EXPLORATORY RESEARCH

D2.1 Development of Use Cases, User Stories and Requirements

Deliverable ID:	**D2.1**
Dissemination Level:	**PU**
Project Acronym:	**SafeOPS**
Grant:	**892919**
Call:	**H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-06-2019**
Consortium Coordinator:	**TUM**
Edition date:	**29 July 2021**
Edition:	**00.02.00**
Template Edition:	**02.00.02**

Founding Members

EUROPEAN UNION EUROCONTROL

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Lukas Beller / TUM	Project Coordinator	23.06.2021
Marco Pfahler / TUM	Project Team Member	23.06.2021
Josef Hartl / TUM	Project Team Member / Cpt (A320)	23.06.2021
Philipp Kurth / DFS	Project Team Member	23.06.2021
Simon Rangger / DFS	Project Team Member	23.06.2021
Pablo Hernandez / INX	WP4 Leader	23.06.2021
Carlo Abate / DBL	WP3 Leader / PhD	23.06.2021

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Carlo Abate / DBL	WP3 Leader / PhD	25.06.2021
Phillip Koppitz / TUM	Project Team Member	28.06.2021
Josef Hart / TUM	Project Team Member / Cpt (A320)	27.06.2021
Philipp Kurth / DFS	Project Team Member	28.06.2021

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Lukas Beller / TUM	Project Coordinator	28.07.2021
Patrick Gontar / DFS	Management Board Member / PhD	29.06.2021
Simon Rangger / DFS	Project Team Member	29.07.2021
Carlo Abate / DBL	WP3 Leader / PhD	03.08.2021
Pablo Hernandez / INX	WP4 Leaser	29.07.2021

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
-	-	-

Document History

Edition	Date	Status	Author	Justification
00.00.01	23.06.2021	Under Review	SafeOPS Consortium	1st version: content transferred from Project Internal Platform

00.01.00	28.06.2021	Initial Release	SafeOPS Consortium	
00.01.01	27.07.2021	1st Revision	Lukas Beller	Update with minor changes, as requested by SJU
00.02.00	29.07.2021	Second Issue Version	SafeOPS Consortium	

History of Changes

Edition	Date	Author	Change
00.01.01	27.07.2021	Lukas Beller	Requested abbreviations have been written-out in the text. All abbreviations are still defined in the text and listed in the index of abbreviations on page 11, as they are still used in illustrations.
00.01.01	27.07.2021	Lukas Beller	The wording remark in the tables of requirements has been replaced by rationale.
00.01.01	27.07.2021	Lukas Beller	Logical ANDs have been added to requirements that use bullet points in their specifications.

SafeOPS

SAFEOPS

This Deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892919 under European Union’s Horizon 2020 research and innovation programme.

Abstract

How big data and artificial intelligence based decision support systems could impact daily air traffic operations has not been explored yet. Over the course of the past five months, the SafeOPS team held several workshops together with air traffic controllers from two major European hubs to elaborate this question in the context of Go-Around handling. Based on these workshops, several scenarios have been identified in which Go-Arounds can lead to complex situations in daily operations. Based on these scenarios potential use cases for a decision support tool able to predict Go-Arounds, have been elaborated. A requirements engineering approach is used to refine the use cases into user stories and finally requirements to shape an initial design proposal. The result of this document are the requirements, that define the further development process over the course of work packages 3 and 4. Additionally to the documented user stories and requirements, this deliverable provide a technical problem statement. This problem statement gives an overview on the state of the art technical methods as well as the challenges which can be derived from the requirements at this stage of the project.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A

A/C	*Aircraft*
ALA	*Approach and Landing Accidents*
ANSP	*Air Navigation Service Provider*
AP	*Airport*
ATC	*Air Traffic Control*
ATCO	*Air Traffic Controller*
ATM	*Air Traffic Management*

C

CW	*Calendar Week*

D

DBL	*Deep Blue SrL*
DFS	*Deutsche Flugsicherung*

F

FC	*Flight Crew*

G

GA	*Go-Around*

I

IMC	*Instrument Meteorological Condition*
INX	*Innaxis*

L

LDG	*Landing*
LOC-I	*Loss of Control in Flight*

M

MA	*Missed Approach*
MAP	*Missed Approach Procedure*

N

NOTAM	*Notice(s) to Airmen*

P

PL	*Tower Controller (german: Platzlotse)*
PLC	*Center Runway Tower Controller*
PLCS	*Center and Southern Runway Controller in personal union*
PLN	*Northern Runway Tower Controller*
PLS	*Southern Runway Tower Controller*
PLW	*Western Runway Tower Controller*

R

RWY	*runway*

S

SID	*Standard Instrument Departure*

T

T/O	*Take-Off*
TFDPS	*Tower Flight Data Processing System*
THR	*Runway Threshold*
TUM	*Technical University of Munich*

U

UA	*Unstabilized Approach*

V

VMC	*Visual Meteorological Condition*

W

WTC	*Wake Turbulence Category*

1 Introduction

SafeOPS investigates how ATC operations can benefit from big data based predictions. Therefore, SafeOPS exemplarily investigates Go-Around (GA) scenarios to elaborate upon the benefits and risks big data based predictions bear in the decision making processes of Air Traffic Controllers (ATCOs) at tower units.

After the management and the dissemination plan, this is the third deliverable of the SafeOPS project team. The overarching goal of this deliverable is to provide an initial set of user stories and requirements for the upcoming development phase of SafeOPS. Therefore, SafeOPS uses a user centric approach that includes principles from the two disciplines resilience engineering and requirements engineering. The described methodological approach has been applied to workshops with ATCOs to develop a mutual understanding of an ATCO's daily work. The workshops in particular focus on complex situations that can arise when Go-Arounds happen. This understanding is used to develop scenarios, use cases and user stories which the development team breaks down into requirements.

The methodology is described in detail in chapter 2. Chapter 3 describes the results that were achieved by applying the methodology to workshops with ATCOs. A detailed description of scenarios, use cases, user stories and requirements is provided.

These results however do not describe how goals are achieved - this remains to be answered by the experts of the respective domain in the upcoming development phase. Rather, the focus lies on what we as a team aim to achieve to provide a benefit to Air Navigation Service Providers (ANSP). The presented results provide the guard rails for a user-oriented development phase and built the basis for finalizing impact assessment, once the development phase has progressed.

Requirements, user stories and use cases are, in consultation with ATCOs, allowed to change over the course of work packages 3 and 4. This ensures that the developments satisfy the end user's needs and is accounted for by the agile management style already outlined in the management plan.

Based on the requirements, chapter 4 provides a technical problem statement, which outlines possible techniques available to implement the requirements. In accordance with the problem statement and requirements, data sources used for SafeOPS are assessed and deployed in DataBeacon, the platform SafeOPS uses for the development phase.

2 Methodological Approach

This chapter describes the methodology used to derive the requirements for the predictive decision support tool to be investigated in SafeOPS. First, resilience engineering is introduced, capturing important aspects for the scope of this project. The concepts of Safety-I and Safety-II are introduced, and some basic terminology is explained. Subsequently, our agile approach to requirements engineering is outlined. Finally, a separate section is dedicated to the workshops, which are an essential part of the requirements and resilience engineering process in this project. Within this last section, a description of how requirements are derived from workshops is included. The requirements derived from our methodology will inform the following development phase of SafeOPS.

2.1 Resilience Engineering

Air Traffic Control (ATC) operations are required to be highly reliable for ensuring a high level of safety. It is therefore required to consider safety, reliability and resilience already during the early stages of development of a new system or tool, as done in the SafeOPS project. Therefore, the necessary resilience engineering ideas and principles will be briefly introduced in the following.

Traditionally, safety is understood as the absence of unwanted outcomes of a system. Resilience engineering puts a wider perspective on safety, using a holistic approach in understanding a system's outcomes, as illustrated in Figure 1. The focus is not solely to prevent a system from producing unwanted outcomes but also to proactively ensure desired outcomes. [1]

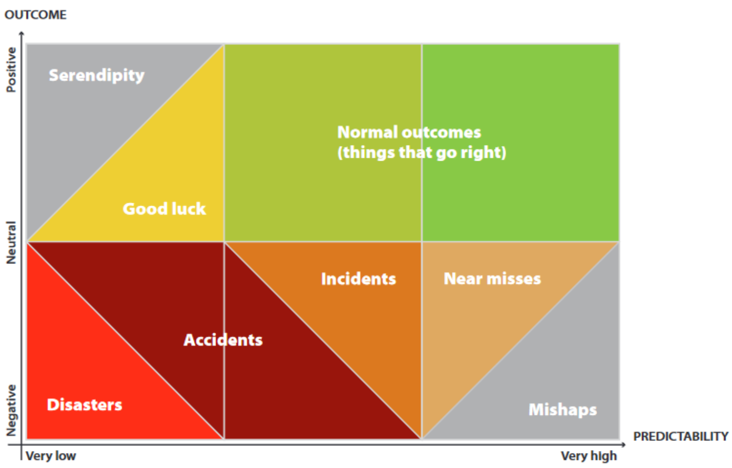


Figure 1: Possible Outcomes from a Resilience Perspective [1]

In resilience engineering, there are two separate approaches to safety, namely Safety-I and Safety-II. According to [2], Safety-II describes the condition of a system operating under expected conditions and producing as many positive outcomes as possible and proactively avoiding negative outcomes. Therefore, the author claims Safety-II provides a consistent definition of safety for systems that are intractable or underspecified. Safety-I on the other hand refers to the more conventional approach to safety, by minimizing negative outcomes and reacting to negative outcomes after they have occurred.

Air Traffic Management (ATM) is a complex socio-technological field and [3] concludes accordingly that the traditional understanding of Safety-I needs to be expanded to also include the Safety-II perspective, by adopting a resilience engineering view in ATM. This is also reflected in the SESAR Safety Reference Material [4] and the respective Guidance Documentation [5]. This material is of particular interest for SafeOPS' Task 2.2, where an evaluation of the impact of the project’s ideas and developments on safety, capacity and resilience on the ATM processes takes place. By combining principles of resilience engineering and the Safety-II approach with the requirements engineering, described in section 2.2, enables us to already account for the challenges of task 2.2 from the beginning of the project. At the same time this approach increases the chances of success of the development beyond the scope of this project's timeline. Furthermore, applying these ideas from the beginning helps the team to further familiarize with these concepts. In this context, the principles of resilience engineering SafeOPS focuses on are:

- Work As Done
- System Thinking

Work as Done

The term *Work as Done* describes the way people work to achieve an assigned task. On the contrary, *Work as Imagined* describes the work as it is envisioned or expected to be done [6]. ATM is a complex task which is assumed to be intractable and cannot be analyzed entirely [3]. Prior to developing a new idea or tool, one must therefore understand the ins and outs of a system. In terms of the current project, the development team needs to gain a thorough understanding of an ATCO's behaviour in different situations, and his/her strategy to mitigate risk. This understanding helps when developing a tool that shall assist an ATCO's daily routines, while at least maintaining the same level of reliability under varying conditions. Based on this understanding, possible use cases for a new system can be developed, which, combined with a requirements engineering approach, is the first step in the system design process.

System Thinking

Understanding Work as Done cannot be achieved from the outside, however. Advise on how to understand and capture Work as Done, the surrounding system conditions and system behavior is described in Eurocontrol's System Thinking for Safety white paper [7]. The principles used at this stage of the project and for this deliverable are summarized below.

- Field expert involvement: It is necessary to include the people who actually do the work in order for the developers to understand the system behavior and eventually also Work as Done.
- Local rationality: System behavior is dependent on the system's location. Tower ATCOs in Airport 1 will handle Go-Arounds differently from ATCOs in Airport 2 due to e.g. local procedure designs or geographical limitations. Therefore, performance of a system must be understood from the perspective of the local experts.
- Demand and Pressure: The performance of a system is measured against the demand from the customers and colleagues. Especially in unusual events it is important to understand demand on the system to be able to understand a systems' performance. "Designing for demand is a powerful system lever. To optimize the way the system works, the system must absorb and cater for variety, not stifle it in ways that do not help the customer." [7]

- Resources and Constraints: To understand the Work as Done and how the demand is met in a system, one needs to consider the available tools and their usage as well as the procedures in place and how they are interpreted and lived.

Furthermore, the white paper summarizes the following practical advises:

- Identify the stakeholders.
- Consider system purposes.
- Explore the system and its boundary.
- Study system behavior and system conditions.

2.2 Requirements Engineering

Requirements Engineering provides a means of systematically defining, managing, and verifying requirements for a project and guides the development team through the entire development cycle of a product. All the requirements combined define the properties, including functionality and quality, of the final product and are therefore essential to a project's success. It is for that reason that product requirements not only need to be defined and analyzed, but also documented, described, and adapted throughout the development cycle. These tracking and analyzing tasks shall ultimately ensure that the final product meets all quality standards demanded by the end user. Furthermore, requirements engineering ensures that all requirements are coherent and thereby set mutual goals for the project team. Despite all requirement engineers facing the same objectives, requirements engineering is unique to each project. Due to the exploratory nature of this project, with no prior experience in developing a Go-Around prediction tool available, a very user focused approach to requirements engineering has been chosen.

A series of workshops is organized by DFS for the development team to gain an in depth understanding of a tower controller's work and daily challenges. Throughout these workshops, several tower controllers (ATCOs) guide us through approach and missed approach patterns and outline potential bottlenecks. Based on this understanding, different (operational) scenarios are created that define a set of events leading to critical situations within proximity of an airport (AP). Based on these scenarios, use cases can be defined, that provide the overall frame to the project. Each use case can again be broken down into user stories. These user stories are formulated in a nontechnical manner and are meant to provide the development team with some guidelines. At the end of this cycle are the requirements, which are derived from user stories. The general process of deriving requirements is depicted in the flow diagram in Figure 2 below.

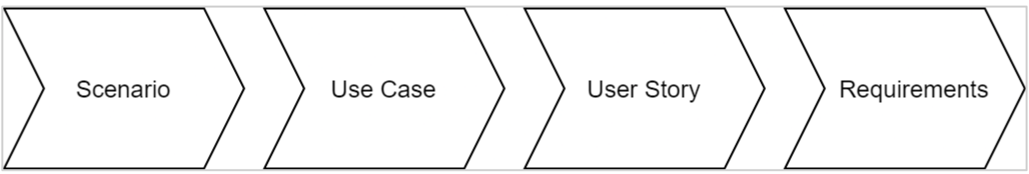


Figure 2: Requirement Engineering Process

The following sections describe the terms "Scenario", "Use Case", "User Story" and "Requirements".

2.2.1 Scenario

Typical requirements engineering literature does not provide a common definition of a scenario. The Oxford Learner's Dictionaries [8] define the term scenario as either "a description of a possible series of events or situations" or a "a written outline of what happens in a film or play". Within the context of this project, we define the term scenario as setup of conditions that lead to situation with increased complexity for the involved actors (ATCOs and Pilots) to handle. Scenarios describe situations in which a Go-Around prediction could yield benefits to the involved actors. Existing local procedures, described in the scenarios, are risk assessed and comply with relevant safety regulations. With the selection of scenarios, we intend to find situations where the impact of a potential tool can best be evaluated. This enables requirements engineers to generalize, and to some extent also simplify, the complexity of the real world.

Such a high-level representation of a situation by a scenario comes with two great benefits to the development team.

1. Scenarios provide a general understanding to when a situation is (not) complex
2. Scenarios help the development team understand how controllers work in different situations.

Each scenario is documented by means of a table, that lists all its relevant aspects, as indicated by Table 1. This simplifies version tracking of scenarios over the course of the project and gives a quick overview of all relating information.

Table 1: Template for Scenario Tables

Scenario ID:			
Scenario Name:		**Version No:**	
Linked Use Case(s):			
Involved Actors:			
Description of Traffic Context:			
Involved Decision-making:			
Effect on ATCO / ATM / Flight Crew:			
Visualization:			
Link to relevant AIP Chart:			

Additionally, a thorough description is provided for each scenario, including a sequence diagram, to provide the reader with all the information needed to gain a good understanding of the overall situation. Scenarios are thought as stages for potential use cases of a decision support tool, which are explained in the next section.

2.2.2 Use Case

A use case is a way of using a system to achieve a particular goal for a particular user. All use cases taken together provides all of the useful ways to use the system. From a developer perspective, they set the boundaries for the system, and help developing a shared understanding of the final product. Visual Paradigm [9]attributes the following properties to a use case:

- A sequence of actions a system performs that yields an observable result of value to a particular user.
- The specific behavior of a system, which participates in a collaboration with a user to deliver something of value for that user.
- Provides the context for a set of related requirements.

For the general understanding of all use cases, it is important to outline the relationship amongst different use cases and the stakeholders involved using some simple principles:

- Keep it simple by telling stories.
- Understanding the big picture.
- Focus on value → elaborate in workshops with ATCOs.
- Build the system in slices → Break down in user stories.
- Deliver the system in increments.
- Adapt to meet the team's needs.

Use cases are elaborated in workshops with the users' (ATCOs) help. Therefore, we first define scenarios of Go-Around situations, in which we try to understand the ATCO's actions in detail. Based on this understanding, use cases will then be worked out. This process also reflects the Work as Done and System Thinking principles defined in section 2.1. In comparison to the scenarios, use cases elaborate on the benefits a predictive decision support tool could provide to the system, to better resolve the conflicts formulated in a scenario. Like scenarios, they shall guide the development team through the development process by refining the general outline set by scenarios. The expected benefits defined for the use case will, in accordance to the SESAR Safety Reference Material Section 7 [4] also serve as basis for the definition of Safety and Resilience Performance Indicators for Task 2.2 of SafeOPS.

Table 2 indicates, how a Use Case will be documented in SafeOPS. Similar to scenarios, a table that lists the actors actions, as well as a sequence diagram which provides a more detailed and accurate representation of the workflow, is provided.

Table 2: Template for Use Case Tables

Use Case ID:			
Use Case Name:		**Version No:**	
Linked Scenario:			
Linked User Stories:			
User(s)/Actor(s):			
Preconditions:			
Brief description:			
Expected Benefit for ATCO / Flight Crew / ATM Operation:			

A table is appended to describe the basic flow taken by all actors. This table's purpose is to provide the reader with a general overview of the sequence in which certain events take place. The sequence of actions may change, according to the situation.

The set of use cases presented in the next chapter do not claim to be the full set of all possible use cases for a Go-Around prediction tool. They are, however, what the development teams portraits as the most likely ones to be implemented.

2.2.3 User Story

A user story is a snippet of a user's need/challenge. It captures what the user does or needs as part of his/her work. The literature is, in comparison to use cases and scenarios, very concise in defining user stories as the lowest level functionality a tool must have to provide a benefit to the user. It cannot be further separated in sublevel functions.

User stories are intentionally kept very short and in simple, plain language. Only a general explanation of a tool feature is given by each story and written from a user's perspective. Part of this explanation is also the incentive behind a story. It shall not only provide the context for the development but also articulate how value is created for the user. Since all stakeholders are encouraged to formulate user stories, they also work as an intermediary between the development team and the remaining stakeholders. A simple guideline to a good user story is given by the "INVEST" principle [10]:

- Independent - user stories can be used in any order, they act as standalone units
- Negotiable - user stories must be flexible and open to change
- Valuable to users or customers
- Estimable - user stories should be reliably measured in terms of the time an effort they take to realize
- Small - simply "Who wants what and why?"
- Testable - There should be a way to check a stories functionality

To define user stories, we rely on the template provided by Visual Paradigm [11]. It outlines the structure of user stories as well as their content.

- Target Audience: Who is it for?
  - Role should be an actual human in a specific position, and not a dev. team.
- Behavior: What it expects from the system?
  - Behavior of the system/product should be described, it is usually unique for each User Story.
- Benefit: Why it is important (optional)?
  - Benefits should be real-word results that are non-fictional or external to a product.

These three properties lead to the blueprint of our user stories.

As a [who], I want to [what], so that [benefit].

As opposed to use cases, user stories focus on individual functions or subsystems. Details of a user story, however, are not documented to the same extend as use cases. User stories intentionally leave out a lot of detail to encourage discussions within the development team. Again, to ease traceability a table format is chosen to document all use cases. However, no further diagrams are provided. The template for user story documentation is given by Table 3.

Table 3: Template for User Story Table

User Story ID:		**Category:**		**Version:**	
Linked Use Case(s):					
Linked Requirement(s):					
User Story	**As a...**	**I want...**		**so that... / in order to...**	

2.2.4 Requirements

Requirements are the description of what a system should be able to do, the services and benefits it provides, as well as its limitations and constraints regarding operational use. A description of the system on a very technical level is necessary to provide the developers with precise goals they have to achieve. In order to gain a common understanding of the terminology we use the definition of requirements from the IEEE Standard Glossary of Software Engineering Terminology:

- A condition or capability needed by a user to master a challenge or achieve an objective.
- A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document.
- A documented representation of a condition or capability as in 1 or 2 (in other words a requirements documentation plan or a formal product requirements document (PRD)).

Much like user stories, requirements are ascribed a set of properties:

- Propriety: Requirement fulfils user's needs
- Uniqueness: Requirement formulation allows for only one interpretation
- Testability: There are tests that can verify a requirement is (not) fulfilled
- Traceability: Root and development of requirement can be traced

All of these properties are solely important for the development team, and are not required to be defined directly by the end user. The development team derives all requirements are derived from user stories. Requirements are formulated in a way, that they fulfil the needs expressed by the users in the user stories. In the literature numerous types of requirements are defined. Since SafeOPS is a comparably small project with a small development team, the requirements are only subdivided into the categories "functional", "non-functional", and "data requirements".

Regardless of the type of requirement, all requirements are formulated and documented using the template shown in Table 4.

Table 4: Template for Requirement Table

Functional Requirement ID:		**Category:**		**Version No:**	
Linked User Story:					
Requirement:					
Brief description:					

2.2.4.1 Functional Requirements

Functional requirements define the functionality of a system or tool. In most cases, these can be derived directly from a user story. Functional requirements define how the user interacts with a system, a thereby impact how a user benefits from a tool. Possible examples of functional requirements are what information shall be provided at what time and how and how a user can interact with a tool. Furthermore, these requirements specify the system's functionality. As SafeOPS investigates data driven functionalities, requirements that specify data handling and data processing will also be listed as functional requirements.

2.2.4.2 Non-Functional Requirements

Non-functional requirements define how a system must perform to provide a benefit to the user. Usually, these describe performance minima that must be achieved by a system. Also, some of the non-functional requirements are hardware and bandwidth limitations, quality requirements or general system requirements (i.e., on which operating system a tool must run).

2.3 Workshops

As format to apply the introduced principles of resilience – and requirements engineering to generate the desired information, SafeOPS organizes recurring workshops with ATCOs from two major European APs. The project partners installed an iterative scheduling for the workshops with ATCOs from Airport 1 and Airport 2 for SafeOPS. This scheduling leaves room for preparing and post-processing each workshop. We chose to organize workshops with 1-3 people from each consortium member with a higher recurrence rate rather than large but sparsely distributed workshops, also because of the Covid enforced video call format.

2.3.1 Planning

ATCO shift planning is done months in advance. The project team started the planning period in January for April. The group decided to limit the workshops to three hours, if possible, twice a week and alternating between the Tower in Airport 2 and Airport 1. Findings and Insights are analyzed between the workshops, allowing the project members to reflect on the new information and further discuss it a couple of days later. After the workshop week, the results are documented and used for the preparation of the following workshop. This process is illustrated in Figure 3.

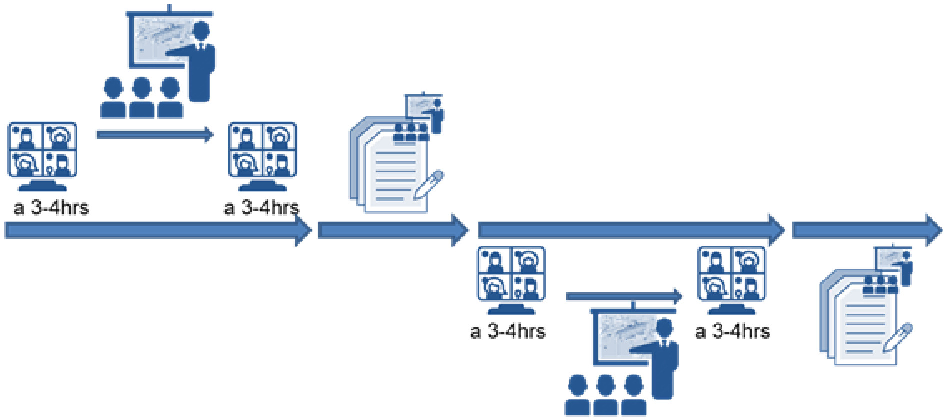


Figure 3: Work Shop Planning

2.3.2 Scheduling

The workshops schedule, illustrated in Figure 4, was developed together with the partners from tower Airport 1 and Airport 2. The dates were chosen with respect to availability of ATCOs with different experience and backgrounds to enhance diversity.

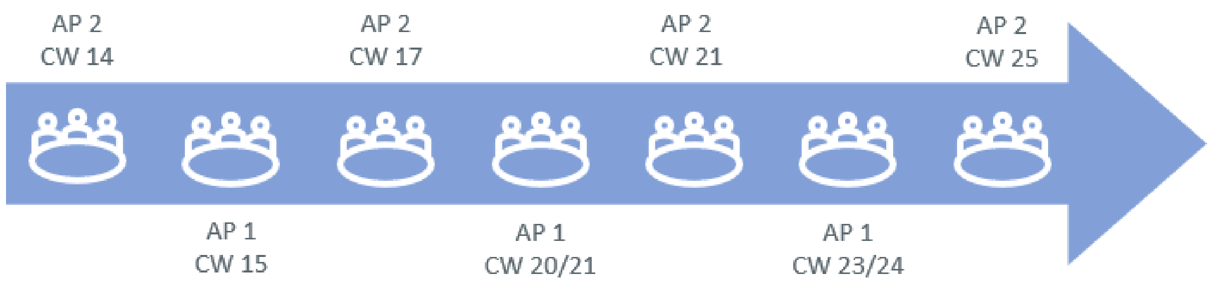


Figure 4: Work Shop Scheduling

2.3.3 Execution

Due to the ongoing pandemic, the preferred option for in-person workshops was not possible. Instead, virtual conference sessions were setup in Microsoft Teams. Members from SafeOPS are participating, including ATCOS and Pilots. Early Workshops from Calendar Week (CW) 14 to 21 focused to understand and retrieve processes as well as develop new use cases and user stories. The workshops from CW20 onwards focused on confirming the developed cases and stories as well as developing and showing ATCOs design options and discussing potential hazards. The process is visualized in Figure 5. Note that with the handover of this deliverable, the process is not finished but will be kept running throughout the development phase, however with reduced occurrence rates of workshops.

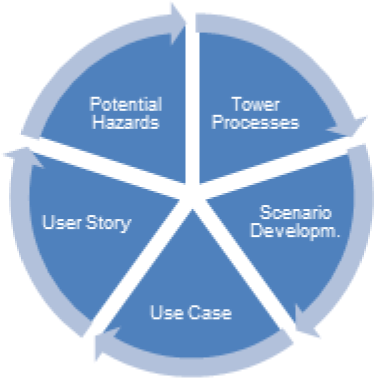


Figure 5: Workshops as Part of a Cyclic Requirement Definition

2.3.4 From Workshop to Requirement

Derivation of user stories begins with detailed planning of the workshops. Objectives for each workshop are defined. As already mentioned previously, the first goal is to define and understand several scenarios in which a Go-Around prediction is expected to provide a benefit to tower controllers. Therefore, it is essential for the developers to gain a thorough understanding of the processes in place at different APs, and how air traffic controllers interact with the air traffic in different situations. A series of aspects need to be investigated:

- Working environment: The workspace, the general equipment and furniture, and the physical environment
- Organization and staffing: Organizational management, people management, and personal factors
- Procedures, roles and responsibilities: Actual/prescribed working methods, positions/functions in the organization and expected tasks.
- Teams and communication: How people work and communicate with each other on shared goals and tasks
- Human in the system: The actions, reactions, and interactions between humans and other system components

Types of questions that help us define scenarios include, but are not limited to:

- What happens?
- What is it that creates a complex situation?
- Why is it a complex situation?
- What are the details of the complex situation?
- How do you resolve the complex situation?

Having a general understanding of different scenarios, use cases are derived, also during workshops. To do so, scenarios are presented to air traffic controllers and complex situations arising in each of the scenarios are evaluated. A focus during this part of the workshops lies on questions such as:

- What makes this situation complex?
- What is the impact of different decisions made in such a situation?
- How can this situation be improved?

As a result, one or more use cases for each scenario may be derived. Also, the same use case may apply to several scenarios. A use case may for instance be "decrease workload of an air traffic controller“. Having derived such a use case, the next step is to derive user stories. The process is much like deriving a use case, but more detailed. Individual tasks within a process are isolated which may be adapted or replaced by a new tool, to help an air traffic controller in each situation. Questions in this part of a workshop may be:

- What information do you need and how do you want it to be presented?
- At what point in time is a given information useful?

Now that the lowest level of functionality is defined, the development team has all information needed to derive the requirements for the project. This part is not carried out during the workshops, but each development team defines those for its area of accountability.

The agile management and development style relied upon in this project encourages us to re-evaluate all use cases, user stories and requirements over the course of the entire project. Hence, the workshops will accompany the development team throughout the entire development process. This shall ensure that the product satisfies the user's needs. Therefore, all results presented in section 3are subject to change, as new knowledge is gained or circumstances change.

3 SafeOPS – Requirement Engineering Results

This chapter presents the results of the requirements and resilience engineering process. They directly stem from the workshops with tower controllers (PLs) from the Airports 1 and 2. Before presenting the results, the different types of Go-Arounds, and the boundaries of this project are introduced.

3.1 Introduction to Go-Arounds

A major cause for approach and landing accidents (ALA) is a failure to recognize the need and execution for a Go-Around. Empirical data obtained by the Flight Safety Foundation in 2017 [12] showed that if Flight Crews (FCs) had decided to go around, 83 percent of runway excursions and 54 percent of all accidents during a 16-year period could have been avoided. However, only approximately three percent of Unstabilized Approaches (UA) resulted in a Go-Around [12]. The Go-Around itself bears a higher risk of loss of control in flight (LOC-I) compared to other phases of flight. Thus, a Go-Around should only be executed if its inherent risk is lower than the risk which comes with an Unstabilized Approach. The same study revealed that noncompliance with Go-Around policies is the main contributor to ALAs, whereas Unstabilized Approaches affect less than half of runway excursions. Since every Go-Around is different, we initially distinguish the ATC and Flight Crew induced Gas, to later specify the system boundaries for the SafeOPS developments.

3.1.1 ATC Induced Go-Arounds

The reasons for discontinued approaches are manifold. But the decision is made either by the Flight Crew or ATC. For both, the missed approach (MA) is a standard procedure which must be expected at any time. The situations where an ATC has to issue a Missed Approach clearance to an arriving aircraft (A/C) derive mostly from separating A/C close to the runway. By procedure A/C are getting closer to each other during landing and departure compared to any other flight phase. The Tower Controller is responsible to assure separation at any time and therefore has to react in case conditions change and a situation develops differently than expected.

The following examples are most relevant for ATCOs to advise a missed approach:

- runway is blocked (by other A/C, by other vehicles) and not clear for landing traffic,
- runway has to be checked on short notice (e.g. due to bird strike or other foreign objects),
- other traffic (e.g. rescue helicopter) requires immediate priority or
- the sequence of approaching A/C is expected to infringe separation minima (between A/C), so the succeeding traffic must conduct a missed approach to prevent a separation infringement.

The workload for a controller increases during Missed Approaches. After transmitting the MA instruction to the pilots, the controller has to observe, if the A/C follows his instructions and often the MA A/C has to be coordinated with the adjacent airspace sectors. Depending on the layout of the AP and the airspace structure around it, coordination procedures could involve different persons. A situation with an unexpected, blocked runway could also result in multiple MAs, if more than one A/C is on final approach.

3.1.2 Flight Crew Induced Go-Around

The chain of events leading to a Flight Crew induced Go-Around often begins at the start of descent from cruise altitude. A diligent and timely approach preparation ensures the A/C, and its crew are ready for the approach. This process includes a landing performance assessment comprising a crosscheck of the runway surface conditions assumed during pre-flight preparation. The crew reviews the A/C's technical status, calculates the approach speed, evaluates the current weather situation and NOTAMs. They determine the approach configuration, the use of automation, and programming the flight management guidance system. To conclude the process, the pilot flying informs the pilot monitoring about the planned course of action in a concise and logical way. During the approach, the Flight Crew continuously updates any changes affecting the lateral and vertical flight path including the Go-Around trajectory [13].

Being fully prepared and minded for a Go-Around is paramount during any phase of the approach. The flight crew should have a proper situational awareness and should be ready to follow the published missed approach procedure (MAP) by application of standard operating procedures, if the criteria for a stabilized approach aren't met, required visual minima cannot be obtained and/or confusion about the use or performance of automation exists. Once a Go-Around is initiated the crew is fully committed to follow the standard missed approach procedure or flight path instructions transmitted by ATC [13].

An approach is considered stabilized when A/C track, flight path angle and air speed are within defined limits at the stabilization height. Recommended minimum stabilization heights are 1000 ft above aerodrome level in instrument meteorological condition (IMC) and 500 ft above aerodrome level in visual meteorological conditions (VMC). At or below the minimum stabilization height any deviation from the flight parameters defined for a stabilized approach should be announced by the pilot not flying/pilot monitoring and a Go-Around should be conducted [14].

An analysis by Flight Safety Foundation [14] categorized factors of Unstabilized Approaches. They are attributed to human performance limitations, meteorological influences, technological reasons, and organizational and training factors. The resulting deviations adversely affect the flight path, state of energy and configuration of the airplane below the minimum stabilization height.

3.1.3 Flight Crew Decision Making in Go-Around Scenarios

Although a Go-Around is considered a normal flight maneuver it carries safety issues such as loss of control inflight (LOC-I), spatial disorientation, reduced separation to other traffic and wake turbulence [15].

Courville [16] claimed that knowing the threats makes them manageable. The crew´s decision to continue or abandon an approach is based on

- Briefed options and decisions
- Updated information from ATC/pilot reports from previous A/C
- Radar display in convective weather
- Computed actual wind on navigation display
- Communication between pilots and ATC

The reaction time to adverse circumstances during an approach continuously decreases as the A/C nears the runway. At or below decision altitude or minimum descend altitude, time pressure is at its highest, see Figure 6. Thus, decision making must be highly efficient. The key to manage time pressure is task sharing between the pilots as stipulated in standard operating procedures and use of technology such as Airbus` Runway Overrun Prevention System or Boeing`s suite of runway situation awareness tools.

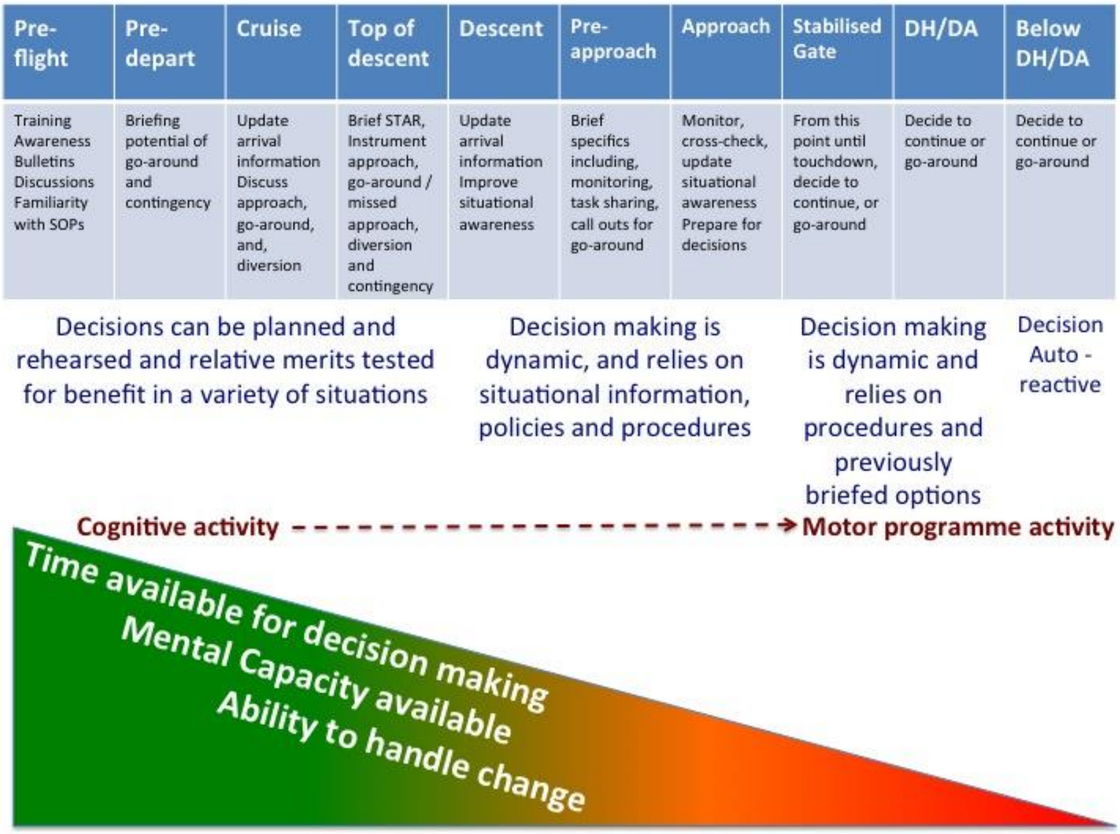


Figure 6: Reaction Time During Approach [17]

3.1.4 The Go-Around Procedure

Each AP publishes Go-Around procedures that must be followed by Flight Crews. Under ideal conditions, no action is necessary from the ATCO in authority, other than clearing the Flight Crew for the procedure and coordinating the flight with adjacent sectors. The Flight Crew on the other hand can fly a Go-Around which it briefed during the approach preparation, and despite an increased workload, this does not pose a great challenge for the crew. However, due to the vast amounts of traffic (prior to Covid-19) and APs operating at their maximum capacity, standard procedures could not always be issued to the flight crew in case of a Go-Around. To ensure (radar or visual) separation between A/C, the ATCO is often required to vector the Missed Approach, rather than simply clearing it for the standard procedure.

3.1.5 Focus on Flight Crew Induced Go-Around

Over the course of this project, the focus will be on predicting A/C induced Go-Arounds. Flight Crew induced are unpredictable to the ATCO. Since ATCOs do not have specific insights on contributing factors, that may lead to a Go-Around decision in the cockpit, as for instance wind shear, this adds to this unpredictability. But even after a Go-Around has been initiated by a pilot, particularly in poor weather conditions, it may take an ATCO over half a minute to realize a Go-Around has been conducted due to a time delay within the radar system and delayed Go-Around reporting by the flight crew, as their first priority is to fly the A/C. It is therefore expected that an ATCO can gain additional insight from a tool based on these data sets. Potentially a prediction tool for A/C induced Go-Arounds positively influences an ATCOs situational awareness and avoids additional stress factors.

3.1.6 System Boundaries for SafeOPS

SafeOPS focuses on the work of Tower Controllers, and therefore some boundaries are defined to a Go-Around prediction tool. The most obvious being that a prediction shall be displayed to Tower Controllers, and even though ATCOs from adjacent sectors will eventually be involved in handling a Go-Around, they are not included in the earliest development process or the requirement generating process of this project. Similarly, no ground or apron operations are considered here. Finally, as elaborated upon in the previous chapter, only A/C induced Go-Arounds will be considered.

Resulting from these boundaries, the project focuses on is the control zone and terminal control area in which a Tower Controller is responsible. The handover from the approach controller to the Tower Controller at approximately 8-12 NM from the runway threshold (THR) yields as entry of A/C, and taxiways as their exit. For departing A/C, affected by a Go-Around, taxiways are an entry for the Go-Around prediction investigation and the boundaries to the adjacent sectors serve as an additional exit.

3.2 SafeOPS – Scenarios

This section documents the scenarios, discussed in SafeOPS during the workshops with the ATCOs. The scenarios will be distributed between Airport 1 and Airport 2. All existing local procedures, described in the scenarios, are risk assessed and comply with relevant safety regulations. With the selection of scenarios, we intend to describe situations where the impact of a potential tool can best be evaluated. Scenarios are generated from discussions in workshops. These discussions are often controversial and multiple opinions on strategies for described traffic contexts exist. The documented versions of scenarios in this section do not claim to be absolute but reflect an agreed summary of the workshop results. Furthermore, all described actions in the flow tables are not strictly ordered and binding but are exemplary and do also vary. For clarity purposes, the tables just provide one example each how the actions could be ordered and when a prediction tool is expected to be supportive to the users.

To be able to understand AP specific procedures mentioned in the scenarios, they will be defined hereafter.

3.2.1 Airport 1 Specific Procedures

Definition Missed Approach Buzzer:

In Airport 1's Tower, the Missed Approach Buzzer is a tool that the tower ATCO applies when a Missed Approach occurs in his area of responsibility. It shall ensure that all other Tower ATCOs are informed about the situation and trigger the necessary coordinative actions.

Definition: Taboo Zone

Depending on the Standard Instrument Departure (SID) of departing A/C from runway (RWY) 25C, compliance with a taboo zone is mandatory for the ATCOs in Airport 1. If the departure is cleared via a Foxtrot or Golf SID, the take-off (T/O) clearance must not be given if a landing A/C is within the taboo zone of RWY 25R (taboo zone R in Figure 7). For all remaining SIDs, the taboo zone R is not applicable.

In such cases, a departure on RWY25C must not be cleared for Take-Off as long as the next landing on RWY25L is in the relevant taboo zone (taboo zone L in Figure 7).

The decisive factor for installing such areas was the commissioning of RWY25R, which resulted in procedural changes. This led to possible situations, where a missed approach would be conflicting with a departure, leaving the controller without sufficient time to separate both A/C properly.

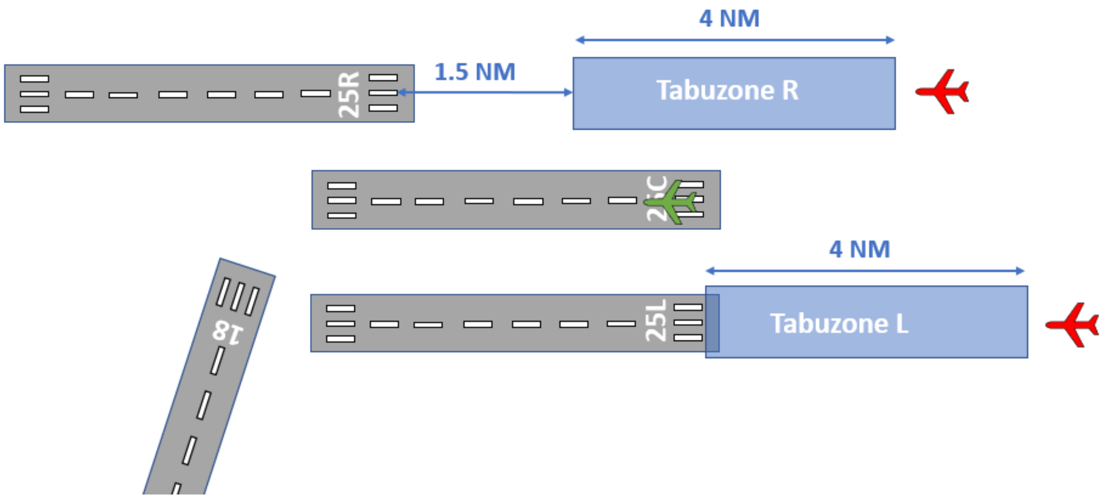


Figure 7: Taboo Zones at Airport 1

Definition: Swing To Depart

The Tower Controller can ask the arriving A/C to perform a swing over from 25L towards 25C, visualized in Figure 8. Therefore, the taboo zone rule described above is not applicable and the A/C on 25C may be cleared for Take-Off in case the arriving A/C agrees to the swing-over maneuver. According to the experience of the local controllers, especially the pilots who are familiar with the procedures at Airport 1 agree since they benefit from shorter taxi times after landing on 25C compared to a landing on 25L.

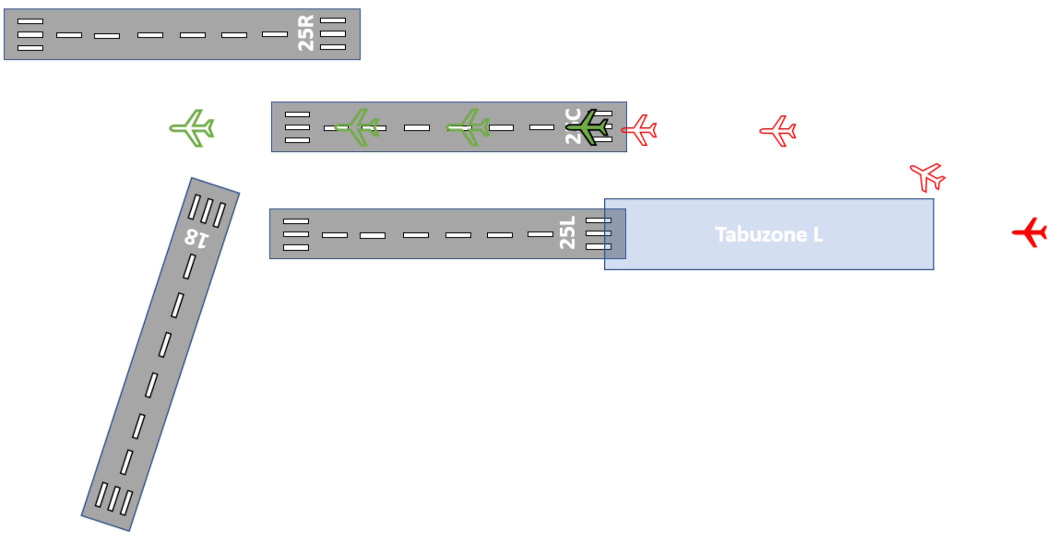


Figure 8: Depiction of Swing-To-Depart Maneuver

Definition Reduced Radar Separation:

For approaches on the parallel runways 07C and 07R or 25C and 25L, a minimum radar separation of 2.5NM applies according to Airport 1 AD 2.22 §2 on following conditions:

- The preceding A/C is of the same or a lower weight category. A/C of the weight category HEAVY, including the B757 as preceding A/C, are excluded from this procedure.
- The exit taxiways of the runway are discernible visually by the tower controller or by means of surface movement radar.
- The runway is dry.

Definition: Parallel Runways Operations:

Following Airport 1 AD 2.22 §4, when using ILS and/or GBAS, independent parallel approaches may be conducted on runways:

- 25L and 25R or 07L and 07R
- 25R and 25C or 07L and 07C

3.2.2 Airport 2 Specific Procedures

Definition: Mixed Mode

Airport 2's AD 2.20 Flight Procedures §3.1.4 implies a mixed mode operation, since arrivals and departures are planned for one runway.

Definition: Independent Parallel Approaches

In Airport 2, under conditions specified in Airport 2's AD 2.22 Flight Procedures §3, both runways can be used in parallel for approaches in all meteorological conditions.

Definition: Reduced Radar Separation

Paragraph 7.11 in the PANS-ATM ICAO Doc 4444 describes the possibility of operating a runway with Reduced Radar Separation Minima, given that a safety assessment shows an acceptable level of safety. In Airport 2, Airport 2 AD 2.22 Flight Procedures §2 states that given the following conditions, a radar separation minimum of 2.5 NM is applied on final between 10NM and touchdown:

- The preceding A/C is of the same or a lower weight category. A/C of the weight category HEAVY, including the B757 as preceding A/C, are excluded from this procedure.
- The turn-off points of the runway are discernible visually or by means of surface movement radar from the control tower.
- The runway is dry.

Definition: Reduced Runway Separation

In general, an arriving A/C may be cleared for landing or a departure may be cleared for Take-Off only, if the preceding departure has overflown the runway end or has initiated a turn, or the preceding landing has vacated the runway. This standard runway separation may be reduced when certain conditions apply, like weather minima and runway condition. Mostly, the reduced runway separation is helpful if a departure flight shall be cleared for Take-Off close in front of the next landing. Instead of full runway length, the distance between the airborne departure and the next landing may be reduced to 2400 m, measured from the time, when the landing is over the THR (or 1500 or even 600m, depending on the category of the A/C involved). Due to the dominant airliner traffic at Airport 2 and Airport 1, almost always the prescribed reduced separation minima of 2400m applies. Compared to the runway length of 4000m in Airport 2, reduced runway separation allows to bring A/C closer around the runway.

Definition: High Intensity Runway Operation (HIRO)

To achieve a high throughput, Airport 2 aims for minimizing the runway occupancy time per A/C. Thus, approaching A/C are requested to use high-speed turn-offs, as defined in Airport 2 AD 2.20 Local aerodrome regulations §3.1.6.1 or earlier ones.

Airport 2 AD 2.20 Local aerodrome regulations §3.2.3 states that pilots of departing A/C should ensure to follow clearances for line-up and take-off without delay. Furthermore, pilots should expect instructions for immediate departure and must inform ATC immediately, in case they are unable to follow.

3.2.3 Airport 1 Scenario 1

In the following, the first Go-Around scenario in Airport 1 is documented. For this scenario, the departure from runway 25C, illustrated in Figure 9, as well as the arrival on runway 25L and especially the Missed Approach Procedure defined thereon, illustrated in Figure 10, are of interest. The standard Missed Approach Procedure for runway 25L is defined with the first instruction to climb straight ahead to 2.5NM DME FRD.

The scenario defined in Table 5 describes the situation if the arriving A/C conducts a Missed Approach. Furthermore, the involved decision-making process from the involved ATCOs to resolve the situation and the potential effects, the resolution strategy has on all involved actors, is described.

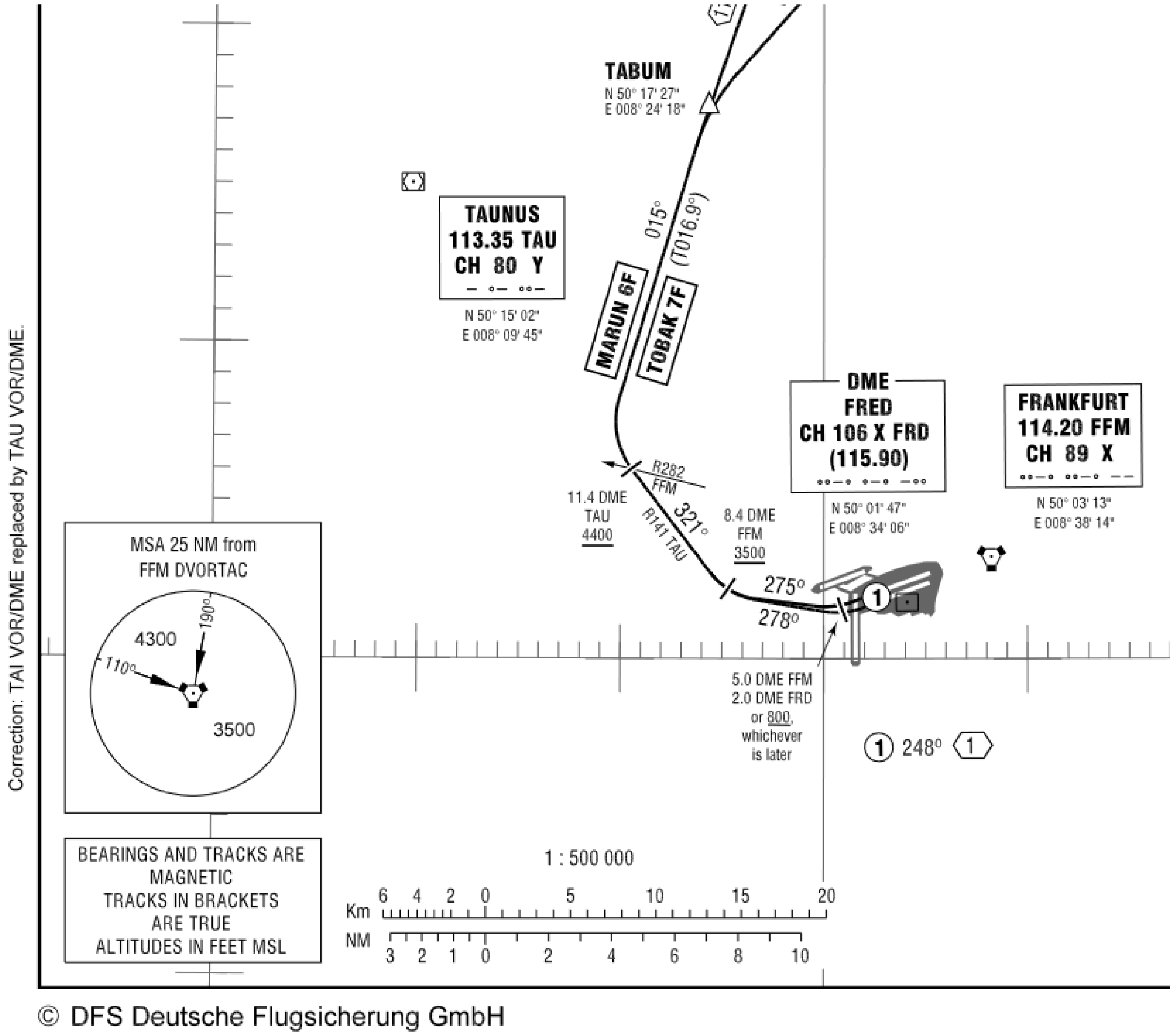


Figure 9: Standard Instrument Departure for Scenario 1 at Airport 1 [18]

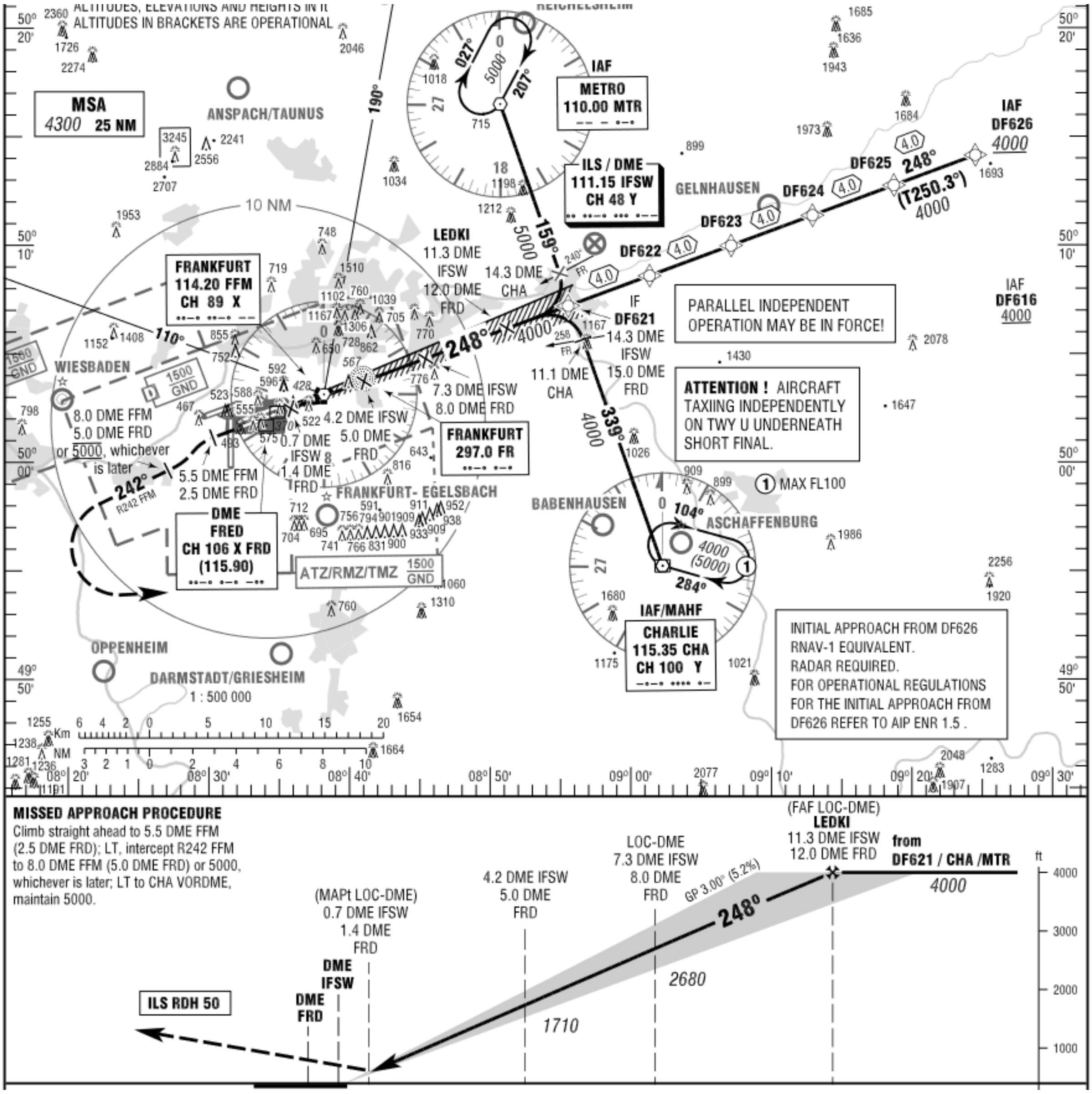
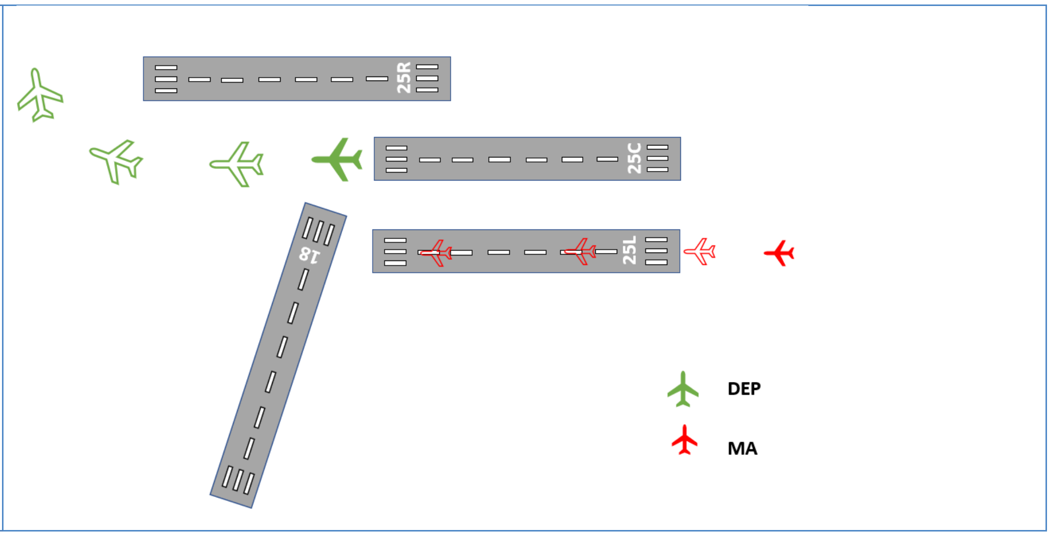


Figure 10: Standard Arrival Route for Scenario 1 at Airport 1 [18]

Table 5: Scenario 1 at Airport 1

Scenario ID:	Scen.Airport 1.1		
Scenario Name:	Northern departure on runway 25C and Missed Approach on 25L	**Version No:**	1.0
Linked Use Cases:	Airport 1 Use Case 1		
Involved Actors:	Center Runway Tower Controller (PLC), Northern Runway Tower Controller (PLN), Western Runway Tower Controller (PLW), Southern Runway Tower Controller (PLS) (or PLCS if Center and South combined), Flight Crew Departure, Flight Crew Arrival		
Description of Traffic Context:	A departing A/C on runway 25C on SIDs FOXTROT and GOLF (NORTH) turns right after 2 NM FRD. In case of a Missed Approach on runway 25L, the A/C on a standard Missed Approach Procedure proceeds also straight ahead until 2.5 NM FRD before turning left. In this situation, the taboo zone procedure RWY 25L is not applicable given the departure's SIDs. Especially if marginal weather conditions prevail, the situation becomes quite complex and is more likely, as during local weather phenomena (Cumulonimbus and thunderstorms) in the southwestern area, all departures, regardless of their A/C category, could be cleared via FOXTROT or GOLF SIDS, whereas during good weather these SIDs are preserved for 2-engine heavy A/C only. The weather conditions could limit the options to turn the A/C in order to reestablish a radar separation quickly. (In poor visibility situations M,H,L departures are used to avoid having to check 2 taboo zones.) This situation is further visualized in the respective row. The green A/C illustrates the heavy 2 engine departure, and the red A/C illustrates the arrival and their projected movement. Note that this illustration is not to scale and just drawn for better a understanding of the situation. The actions taken during this scenario are listed in Table 6. Furthermore, the sequence diagram illustrated in Figure 11 illustrates one sequence in which the actions take place. Note that in the sequence diagram, only the coordination of Southern Runway Tower Controller with Center Runway Tower Controller is shown, thus indicating the 'minimum complexity' of the situation, when Northern Runway Tower Controller and Western Runway Tower Controller do not have interfering traffic.		
Involved Decision-making:	Southern Runway Tower Controller has to coordinate possible turns of the A/C on a Missed Approach with all relevant controllers. Therefore, the other controllers immediately inform Southern Runway Tower Controller what they intend to do with their traffic, after being informed about the Missed Approach on runway 25L.		
Effect on ATCO / ATM / Flight Crew:	A coordination of all tower controllers is necessary, resulting in added workload to generate a solution. The relevant departure controller might be involved as well. Also, the workload of the Flight Crew increases, if an A/C has to perform an unbriefed Missed Approach Procedure. During good weather conditions and/or if separation minima are guaranteed, a standard missed approach procedure is still possible.		

Visualization:



The described actions in the basic flow, listed in Table 6 are not strictly ordered and do also vary. This table just provides one example the actions could be ordered, and a prediction tool is expected to be supportive to the users. The ATCO could also spot the Missed Approach before the pilot communicates it. Also the Missed Approach could be initiated even before the Take-Off clearance is issued, but in this case, the situation wouldn't develop as complex as described in the scenario.

Table 6: Basic Flow Table for Scenario 1 at Airport 1

Basic Flow	
Step	**Actions**
1	Center Runway Tower Controller issues Take-Off clearance for a FOXTROT/GOLF (NORTH) SID departures; taboo zone 25 L not applicable, taboo zone 25R applicable (radar separation with small and medium A/C, additionally wake turbulence separation for heavy departures).
2	Southern Runway Tower Controller issues landing (LDG) clearance
3	Arriving Flight Crew initiates Go-Around
4	Arriving Flight Crew informs Southern Runway Tower Controller about Go-Around,
5	Southern Runway Tower Controller activates the Missed Approach buzzer in the tower (bad weather/thunderstorm: these conditions are leading to more Missed Approach, coordination might become more complex)
6	Center Runway Tower Controller has to react if departure will be airborne shortly → notify Southern Runway Tower Controller (same for Tower Controller 18)
7	Coordinate handling of all conflicting AC
8	Southern Runway Tower Controller deviates from Missed Approach Procedure to establish radar separation asap (e.g. turn Missed Approach left after passing MVA 2000ft → coordination might become necessary)
9	After situation is cleared, coordinate with adjacent sectors

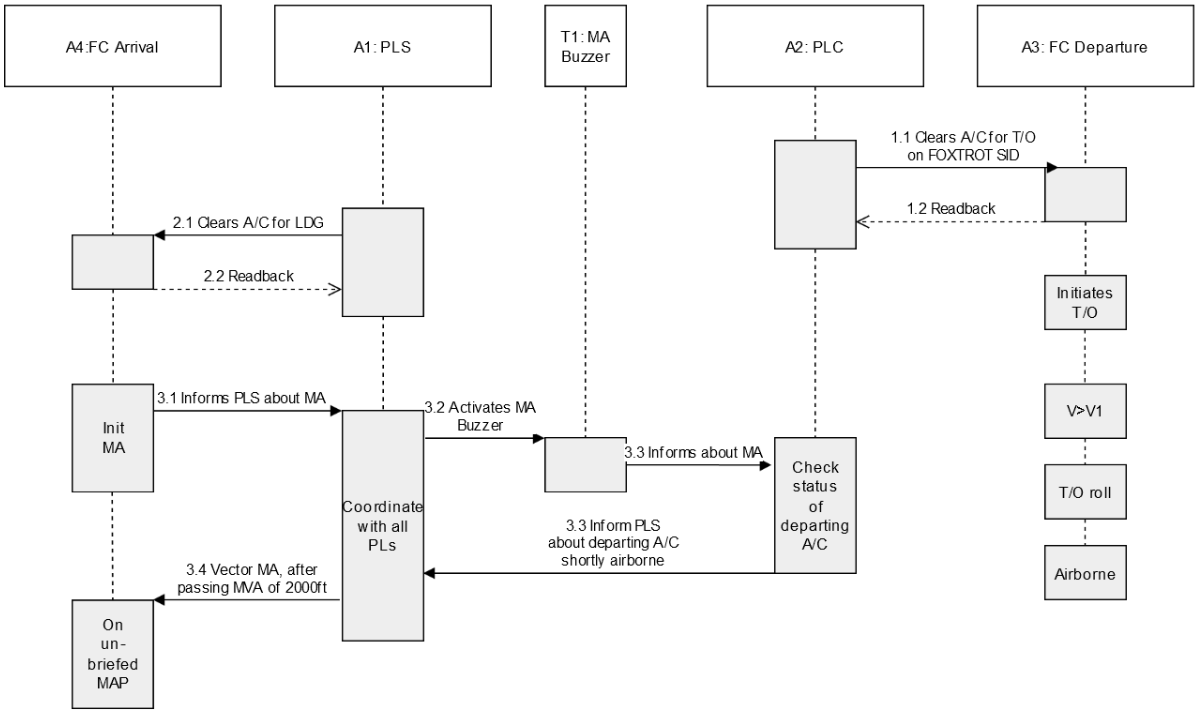


Figure 11: Sequence Diagram for Scenario 1 at Airport 1

3.2.4 Airport 1 Scenario 2

In the second scenario, the departing A/C is lined up on RWY 25C and will follow the FOXTROT/GOLF SID, which was illustrated in Figure 9 in Airport 1 Scenario 1 already, starting with a straight climb to 2NM DME FRD. This departure is used only for two engine, heavy A/Cs in good weather conditions but is open for any other A/C under certain circumstances, e.g. due to weather. In this scenario, an A/C is arriving on RWY 25R. The arrival chart for runway 25R is shown in Figure 12. The standard Missed Approach Procedure on runway 25R starts with a straight climb to 3.5 NM DME FRD before turning right.

Table 7 describes the traffic context, involved actors, involved decision making and the effect on the actors involved. In this scenario, the taboo zone R procedure, defined in section 3.2.1 is applicable. Nevertheless, the longitudinal offset of RWY 25 R compared to RWY 25 C may lead to a traffic constellation where a departure on RWY 25C and a late Missed Approach on RWY 25R have approximately the same altitude.

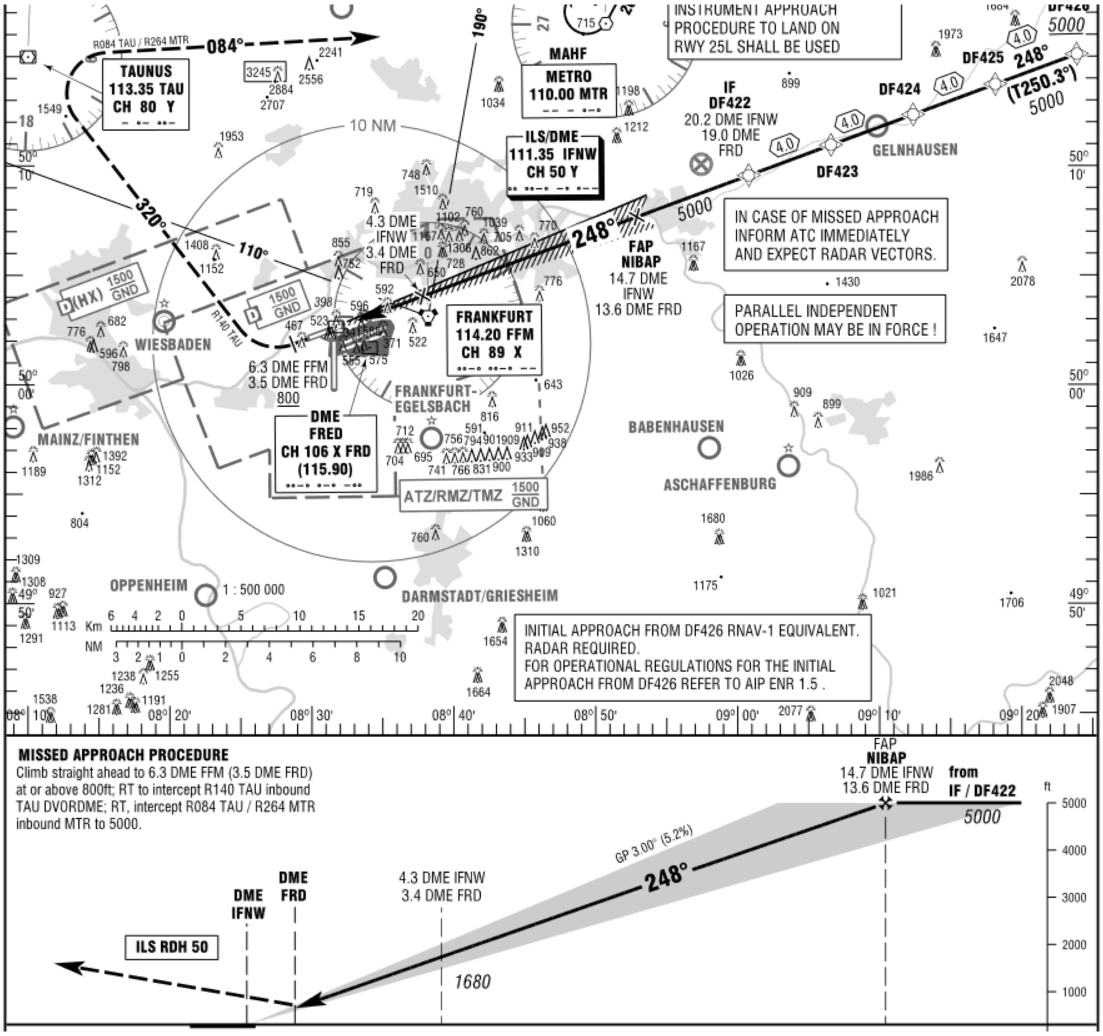


Figure 12: Standard Arrival Route for Scenario 1 at Airport 1 [18]

Table 7: Scenario 2 at Airport 1

Scenario ID:	Scen.Airport 1.2		
Scenario Name:	Dep 25C - Missed Approach Procedure 25R	**Version No:**	1.0
Linked Use Cases:	Airport 1 Use Case 2		
Involved Actors:	Center Runway Tower Controller (PLCS if combined) & Northern Runway Tower Controller, Flight Crew Departure, Flight Crew Arrival		
Description of Traffic Context:	Departing heavy two engine A/C from RWY 25C turns right at 2 NM FRD following 25C/L FOXTROT and GOLF (NORTH) SID. These SIDs are available for all A/C during marginal weather situations also, therefore increased distances between two arrivals would be requested by the tower controller. If an arriving A/C initiates Missed Approach just over THR, the departing A/C is approximately overflying end of RWY 25C. Due to the higher performance of the A/C performing the Missed Approach, both A/C are almost parallel to each other with crossing flight paths. An A/C flying the standard Missed Approach Procedure turns right at 3.5NM FRD, while the departing A/C turns right at 2 NM FRD. If the departing A/C has a higher wake turbulence category (WTC) than the A/C flying the Missed Approach Procedure (or both A/C have WTC heavy), potential WTC separation issues can arise additionally. A list of actions taken by all actors is presented in Table 8. Additionally, the sequence diagram in Figure 13 illustrates an exemplary flow of actions during this scenario.		
Involved Decision-making:	If a late Missed Approach on 25R is initiated between RWY and 1,5 NM from THR (so outside the taboo zone), and the departure from RWY 25C cannot be stopped, radar/WTC separation must be established and resolving action is required. These could be for example (more options exist): • restrict departing A/C in initial climb and demand departing A/C to climb straight ahead on runway track and providing traffic information • vector Missed Approach on an early right turn (the Taunus mountain range lies in this direction, so the A/C climb performance should be considered when choosing this resolution strategy) • a combination of 1. and 2.		
Effect on ATCO / ATM / Flight Crew:	Coordination necessary between Center Runway Tower Controller and Northern Runway Tower Controller. The workload of the Flight Crew of departure increases, due to unbriefed departure procedure and close traffic on the parallel RWY. (In this case it is more likely that departure is kept on RWY heading, rather than turning missed approach due to mountainous terrain)		

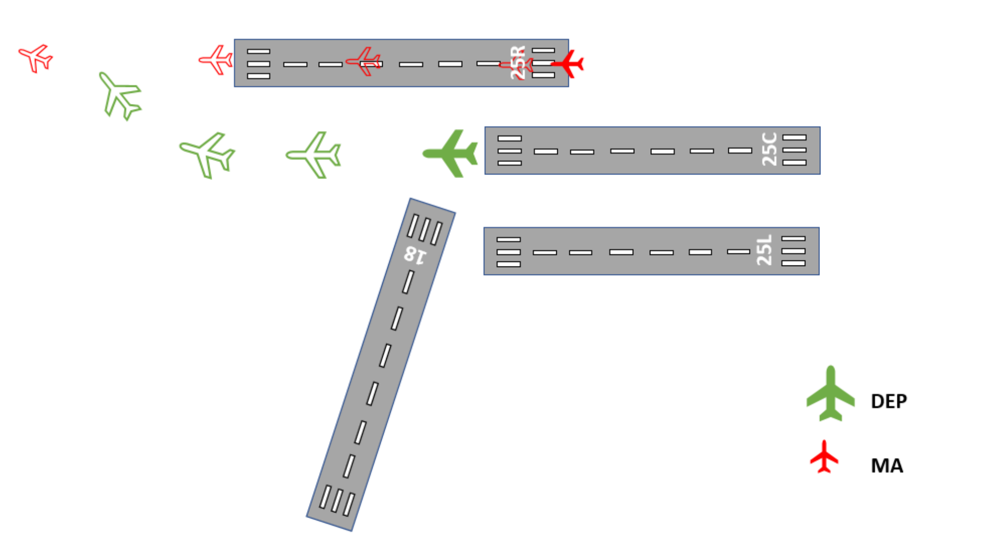
Visualization:	

Table 8: Basic Flow Table for Scenario 2 at Airport 1

Basic Flow Wake Turbulence Context	
Step	**Actions**
1	Center Runway Tower Controller checks taboo zone 25R is clear of arriving traffic
2	Center Runway Tower Controller issue Take-Off clearance for A/C on 25C
3	Northern Runway Tower Controller issues LDG clearance
4	Arriving Flight Crew initiates Go-Around
5	Arriving Flight Crew informs Northern Runway Tower Controller about Go-Around
6	Northern Runway Tower Controller activates Missed Approach buzzer
7	Coordination between Northern Runway Tower Controller and Center Runway Tower Controller
8	Center Runway Tower Controller advises departure to climb on RWY heading, mostly Go-Around cleared for standard missed approach
9	Northern Runway Tower Controller deviates from Missed Approach Procedure with earlier RT to establish (WTC) separation asap

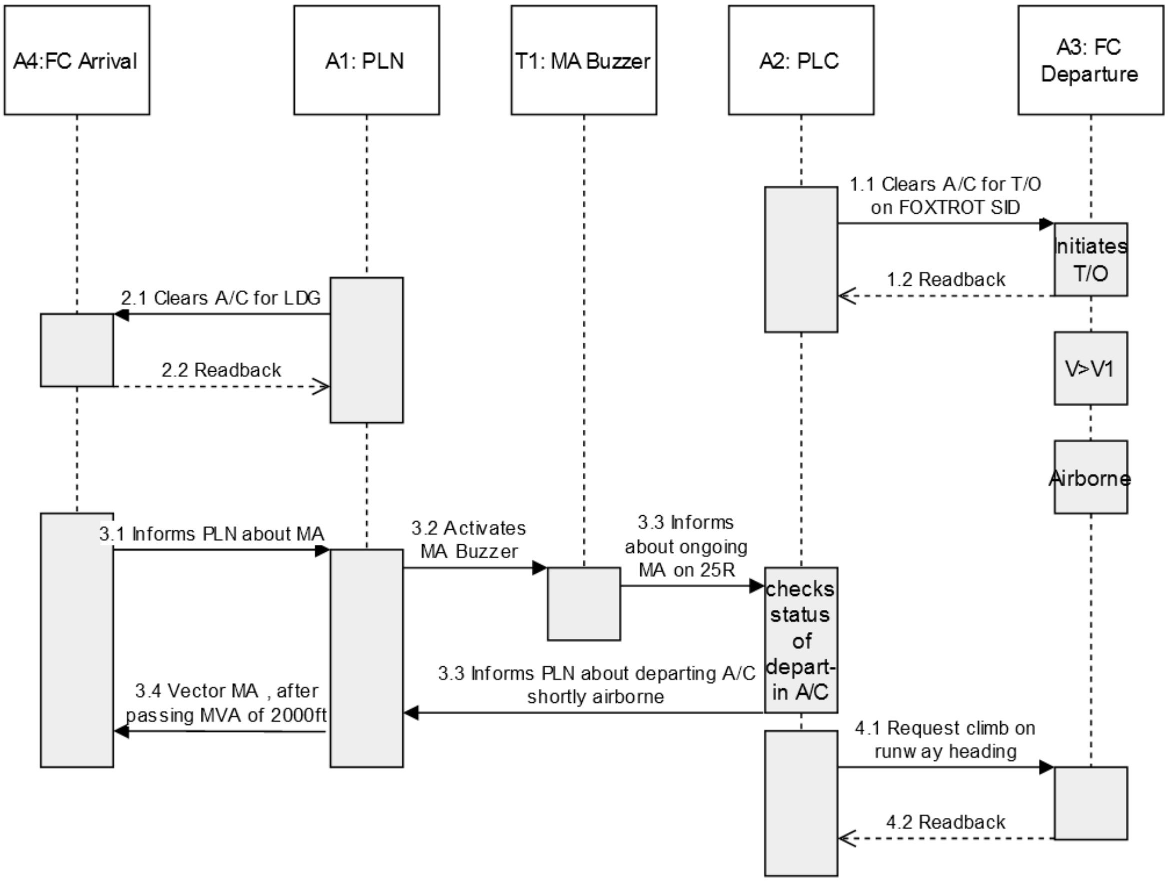


Figure 13: Sequence Diagram for Scenario 2 at Airport

3.2.5 Airport 1 Scenario 3

In this scenario, the swing to depart maneuver, as defined in section 3.2.1 is investigated. Here, the approach charts are not presented as the approach is performed visually and a Missed Approach Procedure is presented by the ATCO when the Swing to Depart is accepted by the Flight Crew. The prerequisite is, that all involved actors agree to perform the swing to depart maneuver. All involved actors could initiate this maneuver, but for clarity reasons, only one example how to initiate a swing to depart is depicted.

Table 9: Scenario 3 at Airport 1

Scenario ID:	Scen.Airport 1.3		
Scenario Name:	Airport 1: Swing to depart	**Version No:**	1.0
Linked Use Cases:	Airport 1 Use Case 3		
Involved Actors:	Southern Runway Tower Controller & Center Runway Tower Controller (or PLCS if one ATCO works both runways), maybe also Northern Runway Tower Controller & Western Runway Tower Controller, Flight Crew Departure, Flight Crew Arrival		
Description of Traffic Context:	To allow a departure from 25C, it is possible to swing a landing A/C from runway 25L to runway 25C. Thereby, the taboo zone procedure of RWY25L can be avoided and departures can be cleared for Take-Off anyway even when the LDG is parallel to the taboo zone, but now approaching runway 25C. Therefore, Southern Runway Tower Controller asks, if the landing Flight Crew agrees on the swing-over and advises a new Missed Approach Procedure and a clearance for a visual approach. Flight Crew usually agree, as with the swing-over comes a shorter taxi time to the stand after landing. As soon as the landing Flight Crew acknowledges the procedure, Center Runway Tower Controller may issue a Take-Off clearance to the departing traffic on 25C. Southern Runway Tower Controller meanwhile sends arriving traffic on to Center Runway Tower Controller's frequency. A new approach clearance at this point of the flight adds additional stress to the CC. The previous briefed checklists have to be adjusted and an established A/C has to be reconfigured for the other RWY. Though, the likelihood of a missed approach increases due to such last minute changes. If the ATCO receives an indication of a possible missed approach for that flight, the controller might not pursue his plan to offer a swing-over, as this would further increase the chance of a missed approach due to the mentioned workload.		
Involved Decision-making:	Center Runway Tower Controller has to consider relevant traffic on RWY 25R and RWY 18 if the Missed Approach Procedure for the visual approach (e.g.: fly straight ahead, climb 5000 ft) doesn't fit anymore. Traffic on 25R, which is also on a Missed Approach and departing traffic on RWY18, restricting possible left turns, would be the worst case. The just departed A/C can be turned slightly to the left to enable the Missed Approach on 25C to climb straight ahead. This only illustrates one of the possible actions to be taken by the actors.		
Effect on ATCO / ATM	The Flight Crew has to deal with additional workload, as the swing-over already created additional workload on them on final approach.		

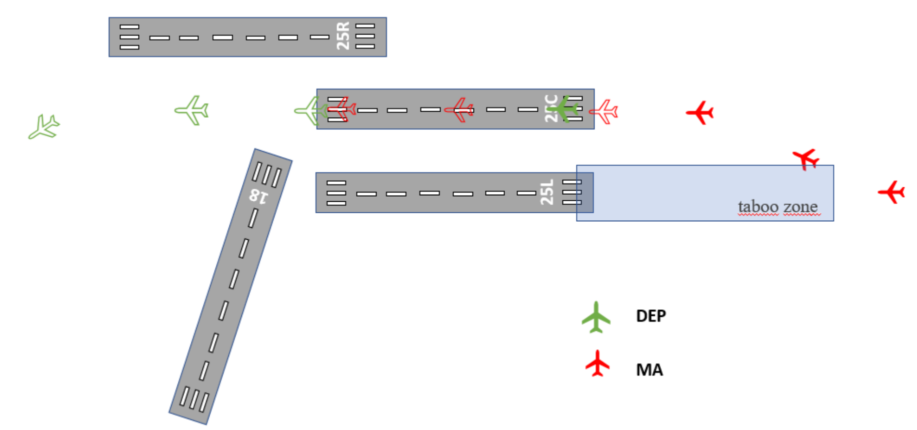
/ Flight Crew:	
Visualization:	

Table 10: Basic Flow for Scenario 3 at Airport 1

Basic Flow	
Step	**Actions**
1	Center Runway Tower Controller asks Southern Runway Tower Controller whether the landing could accept the swing to depart procedure
2	Southern Runway Tower Controller informs landing A/C of the request to swing for a visual approach on RWY 25C
3	Flight Crew accepts procedure
4	Southern Runway Tower Controller clears A/C for visual approach and new Missed Approach Procedure (e.g. straight ahead, 5000ft)
5	Center Runway Tower Controller clears departure on RWY25C for T/O
6	landing Flight Crew calls Center Runway Tower Controller to be on visual approach
7	Flight Crew initiates Missed Approach
8	Flight Crew informs Center Runway Tower Controller about Missed Approach
9	Center Runway Tower Controller pushes the missed approach buzzer
10	a) in regard of other traffic, Center Runway Tower Controller turns the departed a/c for example to the left b) turn the A/C on Missed Approach

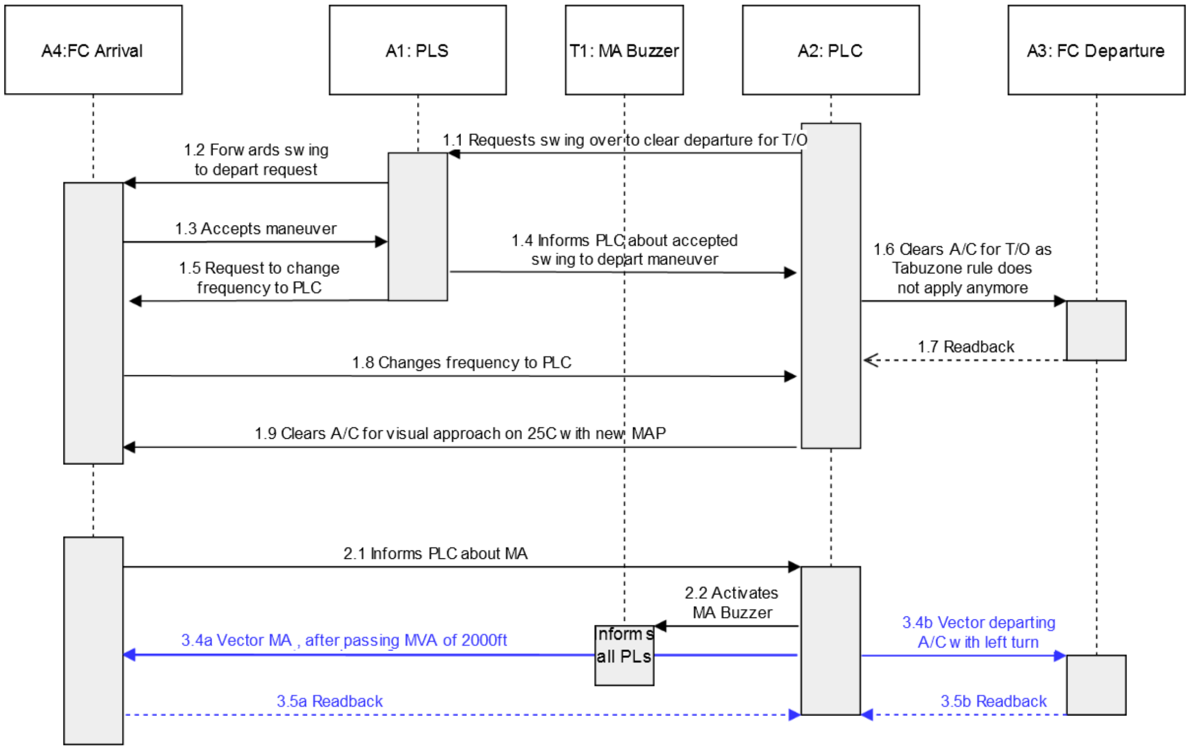


Figure 14: Sequence Diagram for Scenario 3 at Airport 1

3.2.6 Airport 2 Scenario 1

In the following, the first Go-Around scenario in Airport 2 is defined. The scenario is described for the RWY 26L but applies to all four possible runways (08L/08R/26L/26R), as the Missed Approach Procedure are defined similar. (Missed Approaches on 08L and 26R target MIQ NDB whereas 08R and 26L target OTT DVOR after turning from runway heading). The excerpt of the Instrument Approach Chart - ICAO ILS CAT II & III or LOC RWY 26L in Figure 15 shows the standard Missed Approach Procedure for runway 26L:

- Climb straight ahead to 1.0 DME West of DMS or 1900, whichever is later
- Left turn, direct to OTT DVOR/DME climbing 5000

This however will mostly not be used by an A/C performing a Missed Approach in HIRO operation with reduced runway separation, as described in the following scenario.

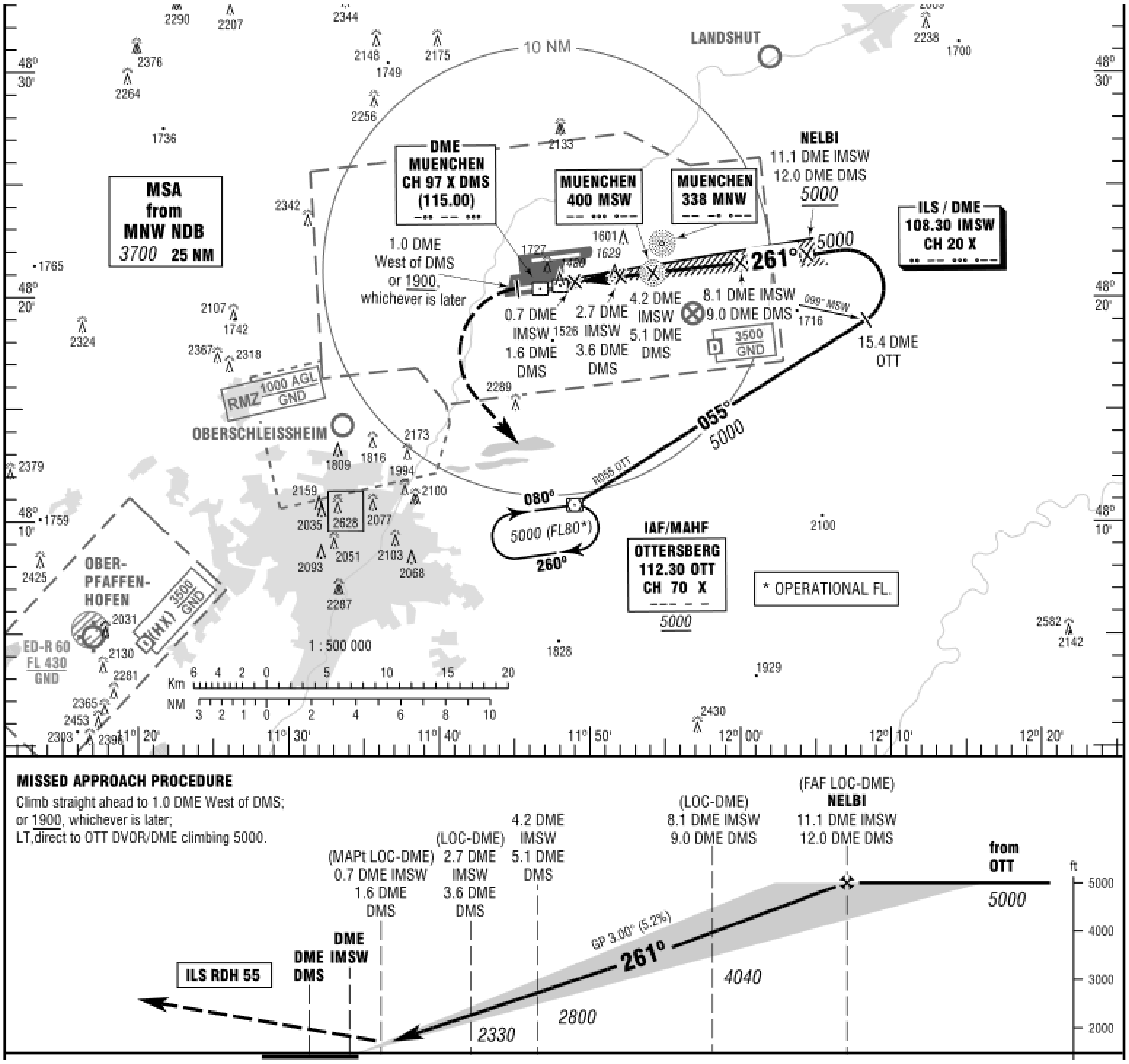
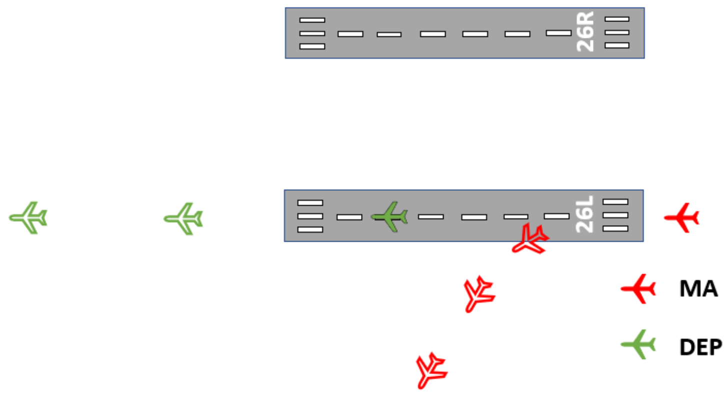


Figure 15: Instrument Approach Chart for Scenario 1 at Airport 2 [18]

Table 11: Scenario 1 at Airport 2

Scenario ID:	Scen.Airport 2.1		
Scenario Name:	Re-establish Radar Separation in HIRO Operation	**Version No:**	1.0
Linked Use Cases:	Airport 2.UC.01		
Involved Actors:	Southern Runway Tower Controller, Flight Crew Departure, Flight Crew Arrival		
Description of Traffic Context:	In case of a Missed Approach during HIRO operation, radar separation is immediately infringed. Due to high performance, the Missed Approach A/C quickly catches up the Departure A/C. Even though separation is still given by visual separation from the PL, it is desired to re-establish radar separation as soon as possible.		
Involved Decision-making:	To re-establish radar separation, the ATCO has to decide whether the Missed Approach Procedure or the departing A/C must be vectored.		
Effect on ATCO / ATM / Flight Crew:	The ATCO has to establish radar separation between the Missed Approach and departing A/C. Therefore, one possible solution is to turn Missed Approach immediately to the left, as depicted in the visualization below. Thereby, the workload of the Flight Crew increases, due to an unbriefed Missed Approach Procedure. Furthermore, Missed Approach has to turn below MVA (Minimum vectoring altitude) and below MSA (Minimum Sector Altitude). This sequence is illustrated in Figure 16. An alternative solution is to request the departing A/C to follow an alternative departure (straight ahead), so the A/C on the Missed Approach can follow the briefed standard Missed Approach Procedure. This option is illustrated in Figure 17.		
Visualization:			

The following Table 12 gives an overview over the actions performed by the different actors in the described scenario. The last action is split up into version a and b, reflecting two possible options the ATCO has to resolve the situation.

Table 12: Basic Flow for Scenario 1 at Airport 2

Basic Flow		
Step	**Actions**	
1	Southern Runway Tower Controller lines up A/C on RWY 26L via R/T	
2	Southern Runway Tower Controller clears departure for TO and marks it in the Tower Flight Data Processing System (TFDPS)	
3	Departure A/C overflown runway end or reduced RWY separation exists	
4	Southern Runway Tower Controller clears Arriving A/C to land	
5	Flight Crew of arriving A/C starts Missed Approach Procedure	
6	Flight Crew Arrival informs Tower Controller about Missed Approach	
7 a|b	Tower Controller vectors Flight Crew Arrival on non-standard Missed ApproachP	Tower Controller vectors Flight Crew Dep on not briefed departure route (climb straight ahead)

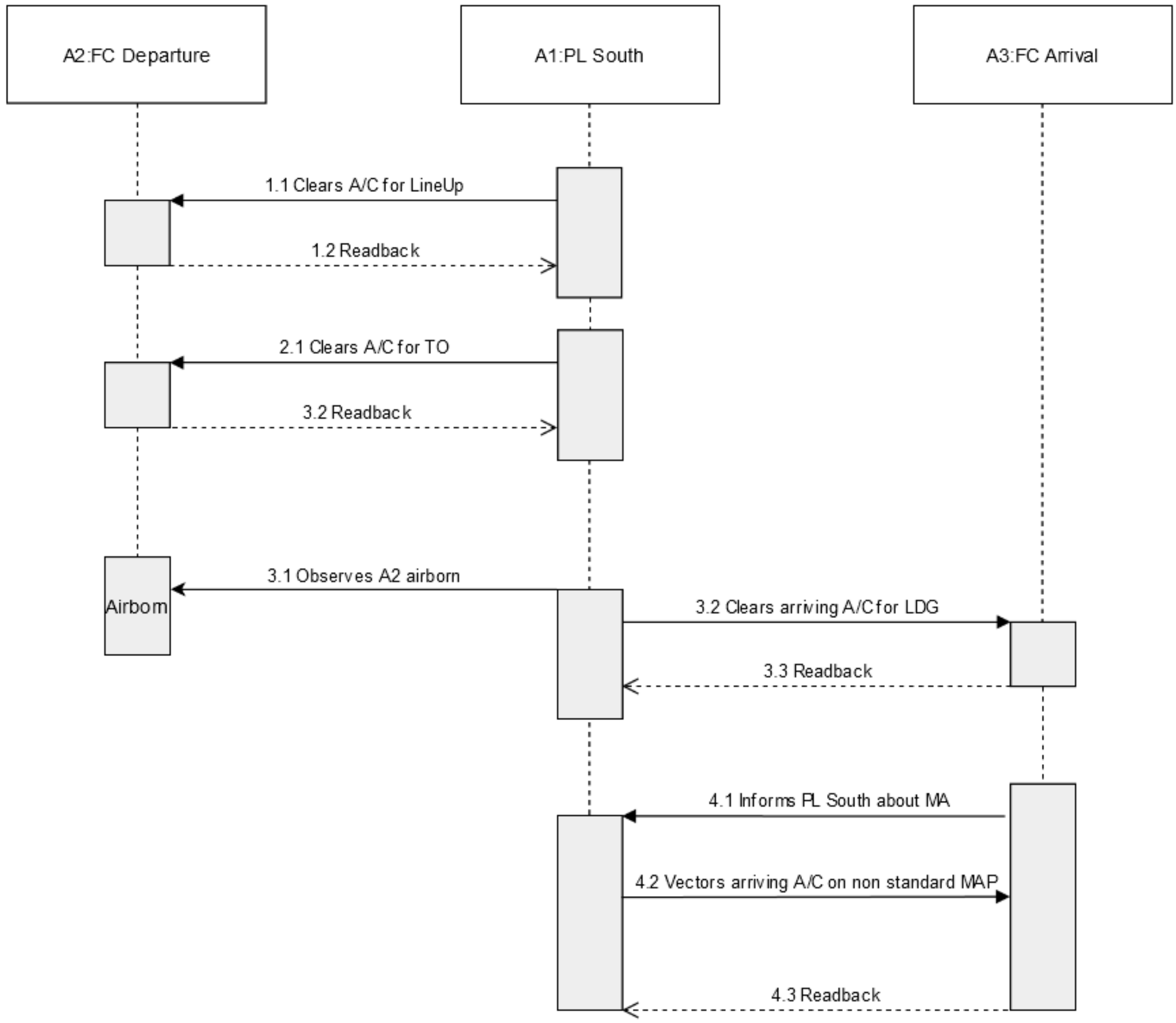


Figure 16: Sequence Diagram 1 for Scenario 1 at Airport 2

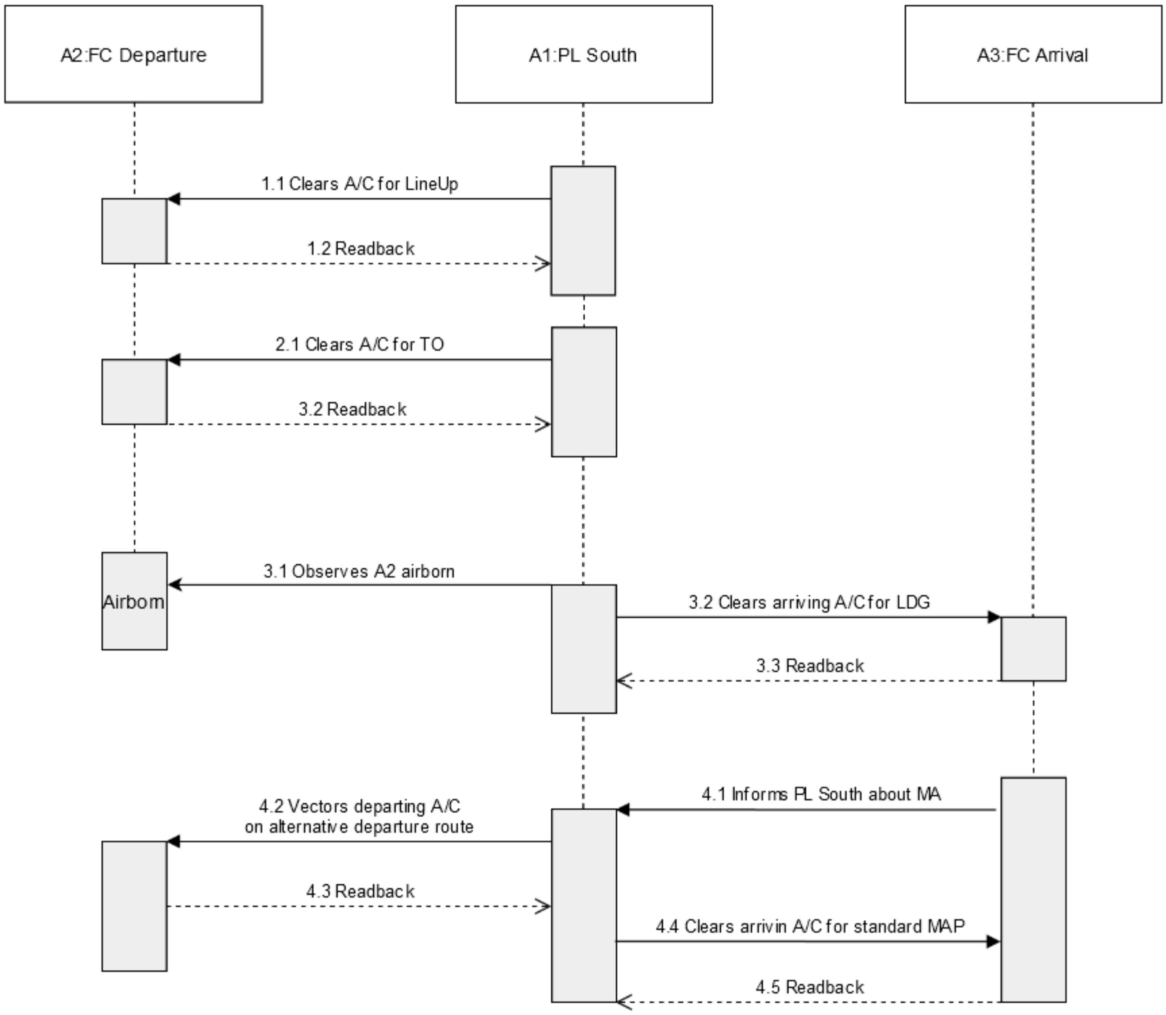


Figure 17: Sequence Diagram 2 for Scenario 1 at Airport 2

3.2.7 Airport 2 Scenario 2

This section describes the second scenario in Airport 2. The focus of this scenario is a potential missed approach (described in Figure 15 in scenario 1) with a preceding departure on the S-SID, as illustrated in the following SID chart for runway 26L. Furthermore, Departures on the W-SID, flying over OTT DVOR/DME will be regarded, as the standard Missed Approach Procedure as defined in Figure 15 also targets OTT.

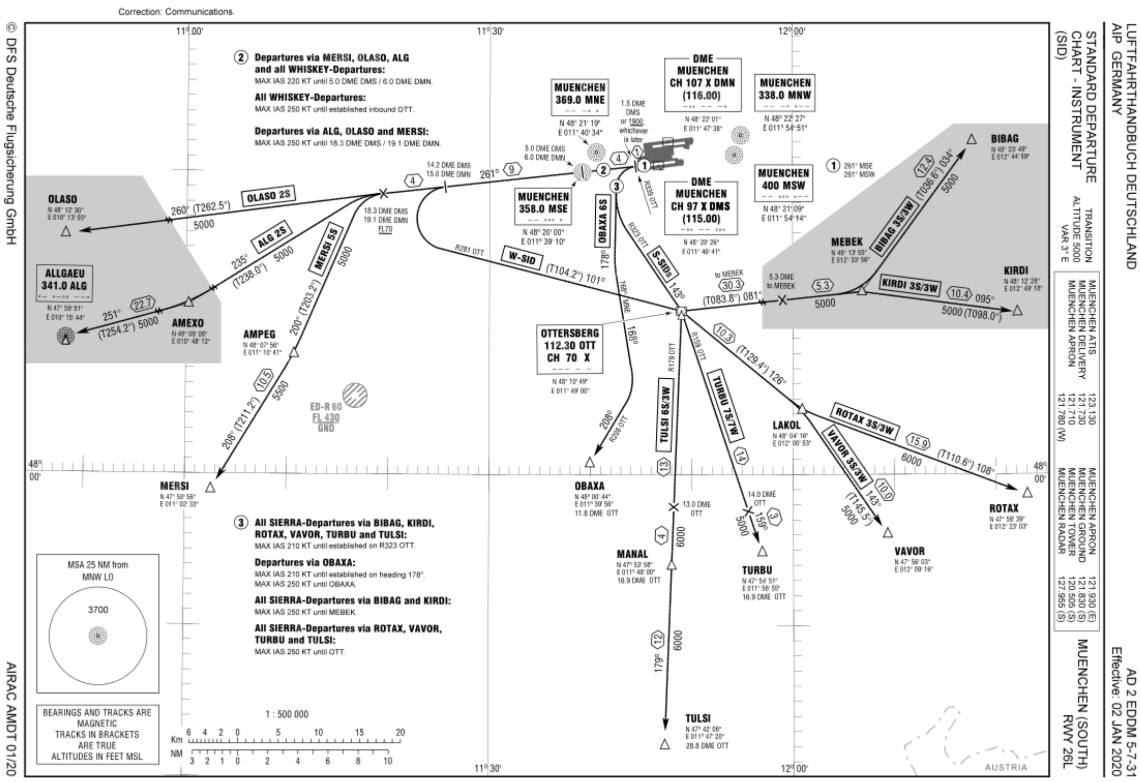
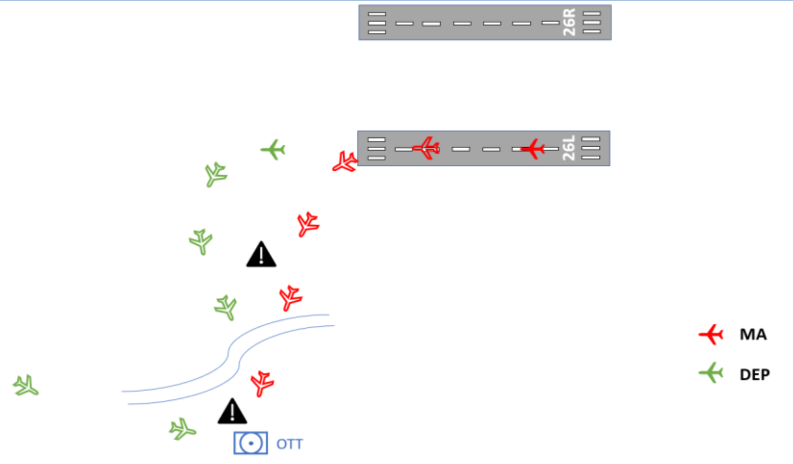


Figure 18: Standard Instrument Departures (South) for Airport 2 [18]

Table 13: Scenario 2 at Airport 2

Scenario ID:	Scen.Airport 2.2		
Scenario Name:	Airport 2 S-SID 26L conflicting with Missed ApproachP	**Version No:**	1.0
Linked Use Cases:	Airport 2 Use Case 2, Airport 2 Use Case 3		
Involved Actors:	Southern Runway Tower Controller, Flight Crew Departure, Flight Crew Arrival		
Description of Traffic Context:	The standard Missed Approach Procedure and the S-SID of RWY 26L are turning to the south early after having overflown the runway end. Therefore, a Go-Around and a preceding departure A/C with a route along S-SID could become a separation conflict, if no action by the ATCO would be taken.		

	Additionally, another crossing point exists overhead OTT DME/VOR by procedure. So a slow climbing dep A/C could be of relevance to an A/C on a Missed Approach, even if the departure departed minutes ago. The different performances may generate a conflict around OTT.
Involved Decision-making:	To clear the traffic situation, the ATCO has to decide how to separate the A/C. Therefore, the ATCO takes one of two options: • to deviate from the SID and to instruct the departing A/C to continue on runway heading • to deviate from Missed Approach Procedure and to turn arriving A/C further away from the departure.
Effect on ATCO / ATM / Flight Crew:	ATCO has to establish vertical separation between Missed Approach and departure A/C. Due to workload in Cockpit, instructions for turns might be read back with delay.
Visualization:	

In the following Table 14, the actors' actions are defined step by step. Furthermore, Figure 19 illustrates the actions taken in a sequence diagram.

Table 14: Basic Flow for Scenario 2 at Airport 2

Basic Flow		
Step	**Actions**	
1	Southern Runway Tower Controller clears Departure A/C (S-SID) for line-up and marks it in TFDPS	
2	Southern Runway Tower Controller clears S-SID Departure A/C for T/O	
3	Flight Crew S-SID Departure takes off	
4	Southern Runway Tower Controller clears Arriving A/C for LDG	
5	Arriving Flight Crew starts Missed Approach	
6	Arriving A/C Flight Crew informs Southern Runway Tower Controller about Missed Approach	
7	Arriving A/C performs standard Missed Approach	
8a|b	Southern Runway Tower Controller restricts Departure A/C to climb straight ahead	Southern Runway Tower Controller restricts Missed Approach A/C to lower Missed Approach altitude

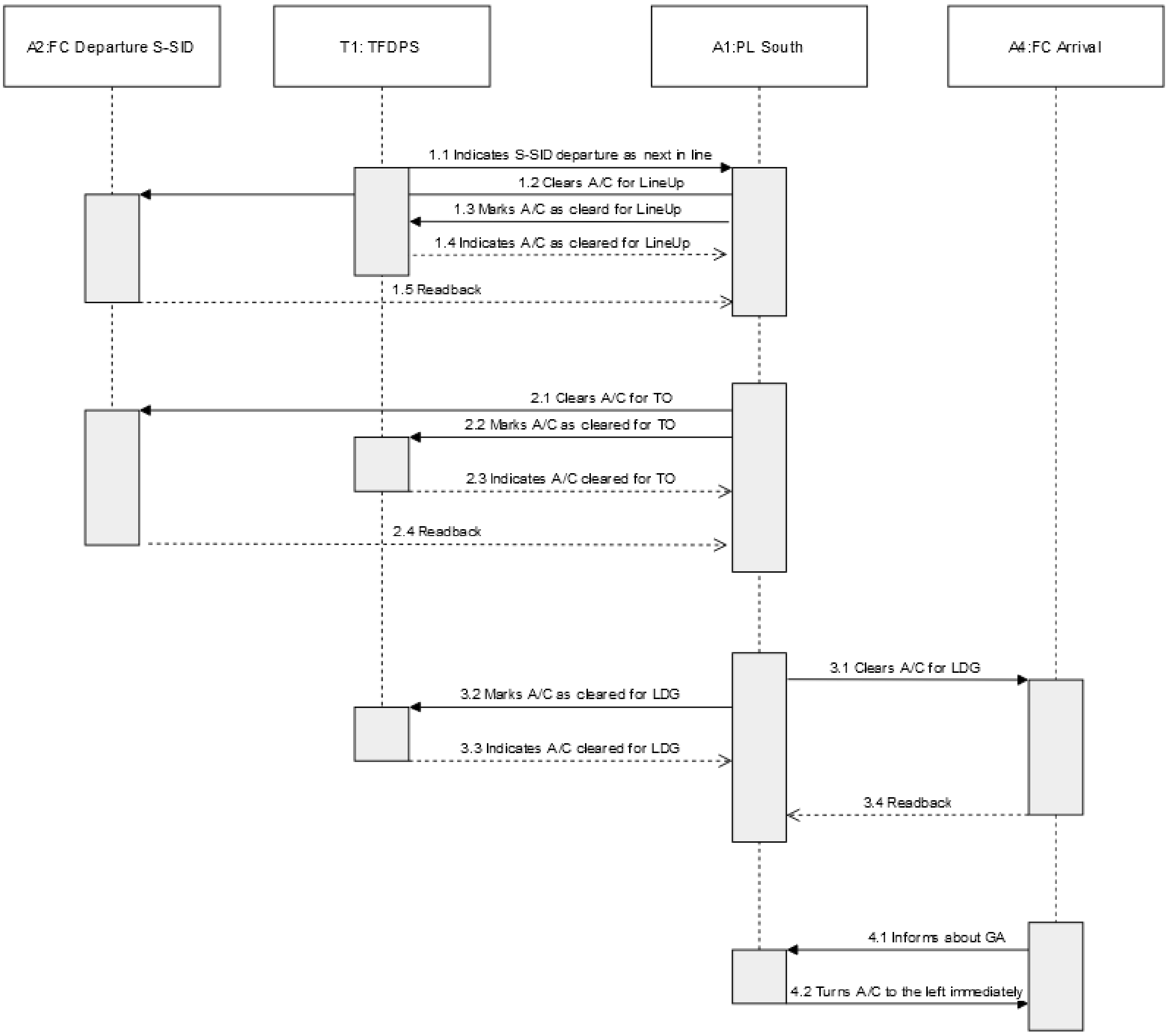


Figure 19: Sequence Diagram for Scenario 2 at Airport 2

3.2.8 Airport 2 Scenario 3

This section describes the third Go-Around scenario in Airport 2. It is exemplarily defined for runway 26L but applies to all runways similarly.

Table 15: Scenario 3 at Airport 2

<table>
<tr><td>Scenario ID:</td><td colspan="3">Scen.Airport 2.3</td></tr>
<tr><td>Scenario Name:</td><td>Establish Wake Turbulence Separation</td><td>Version No:</td><td>1.0</td></tr>
<tr><td>Linked Use Cases:</td><td colspan="3">Airport 2.UC.04</td></tr>
<tr><td>Involved Actors:</td><td colspan="3">Southern Runway Tower Controller, Northern Runway Tower Controller, Flight Crew Arrival</td></tr>
<tr><td>Description of Traffic Context:</td><td colspan="3">In case of Missed Approach (medium or light) after departed heavy/super A/C, wake turbulence separation must be established. Due to high performance, the Missed Approach A/C quickly catches up with the departure A/C. Although the A/C on the Missed Approach over-climbs the departure most probably, the ATCO aims to separate both A/C from each other as fast as possible, because a wake turbulence challenge can arise, as the Missed Approach only climbs to a 5000ft altitude. As vertical separation does not solve the situation completely, because the departing A/C would have to be restricted in climb, the more practical solution is to turn the arriving A/C. For two reasons, mostly the Missed Approach is vectored:
due to the higher performance, the A/C is above the MVA faster
a turn back to the approach leg is understood as a service to shorten the distance flown for another approach.</td></tr>
<tr><td>Involved Decision-making:</td><td colspan="3">To establish wake turbulence separation, the ATCO has to decide how the Missed Approach can be performed.</td></tr>
<tr><td>Effect on ATCO / ATM / Flight Crew:</td><td colspan="3">ATCO has to establish wake turbulence separation between A/C. Therefore, a solution is to turn Missed Approach immediately to the left. The workload of the Flight Crew increases, due to unbriefed Missed Approach Procedure. Furthermore, Missed Approach has to turn below MVA (Minimum vectoring altitude) and below MSA (Minimum Sector Altitude)</td></tr>
</table>

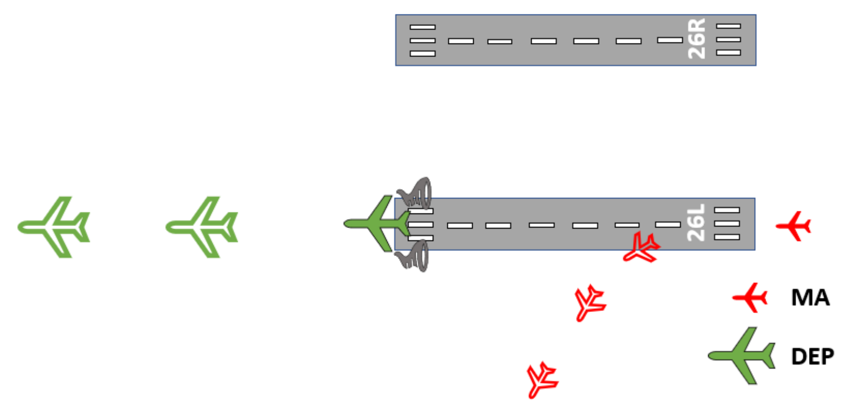
Visualization:	

Table 16: Basic Flow for Scenario 3 at Airport 2

Basic Flow	
Step	**Actions**
1	Southern Runway Tower Controller lines up heavy/super A/C on RWY via R/T and marks it in TFDPS
2	Southern Runway Tower Controller clears departure for Take-Off and marks it in TFDPS
3	Departure A/C is airborne
4	Southern Runway Tower Controller clears arriving A/C to land
5	Flight Crew arriving A/C starts Missed Approach
6	Flight Crew Arrival informs Tower Controller about Missed Approach
7	Southern Runway Tower Controller vectors Flight Crew Arrival on a non-standard Missed Approach Procedure

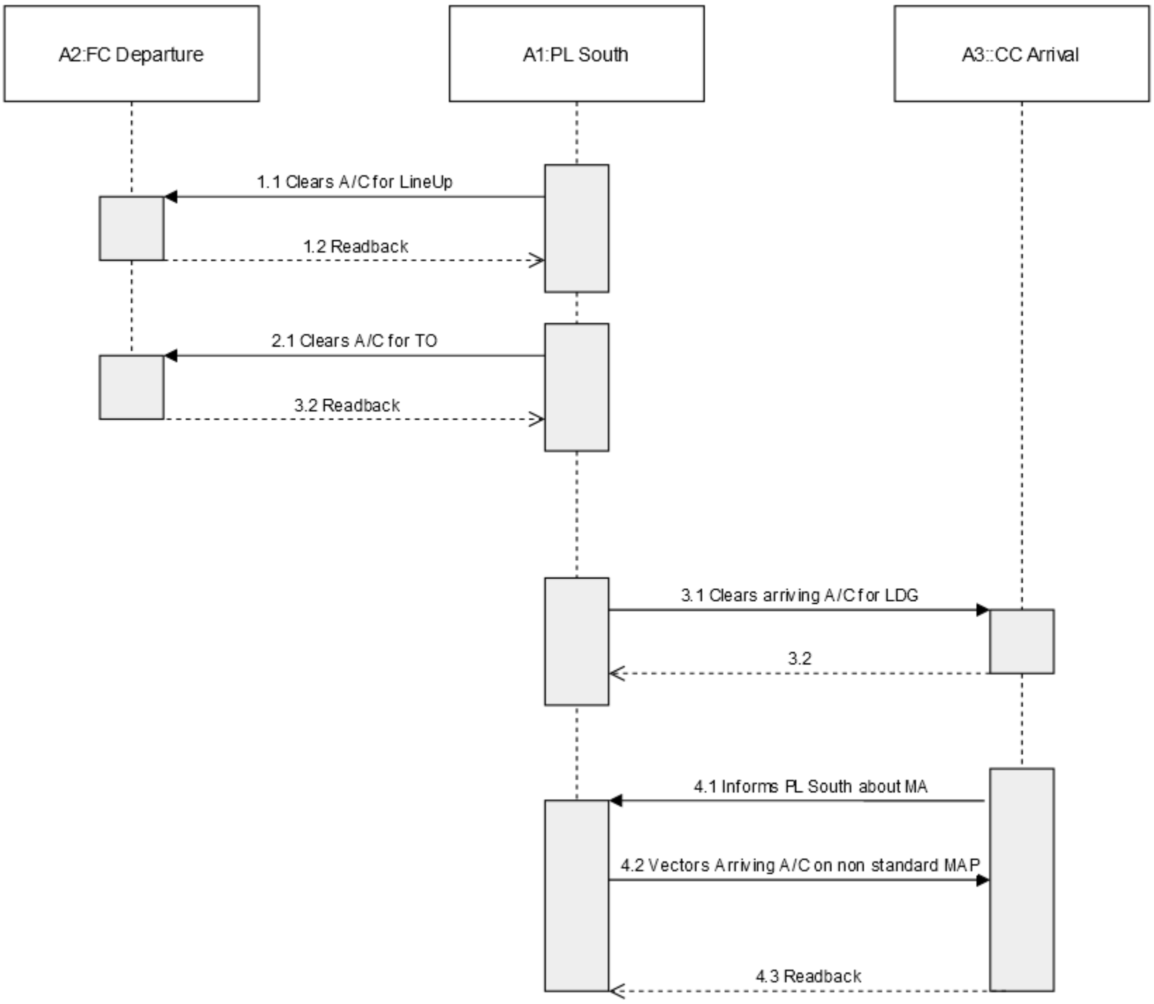


Figure 20: Sequence Diagram for Scenario 3 at Airport 2

3.3 SafeOPS – Use Cases

This section provides the use cases that have been elaborated in the workshops with the ATCOs from Airport 1 and Airport 2. The scenarios defined in section 3.2 serve as foundation for these considerations. The use cases (see section 2.2.2 for a definition) describe, how the handling of the scenarios would be influenced under the premise of time in advance indications that an arriving A/C tends to perform a missed approach. Use Cases are generated from discussions in workshops. These discussions are often controversial, similar to the ones on scenarios, and multiple opinions on benefits in Use Cases, exists. The documented versions of use cases in this section do not claim to be absolute but reflect an agreed summary of the workshop results.

In the discussions on use cases of a predictive decision support tool, the two classes:

- general use cases
- procedure specific use cases

have been separated. The general use cases can be applied AP - or RWY independent and focus mostly on an increase of situational awareness. The procedure specific use cases are tailored to procedures in place for specific APs and runways - in SafeOPS Airport 1 and Airport 2. It is in the nature of this division of use cases, that the general use cases are vaguer in their description and their benefits, than the procedure specific ones. This chapter starts by introducing the general use cases and proceeds with the specific use cases afterwards.

3.3.1 General Use Cases

In the following two subsections, general use cases are defined, that do not depend directly on one specific scenario. These use cases cover a broader, more high level description of a big data based decision support for ATCOs.

3.3.1.1 General Use Case 1

This section documents the first general use case, identified in SafeOPS. It applies to all scenarios and the traffic contexts described therein. Especially in scenarios where cross coordination between ATCOs is necessary (e.g. Airport 1 Scenario 1 and Airport 1 Scenario 2), a time in advance indication of a possible Go-Around could provide additional time for these coordinative actions. Thus, the general use cases do not change the actions as described in the scenarios, nor the way the traffic situation is handled. For Airport 1 Scenario 1, an exemplary version of the sequence diagram for this use case illustrates how necessary coordination actions can be triggered time ahead. Figure 21 is similar to the one from Airport 1 Scenario 1, only that the Go-Around prediction, illustrated additionally in this sequence diagram with the orange box, triggers the coordination actions (here also depicted in orange), instead of the announcement of the pilot, performing the missed approach.

Table 17: General Use Case 1

Use Case ID:	GenUC.01		
Use Case Name:	Go-Around prediction for arriving A/C	**Version No:**	1.0
Linked Scenarios:	Independent (exemplarily visualized for Airport 1 Scenario 1 in Figure 21)		
Linked User Stories:	all General User Stories		
Involved Actors:	PL		
Precondition:	none		
Brief Description:	An ATCO generally includes the possibility of Missed Approach into the traffic planning, by processing information, which from different data sources (Radar, ADS-B, Visual inspection, Mode S) In the case of a time in advance predicted, increased Go-Around likelihood, the controller can be better mentally prepared for a possible Go-Around and the potential arising conflicts. In the sequence illustrated in Figure 21, the coordinative actions can be performed earlier, compared to the sequence illustrated in Figure 11.		

Benefit for ATCO / ATM / Flight Crew:	• Earlier/better **situational awareness** due to provision of Missed Approach likelihood, based on arriving A/C data ○ prediction summarizes/displays information (also from A/C trajectory before Tower Controller observes it) • Improved foundation for decision making due to additional information triggered by prediction (see e.g. User Story GEN.US.14)

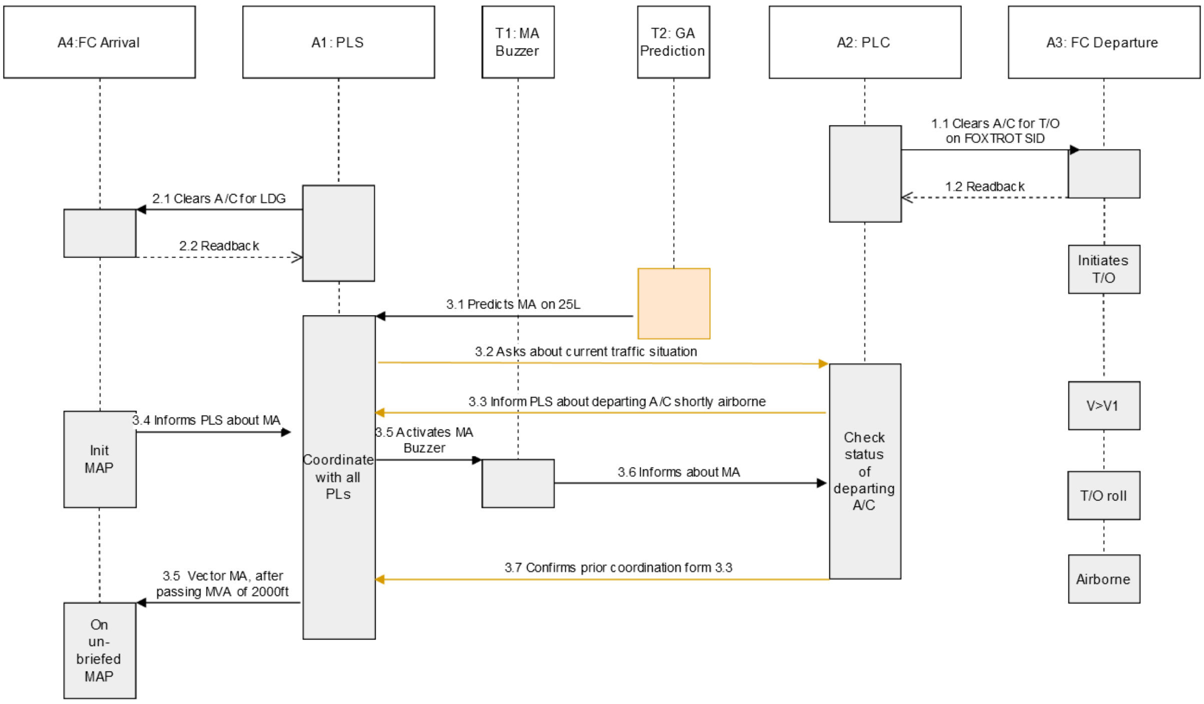


Figure 21: Sequence Diagram for General Use Case 1

3.3.1.2 General Use Case 2

This use case differentiates itself in its nature from the other described use cases. It rather discusses a 'live' indication of Go-Arounds, that are detected when happening instead of predicted time ahead.

Based on the discussions with the ATCOs, a functionality like this is available already, however with the downside that the detection has a serious delay between the initiation of the missed approach and its indication. Pilots also communicate the initiation of a missed approach to the ATCO. Their focus however, after deciding to Go-Around, lies primarily on the A/C. Therefore, the information of an initiated missed approach from the Pilot to the ATCO also underlies a delay.

The idea of a tool which indicates a missed approach to the ATCO with minimum (no noticeable) delay after it is initiated by the pilot was independently brought up by ATCOs during almost every workshop conducted in SafeOPS. We decided to document this use case because of the positive resonance during the workshops.

However, this use case is not completely in line with the SafeOPS project for several reasons:

- This use case is not based on a prediction but a detection of a situation. The subtitle of this project however is "from prediction to decision support".
- Therefore, no probabilistic information of a Go-Around prediction is provided to the ATCO but a deterministic indication of a factual Go-Around happening.

The functionality discussed within this use case would therefore provide a saving in time in-between the initiation of the missed approach and the ATCO taking notice of it. This is something more tangible than the idea of prediction-based decision support and therefore achieved great consent in the workshop with respect to its usefulness, in contrast to the other use cases which have been controversially discussed.

However, further work on this use case would partially deprive the project of the scientific basis, agreed upon in the Grant Agreement. If and how this use case will be pursued in the upcoming development phase is therefore subject to discussion.

Table 18: General Use Case 2

Use Case ID:	GenUC.02		
Use Case Name:	Go-Around detection for arriving A/C	**Version No:**	1.0
Linked Scenarios:	all		
Linked User Stories:	all		
Involved Actors:	PL		
Precondition:	none		
Brief Description:	ATCO generally includes the possibility of Missed Approach into the traffic planning. In the case of an immediate indication of an initiated Go-Around, the ATCO has a time gain over conventional detection of this, such as by observing the radar or visually tracking the flight path.		
Benefit for ATCO / ATM / Flight Crew:	• earlier / better **situational awareness** due to output of ongoing Missed Approach, based on arriving A/C data • earlier basis for decision and reaction time due to immediate information flow triggered by Missed Approach alert ○ time gains of 5 up to 40 seconds have been estimated by pilots and ATCOs during the workshops, if a Go-Around could be detected without delay		

3.3.2 Airport 1 Use Case 1

This section presents the first use case, specifically defined for Airport 1. The use case builds on a scenario defined in Airport 1, Airport 1 Scenario 1 , and describes how the scenario could change, if an early indication of a potential Go-Around could be provided to the ATCOs.

Table 19: Use Case 1 at Airport 1

<table>
<tr><td>Use Case ID:</td><td colspan="3">Airport 1.UC.01</td></tr>
<tr><td>Use Case Name:</td><td>25C northbound vs. Missed Approach Procedure 25L</td><td>Version No:</td><td>1.0</td></tr>
<tr><td>Linked Scenarios:</td><td colspan="3">Airport 1 Scenario 1</td></tr>
<tr><td>Linked User Stories:</td><td colspan="3">all GEN, Airport 1.US.01, Airport 1.US.03</td></tr>
<tr><td>Involved Actors:</td><td colspan="3">Center Runway Tower Controller, Southern Runway Tower Controller, Western Runway Tower Controller</td></tr>
<tr><td>Precondition:</td><td colspan="3">Relevant SID FOXTROT & GOLF (NORTH) is lined up on RWY 25C. For other SIDs, the taboo zone procedure RWY25L has to be applied and the context described in the linked scenario would not occur.</td></tr>
<tr><td>Brief Description:</td><td colspan="3">Airport 1 Scenario 1 describes a situation, in which re-establishing radar separation after a Missed Approach is complex. Coordination between Center Runway Tower Controller, Southern Runway Tower Controller and Western Runway Tower Controller is necessary to provide the best possible solution. The solutions are manifold and depend on the position of other traffic. The number of possible solutions decreases, if Western Runway Tower Controller also has a departure ongoing. Thus, a Missed Approach on 25L requires actions by the controllers involved to separate the A/C. Foremost during marginal weather conditions, actions to clear the situation during the Missed Approach are limited.

Given a time in advance, indication of a potential Missed Approach on RWY25L could impact the PLs in the following:
Southern Runway Tower Controller: the controller checks the traffic constellation and continuously observes it in regard of departures of RWY 18 & 25C
Center Runway Tower Controller: the north-bound departure has to comply with taboo zone on RWY25R, but additionally the controller could take into consideration a missed approach on 25L and the possible vectors and might conclude to delay the Take-Off clearance until the A/C on 25L has landed or initiated a Missed Approach (detailed actions are listed in Table 20 and illustrated in the sequence diagram Figure 22)
Western Runway Tower Controller: departures from RWY18 may restrict left turns for Missed Approach on RWY25L, therefore including a high chance of a Missed Approach into the Western Runway Tower Controller's traffic planning could assist to solve the situation faster when the Missed Approach occurs</td></tr>
<tr><td>Benefit for ATCO / ATM / Flight Crew:</td><td colspan="3">General: Gain time to foster situational awareness
Center Runway Tower Controller could ponder his decision to issue the Take-Off clearance at a suitable moment
Arrival AC can fly the briefed missed approach → safety benefit</td></tr>
</table>

	○ Departing AC might be delayed or a gap for a possible departure is skipped → potential capacity loss • Southern Runway Tower Controller' attention is attracted to check the other PL's traffic earlier

The following table exemplarily describes the actions, summarized in the use case description. It focuses only on the coordination between Southern Runway Tower Controller and Center Runway Tower Controller. The sequence of the actions is also illustrated in Figure 22. The top boxes represent either actors (A) or tools (T) in this use case. Each other box represents an action an actor or tool performs. The arrows indicate how the actions of different actors or tools trigger each other. In contrast to Figure 11, illustrating the sequence for Airport 1 Scenario 1, Figure 22 includes the additional Go-Around prediction tool. In the sequence, the prediction of a likely Go-Around is indicated by the orange box. The green boxes indicate the actions that change compared to the respective scenario, whereas the red transparent boxes indicate the actions that are removed from the sequence of actions, compared to the respective scenario. (Note that the identical color coding will be used throughout all use case sequence diagrams.)

Table 20: Flow for Use Case 1 at Airport 1

Alternate Flow (Compared to Scenario Airport 1 1)		
Step	**User Actions**	**System Actions**
1	Center Runway Tower Controller lines up a/c on RWY 25C	
2		if the Missed Approach-prediction for the next landing RWY 25L is greater than xx%, the tool issues a warning/information about it
3	Southern Runway Tower Controller clears arriving A/C to land	tool updates the prediction for the next landing a/c and displays it constantly
4	Center Runway Tower Controller decides not to clear departure for T/O	
5	Flight Crew initiates Missed Approach	
6	arriving Flight Crew informs ATC about MA	
7	Southern Runway Tower Controller activates Missed Approach buzzer	
8	only coordination between Western Runway Tower Controller and Southern Runway Tower Controller might be relevant, as Center Runway Tower Controller has no traffic airborne	

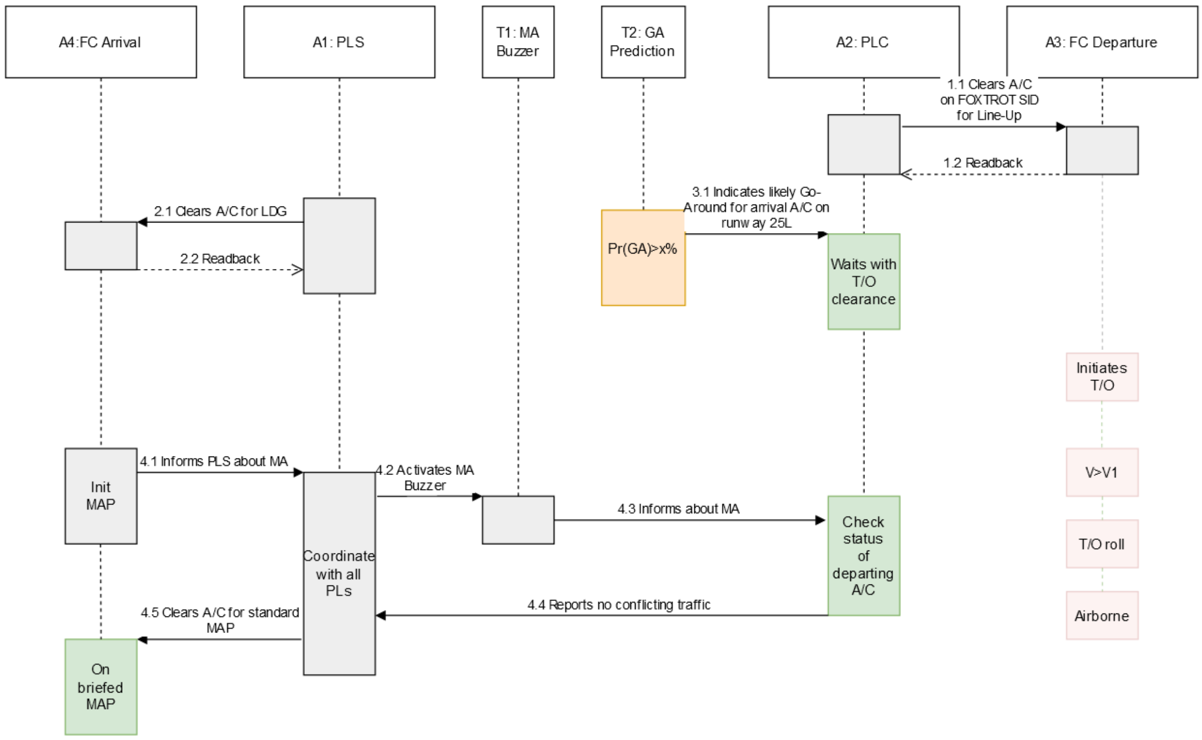


Figure 22: Sequence Diagram for Use Case 1 at Airport 1

3.3.3 Airport 1 Use Case 2

This section describes the second Airport 1 specific use case. It is aligned with the scenario described in Airport 1 Scenario 2. It focuses on how the scenario could be handled if Center Runway Tower Controller gets a time ahead indication of a possible Missed Approach on RWY 25R.

Table 21: Use Case 2 at Airport 1

Use Case ID:	Airport 1.UC.02		
Use Case Name:	25C northbound Departure vs. Missed Approach Procedure 25L	**Version No:**	1.0
Linked Scenarios:	Airport 1 Scenario 2		
Linked User Stories:	all GEN		
Involved Actors:	Center Runway Tower Controller, Northern Runway Tower Controller, Cabin Crew Arrival, Cabin Crew Departure		
Precondition:	Relevant SID FOXTROT (NORTH) is lined up on RWY 25C		
Brief Description:	The taboo zone procedure for runway 25R is relevant for this use case. It is defined in section 3.2.1. Therefore, the landing A/C on RWY 25R must have passed 1.5 NM from THR to issue a Take-Off clearance for FOXTROT and GOLF departures RWY25C. However, if the 25C departing A/C commences its Take-Off when the landing on 25R is at 1.5 NM final, a		

	late initiated Missed Approach just over the THR may result in both A/C in the air, parallel to each other and with converging flight paths.

Benefit for ATCO / ATM / Flight Crew:	In the situation described in Airport 1 Scenario 2, a time in advance indication of a Go-Around could have the following impact on the PLs:
	• **General:** Gain time to foster situational awareness
	• **Center Runway Tower Controller:** could ponder his decision to delay the Take-Off clearance (described in detail in Table 22 and visualized in Figure 23)
	  ○ coordinative actions with Northern Runway Tower Controller become less complex
	  ○ Flight Crew Arrival can perform briefed Missed Approach → safety benefit
	  ○ Flight Crew Departure does not have to diverge from SID → safety benefit
	  ○ Flight Crew Departure gets delayed and a gap for a departure might be skipped → capacity decrease
	• **Northern Runway Tower Controller's**: attention is attracted to check the other Center Runway Tower Controller's traffic earlier

The following table list an exemplary change in actions, triggered by a time ahead indication of a Go-Around in Airport 1 Scenario 2. In contrast thereon, Center Runway Tower Controller decides to not clear a departure for Take-Off in step 4. The consequence is illustrated in the sequence diagram in Figure 23. The orange box indicates the additional Go-Around prediction. The red boxes illustrate the actions from Airport 1 Scenario 2, which are **not** performed within this use case, viz the departing A/C does not get cleared for T/O. The green arrows indicate the changes compared to Airport 1 Scenario 2, which allows the arriving A/C to perform the standard missed approach.

Table 22: Flow for Use Case 2 at Airport 1

Alternate Flow (Compared to Scenario Airport 1 2)		
Step	**User Actions**	**System Actions**
1	Center Runway Tower Controller lines up A/C on RWY 25C	
2		if the Missed Approach-probability of the next landings RWY 25R is greater than xx%, the tool issues a warning/information about it
3	Northern Runway Tower Controller clears arriving A/C to land	
4	Center Runway Tower Controller decides not to clear departure for T/O	
5	arriving Flight Crew initiates Missed Approach on short final	
6	arriving Flight Crew informs ATC about Missed Approach	
7	only coordination but probably no immediate action necessary by PLs	
8	when departure sector is clear from the A/C on a Missed Approach, Center Runway Tower Controller issues Take-Off clearance to his FOXTROT and GOLF departure	

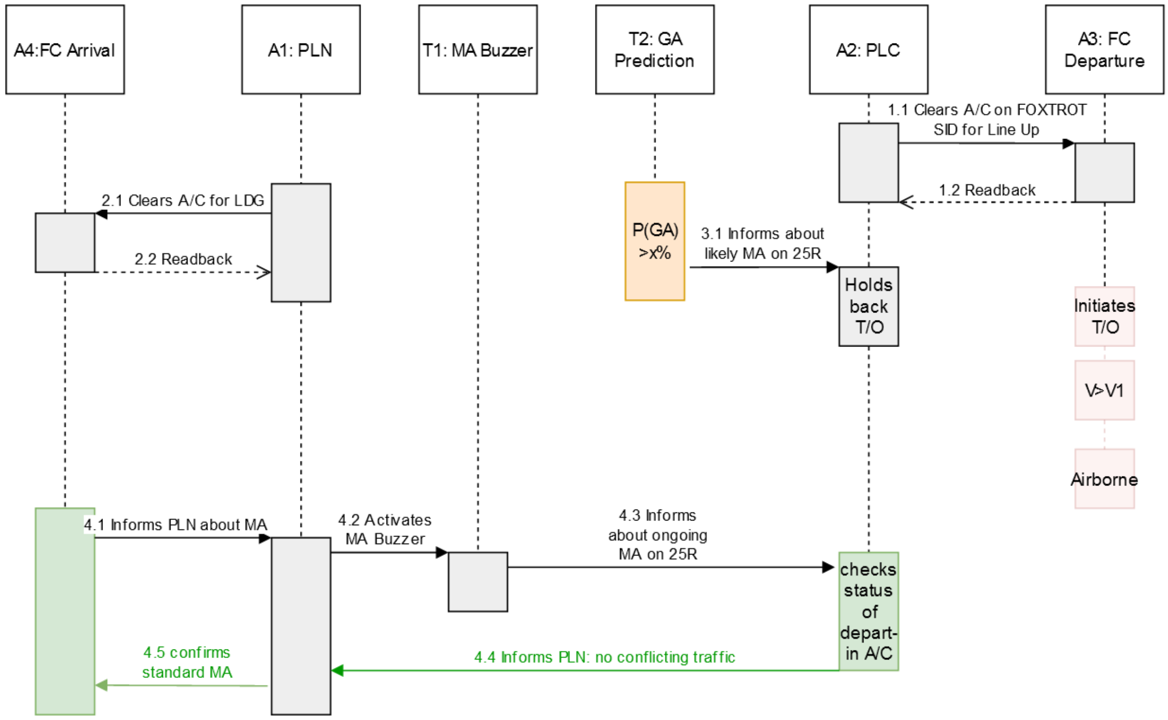


Figure 23: Sequence Diagram for Use Case 2 at Airport 1

3.3.4 Airport 1 Use Case 3

This section documents the third specific use cased identified in Airport 1. It elaborates how a predicted Go-Around could influence the decision of a swing to depart maneuver, as defined in section 3.2.1.

Table 23: Use Case 3 at Airport 1

Use Case ID:	Airport 1.UC.03		
Use Case Name:	Swing-to-depart	**Version No:**	1.0
Linked Scenarios:	Airport 1 Scenario 3		
Linked User Stories:	all GEN, Airport 1.US.02		
Involved Actors:	Center Runway Tower Controller, Southern Runway Tower Controller		
Precondition:	Landing A/C agrees to the swing-over procedure		
Brief Description:	The taboo zone procedure is in place to ensure safe operation, if a missed approach occurs on RWY25L. It prevents that additionally a departure from RWY 25C is in the air parallel to the Missed Approach. To be more efficient and to offer a good service, 25L inbounds may be asked to swing on RWY 25C to allow another departure and also for the benefit of shorter taxi time to the parking stand. However, the change of the RWY implies a change of the Missed Approach Procedure, which has not been briefed by the Flight		

	Crew and would further increase the workload for the pilots. An indication about a potential Missed Approach already during the approach on 25L could induce, that the controller doesn't persist with his plan, as the potential of a Missed Approach would increase due to the complexity of the maneuver.
Benefit for ATCO / ATM / Flight Crew:	Center Runway Tower Controller could refrain from the idea to swing an arrival to allow another departure and stick to the taboo zone procedure • Departing A/C has to wait until arriving A/C has landed → capacity decreases • Departing A/C can, once he has Take-Off clearance follow SID route as planned → safety benefit • Arriving A/C performs Missed Approach Procedure as briefed during approach → safety benefit

This section does not contain a sequence diagram, as with the decision to not perform a swing over, the Take-Off clearance on runway 25C cannot be granted until the arriving A/C either landed or performed the Missed Approach Procedure. Thus the sequence is clear by procedure.

Table 24: Flow for Use Case 3 at Airport 1

Alternate Flow (Compared to Scenario Airport 1 2)		
Step	**User Actions**	**System Actions**
1		a Missed Approach-probability of the next landing RWY 25L is greater than xx%, the tool issues a warning/information about it
2	Center Runway Tower Controller takes the alert into consideration whether to swing the A/C on RWY25C	
3	Center Runway Tower Controller doesn't conduct the swing over	
4	Center Runway Tower Controller is not allowed to issue Take-Off clearance for departure on RWY25C as landing A/C on RWY25L enters taboo zone	
5	arriving Flight Crew initiates Missed Approach	
6	arriving Flight Crew informs Southern Runway Tower Controller	
7	Southern Runway Tower Controller coordinates with Northern Runway Tower Controller, Center Runway Tower Controller & Western Runway Tower Controller	
8	traffic from RWY 25C is no factor, as the procedure secured the separation under such circumstances	
9	Center Runway Tower Controller issues Take-Off clearance as soon as the situation has been cleared	

3.3.5 Airport 2 Use Case 1

This section introduces the first specific use case for Airport 2. It is based on Airport 2 Scenario 1. The use case is summarized in Table 25. Based on the premise of a time in advance indication of a likely Go-Around, three possible branches of actions are described with this use case.

Table 25: Use Case 1 at Airport 2

Use Case ID:	Airport 1.UC.01		
Use Case Name:	Airport 2 - 26L HIRO Operation	**Version No:**	1.0
Linked Scenarios:	Airport 2 Scenario 1		
Linked User Stories:	all GEN, Airport 2.US.01, Airport 2.US.02, Airport 2.US.03		
Involved Actors:	Southern Runway Tower Controller, Flight Crew Departure, Flight Crew Arrival (works for other runways identically)		
Precondition:	HIRO Operations, Reduced Radar Separation, Reduced Runway Separation		
Brief Description:	RWYs in Airport 2 are mostly used in mixed mode, meaning that landings and departures are handled combined at a RWY. During peak times (pre-Corona) in VMC, the closest spacing possible between landing and arriving traffic is necessary to cope with the traffic demand. Providing any additional and reliable information about the Missed Approach-probability of an arrival could induce reactions concerning the traffic planning of the ATCO to diminish complex situations close to the RWY and preventing turning A/C below the MVA as last line of defense. In the following, three possible strategies to handle the complex situation are presented: • Holding back the departing A/C from lining up (see Table 26 and Figure 24 for more detailed descriptions) • Increase the gap between departing A/C and arriving A/C by reducing speed of arriving A/C (see Table 27 and Figure 25 for more detailed descriptions) • Request good speed and climb rate from departing A/C to increase the gap between departing and arriving A/C (see Table 28 and Figure 26 for more detailed descriptions)		
Benefit for ATCO / ATM / Flight Crew:	• Gain time to foster situational awareness. • Refrain Departure A/C from line up → prevent radar separation infringement → prevent unbriefed Missed Approach Procedure (including turns below MVA) • Subsequently, the approaching Flight Crew can perform standard Missed Approach Procedure. → safety benefit		

Table 26: Basic Flow for Use Case 1 at Airport 2

Basic Flow		
Step	**User Actions**	**System Actions**
1		prediction tool displays information about Missed Approach likelihood
2	Tower Controller decides not to use the gap in front of that landing	
3	Tower Controller issues LDG clearance to the arriving A/C	
4	arriving Flight Crew initiates Missed Approach	
5	Tower Controller issues clearance to follow standard Missed ApproachP	

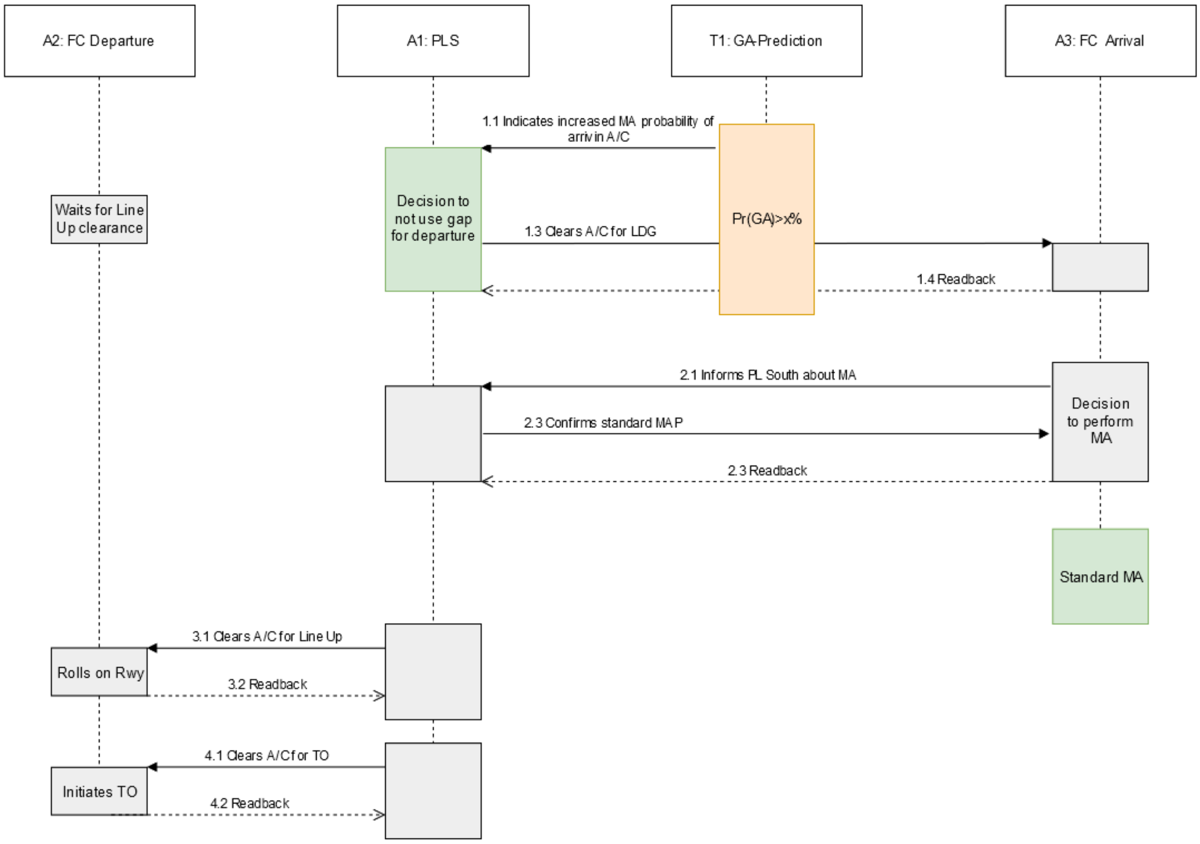


Figure 24: Sequence Diagram for Basic Flow for Use Case 1 at Airport 2

Table 27: Alternate Flow 1 for Use Case 1 at Airport 2

Alternate Flow 1		
Step	**User Actions**	**System Actions**
1		prediction tool displays information about Missed Approach likelihood at a distance of TBD NM
2	Tower Controller instructs ARR Flight Crew to reduce speed in order to gain more mileage between DEP A/C and ARR A/C	
3	Tower Controller lines up DEP A/C on RWY	
4	Tower Controller issues Take-Off clearance for DEP A/C	
5	ARR Flight Crew initiates Missed Approach	
6	ARR Flight Crew informs ATC about the Missed Approach	
7	Tower Controller advises ARR Flight Crew to follow standard Missed Approach Procedure	

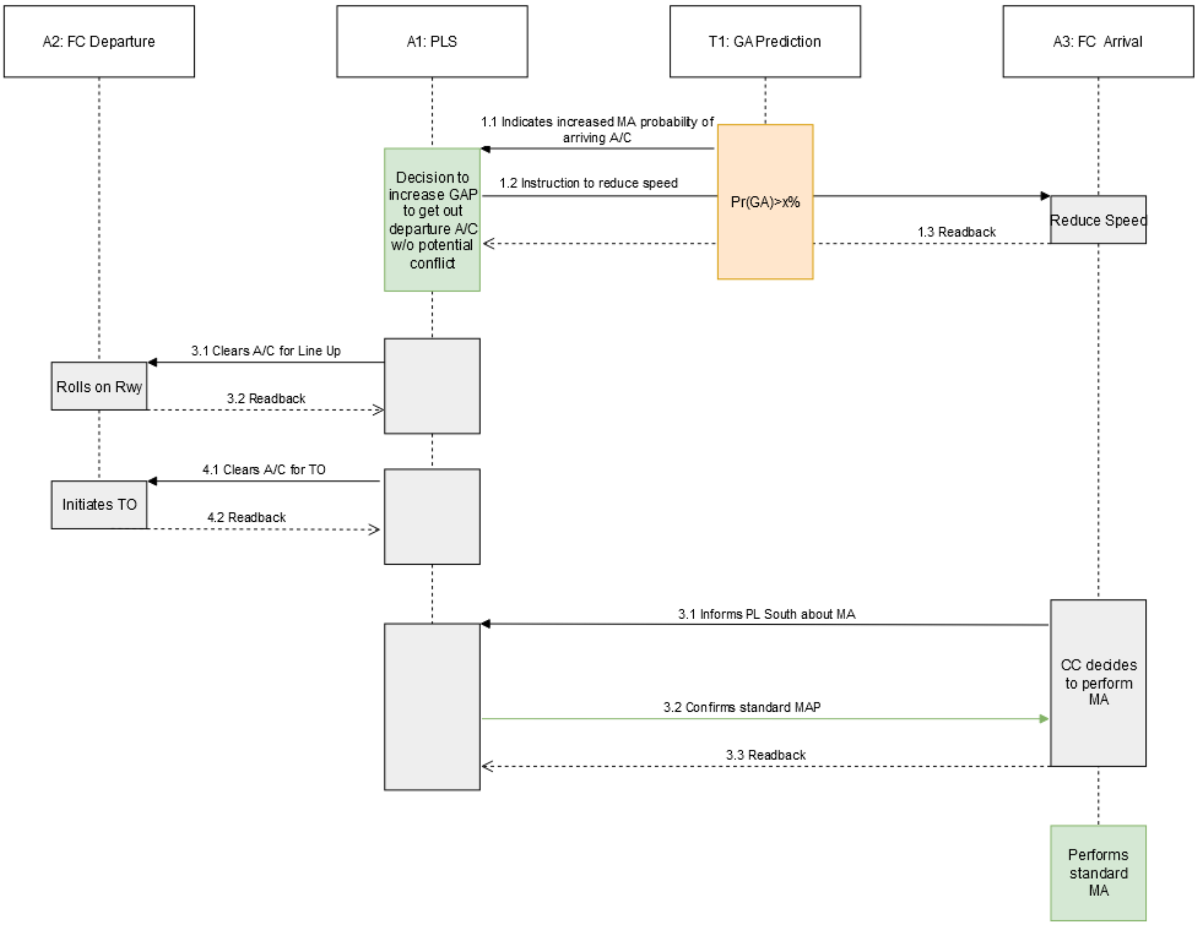


Figure 25: Sequence Diagram for Alternate Flow 1 for Use Case 1 at Airport 2

Table 28: Alternate Flow 2 for Use Case 1 at Airport 2

Alternate Flow 2		
Step	**User Actions**	**System Actions**
1		prediction tool displays information about Missed Approach likelihood at a distance of TBD NM
2	Tower Controller lines up DEP A/C on RWY and informs DEP Flight Crew to provide good speed/good climb rate when airborne or to prepare for a quick departure	
3	Tower Controller issues Take-Off clearance for DEP A/C	
4	ARR Flight Crew initiates Missed Approach	
5	ARR Flight Crew informs ATC about the Missed Approach	
6	Tower Controller advises ARR Flight Crew to follow standard Missed Approach Procedure	

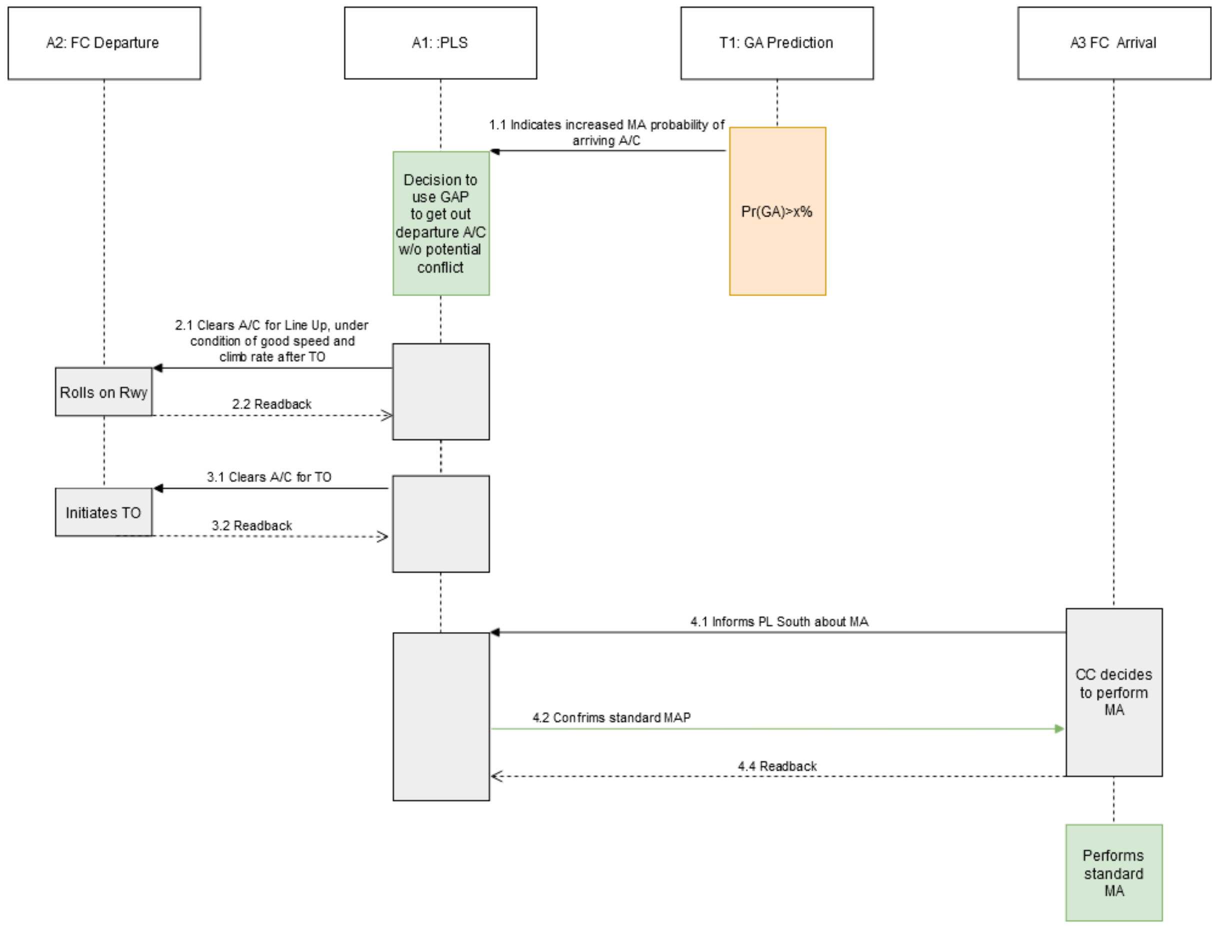


Figure 26: Sequence Diagram for Alternate Flow 2 for Use Case 1 at Airport 2

3.3.6 Airport 2 Use Case 2

This section introduces the second specific use case for Airport 2. It is based on Airport 2 Scenario 2. The use case is summarized in Table 29. Similar to the previous use case, three possible branches of actions, under the premise of a time in advance indication of a likely Go-Around in the context of the respective scenario, are described in more detail.

Table 29: Use Case 2 at Airport 2

Use Case ID:	Airport 1.UC.02		
Use Case Name:	Airport 2 - 26L OTT S-SID A/C in line-up position	**Version No:**	1.0
Linked Scenarios:	Airport 2 Scenario 2		
Linked User Stories:	all GEN, Airport 2.US.01		
Involved Actors:	Southern Runway Tower Controller, Flight Crew Arrival, Flight Crew Departure		

Precondition:	Departure on S-SID has line up clearance
Brief Description:	Several possible departure routes from RWY 26L may conflict with Missed Approaches due to the procedure design. Therefore, providing the information about a possible Missed Approach, when a S-SID departure is marked as line up in the TFDPS. The information shall assist the ATCO's traffic planning and to provide extra time for preparing a plan to separate A/C when the Missed Approach occurs. In the following, three possible strategies to handle the traffic situation described in Airport 2 Scenario 2 are presented: • Demand a higher climb rate from Flight Crew Departure (see Table 30 and Figure 27 for more details) • Demand Flight Crew Departure to deviate from S-SID (see Table 31 and Figure 28 for more details) • Advise non-standard Missed Approach Procedure to Flight Crew Arrival (see Table 32 and Figure 29 for more details)

Benefit for ATCO / ATM / Flight Crew:	• Gain time to foster situational awareness.
	• Early resolution of the potential conflict over OTT
	  ○ Demand higher climb rate from departing A/C → establish vertical separation over OTT with Missed Approach A/C → Flight Crew can perform standard Missed Approach Procedure → safety benefit
	  ○ Demand departure to deviate from S-SID → Flight Crew can perform standard Missed Approach Procedure → safety benefit
	  ○ Advise non-standard Missed Approach Procedure to Flight Crew Arrival in advance, to provide more time

Table 30: Basic Flow for Use Case 2 at Airport 2

Basic Flow		
Step	**User Actions**	**System Actions**
1	Southern Runway Tower Controller clears S-SID departure for line-up and marks it in TFDPS	tool updates only the prediction for the next landing a/c, as the relevant departure is planned in front of this one
2		Tool displays increased Missed Approach probability
3	Southern Runway Tower Controller clears Departure A/C for TO with request of higher climb rate	
4	Southern Runway Tower Controller clears Arriving A/C for ldg	
5	Arriving Flight Crew starts Missed Approach Procedure	
6	Arriving A/C Flight Crew informs Southern Runway Tower Controller about Missed Approach	
7	Arriving A/C performs standard Missed Approach Procedure → OTT	

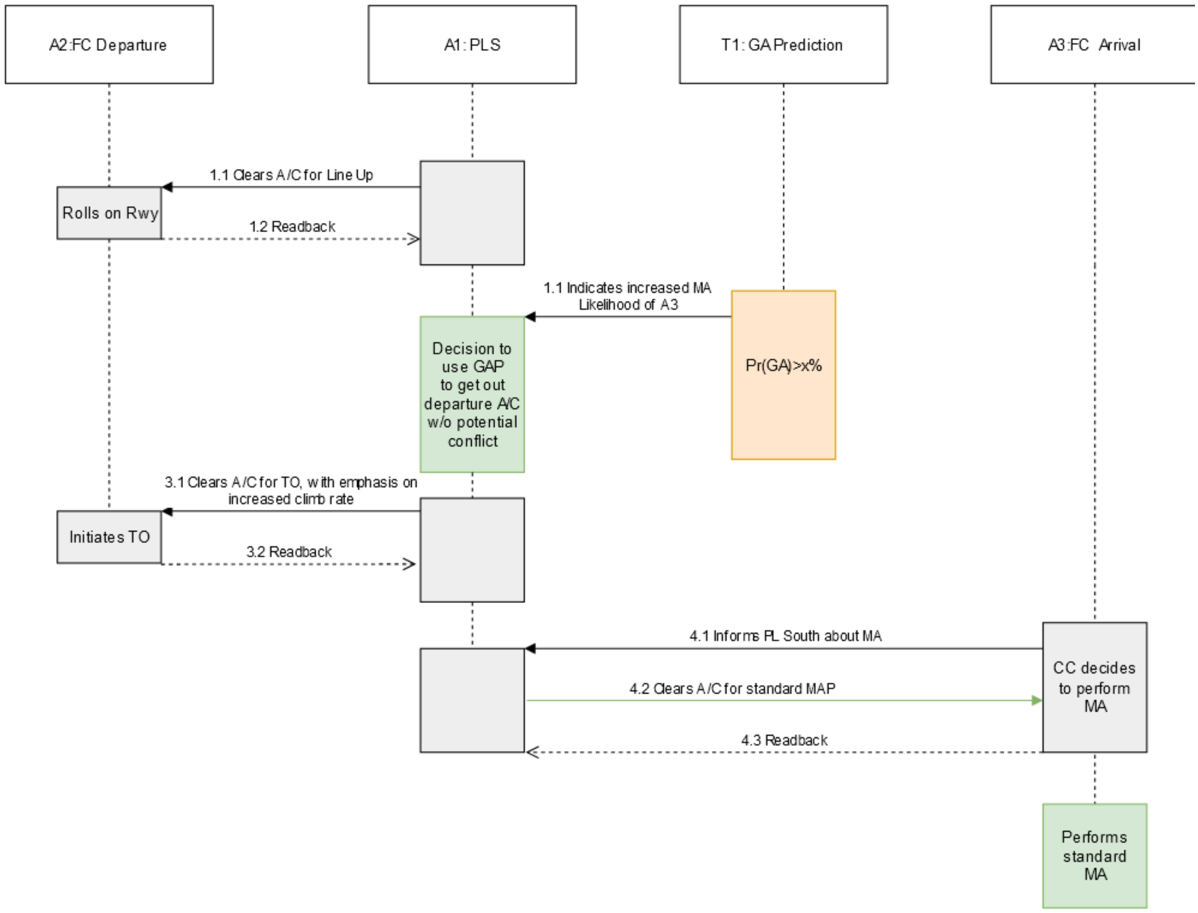


Figure 27: Sequence Diagram for Basic Flow for Use Case 2 at Airport 2

Table 31: Alternate Flow 1 for Use Case 2 at Airport 2

Alternate Flow 1		
Step	**User Actions**	**System Actions**
1	Southern Runway Tower Controller clears S-SID departure for line-up and marks it in TFDPS	tool updates only the prediction for the next landing a/c, as the relevant departure is planned in front of this one
2		Tool displays Missed Approach likelihood above xx THR
3	Southern Runway Tower Controller clears Departure A/C for TO with request to stay on RWY heading	
4	Southern Runway Tower Controller clears Arriving A/C for ldg	
5	Arriving Flight Crew starts Missed Approach Procedure	
6	Arriving A/C Flight Crew informs Southern Runway Tower Controller about Missed Approach	

7	Arriving A/C performs standard Missed Approach Procedure → OTT	

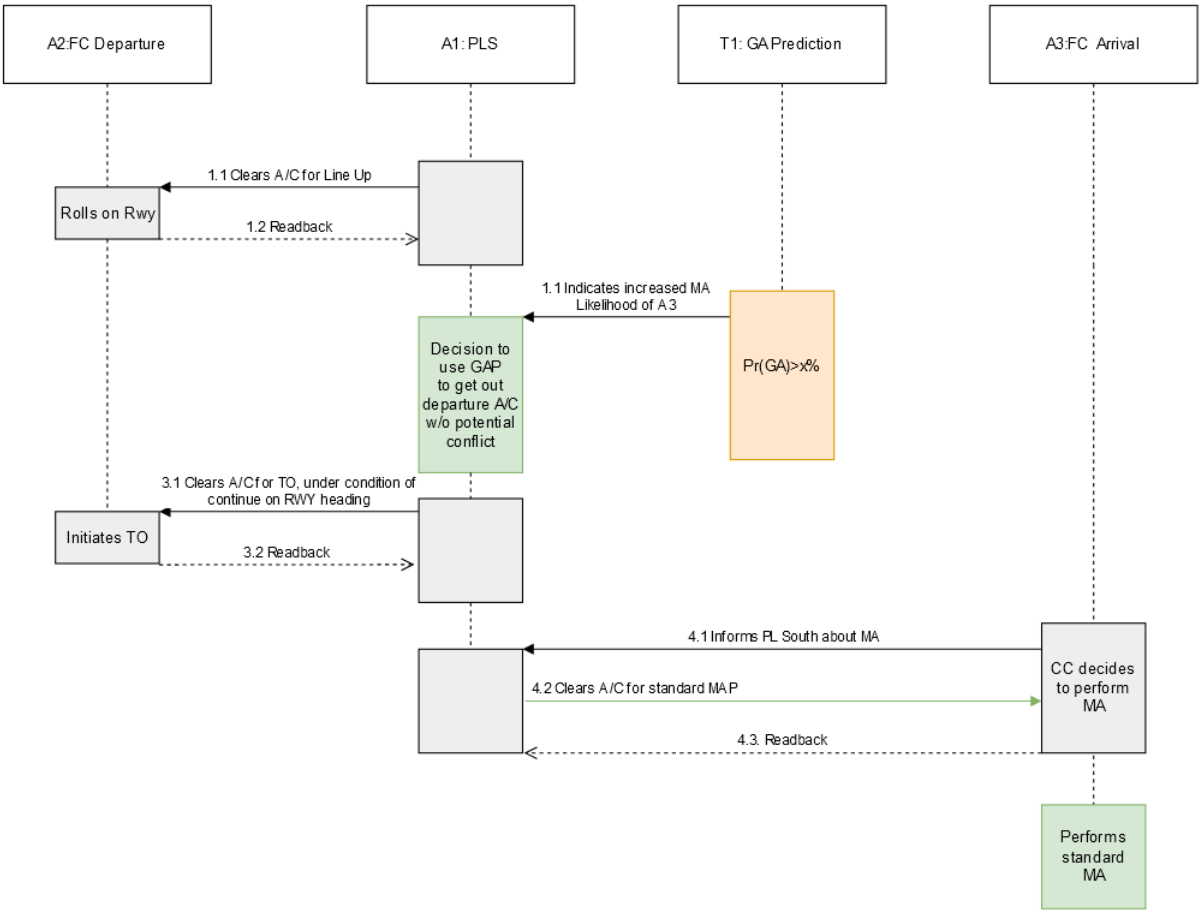


Figure 28: Sequence Diagram for Alternate Flow 1 for Use Case 2 at Airport 2

Table 32: Alternate Flow 2 for Use Case 2 at Airport 2

Alternate Flow 2		
Step	**User Actions**	**System Actions**
1	Southern Runway Tower Controller clears S-SID departure for line-up and marks it in TFDPS	tool updates only the prediction for the next landing a/c, as the relevant departure is planned in front of this one
2		Tool displays Missed Approach likelihood above xx THR
3	Southern Runway Tower Controller prepares arriving Flight Crew with a non-standard Missed Approach Procedure	
4	Southern Runway Tower Controller clears Departure A/C for TO	
5	Southern Runway Tower Controller clears Arriving A/C for landing	

6	Arriving Flight Crew starts Missed Approach Procedure	
7	Arriving A/C Flight Crew informs Southern Runway Tower Controller about Missed Approach	
8	Arriving A/C performs the pre-mentioned non-standard Missed Approach Procedure	

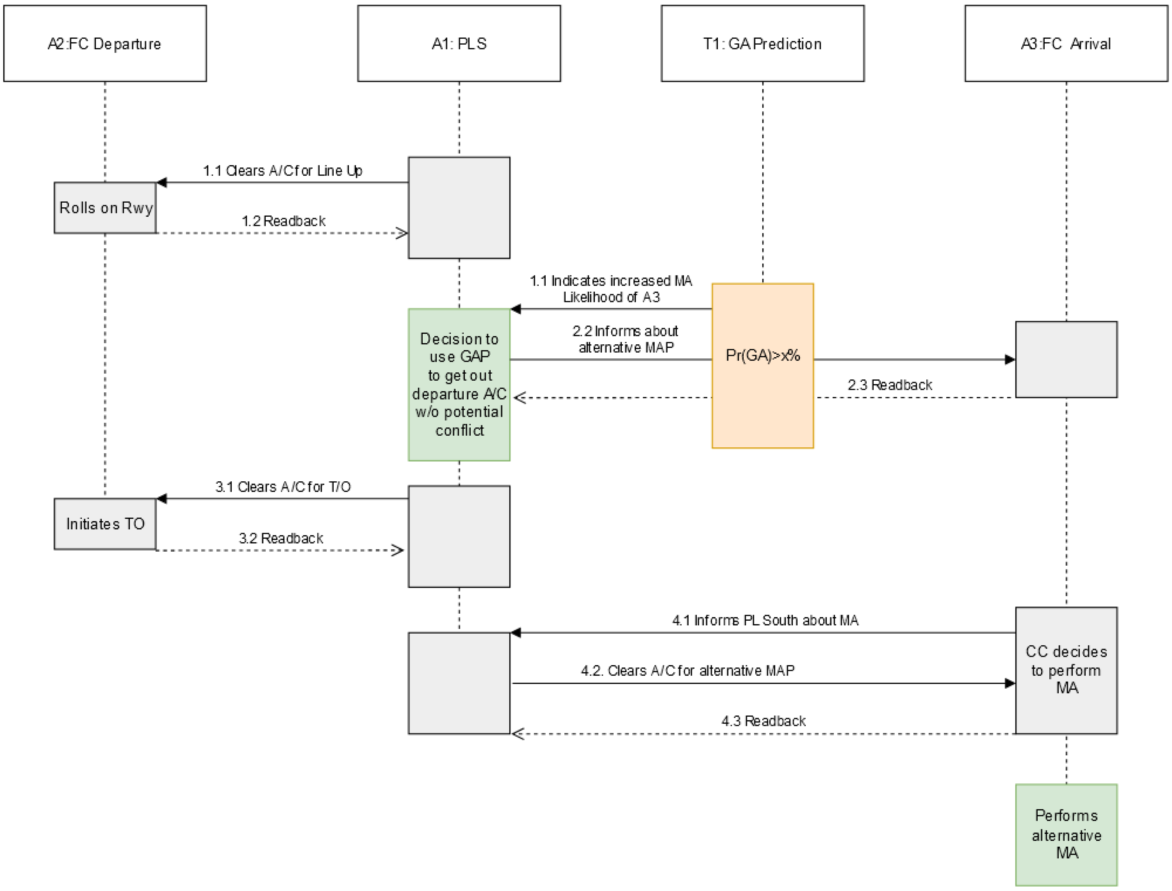


Figure 29: Sequence Diagram for Alternate Flow 2 for Use Case 2 at Airport 2

3.3.7 Airport 2 Use Case 3

This section introduces the third specific use case for Airport 2. It is based on Airport 2 Scenario 2, similar to the previous use case. This use case is, compared to Airport 2 Use Case 2, more radical in its actions as the change of sequence of departing A/Cs is discussed. The use case is summarized in Table 33Table 29. This use case describes two possible branches of actions, under the premise of a time in advance indication of a likely Go-Around in the context of the respective scenario.

Table 33: Use Case 3 at Airport 2

Use Case ID:	Airport 1.UC.03		
Use Case Name:	Airport 2 - 26L OTT no-line up for S-SID	**Version No:**	1.0
Linked Scenarios:	Airport 2 Scenario 2		
Linked User Stories:	all GEN, Airport 2.US.01		
Involved Actors:	Southern Runway Tower Controller, Flight Crew Departure (on S-SID), Flight Crew Arrival, Flight Crew Departure (non S-SID)		
Precondition:	1. Departure on S-SID waiting for line up clearance at the holding point 2. opt. an alternative Departure with different SID is ready for departure and waiting at the holding point		
Brief Description:	Several departures from RWY 26L (S-SIDs) may conflict with Missed Approaches due to the procedure design overhead OTT. Therefore, providing the information about a possible Missed Approach, when a S-SID departure would be next in sequence according TFDPS but hasn't received a line-up clearance yet, should assist the ATCO's traffic and sequence planning. Two possible options are described within this use case: • Hold back the Flight Crew Departure from line up (described in more detail in Table 34 and Figure 30) • Hold back the Flight Crew Departure from line up, whilst clearing an alternative Flight Crew Departure (non S-SID) for line up (described in more detail in Table 35 and Figure 31)		
Benefit for ATCO / ATM / Flight Crew:	• Gain time to foster situational awareness. • resolution of the complex traffic situation in Airport 2 Scenario 2 by early change of traffic sequence • First option will decrease capacity of opeartion		

Table 34: Basic Flow for Use Case 3 at Airport 2

Basic Flow		
Step	**User Actions**	**System Actions**
1	TFDPS shows S-SID departure as number next in sequence	if the Missed Approach-probability of the next landing is greater than xx%, the tool issues a warning/information about it
2	Southern Runway Tower Controller decides to hold the S-SID departure due to possible Missed Approach indication and doesn't use the gap for a departure	
3	Southern Runway Tower Controller issues LDG clearance to approaching A/C	
4	Flight Crew initiates Missed Approach	

5	Flight Crew informs ATC about the Missed Approach	
6	Southern Runway Tower Controller advises Flight Crew to follow standard Missed Approach Procedure	

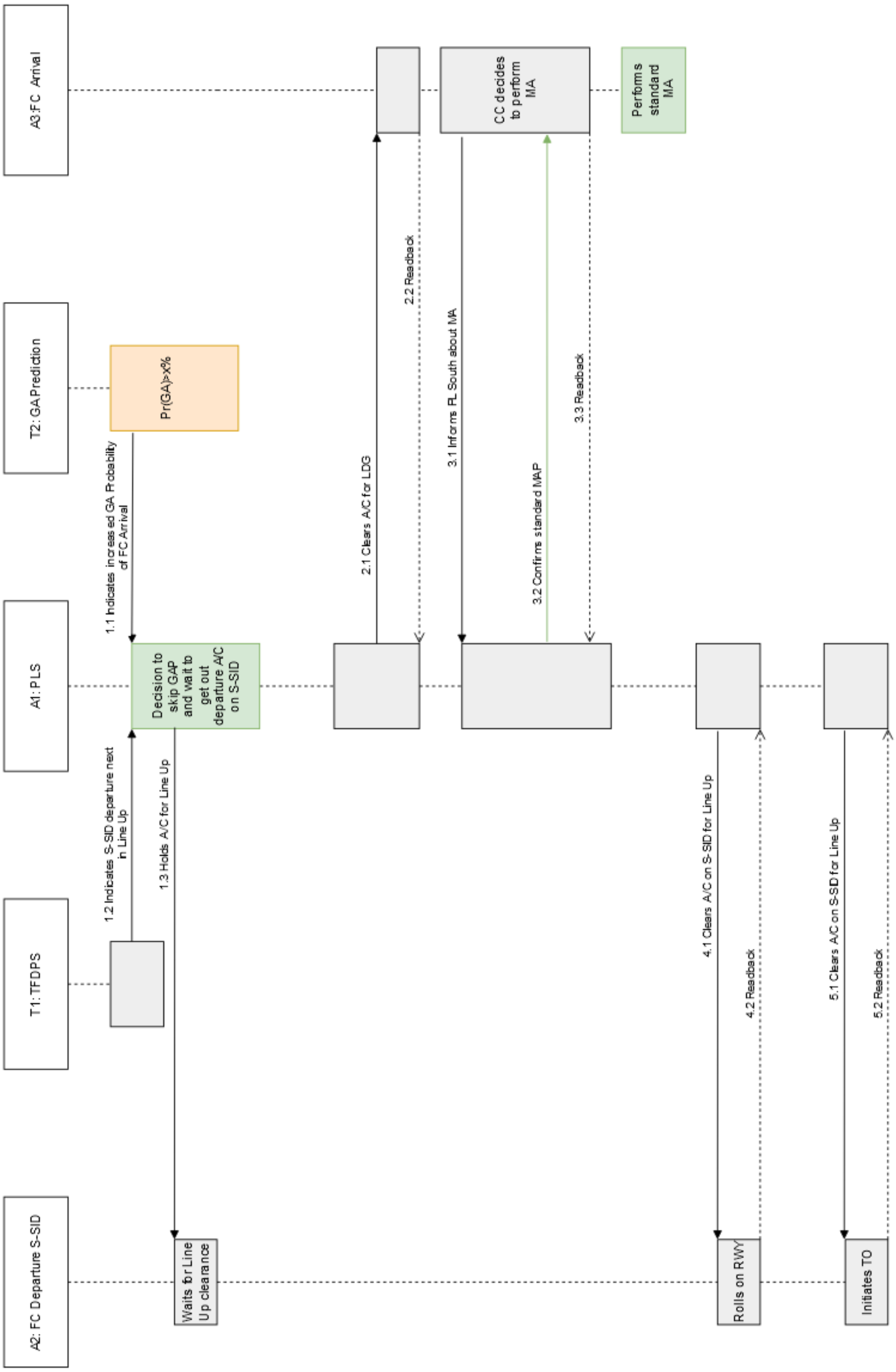


Figure 30: Sequence Diagram for Basic Flow for Use Case 3 at Airport 2

Table 35: Alternate Flow 1 for Use Case 3 at Airport 2

Alternate Flow 1		
Step	**User Actions**	**System Actions**
1	TFDPS shows S-SID departure as number next in sequence	if the Missed Approach-probability of the next landing is greater than xx%, the tool issues a warning/information about it
2	Southern Runway Tower Controller changes the sequence in the TFDPS, S-SID not the next anymore in TFDPS sequence	
3	Southern Runway Tower Controller issues line-up clearance to alternative departure	
4	Southern Runway Tower Controller issues Take-Off clearance to this departure	
5	Southern Runway Tower Controller issues LDG clearance to approaching traffic	
6	approaching Flight Crew initiates Missed Approach and is able to conduct a standard Missed Approach Procedure	
7	Arriving A/C performs standard Missed Approach Procedure → OTT	

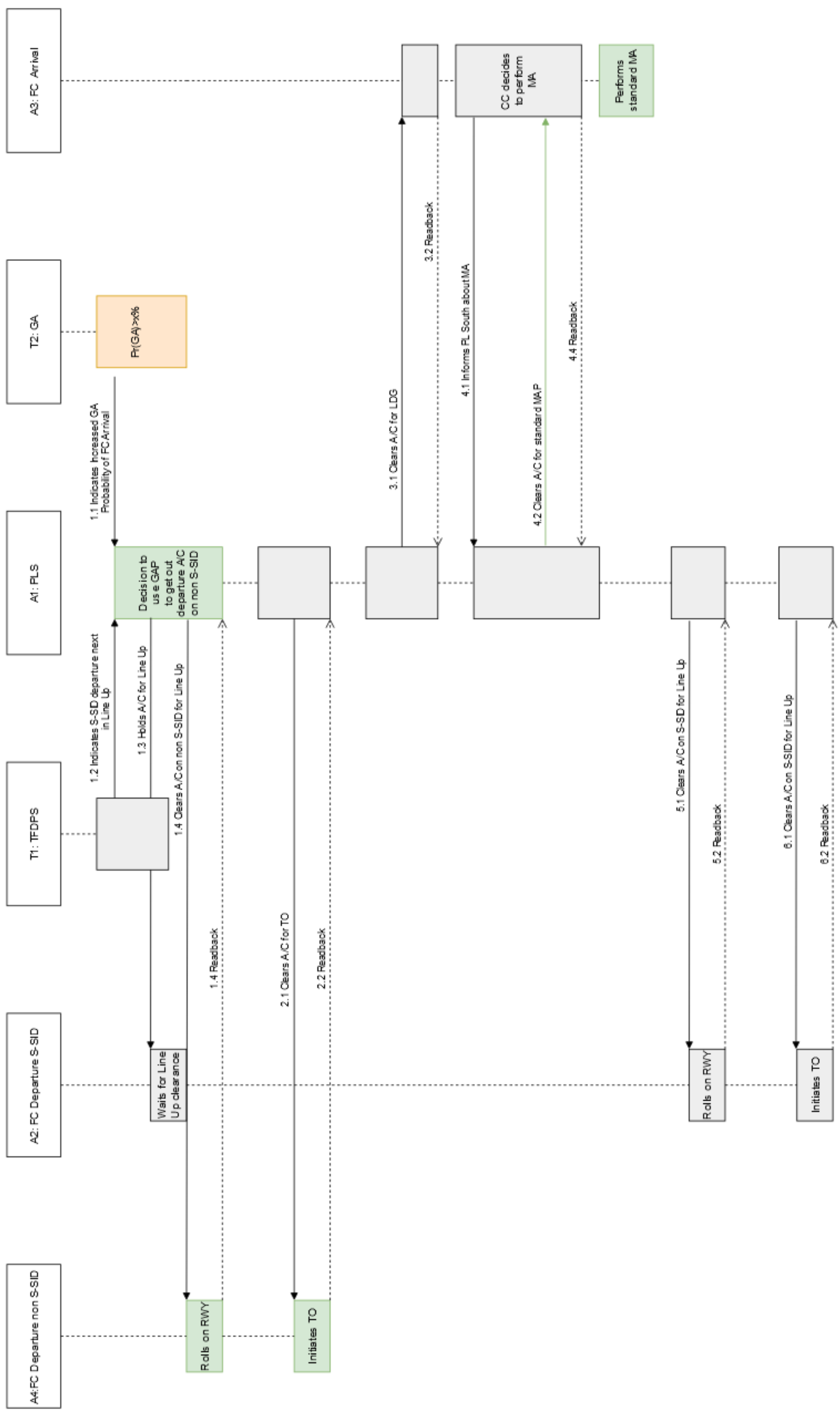


Figure 31: Sequence Diagram for Alternate Flow 1 for Use Case 3 at Airport 2

3.3.8 Airport 2 Use Case 4

This section introduces the fourth specific use case for Airport 2. It is based on Airport 2 Scenario 3. The use case is summarized in Table 36. This use case describes two possible branches of actions, under the premise of a time in advance indication of a likely Go-Around in the context of the respective scenario.

Table 36: Use Case 4 at Airport 2

Use Case ID:	Airport 1.UC.04		
Use Case Name:	Airport 2 all RWY directions WTC challenge	**Version No:**	1.0
Linked Scenarios:	Airport 2 Scenario 3		
Linked User Stories:	all GEN, Airport 2.US.01, Airport 2.US.02, Airport 2.US.04		
Involved Actors:	Northern Runway Tower Controller or Southern Runway Tower Controller , Flight Crew Departure, Flight Crew Arrival		
Precondition:	1. Heavy or super-heavy A/C is waiting for line up clearance at the holding point 2. VMC 3. opt. an alternative Departure with lower WTC is ready for departure and waiting at the holding point		
Brief Description:	During peak hours at Airport 2, in combination with HIRO (High Intensity Runway Operations), ATCO request adjusted gaps in VMC for heavy departures.		
Benefit for ATCO / ATM / Flight Crew:	• gain time to foster situational awareness • provide reliable additional information to support a change in departure sequence. • resolution of WTC separation conflict right after a Missed Approach behind departed heavy/super-heavy A/C		

Table 37: Basic Flow for Use Case 4 at Airport 2

Basic Flow		
Step	**User Actions**	**System Actions**
1	TFDPS shows heavy/super-heavy departure as number next in sequence	if the Missed Approach-probability of the next landing is greater than xx%, the tool issues a warning/information about it
2	Tower Controller decides to hold the departure due to possible Missed Approach indication and doesn't use the gap for a departure	
3	Tower Controller issues LDG clearance to approaching A/C	
4	Flight Crew initiates Missed Approach	
5	Flight Crew informs ATC about the Missed Approach	
6	Tower Controller advises Flight Crew to follow standard Missed Approach Procedure	

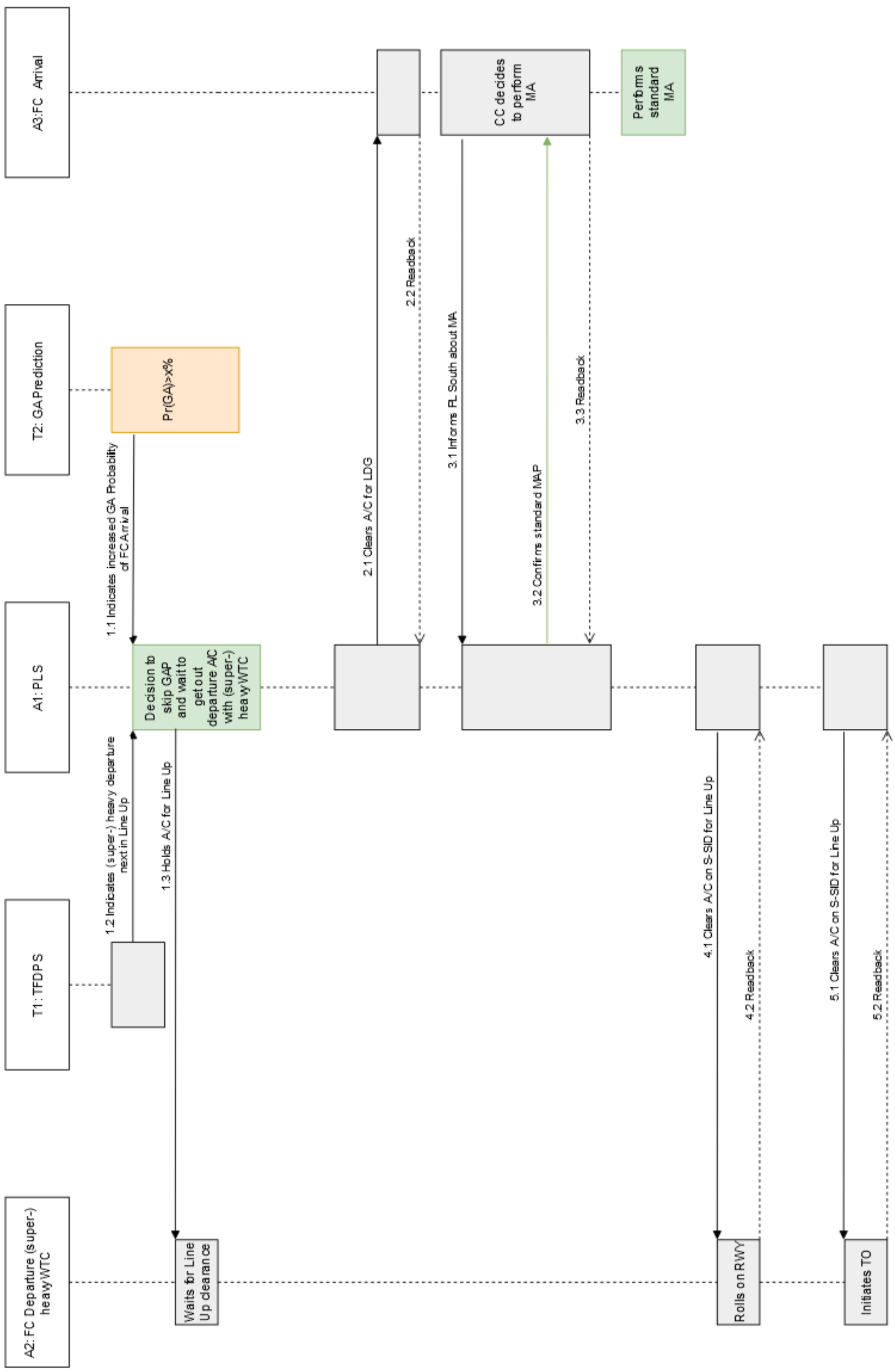


Figure 32: Sequence Diagram for Basic Flow for Use Case 4 at Airport 2

Table 38: Alternate Flow 1 for Use Case 4 at Airport 2

Alternate Flow 1		
Step	**User Actions**	**System Actions**
1	TFDPS shows heavy/super-heavy departure as number next in sequence	if the Missed Approach-probability of the next landing is greater than xx%, the tool issues a warning/information about it
2	Tower Controller changes the sequence of departures, so a A/C with lower WTC is next to line up	
3	Tower Controller issues line-up clearance to alternative departure	
4	Tower Controller issues Take-Off clearance to this departure	
5	Tower Controller issues LDG clearance to approaching traffic	
6	approaching Flight Crew initiates Missed Approach and is able to conduct a standard Missed Approach Procedure	
7	TFDPS shows heavy/super-heavy departure as number next in sequence	

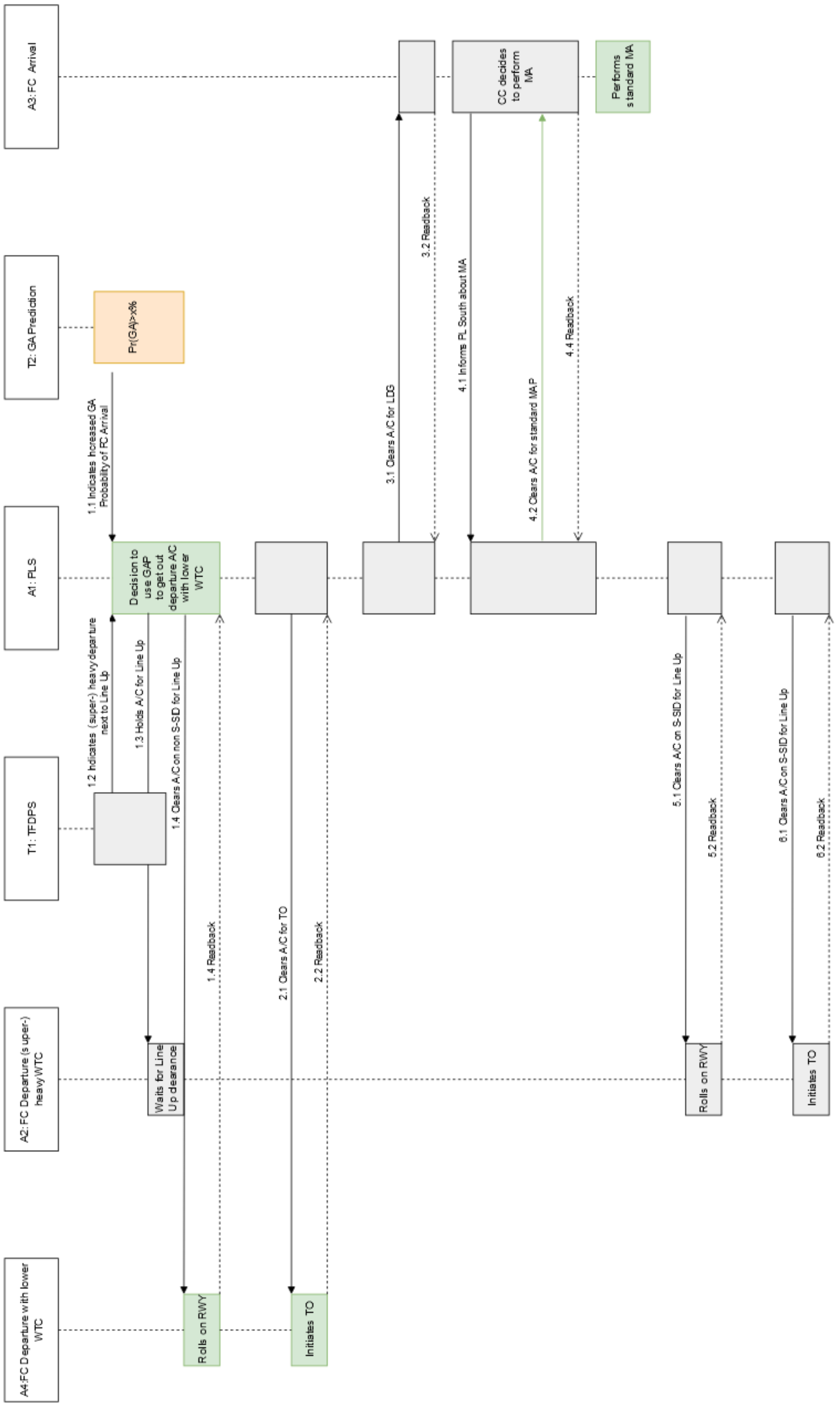


Figure 33: Sequence Diagram for Alternate Flow 1 for Use Case 4 at Airport 2

3.4 SafeOPS – User Stories

This section covers the third step of the requirements engineering process described in section 2.2. Based on the findings and protocols from the workshops conducted with individual ATCOs, user stories are derived. These user stories document statements of the ATCOs that further refine desired functionalities of a Go-Around prediction tool. We thereby tried to extract user stories which fulfil the characteristics of a user story as described in section 2.2, e.g. user stories should be estimable and independent.

After the first set of user stories based on the first workshops were derived, several user stories contradicted each other or certain details differed between ATCOs, e.g. values of prediction horizon or colour codes. For clarification of contradictions the user stories were revised in later workshops. Occasionally, contradicting statements will lead to the need for personalization of the output of the Go-Around prediction.

In the following subsections all user stories are collected and grouped into general user stories in section3.4.1 and use case specific user stories for Airport 1 in section 3.4.2 as well as Airport 2 in section 3.4.3 with one table per user story. Where possible, statements of the ATCOs were combined to one user story with a certain reasoning. If the statements only differed slightly with different reasons for the user story, multiple lines per user story ID are provided. For some of the user stories it depends on the capabilities of the future tool if they can be followed on or not. Also, if contradictions could not be resolved for this deliverable, they were kept for the sake of transparency and traceability. Links provided in the tables indicate the relation between user stories, use cases and requirements.

3.4.1 General User Stories

The general user stories are applicable for all use cases defined in previous sections. They are grouped in different categories, namely "input data", "output data", "user interface" and "reliability". Especially in the category "user interface" the opinions of the ATCOs diverge. Here, the stories contradict each other on different aspects of a future user interface. Some ATCOs prefer a symbol like indication where others like text based displays. Some want to have it on one system's screen (e.g. TFDPS) and others on the other system's screen (e.g. radar). These contradictions need to be resolved to derive requirements. The category "input data" mainly provides guidance for the selection of relevant data sources and parameters. Within the category of "output data" user stories are grouped, that provide insight in the expected information to be gained from the tool. And the category "reliability" is concerned with prediction performance necessary for the acceptance of the tool by the ATCOs.

User Story ID:	GEN.US.01	**Category:**	Input Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.D.01, NF.D.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to take into account the large bandwidth of pilot behavior		the tool can cope with A/C induced Go-Arounds.	

User Story ID:	GEN.US.02	**Category:**	Input Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.D.01, NF.D.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to take into account the meteorological condition (VMC/IMC) and in general weather data		weather as contributing factor is covered and thunderstorms, fog and wind shear can be taken into account.	

User Story ID:	GEN.US.03	**Category:**	Input Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.D.01, NF.D.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to take into account the current type of operation (mixed, single, staggered)		become more reliable.	

User Story ID:	GEN.US.04	**Category:**	Input Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.D.01, NF.D.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to take into account the performance of landing and departing A/C as well as the WTC		it reflects the effect of traffic mix at a commercial AP.	

User Story ID:	GEN.US.05	**Category:**	Output Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to inform me about the development of the current gaps		use gaps more efficiently and to recognize potential conflicts earlier	

User Story ID:	GEN.US.06	**Category:**	User Interface	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.H.01, FR.H.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to show the prediction either in the TFDPS with a color change or the air radar with a color code		attract attention but not distract at the same time	
	ATCO	the tool to indicate information on the radar screen rather than the TFDPS, at least until confirmation. It could be a square around the callsign which changes color to yellow/orange		ATCO's attention is drawn on the information	

User Story ID:	GEN.US.07	**Category:**	User Interface	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.01, FR.H.02, FR.M.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to present the probability of the prediction in a continuous representation		I gain understanding of how a situation evolves to gain situational awareness and to be able to respond in a timely manner	

User Story ID:	GEN.US.08	**Category:**	User Interface	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.H.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to allow the identification of the situation quickly through presenting the information in a short and concise manner		improved preplanning is possible and overloaded screens are not further loaded	
Rationale:	The ATCOs stated a good time frame for the identification of the situation are 1-2 seconds.				

<table>
<tr><td>User Story ID:</td><td>GEN.US.09</td><td>Category:</td><td>User Interface</td><td>Version:</td><td>1.0</td></tr>
<tr><td>Linked Use Case(s):</td><td colspan="5">Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03</td></tr>
<tr><td>Linked Requirement(s):</td><td colspan="5">FR.H.02</td></tr>
<tr><td>User Story:</td><td>As a...</td><td colspan="2">I want...</td><td colspan="2">so that... / in order to...</td></tr>
<tr><td></td><td>ATCO</td><td colspan="2">the tool to have a personalized way of presenting information</td><td colspan="2">fit my working style</td></tr>
</table>

<table>
<tr><td>User Story ID:</td><td>GEN.US.10</td><td>Category:</td><td>User Interface</td><td>Version:</td><td>1.0</td></tr>
<tr><td>Linked Use Case(s):</td><td colspan="5">Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03</td></tr>
<tr><td>Linked Requirement(s):</td><td colspan="5">FR.H.02</td></tr>
<tr><td>User Story:</td><td>As a...</td><td colspan="2">I want...</td><td colspan="2">so that... / in order to...</td></tr>
<tr><td></td><td>ATCO</td><td colspan="2">the information to be presented as colorful vis/ color markers</td><td colspan="2">be easily distinguished from other information</td></tr>
<tr><td></td><td>ATCO</td><td colspan="2">the tool to have a text type/written display, included in an existing system (e.g. TFDPS)</td><td colspan="2">information is easily accessible and understandable</td></tr>
<tr><td></td><td>ATCO</td><td colspan="2">the tool not to use color code when incorporated in radar screen</td><td colspan="2">other information, seen as more important, is not blocked</td></tr>
<tr><td></td><td>ATCO</td><td colspan="2">the tool only to display numbers if incorporated in TFDPS, no color code</td><td colspan="2">already high number of colors is not increased</td></tr>
</table>

<table>
<tr><td>User Story ID:</td><td>GEN.US.11</td><td>Category:</td><td>User Interface</td><td>Version:</td><td>1.0</td></tr>
<tr><td>Linked Use Case(s):</td><td colspan="5">Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03</td></tr>
<tr><td>Linked Requirement(s):</td><td colspan="5">FR.C.03, FR.H.03</td></tr>
<tr><td>User Story:</td><td>As a...</td><td colspan="2">I want...</td><td colspan="2">so that... / in order to...</td></tr>
<tr><td></td><td>ATCO</td><td colspan="2">the information to be presented only when critical values (warnings, cautions) are reached</td><td colspan="2">information overflow is avoided</td></tr>
</table>

User Story ID:	GEN.US.12	**Category:**	User Interface	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.03, FR.M.01, NF.M.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to provide predictions with high precision and high likelihood		the forecast will not be ignored and precautionary action can be taken/tool can be trusted	
Rationale:	A minimum accuracy of 60% was stated by ATCOs.				

User Story ID:	GEN.US.13	**Category:**	Input Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.D.01, NF.D.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to take into account human factors		it becomes more reliable	
Rationale:	This user story is very broad and the currently available data and time will not allow to pursue it in the scope of this project. If this assessment is too conservative and it will be possible to follow up on this issue it will be integrated later.				

User Story ID:	GEN.US.14	**Category:**	User Interface	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.04, FR.H.02, FR.M.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to present the main contributing factor as reasons for high likelihood of Go-Around (importance of explainable AI), e.g. with abbreviations like W for weather S for speed; presenting the currently driving factor		be able to understand the situation better, information is less likely to be ignored and ATCO can act accordingly	
	ATCO	the tool to not always present the main contributing factor, but to allow access by clicking		overload is avoided	

User Story ID:	GEN.US.15	**Category:**	Output Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.04, FR.M.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to be designed to predict defined causes for Go-Around and to be precise on as many causes for Go-Around as possible (e.g. Unstabilized Approach, GS deviation,...)		increase reliability of the tool and to comprehend the tools prediction	

User Story ID:	GEN.US.16	**Category:**	Reliability	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.03, NF.M.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to predict Go-Around with a very low false negative rate as well as a low false positive rate		a false negative does not lead to stress situations with unexpected Missed Approach and a false positive does not lead to unnecessary workload and to loosing trust in the tool	

User Story ID:	GEN.US.17	**Category:**	Output Data	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.T.01, FR.T.02, NF.C.01				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to predict a Go-Around in real time		gain time and improve preplanning / not loose time. Whenever a prediction is issued, the time for the computer to calculate the result shot be short enough so that the result is still relevant	

User Story ID:	GEN.US.18	Category:	Output Data	Version:	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to also detect an actual G/A in real time		be able to respond quicker and to help ATCO to gain time	

User Story ID:	GEN.US.19	Category:	Output Data	Version:	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):					
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	wind shear warnings compared to the once on board		ATCO can act accordingly	

User Story ID:	GEN.US.20	Category:	User Interface	Version:	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.H.01, FR.H.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool not to use acoustic warnings		distraction is avoided	

User Story ID:	GEN.US.21	Category:	User Interface	Version:	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.C.04, FR.H.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to be explainable to controller including description of used data			

User Story ID:	GEN.US.22	**Category:**	User Interface	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04, Airport 1.UC.01, Airport 1.UC.02, Airport 1.UC.03				
Linked Requirement(s):	FR.H.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool not to draw too much attention away from currently existing information		distraction is avoided	

3.4.2 Airport 1 Specific User Stories

For the crucial aspect of a prediction, the time of the prediction, differences between use cases become evident during the workshops. All user stories assigned to Airport 1 belong to the category "timing". The optimal time point of a prediction result available to the ATCO is partly defined as actual time before touch down or as distance to the THR. The mentioned numbers differ but still deliver insight in what the controllers expect. These expectations can then be tried to be fulfilled by the developed tool.

User Story ID:	Airport 1.US.01	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 1.UC.01				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to predict Go-Around in a certain time frame, so it fits the usual pre planning time horizon		I can improve preplanning	
	ATCO	the tool to predict a specific time before Go-Around is performed		time for managing the situation is gained	
Rationale:	The time frame for preplanning purposes are 2-3 minutes according to ATCOS. For the imminent situation the ATCOs mentioned the prediction would be valuable if it happened 10-15 seconds before the Go-Around.				

User Story ID:	Airport 1.US.02	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 1.UC.03				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to warn ATCOs of a high Go-Around probability before the A/C reaches a certain distance from THR		no swing over will be offered to the pilot, which would further increase G/A probability	
	ATCO	the tool to warn ATCOs between within a certain distance interval of a high Go-Around probability because of an event on RWY25L		I can offer a swing over and possibly prevent a G/A	
Rationale:	The ATCOs mentioned that if the tool predicts a Go-Around at a distance of 7NM or more, they would not offer a swing over. If the warning is provided between 4 to 7NM, a swing over could be offered				

User Story ID:	Airport 1.US.03	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 1.UC.01				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to indicate a Go-Around at a short distance before the THR.		taxiing a/c can cross runway	
Rationale:	The distance mentioned during the workshops are 2NM.				

3.4.3 Airport 2 Specific User Stories

Similar to the user stories exclusive to Airport 1, only the user story category "timing" is evident for Airport 2. The user stories in part do not contain unique wishes by the ATCOs. These differences are based on different lines of reasoning by the ATCOs. These differences need to be addressed in the following development phase.

User Story ID:	Airport 2.US.01	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.02, Airport 2.UC.03, Airport 2.UC.04				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to warn latest before a certain point in front of the THR		I can decide to skip a gap for the upcoming departure in time	
Rationale:	The latest point mentioned by the ATCOs is 3NM, where 4NM were preferred.				

User Story ID:	Airport 2.US.02	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01, Airport 2.UC.04				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to warn at a specified distance before the THR		I am still able to try to reduce the chance of a Missed Approach and possible conflicts can be avoided, while not affecting the outbound capacity by reducing speeds of the inbound AC	
	ATCO	the tool to provide early predictions, rather sooner than later		allow to move the decision point further out and to provide additional time for planning	
Rationale:	Here no specific distance is defined.				

User Story ID:	Airport 2.US.03	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.01				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to warn before a certain distance from the THR		information can be used for capacity planning/adjustments	
	ATCO	the tool to predict between a certain point before the THR		departure planning can be influenced	
Rationale:	A distance of 10NM or greater is seen as useful for the capacity planning. For the departure planning a prediction beginning at 4NM is seen as helpful.				

User Story ID:	Airport 2.US.04	**Category:**	Timing	**Version:**	1.0
Linked Use Case(s):	Airport 2.UC.04				
Linked Requirement(s):	FR.T.01, FR.T.02				
User Story:	**As a...**	**I want...**		**so that... / in order to...**	
	ATCO	the tool to warn latest before a certain point,		I can decide to change the sequence of departures to restore WTC compliance	
Rationale:	Here no distance was specified.				

3.5 SafeOPS – Requirements

In the following two subsections, the functional and nonfunctional requirements are specified. They have been developed based on the user stories from the project members that will be responsible for the developmental part in SafeOPS.

3.5.1 SafeOPS – Functional Requirements

The information captured by the User Stories is subsequently translated into a set of Functional Requirements which describe the minimum performance that the system has to achieve to address the User needs.

We organized the collected Functional Requirements in four main categories:

1. **High Level Functionality:**
   These requirements are related with the confidence and explainability of the Go-Around predictions.
2. **Human Machine Interface (HMI):**
   These requirements describe how the information about a possible Go-Around should be provided (e.g. visually, acoustically, etc.).
3. **Timing of the Predictions:**
   This is the instant or time interval at which information about a Go-Around should be delivered, which correspond to distances of the incoming A/C from the THR.
4. **Big Data and Machine Learning Requirements:**
   They refer to the data-storage and data-processing capabilities necessary to analyse incoming data and provide real-time predictions.

3.5.1.1 High Level Functionality

This section describes the functional requirements that describe the high level functionality extracted from the user stories. These high level requirements will be refined to define the functionality in more detail.

Requirement ID:	FR.C.01	**Category:**	Input Output Relation	**Version:**	1.0
Linked User Story(ies):	GEN.US.07				
Requirement:	The system shall output a probability of an approaching A/C performing a Go-Around (also referred to as prediction), given data that describes the A/C's approach and the conditions thereof as input.				
Rationale:	Top level functional requirement, that describes the idea of a data based prediction of Go-Arounds, specifying input-output relationship.				

Requirement ID:	FR.C.02	**Category:**	Input Output Relation	**Version:**	1.0
Linked User Story(ies):	GEN.US.18				
Requirement:	The system shall output detected Go-Arounds, given operational performance data as input.				
Rationale:	When the Flight Crew starts a Go-Around operation, there is normally a gap of 15 to 60 seconds before they can inform the ATCO about it. Unless the visibility is perfect, the ATCO has no chance to see what is happening until the information is communicated by the CC. The immediate information that a Go-Around has started would therefore gain time for the ATCO to decide how to manage the situation.				

Requirement ID:	FR.C.03	**Category:**	Performance Assessment	**Version:**	1.0
Linked User Story(ies):	GEN.US.11, GEN.US.12, GEN.US.16				
Requirement:	The system shall provide quantifiable metrics on the performance quality of the prediction.				
Rationale:	This requirement links to the user stories, demanding an indication of quality of the predictive outcome of the system. A detailed description on how performance assessment can be performed on machine learning algorithms is provided in section 4.3.1.				

Requirement ID:	FR.C.04	**Category:**	Output, Interpretability, Explainability	**Version:**	1.0
Linked User Story(ies):	GEN.US.14, GEN.US.15, GEN.US.21				
Requirement:	The system shall provide information on the contributing factors, responsible for the prediction.				
Rationale:	ATCOs want to know the reason why the tool predicts that a given inbound flight will likely Go-Around to better understand the situation and decide if the best option is to intervene (e.g. directly contacting the pilot), ignore the warning, or get prepared to manage the traffic when the Go-Around effectively occurs. The more information is available about the reasons for the prediction, the more elements the ATCO has to decide how to manage the situation. The explainability and interpretability of ML algorithms is described in detail in section 4.3.2.				

3.5.1.2 Human Machine Interface (HMI)

Requirement ID:	FR.H.01	**Category:**	HMI	**Version:**	1.0
Linked User Story(ies):	GEN.US.20, GEN.US.06				
Requirement:	The system output shall be provided as visual indication.				
Rationale:	The Users want to be able to visualise on a screen the relevant information about possible Go-Around. Acoustic signals are rejected as typically associated with emergency events and are considered distracting.				

Requirement ID:	FR.H.02	**Category:**	HMI	**Version:**	1.0
Linked User Story(ies):	GEN.US.06, GEN.US.07, GEN.US.08, GEN.US.09, GEN.US.10, GEN.US.14, GEN.US.20, GEN.US.21, GEN.US.22				
Requirement:	The content of the visualized indication shall be customizable.				
Rationale:	The User Stories highlight a variety of different opinions the Users have on what information should be shown, in particular: • The way the information can be visualized (e.g. readily on the screen, by actively clicking on the callsign, etc.) AND • The amount of information shown (e.g. just a colored sign, or a sign together with the Go-Around probability, or additionally the main causes that will likely induce a Go-Around, etc.) AND • The amount of accessible information (e.g. whether all the available information is displayed or additional information might be accessed by clicking e.g. on the flight ID) • The screen on which the information is delivered (e.g. on the radar screen, on the TFDPS screen, on a separate screen, etc.) AND • The frequency at which the information is delivered and updated (e.g. continuously, every 1-2 seconds, after the Go-Around probability crosses pre-defined thresholds, etc.) AND • The colors used to deliver the information (e.g. with our without a color coding for different levels of alert) AND • The probability threshold above which information should be displayed (e.g. no threshold, a minimum of 60% probability, 80% probability, etc.) To account for this variety, the HMI should be as customizable as possible, so that it can be adapted to the User's working style and it does not distract the User's attention from				

	their ongoing tasks. If possible, the purpose is to enable the ATCOs to choose the options on how the prediction should be delivered as input of the predictive tool.

Requirement ID:	FR.H.03	**Category:**	HMI	**Version:**	1.0
Linked User Story(ies):	GEN.US.11				
Requirement:	The prediction shall only be presented, if the predicted probability of a Go-Around is above a quantifiable *minimum Go-Around probability* threshold.				
Rationale:	The ATCOs want to avoid nuisance information about flights that are likely to land safely, as this would not only cause an overload of information but also reduce the trust in the prediction. For this reason, they suggest that Go-Around predictions with low probability, for example below a *minimum Go-Around probability* = 60%, should not be displayed. The threshold above which the Go-Around predictions are shown could be customized to fit the preference of the individual user (see also FR.H.02)				

3.5.1.3 Timing of Prediciton

Requirement ID:	FR.T.01	**Category:**	Timing of Prediction	**Version:**	1.0
Linked User Story(ies):	GEN.US.05, GEN.US.17, Airport 1.US.01, Airport 1.US.02, Airport 1.US.03, Airport 2.US.01, Airport 2.US.02, Airport 2.US.03, Airport 2.US.04				
Requirement:	The prediction shall be computed every *prediction update rate* seconds in between a *minimum distance* and *maximum distance* measured from the runway threshold.				
Rationale:	ATCOs need to have a picture of the situation as up-to-date as possible to improve preplanning and prepare at best to the evolving circumstances. The sooner information about a likely Go-Around is given, the more time the ATCO has to decide their next actions. The *prediction update rate*, *minimum distance*, *maximum distance* describe parameters that can be varied. From the discussions in the workshops, the following initial values are set: • *prediction update rate* = 1-2 seconds • *minimum distance* = 0NM from THR • *maximum distance* = 10NM from THR Note that this requirement and FR.T.02 overlap in parts of their functionality.				

Requirement ID:	FR.T.02	**Category:**	Timing of Prediction	**Version:**	1.0
Linked User Story(ies):	GEN.US.05, GEN.US.17, Airport 1.US.01, Airport 1.US.02, Airport 1.US.03, Airport 2.US.01, Airport 2.US.02, Airport 2.US.03, Airport 2.US.04				
Requirement:	The prediction shall be computed at *specified distance increments* in between a *minimum distance* and *maximum distance* measured from the runway threshold.				
Rationale:	ATCOs need to have a picture of the situation as up-to-date as possible to improve preplanning and prepare at best to the evolving circumstances. The sooner information about a likely Go-Around is given, the more time the ATCO has to decide their next actions. A Go-Around prediction within this range can be used by ATCO for capacity planning, that is to decide whether or not to clear an outbound flight for departure, and also to coordinate with the Flight Crew of the inbound to reduce the chances of a Missed Approach Procedure. The *specified distance increments*, *minimum distance*, *maximum distance* describe parameters that can be varied. From the discussions in the workshops, the following initial values are set: • *specified distance increments* = 0.1 NM • *minimum distance* = 0NM from THR • *maximum distance* = 10NM from THR				

3.5.1.4 Big Data and Machine Learning Requirements

This section covers the data and machine learning related functional requirements. The requirements are kept relatively high level. A detailed description on how they can be understood is linked in each description section of the requirements. This link provides the relevant information for the requirement documented in the technical problem statement.

Requirement ID:	FR.D.01	**Category:**	Data Pipeline	**Version:**	1.0
Linked User Story(ies):	GEN.US.01, GEN.US.02, GEN.US.03, GEN.US.04, GEN.US.13				
Requirement:	The data sets available to the system shall be stored in a data lake, where they can be accessed as input for the data pipeline.				
Rationale:	This requirement describes the storage of the data used in SafeOPS. This will be handled using DataBeacon. DataBeacon already provides Data Security and Data Acquisition functionalities that were implemented during the H2020 project SafeClouds.eu. Therefore, implementation of Data Security and Data Acquisition Requirements are not part of the SafeOPS developments. For completeness, these parts are described in the Appendix A.1. A detailed description on data acquired for SafeOPS and deployed in DataBeacon is given in section 4.1.				

Requirement ID:	FR.D.02	Category:	Data Pipeline	Version:	1.0
Linked User Story(ies):):	GEN.US.12				
Requirement:	The system shall contain a data processing pipeline that automates **data cleaning** and **data preparatio**n tasks.				
Rationale:	This requirement will be implemented using DataBeacon. Some functionalities will already be in place, others not. This is further specified in the following requirements on data cleaning and data preparation. A detailed description on these tasks is provided in section 4.2.				

Requirement ID:	FR.D.03	Category:	Data Pipeline	Version:	1.0
Linked Requirement(s):	FR.D.02				
Requirement:	The system shall contain a **data cleaning** process, that automates the following tasks: • outlier detection AND • filtering / missing value handling for the data sets available in the data lake.				
Rationale:	This functionality was implemented into DataBeacon for several data sources during the SafeClouds.eu project. It is describe in deliverable "Data preparation" available on the CORDIS website [19]. In case SafeOPS acquires new data types. These methods have to be implemented accordingly.				

Requirement ID:	FR.D.04	Category:	Data Pipeline	Version:	1.0
Linked Requirement(s):	FR.D.02				
Requirement:	The system shall contain a **data preparation** process, that automates the following tasks: • data fusion AND • target labelling AND • feature engineering for the data sets available in the data lake, and generates training data sets, test data sets and validation data sets.				
Rationale:	These tasks are dependent on the machine learning application they are performed for. This will have to be implemented for SafeOPS. A detailed description on these tasks is provided in section 4.2.				

Requirement ID:	FR.M.01	**Category:**	Model Training	**Version:**	1.0
Linked User Story(ies):	GEN.US.07, GEN.US.12, GEN.US.14, GEN.US.15				
Requirement:	The system shall contain a machine learning **model training** process, that optimizes the prediction of a machine learning model, given a training data set.				
Rationale:	These tasks are dependent on the machine learning application they are performed for. A detailed description on these tasks is provided in section 0.				

3.5.2 SafeOPS – Non-Functional Requirements

This section collects the non-functional requirements, which document performance, quality and boundary condition specifications, left open in the functional requirements.

Requirement ID:	NF.D.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	GEN.US.01, GEN.US.02, GEN.US.03, GEN.US.04, GEN.US.13				
Requirement:	The data set provided as input to the system shall contain information on: • A/C performance AND • meteorological conditions AND • pilot inputs to the A/C AND • WTC of the A/C				
Rationale:	Information on data availability and storage are provided in section 4.1.				

Requirement ID:	NF.C.01	**Category:**	Computational Efficiency	**Version:**	1.0
Linked User Story(ies):	GEN.US.17				
Requirement:	The information about the probability of a Go-Around prediction should be provided in real time (less than 0.5s after provision of input data)				
Rationale:	When notified of a likely upcoming Go-Around event, the ATCOs want to know the probability associated with this prediction to gain a clearer situational awareness.				

Requirement ID:	NF.M.01	**Category:**	Model Training	**Version:**	1.0
Linked User Story(ies):	GEN.US.12, GEN.US.16				
Requirement:	The performance assessment of the system shall include quantifiable metrics on: • true positive, true negative, false positive and false negative ratios AND • accuracy, precision, recall and specificity				
Rationale:	A detailed description of these metrics is provided in section 4.3.1.				

Requirement ID:	NF.M.02	**Category:**	Model Training	**Version:**	1.0
Linked User Story(ies):					
Requirement:	The model training shall be able to cope with imbalanced training data sets.				
Rationale:	As Go-Around occur with a rate of around 0.3% of approaches, the training data set is expected to be highly imbalanced. Therefore, special techniques to deal with this problem must be implemented. A detailed description of this problem is provided in section 4.3.3.				

4 Technical Problem Statement

The following chapter provides an overview over the challenges and methods, that can at this stage of the project already be identified, based on the user stories and requirements defined in sections 3.4 and 0. Furthermore, it describes the data sets acquired for the project at this stage.

4.1 Data Acquisition

The following section contains the description of the currently available data sets in the BeSt platform from DataBeacon for use in the SafeOPS project. This will include a brief description summarizing the main properties, the structure, the data items descriptions for each parameter contained within the data set and the range of available data (temporal and geographical). As previously stated, this data sets used in the project will not be restricted to the ones presented here, but the data catalogue might evolve during the duration of the project (e.g., data might become available).

Table 39: Data Catalogue for BeSt Platform

Data Source	Data Provider	Timeframe	Geographical Coverage	Relevance	Related User Stories
ECTL R&D	EUROCONTROL	2015 - 2018 (Mar, Jun, Sep, Dec)	Europe	Possible use for the estimation of operational parameters such as demand, planned and actual trajectories.	GEN.US.01 GEN.US.03 GEN.US.04 GEN.US.05
ADS-B	OpenSky	20/4/2018 - ongoing	Europe	Detailed actual trajectory flown, useful for detecting and labelling Go-Around situations, feature extraction	GEN.US.01 GEN.US.03 GEN.US.04 GEN.US.05
FDM	Iberia and Pegasus	Jun 2017 - Oct 2018 (Iberia) Jun 2017 - Apr 2019 (Pegasus)	Iberia's network and Pegasus's network	Actual flight performance, useful for detecting and labelling Go-Around situations, provide enhanced features for prediction	GEN.US.01 GEN.US.04 GEN.US.13
METAR	Iowa State University	2017 - 2019	Europe - Large and Medium Airports	Airport weather features	GEN.US.02
ERA5	ECMWF	September 2014	Polygon depending on request	Weather for forecast available at different prediction horizons and could be used as actual (reanalysis) weather for labelling.	GEN.US.02
SIGMET	NOAA – National Weather Service	2020/01 - ongoing	Europe	High disruptive meteorological features	GEN.US.02

TAF	Navlost	September 2018 (potential availability for more months/years)	Europe - Large and medium APs	Airport weather features	GEN.US.02

4.1.1 Description of ADS-B OpenSky

Brief Description

The OpenSky Network is a non-profit association based in Switzerland. It aims at improving the security, reliability and efficiency of the air space usage by providing open access of real-world air traffic control data to the public. The main technologies behind the OpenSky Network are the Automatic Dependent Surveillance-Broadcast (ADS-B) and Mode S. These technologies provide detailed A/C information in real time over the publicly accessible 1090 MHz radio frequency channel.

Automatic Dependent Surveillance - Broadcast (ADS-B) is a surveillance technology in which an A/C determines its position via satellite navigation and periodically broadcasts it, enabling it to be tracked. ADS–B is "automatic" in that it requires no pilot or external input. It is "dependent" in that it depends on data from the A/C's navigation system. The system relies on two avionics components—a high-integrity GPS navigation source and a datalink (ADS-B unit). There are several types of certified ADS-B data links, but the most common ones operate at 1090 MHz, essentially a modified Mode S transponder.

Structure and Size

From ICAO's Annex 10 (Aeronautical Telecommunications) Vol.4 (Surveillance and Collision Avoidance Systems), 3.1.2.8.6 EXTENDED SQUITTER, DOWNLINK FORMAT 17:

Extended squitter format. The format used for the extended squitter shall be a 112-bit downlink format (DF = 17) containing the following fields at the bit positions:

Table 40: Extended Squitter Format – Fields at the Bit Position

MSB (Most Significant Bit)	1	6	9	33	89
content	**DF**	**CA**	**AA**	**ME**	**PI**
LSB (Least Significant Bit)	5	8	32	88	112

- ***DF*** *downlink format: This downlink format field (5 bits long) shall serve as the downlink format descriptor in all Mode S replies. This is "10001" for DF17.*
- ***CA*** *capability: This 3-bit (6-8) downlink field shall convey information on the transponder level, the additional information below, and shall be used in formats DF = 11 and DF = 17.*

Table 41: Capability Description as Part of ADS-B Message

Coding	CA Capability
0	signifies Level 1 transponder (surveillance only), and no ability to set CA code 7 and either airborne or on the ground
1	reserved
2	reserved
3	reserved
4	signifies Level 2 or above transponder and ability to set CA code 7 and on the ground
5	signifies Level 2 or above transponder and ability to set CA code 7 and airborne
6	signifies Level 2 or above transponder and ability to set CA code 7 and either airborne or on the ground
7	signifies there exists a Downlink Request (DR field is not equal to 0) or the Flight Status has an alert or Special Position Identification pulse (SPI) (FS field equals 2, 3, 4 or 5, and either airborne or on the ground)

- ***AA** address, announced: This 24-bit (9-32) downlink field shall contain the A/C address which provides unambiguous identification of the A/C.*
- ***ME** message, extended squitter: This 56-bit (33-88) downlink field in DF = 17 shall be used to transmit broadcast messages. Extended squitter shall be supported by registers 05, 06, 07, 08, 09, 0A {HEX} and 61-6F {HEX} and shall conform to either version 0, version 1 or version 2 message formats as described below:*
  - *a) Version 0 ES message formats and related requirements report surveillance quality by navigation uncertainty category (NUC), which can be an indication of either the accuracy or integrity of the navigation data used by ADS-B. However, there is no indication as to which of these, integrity or accuracy, the NUC value is providing an indication of.*
  - *b) Version 1 ES message formats and related requirements report surveillance accuracy and integrity separately as navigation accuracy category (NAC), navigation integrity category (NIC) and surveillance integrity level (SIL). Version 1 ES formats also include provisions for enhanced reporting of status information; and*
  - c) *Version 2 ES message formats and related requirements contain the provisions of version 1 but further enhance integrity and parameter reporting. Version 2 ES formats separately report position source integrity from the integrity of the ADS-B transmitting equipment. Version 2 ES formats also separate vertical accuracy reporting from horizontal position accuracy, remove vertical integrity from position integrity, and provide for the reporting of the SSR Mode A code, GNSS antenna offset and additional horizontal position integrity values. Version 2 ES formats also modify the target state report to include selected altitude, selected heading, and barometric pressure setting.*
- ***PI** parity/interrogator identifier: This 24-bit (33-56) or (89-112) downlink field shall have parity overlaid on the interrogator's identity code according to parity check at the error protection. For acquisition or an extended squitter, the Interrogator Identifier (II) and the Surveillance Identifier (SI) codes shall be 0.*

Data features

The variables described in Table 42 are the ones provided by the data provider OpenSky.

Table 42: OpenSky Dataset

	Column	Description	Example
1	baro_altitude	Barometric altitude. Depends on factors such as weather.	10744.2
2	callsign	Callsign identifying the flight. Typically, ICAO airline code plus IATA/ticketing flight number, or the A/C registration	UAL22
3	geo_altitude	Altitude in meters determined using the GNSS (GPS) sensors.	10965.18
4	heading	Track angle in degrees	278.11
5	icao24	ICAO 24-bit address	aa8c39
6	last_contact	Time information of last contact (Unix timestamp), when OpenSky received the last signal of the A/C.	1543658039
7	latitude	Latitude in degrees	53.7721
8	longitude	Longitude in degrees	-11.8272
9	on_ground	is A/C on ground	False
10	origin_country	Country of the AP of origin	United States
11	position_source	-	0
12	sensors	-	None
13	spi	Special Position Indicator	False
14	squawk	4-digit octal number. Transporter code used by ATC and pilots for identification purposes and indication of emergencies.	1446
15	time_position	Time information of A/C position (Unix timestamp)	1543658039
16	velocity	A/C velocity over ground (m/s)	196.94
17	vertical_rate	The A/C's vertical rate of climb/descent (m/s)	-0.33

4.1.2 Description of ECTL R&D

Collected data description by EUROCONTROL R&D Data Release - Metadata:

"EUROCONTROL releases flight data for R&D purposes, subject to users agreeing to the terms and conditions. The data source for the flights and their profiles through points and airspaces is flight plans submitted by airlines and other A/C operators to EUROCONTROL Network Manager (NM) and the flight profiles generated by NM's ATFM systems. Flight plans are required to be submitted to NM for all instrument flight rules (IFR) flights in the NM Area. In some cases, the A/C operator value in the flight plan has been updated with more accurate values from EUROCONTROL Central Route Charges Office (CRCO) data. The point and airspace profile data in the 'actual' version of the data includes some updates from radar observation of the flight's path. The data source for the ATFM environment data is the EUROCONTROL Network Manager database of airspace and route structures used by NM's ATFM systems. Each monthly batch of flight and profile data includes flights whose planned departure time occurred in the month delivered. Flight data is filtered to include flights of ICAO flight types 'S' (scheduled) and 'N' (non-scheduled flight), excluding ICAO types General aviation, Military and Other. The data are provided 'as is', with no quality guarantees, and no support beyond the provision of the metadata that is given in this document." [20]

Data Features

The EUROCONTROL R&D data archive is made up of a number of different data sets:

Table 43: FLIGHTS - Flight details from Eurocontrol Network Manager flight plans in PRISME Data Warehouse

Column Name	Description	Format
ECTL_ID	Unique numeric identifier for each flight in Eurocontrol PRISME DWH.	-
ADEP	ICAO code of departure airport.	Four-letter alphanumeric code
ADEP Latitude	Latitude of departure airport.	Decimal degrees
ADEP Longitude	Longitude of departure airport.	Decimal degrees
ADES	ICAO code of destination airport.	Four-letter alphanumeric code
ADES Latitude	Latitude of destination airport.	Decimal degrees
ADES Longitude	Longitude of destination airport.	Decimal degrees
Filed Arrival Time	Time of arrival (UTC) of the last filed FP. Touchdown time in de ADES	%dd-%mm-%YYYY %HH:%MM%SS
Actual Off-Block Time	Off-Block Time (UTC) based on the ATFM-updated flight plan.	%dd-%mm-%YYYY %HH:%MM%SS
Actual Arrival Time	Time of arrival (UTC) based on the ATFM-updated flight plan.	%dd-%mm-%YYYY %HH:%MM%SS
AC Type	ICAO A/C type	Two-, three- or four-character alphanumeric code
AC Operator	ICAO operator code (3 letter).	-
AC Registration	A/C registration (tail number)	-
ICAO flight type	S (Scheduled), N (Non-Scheduled) commercial operation	S – Scheduled, N - Non-scheduled commercial operation
STATFOR Market Segment	1: Business Aviation 2: Military IFR 3: All-Cargo 4: Low-Cost 5: Other Scheduled (ICAO Flight Type "N") 6: Traditional Scheduled (ICAO Flight Type "S")	-
Requested FL	Request cruising Flight Level from the flight plan.	-
Actual Distance Flown (NM)	Distance Flown in NM.	-

Table 44: FLIGHT POINTS - Filed and Actual Flight points

Column Name	Description	Format
ECTL_ID	Unique numeric identifier for each flight in Eurocontrol PRISME DWH.	-
Sequence Number	Numeric sequence number of the points crossed by the flight in chronological order. (Points can be not only known named **waypoints**, **navaids**, etc. but also **intermediate points inserted by NM** profile-generation processes.)	-
Time Over	Time (UTC) at which the point was crossed	%dd-%mm-%YYYY %HH:%MM%SS

Flight Level	Altitude in flight levels at which the point was crossed	-
Latitude	Latitude	Decimal degrees
Longitude	Longitude	Decimal degrees

Table 45: FLIGHT AIRSPACES - Flights entering/leaving FIR airspace (Filed and Actual)

Column Name	Description	Format
ECTL_ID	As in Flights file above	-
Sequence number	Numeric sequence number of the airspace entered by the flight in chronological order	-
FIR ID	The identifier of the FIR	-
Entry Time	Time (UTC) the flight entered the airspace	%dd-%mm-%YYYY %HH:%MM%SS
Exit Time	Time (UTC) the flight exited the airspace	%dd-%mm-%YYYY %HH:%MM%SS

Table 46: FLIGHT AIRSPACES - Flights entering/leaving AUA (Air traffic control Unit Airspace) airspaces (Filed and Actual)

Column Name	Description	Format
ECTL_ID	As in Flights file above	-
Sequence number	Numeric sequence number of the airspace entered by the flight in chronological order	-
AUA ID	The identifier of the AUA	-
Entry Time	Time (UTC) the flight entered the airspace	%dd-%mm-%YYYY %HH:%MM%SS
Exit Time	Time (UTC) the flight exited the airspace	%dd-%mm-%YYYY %HH:%MM%SS

Table 47: ATM ENVIRONMENT DATA - AIRAC

Column Name	Description	Format
External ID	Unique ID of the AIRAC.	YYXX YY: Year in which the AIRAC was published XX: Sequential number of the AIRAC cycle
Date From	Start validity date of the AIRAC.	-
Date to	Expiration date of the AIRAC.	-

Table 48: ATM ENVIRONMENT DATA - Routes

Column Name	Description	Format
Route ID	Unique ID of the Route.	A. The basic designator consists of one letter of the alphabet followed by a number from 1 to 999. The letters may be: 1. A, B, G, R — for routes which form part of the regional networks of ATS routes and are not area navigation routes; 2. L, M, N, P — for area navigation routes which form part of the regional networks of ATS routes; 3. H, J, V, W — for routes which do not form

		part of the regional networks of ATS routes and are not area navigation routes; 4. Q, T, Y, Z — for area navigation routes which do not form part of the regional networks of ATS routes. B. Where applicable, one supplementary letter shall be added as a prefix to the basic designator as follows: 1. K — to indicate a low level route established for use primarily by helicopters;
Sequence Number	Numeric sequence number of a point on the route	-
Latitude	Latitude of the point in the route.	Decimal degrees
Longitude	Longitude of the point in the route.	Decimal degrees.

Table 49: ATM ENVIRONMENT DATA - Flight Information Regions (FIRs)

Column Name	Description	Format
Airspace ID	Unique identifier of the FIR (could also be a UIR, Upper Information Region)	A. The basic designator consists of one letter of the alphabet followed by a number from 1 to 999. The letters may be: 1. A, B, G, R — for routes which form part of the regional networks of ATS routes and are not area navigation routes; 2. L, M, N, P — for area navigation routes which form part of the regional networks of ATS routes; 3. H, J, V, W — for routes which do not form part of the regional networks of ATS routes and are not area navigation routes; 4. Q, T, Y, Z — for area navigation routes which do not form part of the regional networks of ATS routes. B. Where applicable, one supplementary letter shall be added as a prefix to the basic designator as follows: 1. K — to indicate a low level route established for use primarily by helicopters;
Min Flight Level	Minimum vertical boundary of the airspace volume expressed as a flight level, repeated for each point	-
Max Flight Level	Maximum vertical boundary of the airspace volume expressed as a flight level, repeated for each point	-
Sequence Number	Numeric sequence number of a boundary point of the FIR's shape	-
Latitude	Latitude of the point.	Decimal degrees
Longitude	Longitude of the point	Decimal degrees.

4.1.3 Description of ERA5

The ERA5 dataset contains one (hourly, 31 km) high resolution realization (referred to as "reanalysis" or "HRES") and a reduced resolution ten member ensemble (referred to as "ensemble" or "EDA"). Generally, the data are available at a sub-daily and monthly frequency and consist of **analyses** and short (18 hour) **forecasts**, initialized twice daily from analyses at 06 and 18 UTC. Most analyzed parameters are also available from the forecasts.

The data type is gridded and thus stored in GRIB format. It has a global horizontal coverage with a horizontal resolution of 0.25°x0.25° for reanalysis and 0.5°x0.5° for mean, spread and members. The vertical coverage ranges from 1000 hPa to 1 hPa with a vertical resolution of 37 pressure levels. The temporal coverage goes from 1979 until present with a hourly temporal resolution.

Data Variables

The variables described in Table 50 are the ones stored by ERA5 files. When making a data request the desired variables are to be selected, along with the geographical region and the timeframe.

Table 50: Main Variables of ERA5 Dataset

Name	Units	Description
Divergence	s^{-1}	This parameter is the horizontal divergence of velocity. It is the rate at which air is spreading out horizontally from a point, per square meter. This parameter is positive for air that is spreading out, or diverging, and negative for the opposite, for air that is concentrating, or converging (convergence).
Fraction of cloud cover	-	This parameter is the proportion of a grid box covered by cloud (liquid or ice) and varies between zero and one. This parameter is available on multiple levels through the atmosphere.
Geopotential	$m^2\ s^{-2}$	This parameter is the gravitational potential energy of a unit mass, at a particular location, relative to mean sea level. It is also the amount of work that would have to be done, against the force of gravity, to lift a unit mass to that location from mean sea level. The geopotential height can be calculated by dividing the geopotential by the Earth's gravitational acceleration, g (=9.80665 m s-2). The geopotential height plays an important role in synoptic meteorology (analysis of weather patterns). Charts of geopotential height plotted at constant pressure levels (e.g., 300, 500 or 850 hPa) can be used to identify weather systems such as cyclones, anticyclones, troughs and ridges. At the surface of the Earth, this parameter shows the variations in geopotential (height) of the surface, and is often referred to as the orography.
Ozone mass mixing ratio	$kg\ kg^{-1}$	This parameter is the mass of ozone per kilogram of air. In the ECMWF Integrated Forecasting System (IFS), there is a simplified representation of ozone chemistry (including representation of the chemistry which has caused the ozone hole). Ozone is also transported around in the atmosphere through the motion of air. Naturally, occurring ozone in the stratosphere helps protect organisms at the surface of the Earth from the harmful effects of ultraviolet (UV) radiation from the Sun. Ozone near the surface, often produced by pollution, is harmful to organisms. Most of the IFS chemical species are archived as mass mixing ratios [kg kg-1].
Potential vorticity	$K\ m^2\ kg^{-1}\ s^{-1}$	Potential vorticity is a measure of the capacity for air to rotate in the atmosphere. If we ignore the effects of heating and friction, potential vorticity is conserved following an air parcel. It is used to look for places where large wind storms are likely to originate and develop. Potential vorticity increases strongly

		above the tropopause and therefore, it can also be used in studies related to the stratosphere and stratosphere-troposphere exchanges. Large wind storms develop when a column of air in the atmosphere starts to rotate. Potential vorticity is calculated from the wind, temperature and pressure across a column of air in the atmosphere.
Relative humidity	%	This parameter is the water vapor pressure as a percentage of the value at which the air becomes saturated (the point at which water vapor begins to condense into liquid water or deposition into ice). For temperatures over 0°C (273.15 K) it is calculated for saturation over water. At temperatures below -23°C it is calculated for saturation over ice. Between -23°C and 0°C this parameter is calculated by interpolating between the ice and water values using a quadratic function.
Specific cloud ice water content	kg kg^{-1}	This parameter is the mass of cloud ice particles per kilogram of the total mass of moist air. The 'total mass of moist air' is the sum of the dry air, water vapor, cloud liquid, cloud ice, rain and falling snow. This parameter represents the average value for a grid box. Water within clouds can be liquid or ice, or a combination of the two. Note that 'cloud frozen water' is the same as 'cloud ice water'.
Specific cloud liquid water content	kg kg^{-1}	This parameter is the mass of cloud liquid water droplets per kilogram of the total mass of moist air. The 'total mass of moist air' is the sum of the dry air, water vapor, cloud liquid, cloud ice, rain and falling snow. This parameter represents the average value for a grid box. Water within clouds can be liquid or ice, or a combination of the two.
Specific humidity	kg kg^{-1}	This parameter is the mass of water vapor per kilogram of moist air. The total mass of moist air is the sum of the dry air, water vapor, cloud liquid, cloud ice, rain and falling snow.
Specific rain water content	kg kg^{-1}	The mass of water produced from large-scale clouds that is of raindrop size and so can fall to the surface as precipitation. Large-scale clouds are generated by the cloud scheme in the ECMWF Integrated Forecasting System (IFS). The cloud scheme represents the formation and dissipation of clouds and large-scale precipitation due to changes in atmospheric quantities (such as pressure, temperature and moisture) predicted directly by the IFS at spatial scales of a grid box or larger. The quantity is expressed in kilograms per kilogram of the total mass of moist air. The 'total mass of moist air' is the sum of the dry air, water vapor, cloud liquid, cloud ice, rain and falling snow. This parameter represents the average value for a grid box. Clouds contain a continuum of different sized water droplets and ice particles. The IFS cloud scheme simplifies this to represent a number of discrete cloud droplets/particles including cloud water droplets, raindrops, ice crystals and snow (aggregated ice crystals). The processes of droplet formation, phase transition and aggregation are also highly simplified in the IFS.
Specific snow water content	kg kg^{-1}	The mass of snow (aggregated ice crystals) produced from large-scale clouds that can fall to the surface as precipitation. Large-scale clouds are generated by the cloud scheme in the ECMWF Integrated Forecasting System (IFS). The cloud scheme represents the formation and dissipation of clouds and large-scale precipitation due to changes in atmospheric quantities (such as pressure, temperature and moisture) predicted directly by the IFS at spatial scales of a grid box or larger. The mass is expressed in kilograms per kilogram of the total mass of moist air. The 'total mass of moist air' is the sum of the dry air, water vapor, cloud liquid, cloud ice, rain and falling snow. This parameter represents the average value for a grid box. Clouds contain a continuum of different sized water droplets and ice particles. The IFS cloud scheme simplifies this to represent a

		number of discrete cloud droplets/particles including cloud water droplets, raindrops, ice crystals and snow (aggregated ice crystals). The processes of droplet formation, phase transition and aggregation are also highly simplified in the IFS.
Temperature	K	This parameter is the temperature in the atmosphere. It has units of kelvin (K). Temperature measured in kelvin can be converted to degrees Celsius (°C) by subtracting 273.15. This parameter is available on multiple levels through the atmosphere.
U-component of wind	$m\ s^{-1}$	This parameter is the eastward component of the wind. It is the horizontal speed of air moving towards the east. A negative sign indicates air moving towards the west. This parameter can be combined with the V component of wind to give the speed and direction of the horizontal wind.
V-component of wind	$m\ s^{-1}$	This parameter is the northward component of the wind. It is the horizontal speed of air moving towards the north. A negative sign indicates air moving towards the south. This parameter can be combined with the U component of wind to give the speed and direction of the horizontal wind.
Vertical velocity	$Pa\ s^{-1}$	This parameter is the speed of air motion in the upward or downward direction. The ECMWF Integrated Forecasting System (IFS) uses a pressure based vertical co-ordinate system and pressure decreases with height, therefore negative values of vertical velocity indicate upward motion. Vertical velocity can be useful to understand the large-scale dynamics of the atmosphere, including areas of upward motion/ascent (negative values) and downward motion/subsidence (positive values).
Vorticity (relative)	s^{-1}	This parameter is a measure of the rotation of air in the horizontal, around a vertical axis, relative to a fixed point on the surface of the Earth. On the scale of weather systems, troughs (weather features that can include rain) are associated with anticlockwise rotation (in the northern hemisphere), and ridges (weather features that bring light or still winds) are associated with clockwise rotation. Adding the effect of rotation of the Earth, the Coriolis parameter, to the relative vorticity produces the absolute vorticity.

4.1.4 Description of FDM

Flight Data Monitoring (FDM) is an activity which routinely captures, analyses and/or visualizes recorder data generated by an A/C in order to improve flight safety and increase overall operational efficiency by offering the ability to track and evaluate flight operation trends, identify risk precursors, and take the appropriate remedial action. In 1990's only few countries including France and India, mandated FDM as part of their airline safety management system. However, on January 1st, 2005, ICAO introduced Amendment 26 to ICAO Annex 6 – Operation of A/C, which mandated airlines to implement FDM program as part of its accident prevention and flight safety program. The following is an extract of the Amendment 26, ICAO Annex 6 [21]

> *3.6.3: An operator of an aeroplane of maximum certificated take-off weight in excess of 27,000kg shall establish and maintain a flight data analysis programme as part of its accident prevention and flight safety programme.*

Since then, there has been significant adoption of FDM as a mandatory requirement in most countries, with the notable exception of the USA, where FDM has been introduced as one of voluntary safety initiatives. Data recorded within FDM – usually known as FDM data – is often more comprehensive than those of the crash-protected flight data recorder (FDR), due to the increased capacity of its recorder. It contains over 2.000 flight parameters and has an easily removable recording medium, hence technically named Quick Access Recorder (QAR) data. Just like FDR data, QAR data is obtained from the A/C's digital systems by a Flight Data Acquisition Unit (FDAU). The data is then stored in a bit stream and a standardized structure format. Commonly, ARINC standard (573/717 or the latest ARINC 767 standard) is used to pack and organize the bit data stream structure. To read the data in engineering format, data analysts utilizes an FDM software which is capable of decoding binary data format into readable format. The decoded flight data is then analyzed by data analysts to identify and quantify risks in the airline operation. The cycle of flight data monitoring activity in an airline operation is depicted in the following Figure 34.

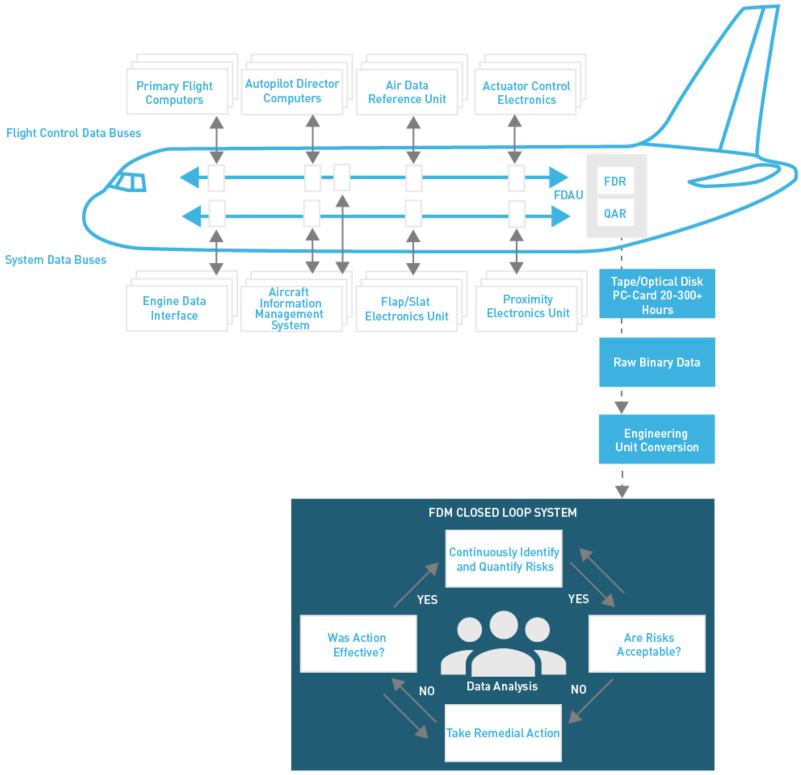


Figure 34: Flight Data Monitoring Cycle in Airline Operation [22]

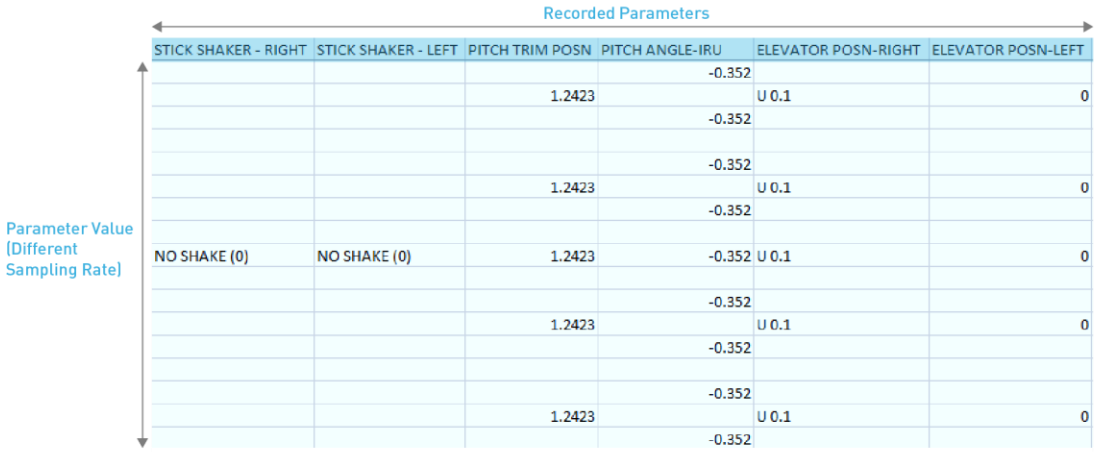
Besides the recorded/measured data, FDM data also contains so-called derived/computed parameters, which are computed during flight or on the ground using algebraic/physical formula. For example, True airspeed is a non-measured parameter, hence it is computed during flight using algebraic relation TAS =a0·M·(√T/T0)

Structure and Size

There are two data formats of FDM data set which can be used further analysis, i.e. decoded data (parameter in engineering units) and raw data (binary format). Both of these data formats are described below.

1. Decoded Data

The decoded data format contains recorded and derived parameters in engineering units. Usually this data format is stored in csv or in tab separated file which obtained from flight data software used by the airline. The size of a single flight of the decoded data format is huge compared to its original raw binary data, e.g. long haul A/C decoded data format might have file size of 1 GB for a single flight while in raw binary format the size of the file is only about 20 MB. The picture below depicts the extract of decoded parameters in tabulated format.



STICK SHAKER - RIGHT	STICK SHAKER - LEFT	PITCH TRIM POSN	PITCH ANGLE-IRU	ELEVATOR POSN-RIGHT	ELEVATOR POSN-LEFT
			-0.352		
		1.2423		U 0.1	0
			-0.352		
			-0.352		
		1.2423		U 0.1	0
			-0.352		
NO SHAKE (0)	NO SHAKE (0)	1.2423	-0.352	U 0.1	0
			-0.352		
		1.2423		U 0.1	0
			-0.352		
			-0.352		
		1.2423		U 0.1	0
			-0.352		

Figure 35: Extract of FDM Decoded Data in Tabulated Format

2. Raw Binary Data

The Flight Data Acquisition Unit (FDAU) takes the A/C sensor inputs and stores the data in a bit data stream for recording onto Quick Access Recorder device. This bit data stream is the so-called raw binary data in which the measured parameters stored in certain pattern. Decoding the binary data into engineering unit requires data frame layout (DFL) documents which depend on the recording system's type. ARINC 717 is one of the data frame layout standard used for QAR data. The document describes:

- The programming method used by the data acquisition system (location of parameters, number of bits used to encode parameters, type and method of encoding)
- The functions used to convert the recorded value into the actual physical value. For each parameter, the conversion function is checked with the calibration of the measuring and processing channel.

ARINC 717 standard classifies bit data stream into 5 categories

Table 51: Bit Data Stream categories - ARINC 717 Standard

bit	smallest unit whose value is either 0 or 1
word	12 bits are packed to one word
subframe	each second one subframe is recorded. One subframe can contain between 64 and 1024 words
frame	4 subframe are combined to a single frame. The frame pattern is repeated after 4 seconds

superframe	16 frames are joined to one superframe. A superframe pattern repeats itself each 64 seconds.

Data Features

The FMD data set/QAR data contains thousands of parameters. However, they can be classified based on the A/C system data buses

- Engine data
  - N1, N2, N3
  - EPR
  - Engine temperature
  - …
- A/C information management system
- Control surface electronic unit
  - Elevator, rudder aileron, slat, flap, THS, spoiler
  - …
- Proximity electronic unit
- Air data reference unit
  - Airspeed, angle of attack, barometric altitude
  - …
- Actuator control unit

Each parameter is recorded in different sampling rate with a maximum sampling rate up to 16 Hz. Table 52 gives some examples of the information that can be obtained from FDM.

Table 52: FDM Parameters

	Parameter	Description
1	Time	GMT (HH:MM:SS)
2	Latitude	Latitude in degrees
3	Longitude	Longitude in degrees
4	Departure Airport	ICAO code of departure airport
5	Arrival airport	ICAO code of arrival airport
6	Date	Start date of flight (dd/mm/yyyy)
7	A/C type	A/C type
8	Flight ID	Flight ID
9	Standard Altitude (MSL)	Standard Altitude (MSL), ft
10	Radio Altitude	Radio Altitude, ft
11	IAS - Indicated Airspeed	Indicated Airspeed, kt
12	CAS - Computed Airspeed	Computed Airspeed, kt
13	GS - Ground Speed	Ground Speed, kt
14	Vapp - Approach Speed	Approach Speed, kt
15	Vertical Acceleration	Vertical Acceleration, g
16	Wind Direction	Wind direction, deg
17	Wind Speed	Wind Speed, kt
18	Flaps Configuration	Flaps Configuration
19	Pitch Angle	Pitch angle of the A/C, deg
20	Pitch Rate	Body Pitch Rate, deg/sec

21	Heading - Yaw Angle	Heading of the A/C, deg
22	Drift Angle	Drift angle (FMC), deg
23	Speed Brake Applied	Speed Brake Applied, -
24	Gross Weight	Gross Weight of A/C, kg
25	TO/GA Switch Pressed	TO/GA Switch Pressed (NOT-PRESSED / PRESSED)
26	Landing Gear Lever Position (Up/Down)	Landing gear selection UP / DOWN
27	N1 (of all engines)	N1 (of all engines), %RPM
28	N2 (of all engines)	N2 (of all engines), %RPM
29	Engine Fuel Flow (of all engines)	Engine Fuel Flow (of all engines), kg/hr
30	TAWS (GPWS/EGPWS) Alert	TAWS (GPWS/EGPWS) Alert

4.1.5 Description of METAR

The Meteorological Terminal Air Report (METAR) is a standard format for reporting weather information near and at a given airport. METAR weather reports are predominantly used by pilots in fulfilment of a part of a pre-flight weather briefing, and by meteorologists, who prepare aggregated METAR information to assist in the planning of the operation.

METARs typically come from airports or permanent weather observation stations. Reports are generated with 30 or 60 minutes time frequency, but if conditions change significantly, a special and unscheduled report (known as SPECI) may be issued. Some METARs are encoded by automated airport weather stations located at airports, military bases, and other sites; on the other hand, some locations still use augmented observations, which are recorded by digital sensors, encoded via software, and then reviewed by certified weather observers or forecasters prior to being transmitted. Observations may also be taken by trained observers or forecasters who manually observe and encode their observations prior to transmission. Due to this, and in spite of being a well-defined and consolidated standard, METAR reports may actually contain errors and misspellings.

Raw METAR is the most common format in the world for the transmission of observational weather data. It is highly standardized through the International Civil Aviation Organization (ICAO), which allows it to be understood throughout most of the world.

Structure and Size

METAR data can be extracted from the raw messages, as described below, which would be saved in CSV files.

The approximate volume would be 3 MB / year / aerodrome.

Data features

The METAR data are originally encoded in a text message, which was historically transmitted through Morse code radio messages. To illustrate, the following text corresponds to the METAR for Madrid Barajas airport, broadcasted on 2017.04.27 at 12:30:

METAR LEMD 271030Z 03011G21KT 360V080 9999 FEW075 11/M03 Q1014 WS R36L WS R36R NOSIG=

Some information that can be highlighted include:

- There is wind coming from 030º, of 11 knots, with gusts of 21 knots. The wind is also variable between 360º and 080º.
- Some clouds are present at FL 075
- The QNH at the airport is 1014.

In order to understand the overall information collected by METAR, we include here below in Table 53 the main items METAR source provides (e.g. the ICAO aerodrome indicator, field D, and the period of validity, fields B and C), besides other meteorological variables.

Additional and more detailed information about the METAR format can be found in the NASA/TM—2014–218385 report by Max Lui [23].

Table 53: METAR Dataset Description

Column	Parameter	Description
A	ID METAR MESSAGE	
B	DATE & HH:MM	UTC hour corresponds to the hour of the last meteorological observation. (Valid from)
C	DATE & HH:MM	(Valid to)
D	OACI ID	OACI AERODROME INDICATOR. It corresponds to the OACI code of the Aerodrome. e.g. LEMD
E	MESSAGE	Type of message, it can be: • METAR: Aerodrome Routine Meteorological Report. Sent every hour. • SPECI: Similar to the METAR message sent not evenly but punctually.
F	WIND DIRECTION (degrees)	Average direction from which the wind blows 10 minutes before the observation rounded to the nearest ten degrees. If: • the velocity is less than 3KT and the variation of the direction is equal or higher than 60 degrees • the velocity is equal or higher than 3KT and the variation of the direction is equal or higher than 180 degrees or and an indeterminate direction Then, the direction is considered as variable and this field is filled out with "VRB" **VARIABLE DIRECTION:** It is filled out with "1" when the last field has a "VRB".
G	WIND INTENSITY (knots)	
H	WIND GUST INTENSITY (knots):	The maximum velocity during the previous 10 minutes to the observation (only if this velocity is equal or higher than the intensity indicated + 10).
I	WIND MINOR DIRECTION of the TOTAL VARIATION OF THE WIND DIRECTION	Used when there are extreme directions, it means, the velocity is equal or higher than 3KT and the variation of the direction is between 60

		and 180 degrees. There are two fields to show the interval.
J	WIND LARGEST DIRECTION of the TOTAL VARIATION OF THE WIND DIRECTION	As above
K	VISIBILITY	• PREDOMINANT (m): Is the one observed from the aerodrome (360 degrees). • PARTICULAR CASES: If: ○ The visibility is higher than 10km the field is filled out with “9999”. ○ The visibility is lower than 50m the field is filled out with “0000”.
L	VISIBILITY	MINIMUN (m): Is the minimum visibility if and when this one is less than 1500 m or 50% than the predominant visibility.
M	CAVOK (Ceiling and visibility OK	This term substitutes the groups of visibility, RVR, significant time and cloudiness or vertical visibility when simultaneously: • Visibility is higher than 10km • There is lack of clouds under 5000 feet or under the highest minimum altitude of the sector when this one is higher than 5000 feet and without cumulonimbus.
N	DIRECTION OF MINIMUN VISIBILITY	Is the direction of the minimum visibility with respect to one of the eight cardinal points (N, NE, E, SE, S, SW, W, and NW).
O	RAW METAR MESSAGE	

4.1.6 Description of SIGMET

SIGMET, or Significant Meteorological Information, is a weather advisory that contains meteorological information concerning the safety of all A/C. There are two types of SIGMETs - convective and non-convective. The criteria for a non-convective SIGMET to be issued are severe or greater turbulence over a 3,000-square-mile (7,800 km2) area, severe or greater icing over a 3,000-square-mile (7,800 km2) area or IMC over a 3,000-square-mile (7,800 km2) area due to dust, sand, or volcanic ash.

This information is usually broadcast on the ATIS at ATC facilities, as well as over VOLMET stations. They are assigned an alphabetic designator from N through Y (excluding S and T). SIGMETs are issued as needed, and are valid up to four hours. SIGMETS for hurricanes and volcanic ash outside the CONUS are valid up to six hours.

A Convective SIGMET is issued for convection over the Continental U.S. Convective SIGMETs are issued for an area of embedded thunderstorms, a line of thunderstorms, thunderstorms greater than or equal to VIP level 4 affecting 40% or more of an area at least 3000 square miles, and severe surface weather including surface winds greater than or equal to 50 knots, hail at the surface greater than or equal to 3/4 inches in diameter, and tornadoes. Severe thunderstorms are characterized by tornado(s), hail 3/4 inches or greater, or wind gusts 50 knots or greater. A Convective SIGMET is valid for 2 hours and they are issued at every full hour + 55 min.

Structure and Size

SIGMET message format:

First line: Location indicator of the ATS dependence, Identification and number of the message series, validity period (UTC) and Meteorological office that sends the message.

Following lines: Indicative + FIR name of the message destination, Meteorological phenomenon and description, Observed and/or predicted, Place and flight levels, Movement, direction and speed, Intensity change.

SIGMET message size: Excel file:1.5MB/year

Data features

Table 54: Data Items in first line of Airspace Configuration Data Set for SIGMET

	Column	Description
1	Location indicator	The ICAO location indicator of the ATS unit serving the FIR or CTA to which the SIGMET refers
2	Message identifier	The message identifier is SIGMET
3	Sequence number	The daily sequence number in the form [n][n]n, e.g. 1, 2, 01, 02, A01, A02, restarts every day for SIGMETs issued from 0001 UTC
4	Validity period	The validity period is given in the format VALID YYGGgg/YYGGgg where YY is the day of the month and GGgg is the time in hours and minutes UTC. The period of validity for a WS SIGMET shall be no more than 4 hours.
5	Issuing Office	The ICAO location indicator of the MWO originating the message followed by a hyphen.

Table 55: Data Items in second line of Airspace Configuration Data Set for SIGMET

	Column	Description
1	FIR/CTA Name	The ICAO location indicator and full name of the FIR/CTA for which the SIGMET is issued in the form CCCC FIR[/UIR] or CCCC CTA
2	Phenomenon	OBSC TS → Obscured thunderstorms EMBD TS → Embedded thunderstorms FRQ TS → Frequent thunderstorms SQL TS → Squall line thunderstorms OBSC TSGR → Obscured thunderstorms with hail EMBD TSGR → Embedded thunderstorms with hail FRQ TSGR → Frequent thunderstorms with hail SQL TSGR → Squall line thunderstorms with hail SEV TURB → Severe turbulence SEV ICE → Severe icing SEV ICE (FZRA) → Severe icing due to freezing rain SEV MTW → Severe mountain wave HVY DS → Heavy dust storm HVY SS → Heavy sandstorm

		RDOACT CLD → Radioactive cloud
3	Observed or forecast phenomenon	Whether the phenomenon is observed or forecast in the form OBS [AT GGggZ] or FCST [AT GGggZ] where GG is hours and gg minutes UTC
4	Location	The location of the phenomenon is provided with reference to geographical coordinates in latitude and longitude in degrees and minutes.
5	Level	The level and vertical extent of the phenomenon: FLnnn or nnnnM or nnnnFT or SFC/FLnnn or SFC/nnnnM or SFC/nnnnFT or FLnnn/nnn or nnnn/nnnnFT or TOP FLnnn or ABV FLnnn or TOP ABV FLnnn.
6	Movement or Expected Movement	Direction and rate of movement of the phenomenon where the direction is given with reference to one of the sixteen points of the compass (using the appropriate abbreviation) and the rate is given in KT (or KMH)
7	Changes in Intensity	The expected evolution of the phenomenon's intensity as indicated by: INTSF or WKN or NC
8	Forecast time and forecast position	The forecast position of the hazardous phenomena at the end of the validity period of the SIGMET message in the form FCST AT Z .

4.1.7 Description of TAF

In meteorology and aviation, terminal aerodrome forecast (TAF) is a format for reporting weather forecast information. A TAF is distinguished from a METAR by its multiple date/time groups. Once published, the TAF remains constant until its next publication, with the only exceptions being significant traffic shifts by major airlines or a significant data error. Forecasts are prepared for major users of air carrier, air taxi/commuter, general aviation, and military. The forecasts provide information for use by state and local authorities, the aviation industry, and the public.

In the United States the weather forecasters responsible for the TAFs in their respective areas are located within one of the 122 Weather Forecast Offices operated by the United States' National Weather Service. In contrast, a trend type forecast (TTF), which is similar to a TAF, is always produced by a person on-site where the TTF applies. In the United Kingdom most TAFs at military airfields are produced locally, however TAFs for civil airfields are produced at the Met Office headquarters in Exeter.

Structure and Size

TAF data can be extracted from the raw messages, as described below, which would be saved in CSV files.

The approximate volume would be 150 MB / month / main European aerodromes.

Data features

TAFs complement and use similar encoding to METAR reports. They are produced by a human forecaster based on the ground. For this reason there are considerably fewer TAF locations than there are airports for which METARs are available. TAFs can be more accurate than Numerical Weather Forecasts, since they take into account local, small-scale, geographic effects.

Table 56: TAF Info

	Column	Description
1	Type of Report	The report type header will always appear as the first element in the TAF forecast
2	ICAO Station Identifier	The TAF code uses the ICAO four-letter location identifiers
3	Date and Time of Origin	This element is the UTC date and time the forecast is actually prepared. The format is a two-digit date and four-digit time followed, without a space, by the letter z
4	Valid Period Date and Time	The UTC valid period of the forecast is a two-digit date followed by the two-digit beginning hour and two-digit ending hour.
5	Wind	The wind group includes forecast surface winds. The surface wind is the expected wind direction (first three digits) and speed (last two or three digits if 100 knots or greater)
6	Visibility	The expected prevailing visibility is forecast in statute miles and fractions of statute miles followed by SM to note the units of measure
7	Weather	The expected weather phenomenon or phenomena is coded in TAF reports using the same format, qualifiers, and phenomena contractions as METAR reports
8	Sky Condition	TAF sky condition forecasts use the METAR format
9	Temperature Forecast	Informs the atmospheric temperature range on that airport during the validity period of the TAF. The maximum temperature is indicated in TX group, while the minimum temperature is informed in TN group.
10	Probability Forecast	The probability or chance of thunderstorms or other precipitation events occurring, along with associated weather conditions (wind, visibility, and sky conditions).
11	Forecast Change Indicators	Change indicators are used when either a rapid, gradual, or temporary change is expected in some or all of the forecast meteorological conditions. Each change indicator marks a time group within the TAF report.

4.2 Data Processing Pipeline

Data processing pipelines are key components of any data-driven projects. They describe the flow of data from collection, data lake, cleaning and preparation ready to be use in an analysis. It is especially necessary to develop robust pipelines when working with handling messy, inconsistent or un-standardized data and when there is a need of combining data from multiple sources. Another of the main needs for establishing a data processing pipeline is to be able to audit, verify and fulfil the requirement of identifying and correcting errors and/or anomalies in the data. They can be implemented using a myriad of different technologies and techniques, but a data pipeline in general should address the following challenges:

- How do we integrate data from disparate data sources, some of them possibly redundant?
- How do the data look like? Do they need any transformation, to optimize their format and reduce the computational and storage overhead?
- How clean is the data set? Does the system require specific algorithms to handle missing values, and to detect and filter outliers?
- How do we handle temporal (time evolving) data?

In SafeOPS the usage of the BeSt data infrastructure will address these challenges. It will help to avoid redundancies, optimizing workflow and minimize the overhead in collection, storage and processing of the data. Moreover, it will also provide scalable computational power, if needed, to speed up processing tasks. The focus will be on those tasks we foresee as needed in the data processing and machine learning pipelines that are going to be designed.

Figure 36 below presents an overview of the data processing pipeline with examples of expected tasks and actions. The thing to note here is that the main activities of data cleaning and data preparation are iterative processes and not one off actions. They can ,and should be repeated, if the problem conditions change. for example if new de-identification requirements arise, new features are needed or if any errors are detected.

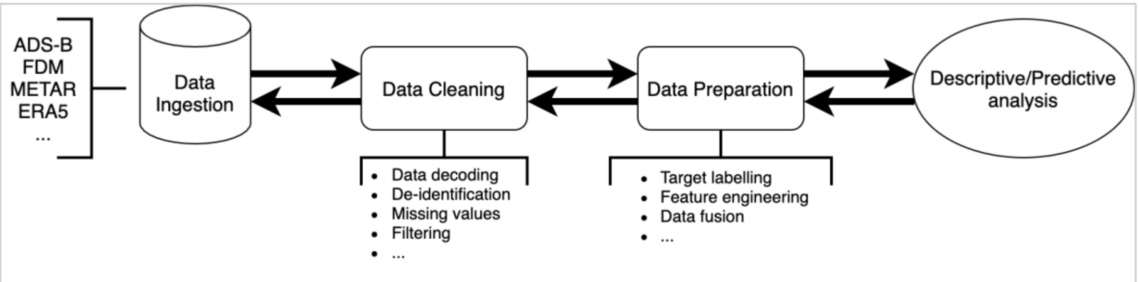


Figure 36: Workflow of Data Processing Pipeline

4.2.1 Data Cleaning

The Data Cleaning step addresses mainly the fixing or removing of **incorrect, corrupted, incorrectly formatted, duplicate**, or **incomplete** data within a dataset to obtain a minimum readable dataset that enables basic data analytics. Usually, a **data cleaning module** should be constructed for each data source (and over the full scope of the data needed for a specific use case, temporally and geographically) that checks that no data field contains erroneous values, applying corrections such as dropping or substituting the values if possible. This is because real world data often contain corrupted data with noise, either systematic or random. The ability to detect unreal, noisy or biased data before making a prediction is essential to ensure the reliability of results. Although there is no unique way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. Exhaustive experience on data cleaning, also for data sets used in SafeOPS was gained during the H2020 project SafeClouds.eu [19] and is available to the SafeOPS team. In case new data sources are used, similar processes have to be implemented for these data sets. Some of the main tasks to be performed are:

- **Remove duplicate/irrelevant data:** When merging different data sources certain data points may be duplicated and these should be eliminated so as not to change the actual distribution of the data. Removing duplicate data can be one of the largest areas to be considered in the data cleaning process. Irrelevant data refers to data points that do not fit into the specific problem. For example, if you want to analyse data relating to approaches to a specific airport, but the dataset includes operations in a region, it may be necessary to remove flights that do not land at the selected airport. Although less necessary than the removal of duplicates, the removal of irrelevant duplicates can help streamline the analysis.
- **Data structural errors**: It is very important to find and fix data inconsistencies (e.g. naming conventions, typos,...) which can cause mislabelled categories or classes.
- **Data outliers**: Outliers are data points significantly different from other data points in a data set. Detecting outliers can be a somewhat subjective practice. If an outlier proves to be irrelevant for analysis or is a mistake, consider removing it as outliers increase the variability

in your data and tend to decrease performance. It is essential to understand how outliers occur in a specific data set and whether they the possibility of them appearing again is expected.

- **Missing data:** Handling missing data is important because many machine learning algorithms do not accept missing values. There are two main ways to deal with missing data. The first one is to drop all data points with missing values with the drawback that this may cause loss of information. The second one is to input missing values based on other data points with the drawback that this may cause loss of data integrity.

4.2.2 Data Preparation

The Data preparation step addresses the final manipulation tasks to transform a raw data set into the correct form to be used for descriptive and predictive analytics. As with the data cleaning step, there is no unique way to prescribe the exact steps in the data preparation process because the processes will vary from dataset to dataset. Some of the main tasks to be performed are:

- **Data transformation:** Data transformation actions are used to change the type or distribution of data variables in a data set. Data values may have one of a few types, such as numeric or categorical, with subtypes for each, such as integer and real-valued for numeric, and nominal, ordinal, and boolean for categorical. For example, for most classifications model it is necessary to encode a categorical variable as integers or Boolean variables (e.g. ordinal transform or one-hot transform). Also, for most machine learning models if data has a Gaussian probability distribution performance is increased if the data to shift to a standard Gaussian with a mean of zero and a standard deviation of one (e.g. Normalization transform or Standardization transform).
- **Data labelling:** This is a key part of the data processing pipeline as it specifies which parts of the data the model will learn from in supervised learning algorithms. For example, to be able to develop a Go-around prediction tool it is necessary to identify not only those flights that do perform a Go-around maneuver but also the exact data point at which this maneuver is performed.
- **Feature engineering:** It refers to the process of creating new input variables from the available data to enhance it. Feature engineering allows you to identify and define the most important information in your data set and utilize domain expertise to get the most out of it. This might mean breaking data into multiple parts to clarify particular relationships or defining features that better represent patterns for your machine learning model.
- **Data fusion**: Machine learning model only admit one final data structure to be trained on. All the information from various data sets and data sources needs to be fused so that all the recordings for one observation are properly delivered in that single data format.
- **Validate and QA:** At the end you should be able to determine the quality of the data. Bad data can lead to erroneous conclusions in the decision-making process and compromising the safety of operations. Some characteristic to assess the quality of the data are:

    - Validity: The data conforms to the constraints of the problem
    - Accuracy: Data is a close to true values as possible
    - Completeness: All required data is known
    - Consistency: Data needs to be consistent across all used data sets
    - Uniformity: Data across data sets is specified using the same unit of measure

4.3 Model Training and Validation

4.3.1 Performance Assessment

Correctly evaluating the performance of an algorithm is one of the most delicate actions in a machine learning project. Classical performance metrics such as accuracy are usually not enough and can be dangerously misleading. Because of this, we will review some of the existent performance metrics and their utility in data mining problems.

4.3.1.1 Bias-Variance Trade-off

To understand the challenges involved when evaluating a machine learning algorithm, it is important to start by understanding the concepts of bias and variance. These two metrics are crucial for understanding the reasons of failure of our model and the potential measures to improve the data fitting process to obtain a more accurate data mining model.

- **Bias** measures the disparity between the prediction and the correct value (Accuracy). Therefore, an error due to bias is explained as the difference between the expected prediction of our model and the correct value which we are trying to predict.
- **Variance** measures how much the predictions for a given point vary between different realizations of the model (Precision). Therefore, the error due to variance is taken as the variability of a model prediction for a given data point.

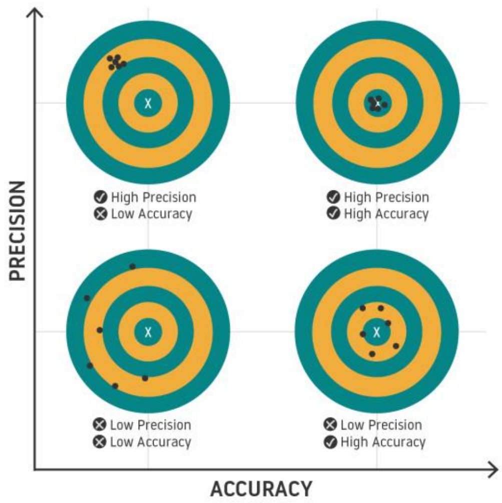


Figure 37: Precision vs. Accuracy [24]

Solving the bias-variance trade-off is not straightforward and is one of the main challenges of any machine learning project. It is important to understand that the bias and variance are equally important especially in solutions such as the one explored in the SafeOPS project. Workshops with the ATCOs (**FR.C.03**) have shown how Go-Around predictions, correct or incorrect, can have a direct impact on the safety meaning that most probably a solution that does not have high levels of precision and accuracy can be viable.

Bias and variance also are closely related with **over-fitting** and **under-fitting**. In simple models, our main concern is a low bias related with under-fitting, meaning the model is too simple to explain the data. For more complex models, bias is reduced and variance is increased because our model fits all the data. However, if our model gets "too complex", it might fit the data and some noise, increasing variance too much. This is what we understand as over-fitting.

4.3.1.2 Binary classification problem

Binary classification problems are a type of **supervised learning** model in which the objective is simply to predict the class of given data value. The main performance metric used in this type of problem is the so-called **confusion matrix**.

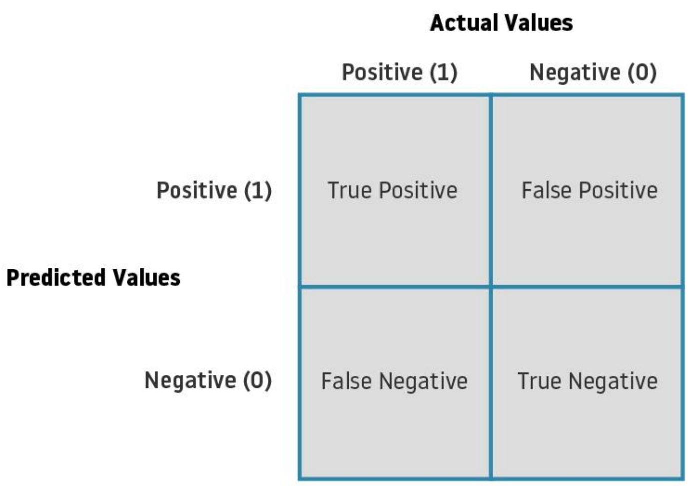


Figure 38: Example of Confusion Matrix [25]

A Confusion matrix is simply a table that describes the performance of a classification model (outputs can be of two or more classes). Four different values can be found:

- **True Positive (TP)**: When the model predicts a “1” and the actual data is also a “1” → A Go-Around is predicted and the flight does so.
- **True Negatives (TN)**: When the model predicts a “0” and the actual data is also a “0” → A flight that lands and was predicted to do so.
- **False Positive (FP)**: When the model predicts a “1” and the actual data is a “0” → A Go-Around is predicted but it actually lands. Also known as **Type I Errors**.
- **False Negative (FN)**: When the model predicts a “0” and the actual data is a “1” → A flight is predicted to land but actually a Go-Around occurs. Also known as **Type II Errors**.

A perfect model would be that in which False Positives and False Negatives are both zero although this is practically impossible in reality. When assessing the results, there is no standard action plan to follow on what should be minimized. This would depend completely on the scenario. For example, it could be concluded that as an error Type II in the prediction of Go-Arounds has a direct effect on safety while a Type I error has an effect in the capacity the former is more critical and should therefore be minimized. Other of the most commonly used metrics for the evaluation of a binary classification model include:

- **Accuracy**: the correct number of predictions made over all the predictions made by the model. It is a metric that must be used with a lot of caution as it can be misleading if used solely and provides very **limited information** in imbalanced data distributions.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- **Recall**: this metric informs us of how well the model predicts the positive (1) events of the data. In our example, it would show us the proportion of flights that perform a Go-Around and that were predicted correctly by the model. The main limitation is that they are insensitive to imbalanced data distributions. Used alone, they are not enough for imbalanced data assessment.

$$Recall = \frac{TP}{TP + FN}$$

- **Precision**: this metric informs us of how often is the model correct when it predicts a positive (1) event. When we predict that a flight is going to perform a Go-Around, how often does it get it right. As with the recall, the main limitation is that they are insensitive to imbalanced data distributions. Used alone, they are not enough for imbalanced data assessment.

$$Precision = \frac{TP}{TP + FP}$$

- **Specificity**: is the complete opposite of Recall. This metric tells us how well is our model in predicting the negative (0) events of our data.

$$Specificity = \frac{TN}{TN + FP}$$

- **F1-score**: is a metric that combines both Precision and Recall. This combination is made using the Harmonic mean. The main limitation is that it gives the same relevance to Precision and Recall (False Positives and False Negatives). In reality, as mentioned before, there can be different costs for the possible misclassifications.

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

- **Kappa (Cohen's Kappa)**: The kappa statistic is an observed accuracy with an expected accuracy. Kappa, as a scalar performance metric, reflects prediction performance of smaller classes (normally hidden behind the prediction performance of larger classes). Using a metric like Kappa to measure performance will not necessarily increase how a model fits to the data. This metric might be used for selecting a best suited model type and hyper-parametrization amongst multiple choices for a **very imbalanced problem**.
- **Receive Operating Characteristics (ROC) curve**: arguably is the most common way of visualizing how well a classifier works. It is graph where True Positives and False Positive values are plotted for all possible classification thresholds [0…1]. The most widely used metric for performance is **area under the curve (AUC)**. As the name suggests, this metric is nothing more than the value of the area that is under the ROC of the model. The AUC can be used to compare the performance of two or more classifiers. A single threshold can be selected and the classifiers' performance at that point compared, or the overall performance can be compared by considering the AUC. However, when analyzing imbalanced datasets, the ROC does not vary compared to a balanced setting.
- **Precision-Recall (PR) curves:** Plots the precision rate over the recall rate. A curve dominates in ROC space (resides in the upper-left hand) if and only if it dominates (resides in the upper-

right hand) in PR space. PR space has all the analogous benefits of ROC space. Additionally, the PR curves provide informative representations of **performance assessment under highly imbalanced data.**

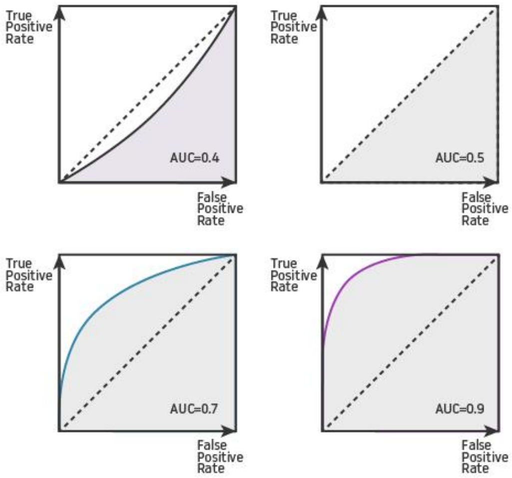


Figure 39: Example ROC-AUC Curves [25]

4.3.1.3 Training and Cross-Validation

Equally as important as selecting the right metrics and models is the process of training, testing and validation of data. The objective being able to assess how well the trained model will generalize on independent/unseen data. Training a model and testing it on the same data may cause the problem of overfitting. In order to avoid this issue, it is standard practice when performing a (supervised) machine learning experiment to hold out part of the available data as a **test set**. Figure 40 provides a flowchart of typical cross validation workflow in model training. Usually, techniques such as random search or grid search are used to determine the best parameters and different types of cross validation methods are typically conducted during the process of training the dataset

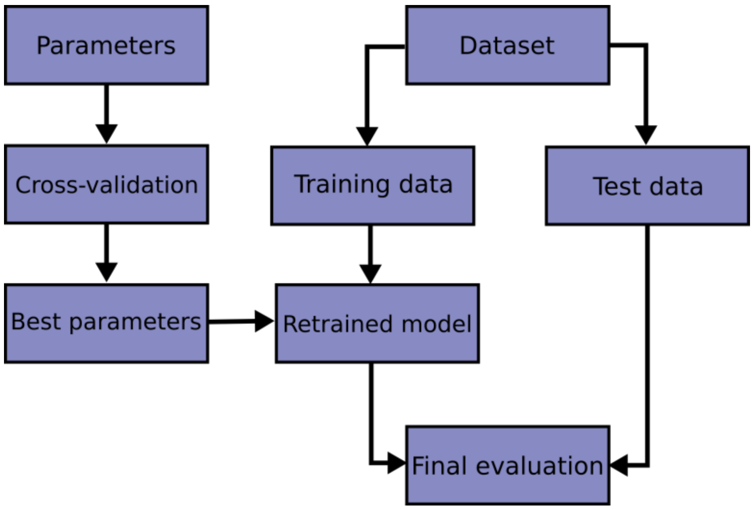


Figure 40: Cross-Validation Workflow [26]

Cross validation mainly involves partitioning a sample of data into complementary subsets for training and testing purposes. In order to reduce the variability, in most of the methods multiple rounds of cross-validation are performed using different partitions. The validation results are then combined (e.g., averaged) over the rounds to obtain an estimation of the model's predictive capabilities. We can perform cross-validation using different strategies:

- **Holdout method**: Consists in removing a part of the training data and using it to get predictions from the model trained using the remaining data - e.g., splitting a data set into two data sets, with one being a "training" dataset containing 70% of the data and another being a "test" dataset containing 30% (the remaining samples). Although this method is better than traditional model validation, it suffers from **high variance**. This is because, at certain points, it is unclear which data points will end up in the validation set. The result might be very different for different sets. Also, by reducing the training data, we risk losing important patterns and trends in the data, which will increase the error due to the bias.
- **K-Fold Cross-Validation**: The data is divided in **k subsets** and the holdout method is repeated k times. Each time one of the k subsets is used as the test dataset and the other k-1 subsets are used as the train dataset. The error estimation is averaged over all k trials to calculate the total effectiveness of the model. This methodology **reduces bias** because we use most of the data for fitting, and also **reduces the variance** because most of the data is also used in the test dataset. **Stratified Cross-Validation** is a variation in the K-Fold Cross-Validation useful when dealing with **imbalanced datasets**, such that each fold contains approximately the same percentage of samples of each target class or the mean response value is approximately equal in all the folds.
- **Leave-P-Out Cross-Validation**: This approach **leaves p data points out** of our training data. I.e., if there are n data points, we will use n-p for training and p for testing. This is repeated for all combinations in which the original sample can be separated using this proportions. Then, the error is calculated as the average of all. The drawback of using this method is obvious as it needs to train and validate **all combinations** and can be infeasible for a large p value. A particular case of this method is when p=1, which is known as **leave-1-out cross-validation**. It yields a number of combinations equal to the number of points in the original dataset.

4.3.2 Explainability and Interpretability

Over the last decade there has been a vast interest in the use of machine learning models. What was previously relegated to academia has made its way for use in real-world problems. But on many occasions the use of powerful machine learning models has been held back by problems related with **explainability** and **interpretability**. There are some domains in the industry where data scientists often end up having to use more traditional machine learning models (linear or tree-based). The reason being that model interpretability is very important for the business to explain each and every decision being taken by the model even at the expense of sacrificing performance. Complex models like ensembles and neural networks typically provide better levels of performance but tend to be less interpretable. Throughout the workshops, the issues of explainability and interpretability is something that has emerged among the ATCOs **(FR.C.04)** as very relevant in order to ensure their **confidence** and **trust** in the tool. So it will be an important aspect of the SafeOPS project already foreseen in WP3 and WP4.

4.3.2.1 Definition of Explainability and Interpretability

In machine learning, the terms explainability and interpretability are often used interchangeably. While in practice they are very closely related it is important to understand it's differences.

Explainability: refers to the extent to which the **internal mechanics** of a machine or deep learning system can be explained in human terms. It provides the reasons for the behavior of a model or produce some insights about the causes of the models decisions. Currently, most techniques for explaining models are "post-hoc," in that they are augmentations to models that are not generally scrutable.

Interpretability: refers to the extent to which a **cause and effect** event can be observed within a machine learning system. Interpretable models are those which describe the internals of a system in such a way that is understandable by humans. The main benefit of this type of models is that it is easy to trace what the model is doing and identify unintended or inappropriate decision-making factors. The key objective of model interpretation is **transparency** and **understanding model** decisions by humans. The most important aspects of model interpretation can be summarized as:

- The ability to find out latent feature interactions to identify which features are more important in the decision-making of the model. This helps to ensure **fairness** in the model.
- The ability to validate and justify why features are ranked as they are in the decision-making of the model. This helps to ensure **reliability** in the model.
- The ability to evaluate and validate any data point in the data set and how the model takes a decision based on it. This helps ensure **transparency** in the model.

4.3.2.2 Interpretability vs accuracy trade-off

As mentioned above, generally, the less complex a model is the more interpretability it has. This may mean that if an application needs of a high level of interpretability there is trade-off between lower accuracy and performance. This is because in most cases, simpler models tend to provide less accurate predictions while deep learning techniques tend to provide the highest performance at the expense of being very difficult and complex to interpret. In the end the type of model selected will depend on the problem one is trying to solve and the expected level of transparency requested by the users.

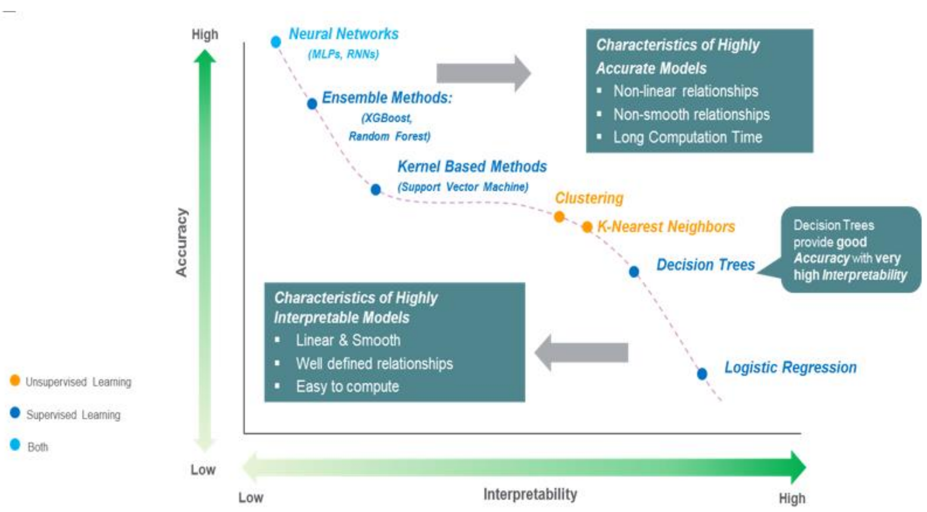


Figure 41: Interpretability vs Accuracy [27]

4.3.2.3 Machine Learning Model Interpretation Methods

There is not a clear consensus in the criteria for classifying interpretable methods since this is still an emerging field. However, a few specific criteria can be used for categorizing model interpretation methods such as:

- **Intrinsic or post hoc:** Intrinsic interpretation methods leverage on machine learning model which are already intrinsically interpretable in nature (e.g. linear models or tree based models). Post hoc interpretation methods are used when training a complex model (e.g. ensemble methods or neural networks) and applying interpretability methods after the training (e.g. feature importance).
- **Model-specific or model-agnostic:** Model-specific interpretation methods are specific to intrinsic model interpretation methods as they depend purely on the capabilities and features on a per-model basis (e.g. p-values, rules from a decision tree). Model-agnostic tools are more relevant to post hoc methods and can be used on any machine learning model. These agnostic methods work by analyzing feature input and output pairs but do not have access to any model internals like hyperparameters.
- **Local or global:** This classification relates to if the method is able to explain a single prediction or the just the entire model behavior.

4.3.3 Imbalanced Data

The reality in many machine learning supervised learning applications is that there is a significant difference between the prior probabilities of different classes. This is especially true in safety related events which are rare and, in some cases, extremely rare. It is estimated that **Go-Arounds occur with an average rate of 1-3 per 1000 approaches** [12] making it a very rare situation. This imbalance have an important impact in the performance of learning algorithms as the standard classification learning algorithms are often biased towards the majority class.

4.3.3.1 Types of Imbalance

Due to the inherent complex characteristics of imbalanced data sets, learning from such data requires new understandings, principles, algorithms, and tools to transform vast amounts of raw data efficiently into information and knowledge. Imbalance and can be found in different forms:

- **Between-class** and **within-class imbalance**: The former is the one expected in rare safety related events where "bad" occurrences are expected to be in the minority. Within-class imbalance occurs when the dataset has balanced between-class but one of the classes is not representative in some regions, i.e. the case that some specific "bad" event is even more rare than the others.
- **Intrinsic** vs **extrinsic imbalance**: Intrinsic imbalance is caused by the nature of the dataset, while extrinsic imbalance is caused by time, storage and other factors that limit the dataset or the data analysis. In SafeOPS it is expected that only intrinsic imbalance is found although no possibility should be ruled out.
- **Relative imbalance** vs **absolute rarity**: Sometimes the minority class may be outnumbered, though this is not necessarily rare. Therefore, data can be accurately learned with little disturbance. Note that in this case, although the data present imbalance, it is not necessarily bad (and could be even good when using certain classifiers). It is very important to determine whether the imbalance is relative or an absolute rarity.

- **Small sample size imbalance**: Datasets with high dimensionality and small sample size are quite normal in data science problems (e.g. face recognition, gene expression, etc.). This can also cause specific issues with certain machine learning algorithms, such as the failure of generalizing inductive rules.

In addition, as already mentioned in the previous section, when dealing with imbalanced datasets, one must be careful in deciding which performance metrics to use, because depending on which one is used misleading results can be obtained that can lead to suboptimal solutions. Among the best performance metrics to use when working with imbalanced datasets we can find:

- **Precision-Recall (PR) curves**
- **Receive Operating Characteristics (ROC) curves**
- **F-Score in combination with Kappa (Cohen's Kappa)**: for imbalanced data mining they are solid performance metrics, but they should not be used alone.

4.3.3.2 Balancing Data Techniques

The existence of imbalanced in datasets is a well-known and documented problem. There are plenty of available algorithms and implementations to tackle the problem. The three main types of imbalanced classification techniques are:

Data Sampling Algorithms

Data sampling algorithms basically consist of **changing the composition of the training data** set to modify the distribution of the data to create a "new" balanced data set and improve performance. There are three particular methodologies for this:

- **Data Oversampling** involves expanding the minority class through duplicating examples or by creating synthetic ones. The main limitation is that it can very easily cause overfitting due to the creation of multiple similar instances. Among the main methods are:
  - Random Oversampling
  - SMOTE (Synthetic Minority Oversampling Technique)
  - ADASYN (Adaptive Synthetic Sampling)
- **Data Under sampling** involves reducing the majority class randomly or using an algorithm. The main limitation is that it might cause loss of relevant information. Among the main methods use are:
  - Random Under sampling
  - Tomek Links
  - Edited Nearest Neighbors
- **Combined Oversampling and Under sampling**: Most oversampling method can be combined with most under sampling techniques. If implemented correctly this will provide a well-rounded approach. Among the main methods use are:
  - SMOTE and Random Under sampling
  - SMOTE and Tomek Links
  - SMOTE and Edited Nearest Neighbors

Cost-Sensitive Algorithms

Cost-sensitive algorithms work as modified versions of standard machine learning algorithms that incorporate the **costs of misclassification** when fitting the model on the training dataset. The main goal of these algorithms is to minimize the overall cost. The overall cost is usually measured by the Bayes conditional risk. These algorithms can be very effective when used on imbalanced datasets by configuring the cost of misclassification to be inversely proportional to the distribution of examples in the training dataset. Some machine learning algorithms can be configured to use cost-sensitive training such as, but not limited to:

- Support Vector Machines
- Artificial Neural Networks
- Bagged Decision Trees
- Random Forest

One-Class Algorithms

In some cases, instead of trying to balance the dataset it can be better to tackle the problem from a different perspective. Algorithms used mainly **outlier detection and anomaly detection** can be used successfully in imbalanced classification problems. Because the usual problem with imbalanced datasets is that there is very low occurrence in some classes, we could frame the detection as a rare events detection solution. Examples of one-class classification algorithms are:

- Elliptic envelope methodology
- Local outlier factor (LOF)
- Isolation Forests
- High dimensional outlier detection

4.3.4 Data Mining Algorithms

This section will present an initial evaluation of some of the current **state-of-the-art machine learning algorithms**. This evaluation will focus primarily on the main characteristics and performance of these algorithms as well as based on some of the initial requirements identified in the ATCOS's workshops (**FR.C.01, FR.C.04**). It should be noted that this evaluation should be taken with care because there are different factors that may influence the final performance of an algorithm. However, it provides an initial basic understanding of each algorithm and may be useful for the future development of the predictive layer. Finally, it is also worth mentioning that the algorithms tested during the development of the predictive layer will not necessarily focus solely on the algorithms shown here.

Table 57: State-of-th-Art Machine Learning Algorithms

Algorithm	Classi-fication / Regression	Training Speed	Average Perfor-mance	Parameter Tuning	Data Quality	Probability of class membership	Level of explain-ability
Linear regressions	Regression	Fast	Low	None	Bad performance with non-linear hypothesis Unable to handle a large number of features	N/A	High

Naïve Bayes	Classification	Fast	Low	Low	Handles well large datasets. Dependencies between features affects the classification performance. It assumes the normal distribution of numeric attributes.	No	High
KNN	Both	Fast	Low	Low	Performs poorly on high dimensional datasets Can handle noisy instances or instances with missing attribute values.	Yes	Medium
Decision trees	Both	Fast	Low	Medium	Can create biased trees if heavy class imbalance Performs well on large datasets Can learn complex functions	Not directly	Medium
SVM	Both	Medium	Medium	Medium	Bad performance with large datasets. Can handle multiple feature spaces. Bad performance with noisy data	Not directly	Medium/Low
Ensemble methods (e.g. Random forest, Gradient boosting)	Both	Slow	High	Medium / high	Robust to missing data Can learn non linear functions Good performance handling lots of features Good with imbalanced datasets	Not directly	Medium/Low
ANNs (e.g. MPL, CNN, LSTM)	Both	Slow (High resource consumption)	High	High	Very good using non linear data with large number of inputs Reliable handling many features Requires high volumes of training data and cases Good with imbalanced datasets	Not directly	Low

Linear regressions	Regression	Fast	Low	None	Bad performance with non linear hypothesis Unable to handle a large number of features	N/A	High
Naïve Bayes	Classification	Fast	Low	Low	Handles well large datasets. Dependencies between features affects the classification performance. It assumes the normal distribution of numeric attributes.	No	High
KNN	Both	Fast	Low	Low	Performs poorly on high dimensional datasets Can handle noisy instances or instances with missing attribute values.	Yes	Medium
Decision trees	Both	Fast	Low	Medium	Can create biased trees if heavy class imbalance Performs well on large datasets Can learn complex functions	Not directly	Medium

5 Conclusion and Outlook

This deliverable documents the SafeOPS project's initial set of scenarios, use cases, user stories and requirements, which build the fundament for the upcoming development phase. Therefore, the document first described the methodological approach, applied during the first phase of the project. Thereafter, the resulting documentation is presented. Based on the derived requirements, a technical problem statement, indicating the challenges posed by the requirements as well as an initial set of techniques to address these is provided.

The scenarios and use cases described in this document have been developed in workshops with ATCOs and pilots where techniques from resilience engineering and requirements engineering have been applied. The description of the scenarios and the potential use cases is still on a high-level basis at this point of the project. Nevertheless, they proved useful to discuss and derive user stories, and subsequently requirements for an initial development of techniques for a predictive decision support tool for ATCOs.

Next to the seven airport specific use cases, two general use cases have been created. One of the use cases does not include the idea of a prediction of Go-Arounds, but an (live) indication of Go-Arounds when they are initiated by the FC. This use case has independently been brought up by ATCOs in the workshops~~,~~ after introducing them to the idea of a predictive tool. Whilst this idea is not completely in line with the research agreed upon in the Grant Agreement for this project, we think it is a fruitful one and therefore worth including in this deliverable. The focus of SafeOPS is on the Go-Around prediction and the provision of resulting probabilistic information to ATCOs. Nevertheless, an automated detection of Go-Arounds in operational flight data (e.g. ADS-B / Mode S / FDM) will be part of the developmental phase, as it is necessary for creating a training data set for a predictive tool. The discussion of whether this use case should also be included in the impact evaluation task 2.2 can be fostered with the SJU until task 2.2 will be kicked off.

Additionally, the defined use cases will be fundamental for task 2.2 of SafeOPS, which will evaluate the impact of a decision support system on ATM, given the achievements of the developmental work, following this deliverable.

Based on the derived user stories, an initial set of requirements has been defined. For the identified requirements, a technical problem statement has further been provided, indicating the techniques and challenges which could be derived from the requirements. Note that the set of requirements presented now is not static for the rest of the project. Based on the achievements and problems, that will be identified during the upcoming work, a refinement and adaption process of the given requirements will be unavoidable. This process will include the users (ATCOs) through continuing the workshops to ensure alignment with their needs.

A further aspect of the continued workshops will be the prioritization of work during the developmental phase. Additionally, to the input of the ATCOs, the external input from the Associated Partners Workshop in July will be considered for the prioritization of use cases as well as the linked user stories and requirements.

Lastly, a set of data which is available for SafeOPS in the DataBeacon platform and how they address the demands of the user stories is presented.

6 References

[1] J. Leonhardt, E. Hollnagel, L. Macchi und B. Kirwan, „A White Paper on Resilience Engineering for ATM," Eurocontrol, Brussels, 2009.

[2] E. Hollnagel, Safety-I and Safety-II: The Past and Future of Safety Management, 2014.

[3] E. Hollnagel, J. Leonhardt, T. Licu und S. Shorrock, „From Safety-I to Safety-II: A White Paper," Eurocontrol, Brussels, 2013.

[4] B. Rabiller, N. Fota und L. Carbo, „SESAR Safety Reference Material," Eurocontrol, Brussels, 2018.

[5] B. Rabiller, N. Fota und C. Laura, „Guidance to Apply SESAR Safety Reference Material," Eurocontrol, Brussels, 2018.

[6] E. Hollnagel, Safety-II in Practice: Developing the Resilience Potentials, London: Routledge, 2017.

[7] S. Shorrock, J. Leonhardt, T. Licu und P. Christoph, „Systems Thinking for Safety: Ten Principles," Eurocontrol, Brussels, 2014.

[8] Oxford University Press, „Oxford Learner's Dictionaire," [Online]. Available: https://www.oxfordlearnersdictionaries.com/definition/academic/scenario. [Zugriff am 21 6 2021].

[9] Visual Paradigm, Visual Paradigm, [Online]. Available: https://www.visual-paradigm.com/guide/use-case/what-is-use-case-specification/. [Zugriff am 21 6 2021].

[10] M. Cohn, User Stories Applied: For Agile Software Development, Boston: Addison-Wesley Professional, 2004.

[11] Visual Paradigm, „What is User Story?," Visual Paradigm, [Online]. Available: https://www.visual-paradigm.com/guide/agile-software-development/what-is-user-story/. [Zugriff am 21 6 2021].

[12] T. Blajev und C. W., „Go-Around Decision Making and Execution Project," Flight Safety Foundation, 2017.

[13] Airbus Custumer Service, „Descend Management. Being Prepared for Go-Around," Airbus, 2008.

[14] Flight Safety Foundation, „FSF ALAR Briefing Note 7.1 Stabilized Approach," Flight Safety Foundation, 2009.

[15] IATA, „Unstable Approaches: Risk Mitigation Policies, Procedures and Best Practices, 2nd Edition," IATA, Montreal, Geneva, 2017.

[16] B. De Courville, „Go-around Decision Making,“ Go-around Safety Forum, Brussels, 2013.

[17] Skybrary, „Go-around Decision Making,“ [Online]. Available: https://www.skybrary.aero/index.php/Go-around_Decision_Making. [Zugriff am 21 6 2021].

[18] Deutsche Flugsicherung GmbH, „AIP Deutschland,“ [Online]. Available: https://aip.dfs.de/basicIFR. [Zugriff am 21 6 2021].

[19] SafeClouds.eu, „CORDIS,“ European Union, [Online]. Available: https://cordis.europa.eu/project/id/724100/results/de. [Zugriff am 11 6 2021].

[20] European Organisation for the Safety of Air Navigation, „R&D Data Release -Metadata,“ 26 02 2020. [Online]. Available: https://www.eurocontrol.int/sites/default/files/2020-09/rnd-data-description-2020.pdf. [Zugriff am 28 06 2021].

[21] International Civil Aviation Organization, „Annex 6 - Amendment 26“.

[22] Civil Aviation Authority, „CAP 739 - Flight Data Monitoring,“ Civil Aviation Authority, Norwich, 2013.

[23] M. Lui, „Complete Decoding and Reporting of Aviation Routine Weather Reports (METARs),“ NASA Armes Reserach Center, Moffet Field, California, 2014.

[24] I. Gómez, „Innaxis' DataScience.aero,“ 20 02 2019. [Online]. Available: https://datascience.aero/machine-learning-aviation-taking-off/. [Zugriff am 28 06 2021].

[25] I. Gómez, „Innaxis' DataScience.aero,“ 05 04 2019. [Online]. Available: https://datascience.aero/metrics-evaluating-machine-learning/. [Zugriff am 28 06 2021].

[26] „scikit learn,“ [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html. [Zugriff am 28 06 2021].

[27] S. Rane, „towards data science,“ 3 12 2018. [Online]. Available: https://towardsdatascience.com/the-balance-accuracy-vs-interpretability-1b3861408062. [Zugriff am 28 06 2021].

[28] SafeClouds.eu, „Cordis,“ 09 07 2018. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bc219e2d&appId=PPGMS. [Zugriff am 28 06 2021].

Appendix A DataBeacon Platform

A.1 Data Acquisition Requirements

This section gathers the main requirements related to data acquisition from the data sources described in section 4.1. These requirements are related to the technical means or protocols to capture data in case of stream processing, the data uploading in case of batch processing, the identification of appropriate parsers, performance needs to guarantee an optimal data acquisition (e.g. asynchrony) or the requisites for different data sources traceability among others. Each requirement is classified per data set currently available, according to a security level (i.e. anonymized or public data) and according to the current compliance by the BeSt data platform.

A.1.1 ADS-B Data Acquisition Requirements

Requirement ID:	DA.ADSB.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	ADS-B OpenSky				
Requirement:	The platform should be able to support the collection and transmission of ADS-B data captured by the OpenSky network				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.ADSB.02	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	ADS-B OpenSky				
Requirement:	The platform should be able to support the transformation and traceability of data received from Weather data source and other sources to allow data merging.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

A.1.2 EUROCONTROL R&D Data Acquisition Requirements

Requirement ID:	DA.ECTLRD.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	EUROCONTROL R&D Data archive				
Requirement:	The platform should be able to support batch collection and processing data from EUROCONTROL R&D Data archive data sources in bulk				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.ECTLRD.02	Category:	Input Data	Version:	1.0
Linked User Story(ies):	all				
Data Set:	EUROCONTROL R&D Data archive				
Requirement:	The platform should be able to support support structured data file (CSV format) parsing				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.ECTLRD.03	Category:	Input Data	Version:	1.0
Linked User Story(ies):	all				
Data Set:	EUROCONTROL R&D Data archive				
Requirement:	The platform should be able to support the transformation and traceability of data received from Weather data source and other sources to allow data merging.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

A.1.3 ERA5 Data Acquisition Requirements

Requirement ID:	DA.ERA5.01	Category:	Input Data	Version:	1.0
Linked User Story(ies):	all				
Data Set:	ERA5 - ECMWF Reanalysis v5				
Requirement:	The platform should be able to support batch collection and processing data from ECMWF data sources in bulk				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.ERA5.02	Category:	Input Data	Version:	1.0
Linked User Story(ies):	all				
Data Set:	ERA5 - ECMWF Reanalysis v5				
Requirement:	The platform should be able to support semi-structured data file (TXT or CSV format) parsing, and text data extraction for ECMWF data acquisition.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.ERA5.03	Category:	Input Data	Version:	1.0

Linked User Story(ies):	all
Data Set:	ERA5 - ECMWF Reanalysis v5
Requirement:	The platform should be able to support the transformation and traceability of data received from Weather data source and other sources to allow data merging.
Security Level:	Public Data
Status BeSt Platform:	Implemented

A.1.4 FDM Data Acquisition Requirements

Requirement ID:	DA.FDM.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	FDM - Flight Data Monitoring Data				
Requirement:	The platform should be able to support batch processing to collect data from FDM data source				
Security Level:	Anonymised Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.FDM.02	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	FDM - Flight Data Monitoring Data				
Requirement:	The platform should be able to support FDM binary data (raw data) and has capability to decode into engineering format.				
Security Level:	Anonymised Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.FDM.03	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	FDM - Flight Data Monitoring Data				
Requirement:	The platform should be able to support the de-identification of FMD data to guarantee the anonymization of the data.				
Security Level:	Anonymised Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.FDM.04	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				

Data Set:	FDM - Flight Data Monitoring Data
Requirement:	The platform should be able to support FDM structured data file (TXT, CSV, TSV) / decoded FMD data and has capability to compress/uncompress the data.
Security Level:	Anonymised Data
Status BeSt Platform:	Implemented

A.1.5 METAR Data Acquisition Requirements

Requirement ID:	DA.METAR.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	METAR - Meteorological Aerodrome Report				
Requirement:	The platform should be able to support batch collection and processing data from METAR data sources in bulk				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.METAR.02	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	METAR - Meteorological Aerodrome Report				
Requirement:	The platform should be able to support semi-structured data file (TXT or CSV format) parsing, and text data extraction for METAR data acquisition.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.METAR.03	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	METAR - Meteorological Aerodrome Report				
Requirement:	The platform should be able to support the transformation and traceability of data received from Weather data source and other sources to allow data merging.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

A.1.6 SIGMET Data Acquisition Requirements

Requirement ID:	DA.SIGMET.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				

Data Set:	SIGMET - Significant Meteorological information
Requirement:	The platform should be able to support batch collection and processing data from SIGMET data sources in bulk
Security Level:	Public Data
Status BeSt Platform:	Implemented

Requirement ID:	DA.SIGMET.02	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	SIGMET - Significant Meteorological information				
Requirement:	The platform should be able to support semi-structured data file (TXT or CSV format) parsing, and text data extraction for SIGMET data acquisition.				
Security Level:	Public Data				
Status BeSt Platform:	Pending				

Requirement ID:	DA.SIGMET.03	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	SIGMET - Significant Meteorological information				
Requirement:	The platform should be able to support the transformation and traceability of data received from Weather data source and other sources to allow data merging.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

A.1.7 TAF Data Acquisition Requirements

Requirement ID:	DA.TAF.01	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	TAF - Terminal Aerodrome Forecast				
Requirement:	The platform should be able to support batch collection and processing data from TAF data sources in bulk				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

Requirement ID:	DA.TAF.02	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	TAF - Terminal Aerodrome Forecast				

Requirement:	The platform should be able to support semi-structured data file (TXT or CSV format) parsing, and text data extraction for TAF data acquisition.
Security Level:	Public Data
Status BeSt Platform:	Pending

Requirement ID:	DA.TAF.03	**Category:**	Input Data	**Version:**	1.0
Linked User Story(ies):	all				
Data Set:	TAF - Terminal Aerodrome Forecast				
Requirement:	The platform should be able to support the transformation and traceability of data received from Weather data source and other sources to allow data merging.				
Security Level:	Public Data				
Status BeSt Platform:	Implemented				

A.2 Data Protection

The protection of datasets according to the confidentiality requirements in aviation, as well as other domains, is key for the correct development of a data-centred project. There are different ways in which sensitive information can be shared such as removing confidential fields that could cause leakage of some private data or that could allow to individually identify a flight as well as data sharing in an aggregated manner. However, for most, if not all, data mining applications, reducing or restricting access to data could severely decrease the quality of the trained models. In particular it may affect data fusion or merging which is a vital part for these projects. That said, it is worth mentioning that all confidential data sets in SafeOPS are subject to the General Provisions agreed under the Data Protection Agreements that are or will be made with respective data owners.

The challenge is to maintain the privacy while allowing the fusion among datasets. This will require sophisticated techniques that go beyond simply deleting sensitive data (which would impede merging datasets). For this reason, the SafeOPS project will use **de-identification techniques** applied to any sensitive data fields identified by the data owners in conjunction with **Smart Data Fusion** techniques developed in the project SafeClouds. This will enable the cross reference of information from different sources without compromising privacy integrity of the data.

A.2.1 Data de-identification and Smart Data Fusion

The combined solutions of data de-identification and subsequently Smart Data fusion are already implemented in the BeSt data platform form DataBeacon which the project will rely on. [28] For data de-identification of sensitive data state-of-the-art hashing operations are used. These hashing operations codify sensitive data such as callsigns or dates into alphanumeric strings. It should be noted that for security reasons these operations are only "one-way" operations which cannot be reversed.

However, it stills allows to identify whether two data sample belong to the same date (if it has been de-identified) without revealing the exact dates (because they will share the same hash code). The de-identification imposes some security limitations on the data filtering - for example, if a sufficiently concrete period is known (e.g., week, month, year), the whole de-identification process might be compromised. In order to overcome these limitations, the data platform BeSt incorporates Secure Data Fusion (SDF) technology. For more details on Smart Data Fusion, revert back to section 4.2.

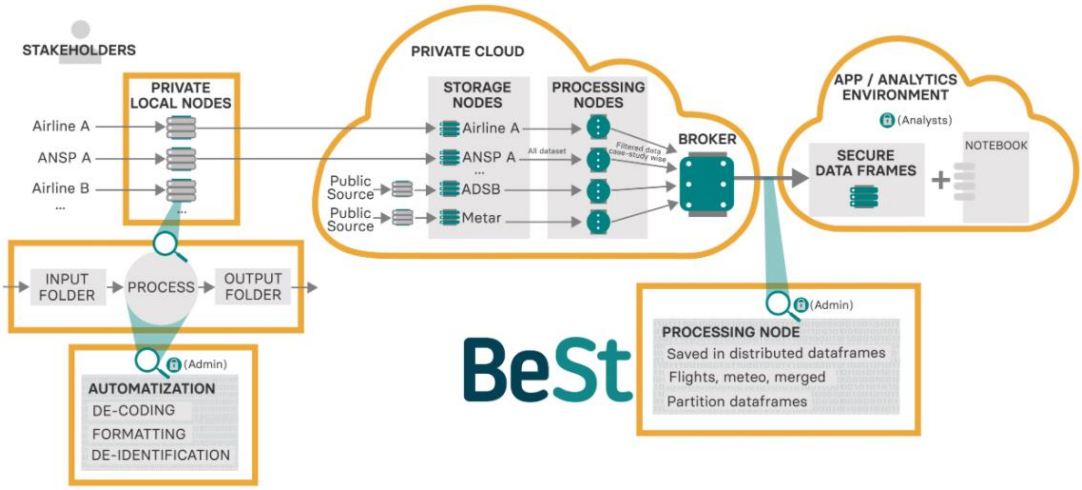


Figure 42: Overview BeSt Data Platform

BeSt data platform uses a multi-layer architecture for data storage developed to preserve data privacy, protection and accessibility:

- From the **data providers** point of view, the first layer consists of a series of so called "**private local nodes**" one for each data provider and where they have full control. These are dedicated to c**ollect and store raw, identified data**. Data is pre-processed directly from input sources, and it is then protected and de-identified, as required, and pushed to the next layers of the architecture. This layer is only accessible to the data engineer role (explained below) in the project.
- The second layer consists of several storage and **processing nodes**; one for each data provider. Data stored in this layer is now **de-identified, cleaned and standardised**, hosted on isolated private cloud environments. The access to these de-identified transformed data sets is based on unique credentials given to individuals who have data analyst or data scientist role (explained below) in the project.

Finally, to make data accessible and useful to data analysts and other developers, BeSt implements and offers **secure sandboxed environments** in which analytics take place. It consists of a **high availability, on-demand cloud computing platform**. After secure login, a complete data science development environment is launched, which includes popular data science tool sets such as Python, Hadoop, Spark and the Anaconda ecosystem.

Although the de-identification is a requisite for private datasets, the data elements from public datasets (e.g. METAR, ADS-B, etc.) might also require to be de-identified for cross-reference and data fusion needs and to ensure the effectiveness of the data protection method. In that sense, all the information about the time or the identification of the trajectories might have to be de-identified across all datasets after ensuring that we will be able to fuse their respective information (e.g. accessing the meteo information of a flight of a given day). In some cases, the de-identification procedures may have restrictive effects on subsequently applied data mining approaches although it is expected little to no effect of the de-identification on the approaches that will be performed.

A.2.2 Roles and Data Privacy

Data privacy issues introduces additional complexity into the data processing pipeline and requires a careful sequencing of various parts of data processing to maximise the efficiency of the whole process and seamless collaboration between various teams of data analysts, engineers and data scientists. In order to meet these requirements and to comply with the Data Protection Agreements (DPAs) that are (or might be) agreed upon SafeOPS will adopt the following role division:

- **Data engineer/manager**. Role in charge of maintaining and developing the data infrastructure BeSt, and serves as the technical point of contact with third parties and data providers. Has access to all the raw data and administrator rights on BeSt systems, but does not run any analysis. The main tasks will include protecting (via de-identification or anonymization) of the private data fields and merging and preparing the data from various sources. Consequently, a person in this role is also in charge of granting data access to the individuals who serve as data scientist/analyst to the data transformed and prepared for analysis.
- **Data scientist/analyst**. The main tasks include the development and deployment of the (machine learning) algorithms defined by the case studies. The individual in this role does not have access to protected raw data sets or identifiable data sets, it only has access to the transformed and prepared data which they use to develop the predictive algorithms according to the defined scenarios.

This dual division of roles serves to streamline the various tasks in a machine learning or data processing pipeline while also protecting private data.