” FAA, Tech. Rep., 1993. [Online]. Available: http://www.tc.faa.gov/acb300/techreports/RD-91-5.pdf

[63] L. Speijker, M. Couwenberg, and H. Kleingeld, “Collision risk related to the usage of parallel runways for landing ,” NLR, Tech. Rep., 1997. [Online]. Available: https://www.skybrary.aero/bookshelf/books/1763.pdf