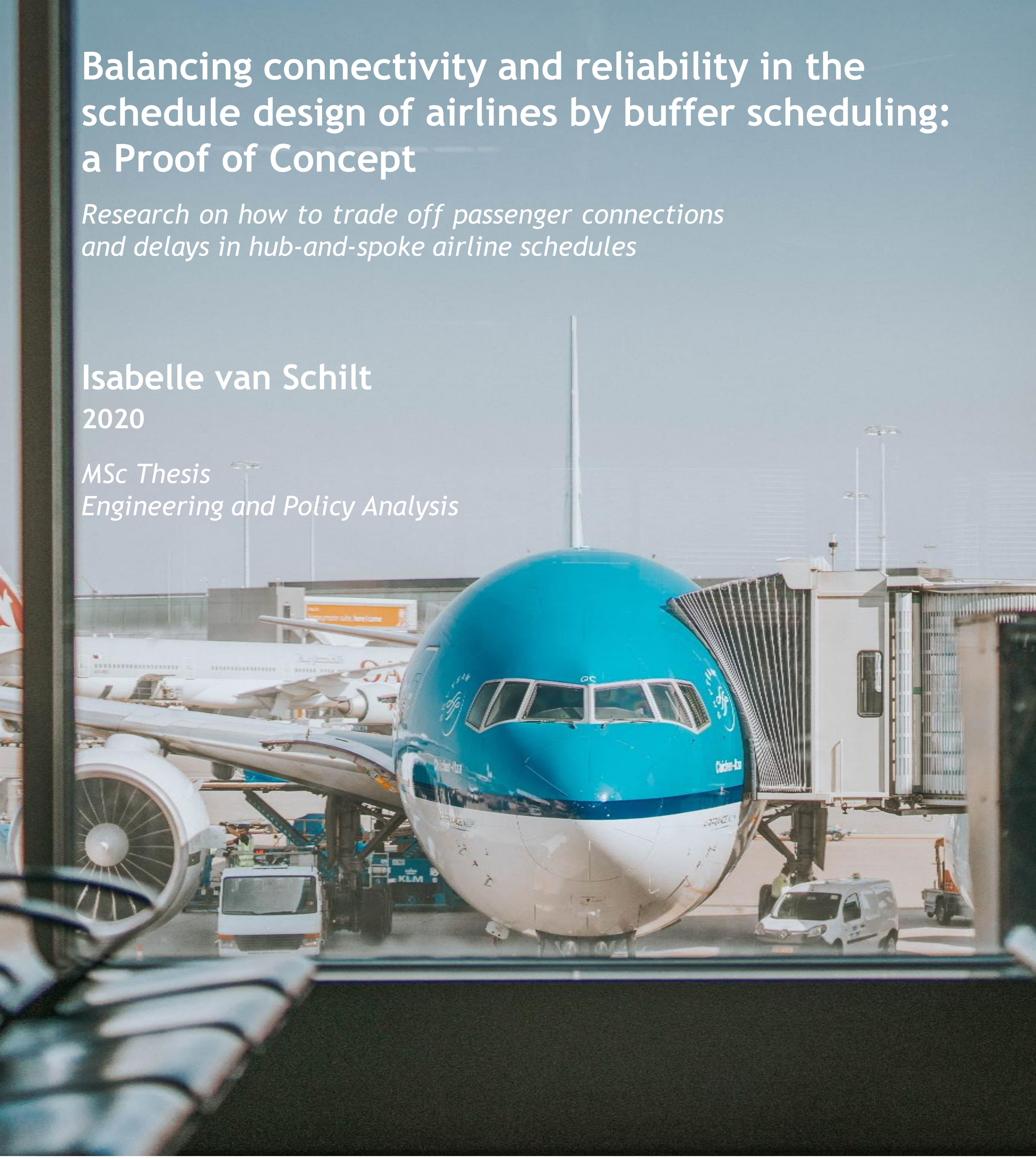
Balancing connectivity and reliability in the schedule design of airlines by buffer scheduling: a Proof of Concept

Research on how to trade off passenger connections and delays in hub-and-spoke airline schedules

Isabelle van Schilt

2020

MSc Thesis
Engineering and Policy Analysis



KLM Royal Dutch Airlines

TU Delft

This page is intentionally left blank

Balancing connectivity and reliability in the schedule design of airlines by buffer scheduling: a Proof of Concept

Research on how to trade off passenger connections and delays in hub-and-spoke airline schedules

by

Isabelle van Schilt
Student number: 4281020

Chair:	Prof.dr.ir. A. (Alexander) Verbraeck Policy Analysis
First Supervisor:	Dr.ir. I (Iulia) Lefter System Engineering
Second Supervisor:	Dr.ir. J.H. (Jan) Kwakkel Policy Analysis
External Supervisor:	Jonna van Kalker KLM Royal Dutch Airlines

ACKNOWLEDGEMENTS

Six months ago, I started my MSc thesis project at KLM Royal Dutch Airlines. During this period, I've learned so much about conducting research, programming, the aviation sector, the company, and about the people.

First, I would like to thank my graduation committee of the Delft University of Technology for their supervision, knowledge sharing, and interesting discussions. I would like to thank Alexander Verbraeck for the great supervision, for helping me to lift the report to a higher level and for continuously challenging me; not only during my MSc thesis but also during my project at MIT and my Bachelor thesis. I would like to thank Jan Kwakkel for helping me with my programming errors and all the other question regarding optimization and therewith, the quick replies on my email requests. I would like to thank Iulia Lefter for her useful insights on overarching goal of the research and for the feedback from another perspective.

I would also like to thank my external supervisor from KLM, Jonna van Kalker, for all our update and worksessions on my thesis, for the enthusiasm and support that you have for this project and for the great guidance and feedback. It was a great pleasure to work with such a supportive and dedicated supervisor. Next to that, I would like to thank the Delphi team of ODS for being able to experience working in a team with great people. It was a valuable and enjoyable experience to be part of the team and get to know all of you. The team outings, Ibis or Wings drinks, daily lunches and coffee breaks, made my internship period a lot of fun.

Lastly, I would like to thank my family, in particular my parents, for supporting and encouraging me throughout this period and throughout my entire study. This accomplishment would not have been possible without them. I would also like to thank my roommates for keeping up with me during my thesis, especially during this intelligent lockdown. Next, I would also like to thank my friends for always lending me an ear when needed or distracting me when necessary.

The past six months were filled with a lot of hard work, long days and most of all, a lot of fun. Many thanks to all!

Isabelle van Schilt
2 April 2020

ABSTRACT

In the last decades, the environmental impact of aviation has been an emerging topic of discussion worldwide. With an expected CO_2 emission increase of 21% by 2040 for air traffic, there is an urgency to reduce the negative impact on the environment (European Commission, 2019). On top of that, the number of passengers is expected to double within the next 20 years (International Air Transport Association, 2018). This challenge requires intervention from governments, airlines and passengers. Creating a more reliable flight planning would reduce the fuel consumption and costs for airlines. However, for airlines with a hub-and-spoke network, the schedule design relies on maximizing the connecting passengers. A more reliable planning with embedded buffers could lead to a lower offering of connecting flights. The goal of this research is to design a novel optimization model for balancing reliability and connectivity by buffer scheduling in the flight schedule, and to evaluate the impact on environmental sustainability. This research provides a Proof of Concept with a case of KLM Royal Dutch Airlines. The optimization model presents a Pareto optimal front that makes the trade-off between reliability and connectivity explicit. Main findings of this research are that (i) the model is suitable for supporting decision-making on a basic level, (ii) the stochastic variance of the model should be limited and (iii) the flight schedule presented by the model reduces the CO_2 emissions significantly. Further research is required on how to optimize schedules with a higher level of complexity for the methodological elements and on the business side.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ADC All Doors Closed

ADO Any Door Open

AF Air France

AMS Amsterdam Schiphol Airport

ATA Actual Time of Arrival

ATC Air Traffic Control

ATD Actual Time of Departure

CDM Collaborative Decision Making, i.e. joint decisions between ATC, Schiphol and airlines on the day of operation to improve the departure process

CDG Charles de Gaulle Airport

CF Completion Factor

CORSIA Carbon Offsetting and Reduction Scheme for International Aviation

CTOT Calculated Take-Off Time, i.e. the take-off time set by EuroControl

EUR European flights

FIR Flight Information Region

HSF High Speed Flying

ICA International flights

MACT Maximum Acceptable Connecting Time

MCT Minimum Connecting Time

MTT Minimum Turnaround Time

OTP On-Time Performance

SPT Standard Processing Time

STA Scheduled Time of Arrival

STD Scheduled Time of Departure

TOBT Target Off-Blocks Time, i.e. the time an aircraft is expected to be ready for off-blocks

TSAT Target Start-up Approval Time, i.e. the target time in which an aircraft can leave for departure set by air traffic control

1 INTRODUCTION

In this chapter, the main problem of this research is described. Next, the concepts, definitions and other work in this field are examined by means of a literature review in Section 1.2. Following, the research design including the knowledge gaps, scope of the research, research questions and case selection are presented in Section 1.3. Lastly, the research approach and research methodology per subquestion is elaborated on in Section 1.5.

1.1 PROBLEM DESCRIPTION

This section first discusses the general problem of this research. Thereafter, the problem from the perspective of the airlines is described.

1.1.1 Problem Explanation

In recent years, the environmental impact of air traffic has been one of the fastest emerging topics in the airline industry. In the 1990s, the first visible negative environmental effects of aviation were pointed out (Dessens et al., 2014; Price & Probert, 1995). The share of carbon dioxide (CO_2) emissions of aviation has increased rapidly; in the European Union (EU), the emissions increased with 87% in the period between 1991 and 2003 (Rothengatter, 2010). Between 2013 and 2019, CO_2 increased with 28% (Air Transport Action Group, 2020; Environmental and Energy Study Institute, 2019). Next to this, CO_2 and nitrogen oxide (NO_x) emissions caused by aviation are expected to increase with at least 21% and 16%, respectively, by 2040 (European Commission, 2019).

Globalization, i.e. the interaction and integration of people, governments and companies worldwide, is one of the main causes for this growth. The number of passengers has increased rapidly in the last decades and is expected to continue to grow with 3.5% each year, adding up to approximately 200% in the next 20 years (International Air Transport Association, 2018; Pels, 2008). Aviation is essential to globalization by providing a worldwide physical connectivity (Button, 2008). Thus, on the one hand the need to travel among people increases due to globalization while on the other hand, the emissions of air traffic urgently need to decrease in order to reduce the negative impact on climate change. This contradiction asks for action from the government, airlines and passengers.

In Europe, the Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) has been implemented with the aim to stabilise the CO_2

emissions at 2020 levels. With CORSIA, airlines are required to offset the growth of their emission after 2020 by monitoring international routes and compensating emission for all routes (European Commission, 2019). Many governments have agreed to pursue emission reductions for air traffic via the International Civil Aviation Organization (ICAO) (Gill, 2015). Next to this, airlines take action to reduce emissions by for example sustainable fuels, no holding patterns, i.e. no circling around the airport before landing, and emission-free airports (Air Transport Netherlands, 2019).

The first theme defined by CORSIA in terms of sustainability is the efficient use of fuel by optimizing the flight routes and procedures (Air Transport Netherlands, 2019). An important factor that should be taken into account for optimizing flight procedures is the probability of delays. Flight delays do not only affect the passengers, airports and airlines; they also affect the environment significantly. Both airborne delays and ground delays have a negative effect on environmental sustainability due to the increase of fuel consumption and gas emissions which results from the higher speed required to catch up with the delay (Carlier et al., 2007; Sternberg et al., 2017). Creating a more robust and predictable flight planning therefore contributes to the reduction of the fuel consumption and is expected to contribute to 20% of the CO_2 reduction in aviation (Air Transport Netherlands, 2019; Ryerson et al., 2014).

As airlines and governments want to tackle the negative impact on aviation of climate change urgently, they are dedicated to create a more robust and predicable flight planning. Therefore, it is valuable for airlines and government to focus on designing a more reliable flight planning to reduce CO_2 emissions.

1.1.2 Problem Explanation: Airline Perspective

The design of a more reliable flight planning would also be beneficial for airlines in terms of cost. Delays (for example due to bad weather, technical problems, crew unavailability) are costly for both the airline and the passengers (Peterson et al., 2013; C.-L. Wu, 2008). In particular for hub-and-spoke networks (i.e. locations called spokes connected through an intermediary location called a hub), delays are problematic due to the high intensity of connecting flights, passengers and crews (Achenbach & Spinler, 2018; Hansen et al., 2001). Therefore, it would be valuable for airlines to incorporate the probability of delays in the design of the flight schedule (Thengvall et al., 2000). In this manner, the flight planning is less likely to be negatively affected by delays on the day of operation (L. H. Lee et al., 2007).

However, a reliable flight schedule could ask for more resources or more time which is also costly for airlines. This could lead to a trade-off between reliability and costs (Clausen et al., 2010). For hub-and-spoke airlines, costs are primarily determined by the connecting passengers. An example of this trade-off is when an extra five minutes is added to the ground time of a flight to increase reliability, however this leads to missing a connecting flight for almost 30% of the passengers. The question arises if it is the most beneficial to add this buffer or not in terms of cost. Currently, this decision is made manually based on case-by-case analyses. Nevertheless, airlines still do not have an easy way to analyze these trade-offs and no clarity on what the actual effects of these decisions are (Wong & Tsai, 2012).

Additionally, there are many other factors that should be dealt with when creating an airline schedule, such as crew scheduling, fleet assignment, slot constraints, connections, and market share. This makes the trade-offs for optimal buffer allocation increasingly complex (Etschmaier & Mathaisel, 1985).

Airlines create optimization tools to analyze a part of this complex trade-off, for example maximizing on connectivity; parts that can be directly implemented. However, all these optimizing tools work sequentially and not simultaneously which makes it difficult to actually balance reliability and connectivity. Next to that, airlines try to trade-off reliability and connectivity in terms of costs. They want to convert delay minutes to cost and compare this with the connectivity value. However, it is challenging and nearly impossible to accurately convert delay minutes to cost. Due to the focus on subparts of the main trade-off and the focus on converting delay minutes to cost, a way to optimize multiple processes simultaneously for supporting decision-making has not been developed yet. Therefore, it would be valuable for airlines to investigate how they can weigh reliability against connectivity in a clear and workable manner when creating a flight schedule.

1.2 LITERATURE REVIEW

In this section, existing literature is examined in order to explain important concepts, to acknowledge existing research and to identify knowledge gaps. First, literature on airline scheduling is discussed. Hereafter, literature on flight delays and buffer scheduling in aviation is presented.

1.2.1 Airline Scheduling

Airline scheduling is one of the most challenging and important operations for airlines. Four core problems can be distinguished in schedule planning at airlines, namely (Barnhart & Cohn, 2004; C.-L. Wu, 2006):

1. *Schedule design:* Determine the markets to serve, at what frequency and how to schedule the flights.
2. *Fleet assignment:* Assign the aircraft to each flight.
3. *Aircraft maintenance routing:* Route the aircraft such that the maintenance requirements are satisfied.
4. *Crew scheduling:* Assign the crew to flights such that costs are minimized.

As described in Section 1.1, the main challenge is to create a more robust flight schedule in terms of reliability and connectivity. Therefore, an optimization model should be designed.

Ideally the four core problems are solved simultaneously in an optimization model, however this becomes an extremely complex and large problem. Therefore, most research focuses either on integrating two core problem or extending one problem (Barnhart & Cohn, 2004).

Several optimization models for airlines scheduling have been defined. Barnhart & Cohn (2004) describes optimization approaches which might be applicable to these four core problems. These techniques are for example the linear programming technique of branch-and-bound and branching. This is used to construct a completely new schedule. Lessons are drawn on how to create a solvable model. With linear programming, the main solution for creating a solvable model is to relax the model, thus limit the number of constraints. Ageeva (2000) also describes a linear programming model for the development of schedules that are robust to disruptions and how airline schedules could be evaluated to increase robustness, with focus on the aircraft maintenance routing. Another optimization technique is used by Şafak et al. (2017) namely multi-objective optimization. The multi-objective optimization is a good way to optimize simultaneously. They link robust fleet assignment, sustainability and passenger cost in a non-mathematical optimization model. There are in total two objectives namely minimizing the operational cost and maximizing the service level of the passenger connections. In the operational cost, they mainly focus on fuel cost. Moreover, Achenbach & Spinler (2018) also optimizes fuel cost and wage cost. They proposes an arrival time prediction with a cost index optimization for short-haul flights based on data from a European carrier. Both papers show that fuel has a large impact on cost and also on sustainability. Thus, fuel is an interesting variable to take into account.

An optimization model for airline scheduling can either be integral with multiple core problems or focused on one single core problem; it can either be defined as a linear programming model or a multi-objective model and fuel is seen as an important variable for cost.

1.2.2 Flight Delays in Aviation

The optimization model for airline scheduling focuses on incorporating delay and connectivity in the schedule design. Therefore, literature on airline delay management is relevant for this research.

Flight delays are an unavoidable and crucial element in the context of aviation; mostly for economic and environmental reasons (Sternberg et al., 2017). Airlines need to incorporate the possibility of delays by disruption such as weather conditions, mechanical problems, and capacity constraints in their scheduling.

Much research has been performed on the airline delay management for the day of operation. Santos et al. (2017) presents a linear programming approach to solve the daily airline delay management problem with capacity constraints and to make decisions on the spot. Jarrah et al. (1993) creates a decision support framework for flight delays and cancellations during the day of operation.

Instead of solving delays during the daily operations, delays could also be prevented. Montlaur & Delgado (2017) shows optimization techniques to minimize the flight and passengers delay by including or excluding reactionary delays. Sternberg et al. (2017) shows a review of approaches that are build to predict the flight delays and how machine learning relates to this. In their paper, there is a distinction between root delay, i.e. local delays, or cancellations and delay propagation, i.e. delays in a flight causes a delay in the subsequent flights. Propagated delays are mostly caused by the connected resources within the airline schedule such as the aircraft, crew, passenger and airport resources (Kafle & Zou, 2016).

Propagated delay is a well-known phenomena in aviation and its impact has been researched extensively. Kondo (2011) compares the impact of propagated delays between hub and point-to-point airports. Churchill et al. (2010) examines the effect of propagated delays on the daily planning. Moreover, Qin et al. (2019) investigates how to optimize the delay propagation in a Chinese aviation network by rescheduling flights. Thus, delay propagation is a crucial element for airline scheduling, especially hub-and-spoke networks (Achenbach & Spinler, 2018).

Next to this, delays can occur either on the ground, i.e. ground delay, or in the air, i.e. en-route delay (Carlier et al., 2007). Ground delay can be defined as the delay during the turnaround of the airplane. En-route delay can be defined as the delay of an airplane between departure (off-blocks) and arrival (on-blocks); this contains taxi and airborne time. En-route delay is also known as block delay (Fricke & Schultz, 2009).

1.2.3 Buffer times in Schedule Design

Much research has shown that embedding buffers would improve the reliability of a flight schedule. With the help of a simulation study, C.-L. Wu (2005) states that schedules can become more robust when buffer times are embedded. Baumgarten et al. (2014) shows that using buffer times in an airline schedule would absorb unexpected delays. Ahmadbeygi et al. (2010) shows that re-allocating existing slack would improve the schedule without any increased planned cost. C.-L. Wu (2006) evaluates the effectiveness of embedded buffer times on reliability in draft airline schedules of a small airline network. Fricke & Schultz (2009) creates a model which optimizes the buffer time size with respect to the expected average delay. However, it is still unknown in which manner these embedded buffers should be scheduled to minimize delays and what related vulnerabilities are.

Pursuing this further, the design of a schedule is a multi-objective process for airlines. For hub-and-spoke networks, connecting passengers and flights are the most important objectives of the design at this moment. Research on the integration of connections and reliability is limited. Dunbar et al.

(2012) presents a new approach on how to integrate aircraft routing and crew scheduling to minimize propagated delay. They focus on the delays caused by missed connection of the crew. Jacquillat & Vaze (2017) designs and assesses a novel approach for scheduling the air traffic congestion of an airport. They want to increase reliability, i.e. mitigate the air traffic congestion at the airport, with network connectivity as a constraint. However, it has not been investigated yet how to handle the trade-off between connecting passengers and reliability in the flight schedule of an airline.

1.3 RESEARCH DESIGN

This section describes the design of the research. First, the knowledge gaps of this research is presented. Hereafter, the scope of the research is defined. Next, the general research method is described. Lastly, the case selection for the Proof of Concept is discussed.

1.3.1 Knowledge Gaps

From literature review, the following knowledge gaps are identified:

- How should embedded buffers be scheduled to minimize delays in aviation.
- How should the trade-off between connectivity and reliability be handled in the schedule design of an airline.

Next to these knowledge gaps from literature, the airline industry emphasizes the need for ways to incorporate the probability of delays in the schedule design such that the gap between the schedule and actual day-to-day operations becomes smaller. However, important connections between flights can be lost when modifying the schedule to make it more robust. Especially for hub-and-spoke airlines, the connecting flights is the most important element in scheduling at the moment. Airlines want these two factors - reliability and connectivity - to both be part of the schedule design in order to have a reliable schedule with maximized connecting flights. However, airlines do not know yet how to optimally align these factors.

1.3.2 Scope

The scope of this research is defined as follows:

- *Focus on hub-and-spoke network*
  As mentioned in Section 1.2.2, incorporating the probability of delays is crucial for designing the schedule especially in hub-and-spoke networks. The high intensity of connecting flights, passengers and crews in hub-and-spoke networks make delays costly and important

to mitigate (Achenbach & Spinler, 2018; Hansen et al., 2001; Lederer & Nambimadom, 1998).

- *Focus on schedule design*
  Section 1.2.1 defines four stages of airline scheduling, namely schedule design, fleet assignment, aircraft maintenance routing and crew scheduling. In the schedule design phase, the planning of flights is developed and this is where buffer scheduling is done. In this manner, it is possible to minimize delays with scheduling in advance instead of only solving delays during the day of operation, as described in Section 1.1.2. Next to this, the flight schedule is designed for two seasons namely summer and winter. Thus, this research also makes a distinction between the airline schedule of these two seasons.

- *Focus on schedule of European short haul flights*
  The short haul flights within Europe have a high intensity compared to the long haul flights. This makes optimal scheduling and handling delays more challenging and difficult for European short haul flights than for long haul flights. Due to the tight turnaround windows and high frequency in this part of the network, there is an urgent need for minimizing delays with buffers and integrating this with connectivity. Long haul flights are taken into account for the connectivity aspect of the short haul flight schedule. However, the long haul flight schedule itself is not in the scope of this research.

- *Focus on ground delay at the hub*
  This research focuses on ground delay for two reasons. First, airlines already try to add buffer to block times to reduce airborne delays. Second, tackling ground delays is also one of the most urgent matters for airline. Delays occurs more often on the ground (75% of the time) than in the air (25% of the time). To tackle delays at their origin, it is the most suitable to add buffer to the ground time. Therefore, this research focuses on modifications in ground times with respect to buffer scheduling.

1.3.3 General Research Method

The research objective is to support decision makers, in this case network planners, in optimally balancing reliability and connectivity when scheduling buffer time. This objective is reached by means of a quantitative research approach.

An optimization model is developed for simultaneously optimizing the schedule in terms of reliability and connectivity. The optimization model helps to understand and visualize how elements interact within a complex socio-technical system and what the impact is of system interventions (Creswell & Creswell, 2017; Shannon, 1998). For this case, system interventions are the various ways of buffer scheduling. With the help of visualizations, these insights could be communicated effectively to the airlines. This research tries to understand and communicate how to balance reliability and connectivity which is in line with a modeling approach.

A case study on a real airline is conducted to evaluate how this model would work in practice. This is used as a Proof of Concept, i.e. the model proves to work with real data.

Concluding, this research is tackled by designing an optimization model as quantitative research approach and provides a Proof of Concept.

1.3.4 Proof of Concept Selection

This research provides a Proof of Concept by KLM Royal Dutch Airlines, also known as KLM. KLM is a suitable airline for this Proof of Concept since it is a large carrier, has a leading role in the European aviation network, operates as a hub-and-spoke network, has a strong focus on transfer passengers, actively stimulates sustainable growth and has high data availability.

Firstly, KLM is a large carrier which makes it an interesting airline to investigate. KLM Royal Dutch Airlines is part of the Air France-KLM Group airline holding company. Together, this company is leading in international cargo and passengers traffic from Europe. With 2.300 daily flights, Air France-KLM groups operates from the two hubs at Paris-Charles de Gaulle (CDG) and Amsterdam-Schiphol (AMS) (Air France-KLM, 2020). Air France-KLM is also member of the SkyTeam alliance with 19 member airlines. These airlines together serve more than 1.150 destination in 175 countries. Airlines that are also in the alliance, are for example Delta, China Airlines and Kenya Airways (SkyTeam, 2020). KLM itself is one of the oldest and largest international carriers operating passengers and cargo to 162 destinations. The large size of the company also makes the network more complex and more challenging. It is interesting for this research to investigate a large international carrier with a high complexity.

Secondly, KLM plays a leading role in the European air industry together with Air France (KLM, 2019). Due to the focus of this research on European short haul flights, KLM would be a well-suited case for this research.

Thirdly, KLM operates via a hub-and-spoke network. Amsterdam Airport Schiphol functions as a hub within KLM's network. Via this location, many interconnections are made which allow passengers to efficiently access major destinations in the world.

Fourthly, passenger connections are one of the most important elements for KLM in their hub-and-spoke network. The focus of this research is on finding a clear balance between minimizing delay and therewith the need for high-speed flying, while at the same time maximizing passengers connections. Thus, connections play an important role in this research and are an essential part of the airline.

Fifthly, KLM is actively stimulating sustainable growth. Since the higher purpose of this research is to reduce the negative environmental impact, it is valuable to have an airline as case that support this.

Lastly, there is much data available on flights at KLM. The advantage of a large carrier is that many flight are executed during a period. Thus, there are sufficient data points available to perform a trustworthy analysis.

1.4 RESEARCH QUESTIONS

Combining the knowledge gaps from literature and industry and the scope, the main research question is:

How can airlines use buffer scheduling to ensure the optimal balance between the reliability of the flight schedule and the value of passenger connections, and what is the impact on environmental sustainability?

The following subquestions are needed to answer the main research question:

1. How are reliability and connectivity operationalized in aviation?
2. What is the current state of reliability and connectivity of the short haul flight schedule of KLM Royal Dutch Airlines?
3. Which optimization model is suitable for creating a reliable schedule by buffer scheduling while trading off passengers connections?
4. How can reliability and connectivity of the short haul flight schedule of KLM Royal Dutch Airlines be improved by buffer scheduling with the optimization model from sub question 3?
5. What is the impact of this improvement of the short haul flight schedule of KLM Royal Dutch Airlines on environmental sustainability?

The goal of this research is to investigate how a reliable flight schedule can be created to minimize delays and how this is balanced with the connectivity of the network.

1.5 RESEARCH METHOD PER SUBQUESTION

In this section, the research methodology per subquestion is discussed and presented in a research flow diagram.

The research methodology that fits the research question varies per subquestion. Figure 1.1 provides an overview of the suitable research methodology for each subquestion by means of a research flow diagram.

Subquestion 1

Subquestion 1 quantifies the relevant concepts of the main research question namely reliability and connectivity. For this research, it is important that these main concepts are quantified and clarified for further analysis. Literature research is conducted to examine which information and metrics

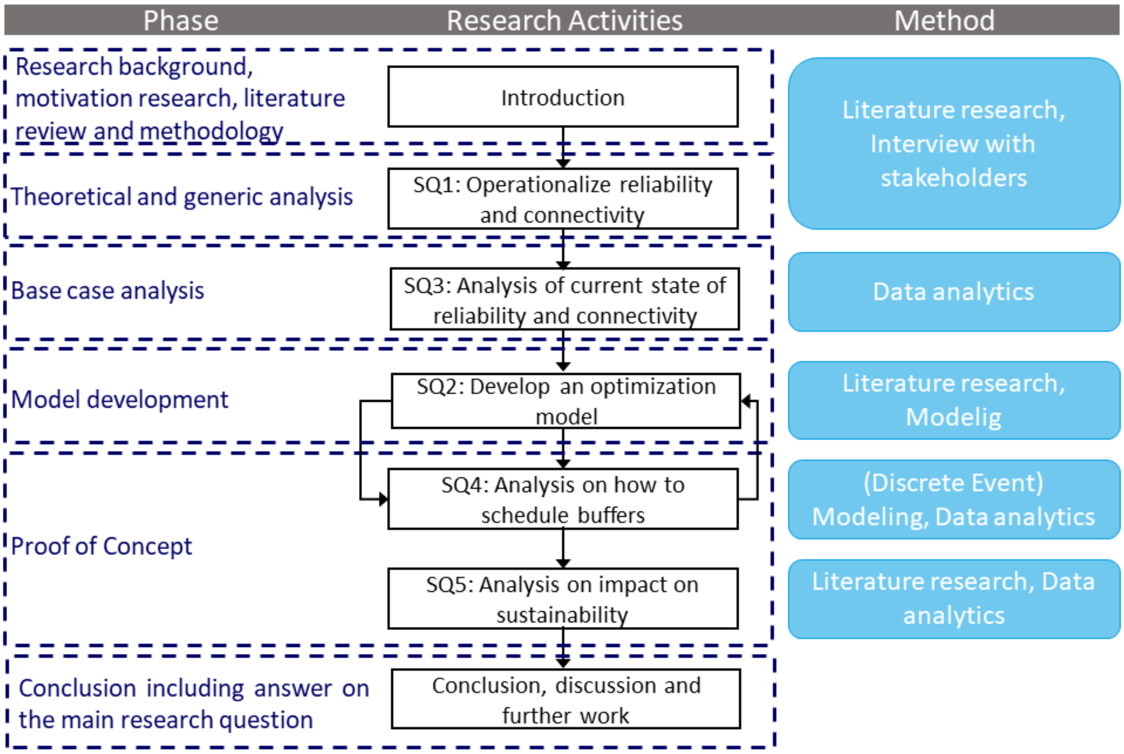


Figure 1.1: Research Flow Diagram

already exist on these topics in aviation. Combining this literature research with interviews at KLM, reliability and connectivity are operationalized. Literature research is conducted by means of search engines such as Google Scholar and Science Direct.

Subquestion 2

Based on subquestion 1, the current state of reliability and connectivity are evaluated. Subquestion 3 is answered by means of data analytics. The current situation of the reliability and connectivity of a short haul flight schedule is examined as a base case. First, the flight procedure and schedule planning should be understood and conceptualized with the help of models such as IDEF0. Hereafter, data on the current situation needs to be collected such as the schedule of previous years, the actual performance during the day, delay minutes and more. Depending on the type of data, data is gathered by means of interviews with experts, literature research and extracted from databases of the airline (Law, 2008). This data is analyzed in order to examine the current state of schedule reliability.

Data on connecting passengers is also required. This data is, however, one of the limitations of this subquestion. The real-time data on connectivity is difficult to access since it is highly confidential and sensitive, even within the airline. Therefore, a dummy dataset can be obtained on the value of connectivity. Additional information on data is collected based on interviews with commercial managers, network planners and other stakeholders that have experience in this field. This gives an approximation of the actual data on connecting passengers and flights. Historical data on connecting passengers and the price of the transfer is available for this research.

Subquestion 3

In subquestion 3, a novel multi-objective optimization model is developed which incorporates the relationship between reliability and connectivity for the schedule design as quantified in subquestion 1. The optimization model is based on the flight procedure as defined in subquestion 2. Literature research and interviews with stakeholders is conducted to create this optimization model. The biggest pitfall for this subquestion is that a model will be created which will not be realistic for airlines and thus, not accepted. Therefore, it is necessary to discuss and evaluate the characteristics of this model with experts such as network planners and commercial managers.

Subquestion 4

This subquestion investigates how buffers can be embedded to ensure a more reliable and highly connected schedule by the optimization model of subquestion 3. The optimization model proposes new schedules in which buffers are re-allocated for minimizing delays and maximizing passenger connections.

The impact and reliability of the schedule is tested by means of a Discrete Event Simulation (DES) model. This model evaluates the output of the optimization model on reliability and other key performance indicators. In this research, the system operates on flight level, not on passenger or baggage level. For this model, additional data on flights and distributions of delays is required. This data is gathered via interviews with experts such as network schedulers and data analysts, literature research and databases for the airline. Hereafter, the model is verified, validated and experiments are performed. Next to this, the simulation model is also used to determine the reliability from a buffer allocation solution of the optimization model. The limitation for this step could be the amount of data that is available and the quality of the data (Robinson, 2010). If there is either a lack of or a (very) small set, then an approximation of the data or distribution should be made in consultation with experts. Since subquestion 2 already requires data collecting in an exploratory manner, many issues with data could be tackled there. The simulation-optimization model is developed in the software Python, an open source programming language.

This analysis focuses on helping airlines to find the optimal balance of reliability and connectivity in their schedule design. From this analysis, it is expected that the trade-off between reliability and connectivity becomes visible from the optimal solutions. The optimization model that is defined in subquestion 3, is used to handle and analyze this trade-off for KLM.

Schedule design within aviation is a very complex and lengthy process (Etschmaier & Mathaisel, 1985). Due to time limitations and the complexity of the trade-off, it is not possible to re-schedule the entire Europe network planning. Therefore, the optimization model is tested on a small subset of the Europe network. Later on, this could extended to a larger subset or even the entire network. Here, the pitfall could be that the optimization model does not work well with larger networks and is not realistic. In close collaboration with network planners, this model should be validated repeatedly and limitations should be made clear.

Subquestion 5

Subquestion 5, the impact of the improvements on environmental sustainability, uses the results from the simulation-optimization study from subquestion 4. These results are translated into a measure for sustainability. Literature research on aviation and sustainability is conducted to find a suitable manner for translating the results in a sustainability value. With the help of data analytics, the impact of buffer re-allocation resulting from the optimization model on sustainability is assessed.

2 OPERATIONALIZATION OF RELIABILITY AND CONNECTIVITY

This chapter quantifies the main concepts of this research namely reliability and connectivity. Previous studies are evaluated and combined to obtain a suitable operationalization of reliability and connectivity. First, the operationalization of reliability is described and following, connectivity.

2.1 RELIABILITY

Reliability can be defined as how closely the actual day of operation matches the flight schedule on time. However, it is infeasible as airline to actually influence the unexpected disruptions on the day of operation. Therefore, the flight schedule should match the day of operation closely to ensure reliability.

The purpose of reliability in this research is to create a flight schedule that matches the day of operation closely, given the uncertainties and disruptions, such that executing flights becomes more reliable.

2.1.1 Previous Theories

Schedule reliability has been studied by several researchers. In most studies, schedule reliability is related to the on-time performance of the schedule which is in line with the definition of airlines (L. H. Lee et al., 2007). Airlines define reliability as the on-time performance, i.e. on time compared to the time standards. For this, the punctuality Key Performance Indicators (KPIs) are defined for many different processes. For the reliability of the schedule, the arrival and departure punctuality are the most important. Network planners mostly focus on the arrival punctuality at the hub.

C.-L. Wu (2005) differentiates arrival and departure reliability of a schedule. It defines reliability by the expected arrival and departure delay:

$$reliability^{D} = \frac{expected\ delay^{D}}{actual\ delay^{D}} \tag{2.1}$$

and

$$reliability^{A} = \frac{expected\ delay^{A}}{actual\ delay^{A}} \tag{2.2}$$

where

D = departing flights

$A =$ arriving flights

They assume that airlines expect a certain delay in advance. However, when creating the schedule, most airlines do not expect any delay and this could give an incorrect calculation of reliability.

L. H. Lee et al. (2007) and Sohoni et al. (2011) base the on-time performance of a flight on the aircraft arriving at the gate within 15 minutes of its originally schedule arrival time. This is an important operational measure for schedule reliability according to these studies.

Burke et al. (2010) describes departure reliability for schedule A, $R(A)$, as the probability p_i that the next flight $(f_i + 1)$ of an aircraft can leave on time, given the time that is allocated to its previous flight (f_i) and its turn around operations. This is a measure of schedule's ability to absorb the effects of the operating environment. p_i is estimated based on stochastic distributions from historical data. There is a penalty when the minimum reliability, p_{min} is violated. The scaling parameter y is introduced to manipulate the quadratic shape of the function and influence the ratio for less reliable connections versus more reliable connections. This gives the following calculation of the reliability per flight:

$$p_i > p_{min} : R(A) = f(p_i) = (1 - p_i)^y \tag{2.3}$$

$$p_i \leq p_{min} : R(A) = f(p_i) = (1 - p_i)^y + P(p_{min} - p_i) \tag{2.4}$$

And the reliability of schedule A:

$$R_i(A) = \sum_{i=0}^{|f|} R_i(A) \tag{2.5}$$

Their formulation mostly focuses on the reliability of aircraft handling on the ground and does not include the airborne punctuality of flights.

These studies have in common that they relate schedule reliability with delay minutes of a flight. In other sectors such as railway, studies also define delay minutes or hours as a measure for punctuality (Olsson & Haugland, 2004; Veiseth et al., 2007). Delay minutes can also be defined as an indicator for traffic congestion namely the weighted average of total delay minutes for each road link with length as weight (Christidis et al., 2012). In addition, Chen et al. (2003) notes that the average travel time and travel variability are important measures for freeway performance.

2.1.2 Quantification

Combining the insight of previous theories and sectors, the reliability of flight i can be measured with delay minutes, thus how many minutes a flight has "lost" compared to the scheduled time.

The reliability of flight i can be defined by the average of delay minutes of that particular flight in schedule ξ. It is assumed that flights with the same flight number and on the same day, for a given time window W, is

$i = f_1, f_2, ..., f_n$ [1]. This gives the following formula for reliability of flight i:

$$R_a(i) = \frac{\sum_{f \in i} max(0, ata_f - sta_f)}{i} \tag{2.6}$$

and

$$R_d(i) = \frac{\sum_{f \in i} max(0, atd_f - std_f)}{i} \tag{2.7}$$

where

R_a = arrival reliability

R_d = departure reliability

ata_f = actual time of arrival of flight $f \in i$

sta_f = scheduled time of arrival of flight $f \in i$

atd_f = actual time of departure of flight $f \in i$

std_f = scheduled time of departure of flight $f \in i$

Reliability of the schedule is measured by the summation of all flights in the schedule.

$$R_a(\xi) = \sum_{i \in \xi} R_a(i) \tag{2.8}$$

and

$$R_d(\xi) = \sum_{i \in \xi} R_a(i) \tag{2.9}$$

where

ξ = flight schedule

Airlines mostly focus on the arrival punctuality as on-time performance and therefore, this paper only uses arrival reliability as a measure of reliability.

2.2 CONNECTIVITY

The temporal configuration of an airline network, i.e. organisation of the flight schedule such that a given number and quality of indirect connections is offered at a station, is one of the main features of a hub-and-spoke network (Burghouwt & de Wit, 2005). The other main feature is the concentration of air traffic in space, i.e. spatial configuration (Reynolds-Feighan, 2001). Spatial configuration is less relevant in this research since the amount of air traffic and the spatial aspect of the network will be fixed.

1 For example, flight (i) KL1009 to London is scheduled on Monday. Every individual flight (f) KL1009 on Monday during Summer 2019 (W) is part of the collection of flights i to average over.

Indirect connectivity is often associated with hub-and-spoke networks. Hub connectivity refers to the number and quality of indirect flights available to passengers via an airline hub (Burghouwt & de Wit, 2005; Danesi, 2006; S. Y. Lee et al., 2014). Quality can be defined in various ways for example, Burghouwt & de Wit (2005) defines it as the attractiveness of a connection for a passenger and S. Y. Lee et al. (2014) as the convenience level of a connection for a passenger. According to Danesi (2006), hub connectivity depends on three elements namely

1. number of markets linked to the hub with direct services,
2. service frequencies, and
3. times of scheduled flight arrivals and departures at the hub.

A way to enhance the hub connectivity is to adopt a wave-system in the network. Currently, hubs operate more often with waves of flights. Purpose of a hub wave-system is to maximize its connectivity. A wave-system structure of an airline hub has the following elements namely, (Burghouwt & de Wit, 2005; Danesi, 2006)

i. the number of flight waves,
ii. the timing of the waves and the time interval between the same points of consecutive waves, also referred to as the "hub-repeat-cycle", and
iii. the structure of the individual waves.

The structure of an individual wave is determined by the minimum connecting time for (inter)continental flights, the maximum acceptable connecting times and the maximum number of flights that can be scheduled in a time period.

2.2.1 Previous Theories

Hub connectivity has been quantified by several studies, mostly by means of indices.

Burghouwt & de Wit (2005) defines the hub connectivity as the number and quality of the indirect connections generated by the existing flight schedule. In their case, the quality of the indirect connections refers to the attractiveness of the connections. They specify attractiveness as the perceived transfer time and the in-flight time compared to the direct flight time.

A weighted indirect connectivity index (WNX) for a schedule is created namely,

$$WNX = \sum(WI) = \sum\left(\frac{2.4 * TI + RI}{3.4}\right) \tag{2.10}$$

with TI the transfer index which refers to the quality of the connection, and RI the routing index which refers to the quality of the indirect flight compared to the direct flight. With this index, it is assumed that the passengers perceive that their transfer time is 2.4 times longer than the in-flight time.

Danesi (2006) suggests a novel Weighted Connectivity Ratio (WCR) consisting of the weighted indirect connection number and the approximate number of weighted connections in a purely random situation, during time period T. The weighted indirect connection number at the hub during time period T is by the following three matrices:

i. temporal connectivity matrix, $\tau_{i,j}$

ii. spatial connectivity matrix, $\delta_{i,j}$

iii. weighted connectivity matrix, $w_{i,j} = \tau_{i,j}\delta_{i,j}$

where $i = 1, ..., f_a$ is any arriving flight during T and $j = 1, ..., f_d$ is any departing flight during T.

The number of weighted connections offered at the airline hub during time period T is:

$$WN_c = \sum_i \sum_j w_{i,j} = \sum_i \sum_j \tau_{i,j}\delta i, j \tag{2.11}$$

This measure has a discrete character, meaning that it only has four different states.

For the WCR, it is needed to calculate the approximate number of weighted connections in a purely random arrival and departure timetable, WN_r. With this, WCR can show whether the viable weighted connections of the airline are more or better than purely random. WCR is defined as

$$WCR = \frac{WN_c}{WN_r} \tag{2.12}$$

It is concluded that WCR is quite precise for the evaluation procedure of the hub connectivity and WN_c can be seen as an acceptable hub connectivity measure.

Kim & Park (2012) presents a connectivity index that measures the relationship between arrivals and departures of flights in one day. Their research is mostly focused on the freight connectivity within airline industry. In their research, the quality of an indirect connection is measured by the difference of indirect flight time versus direct flight time. The following connectivity index (CI) for schedule is defined

$$CI = \sqrt{\sum_{t=1}^{24} \left\{ \left(\frac{A_t}{AA} \right) - \left(\frac{D_{t+LUT}}{AD} \right) \right\}^2} \tag{2.13}$$

where A_t is the number of flights that have yet to arrive at the hub at time t, D_{t+LUT} is the number of flights that have yet to depart at the hub at time $t + LUT$, LUT is the loading and unloading time, t is the time slot that varies from 00:00 hour to 24:00 hour for a given day, AA is the average flights of all A_ts and AD is the average flights of all D_ts.

Following the WCR of Danesi (2006), S. Y. Lee et al. (2014) developed the Continuous Connectivity Index (CCI) for hub-and-spoke operations consisting of:

i. temporal connectivity index, $\tau_{i,j}$

ii. spatial connectivity index, $\delta_{i,j}$

iii. relative intensity index, $\beta_{i,j}$

iv. weighted connectivity index, $w_{i,j} = \tau_{i,j}\delta_{i,j}\beta_{i,j}$

The main difference between CCI and the weighted hub connectivity measure WN_c of Danesi (2006) is that CCI has a continuous character and has an extra weighted element namely the relative intensity to reflect the effect of direct flight frequency on transfer routes.

Lastly, O'Connell & Bueno (2018) gives an overview of the different measurements for hub connectivity. This overview distinguishes the temporal coordination, routing factor and the character of the connection quality (None, Binary, Discrete or Continuous) of various hub connectivity indices. In the end, they apply the weighted connectivity ratio of Danesi (2006) in their research.

2.2.2 Quantification

Previous connectivity indices and studies have been combined to define a suitable hub connectivity measure for this research.

As mentioned, hub connectivity refers to the number and quality of the indirect connections at the hub. In this research, the quality of the indirect connections is determined by the transfer time for the passengers and the revenue of a connection for the airline.

In an airline with a hub-and-spoke network, schedule planners mostly focus on the entire rotation when scheduling. A rotation, r, consist of two or more flight legs and is seen as one block within the network schedule. In a hub-and-spoke network, one rotation mostly consist of two flight legs namely a flight from the hub to an outstation as outbound flight, j, and a flight from the outstation to the hub as inbound flight, i.

With the help of the existing discrete temporal connectivity matrix of Danesi (2006) and continuous temporal connectivity index of S. Y. Lee et al. (2014), the temporal connectivity index for this research is operationalized. For an airline, the actual number of passengers catching their transfer is interesting for connectivity. The percentage of passengers that make the transfer is a function of the transfer time, $tt_{i,j} = std_j - sta_i$, and can be extracted from real airline data. This can be translated to the probability that a passenger will actually be boarded on the transferring flight given the scheduled transfer time, $p_{tp,k}(tt_{i,j})$. This probability differs per transfer type k, for example it is faster to transfer from an Europe flight to another Europe flight than an international flight since the distance between gates for Europe flights is often smaller. With this, the following holds for this continuous temporal connectivity index for all possible connections between any flight arriving at the hub during time period T, $i = 1, ..., f_n$, and any flight departing at the hub during time period T, $j = 1, ..., f_n$:

$$\tau_{i,j} = \begin{cases} p_{tp,k}(tt_{i,j}), & \text{if } MCT_k \leq std_j - sta_i \leq MACT_k \\ 0, & \text{otherwise} \end{cases} \tag{2.14}$$

where

MCT_k = minimum connecting time for a passenger with transfer type k

$MACT_k$ = maximum acceptable connecting time for a passenger with transfer type k

The revenue of a passenger connection for an airline is included to measure hub connectivity. The expected number of passengers connecting from flight i to flight j, $E[tp]_{i,j}$, is taken as an approximation for earnings, $\rho_{i,j}$, of a connection for an airline. This is multiplied by the revenue of one passenger connecting between flight i and j, $r_{i,j}$, defined by the commercial branch. The following holds for all possible connections between arriving flights i and departing flight j:

$$\rho_{i,j} = E[tp]_{i,j}\, r_{i,j} \tag{2.15}$$

The weighted connectivity matrix for a connection can be defined as:

$$w_{i,j} = \tau_{i,j}\rho_{i,j} \tag{2.16}$$

This results in the following continuous weighted hub connectivity measure, thus the number of weighted connections offered at the airline hub during time period T:

$$C = \sum_i \sum_j w_{i,j} = \sum_i \sum_j \tau_{i,j}\rho_{i,j} \tag{2.17}$$

The continuous weighted hub connectivity measure only describes the outbound connectivity of the hub, thus which inbound flights connect to the outbound flights at the hub during period T. The revenue of a connection differs for inbound and outbound connections of a flight and they should both be taken into account. Thus, it is necessary to distinguish the inbound and outbound connectivity on rotation level in this research. Therefore, two elements are defined for a rotation namely inbound connectivity, C_{in}, and outbound connectivity, C_{out}. This follows:

$$C_{in} = \sum_j \sum_i \tau_{i,j}\rho_{i,j} \tag{2.18}$$

and

$$C_{out} = \sum_i \sum_j \tau_{i,j}\rho_{i,j} \tag{2.19}$$

Figure 2.1 shows the inbound and outbound connectivity related to a rotation.

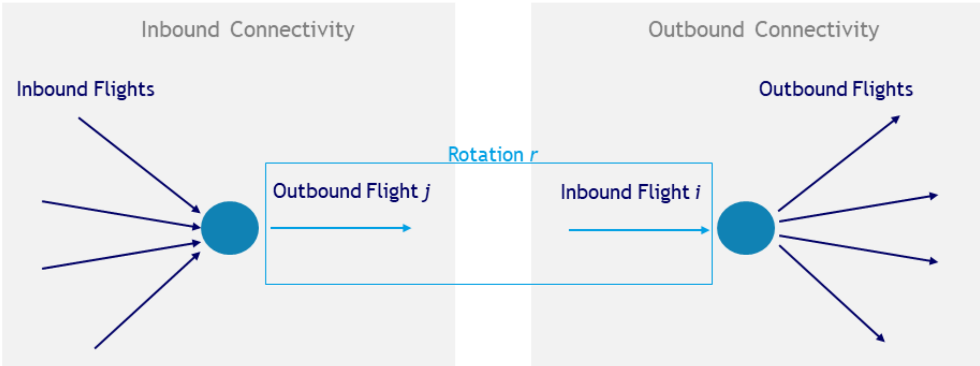


Figure 2.1: Inbound and outbound connectivity

Note that Danesi (2006) defines the weighted connectivity ratio to show whether the viable weighted connections during time period T are more than purely random for optimal hub coordination. Since this research only focuses on the hub connectivity and not the hub coordination, it is not necessary to incorporate this step. This could be interesting to include when testing whether the hub connectivity is improved for further research.

3 CURRENT STATE OF RELIABILITY AND CONNECTIVITY

This chapter gives insight in the current state of reliability and connectivity of the flight schedule of KLM Royal Dutch Airlines. First, the flight procedure and the process of the schedule design are explained. Next, the current flight schedule of Summer 2019 is evaluated.

3.1 CONCEPTUALIZATION OF FLIGHT OPERATIONS

In order to understand how the flight schedule is created, it is important to get familiar with the flight procedure. Thus, the process of executing flights is discussed first. Hereafter, it is explained how the schedule is currently designed.

3.1.1 Flight Procedure

A flight is a trip between origin and destination in the air. Every flight has its own flight number, origin and destination. A flight is scheduled on a specific day and time, and assigned to a specific aircraft type. The scheduled times are for departing from the origin, Scheduled Time of Departure (STD), and for arriving at the destination, Scheduled Time of Arrival (STA).

Before a flight can be executed, preparations are needed such as cleaning the aircraft, safety checks, boarding passengers and baggage. The activities on the ground that happen prior to take-off, are called ground processes. The combination of these preparatory ground processes and the flight, also known as airborne time, define a leg. Figure 3.1 visualizes these definitions.

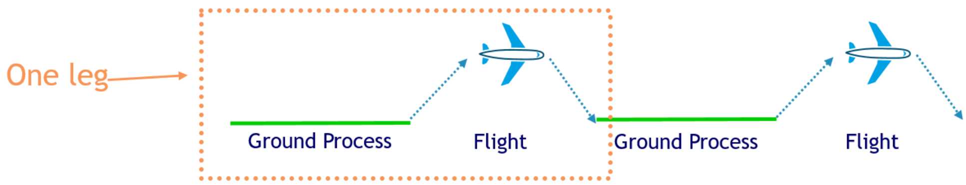


Figure 3.1: Conceptualization of a leg

During the day of operation, the scheduled flights are executed. A flight is assigned to an aircraft and leaves the origin on the Actual Time of Departure (ATD). This means that the aircraft is actually leaving the gate, also called blocks-off. The aircraft is flying towards their destination, which is called airborne or flight time, and arrives at the destination on the Actual Time of

Arrival (ATA) at the gate. ATA is defined as the time that the aircraft is at the gate and the blocks are on. Preferably, ATD and ATA are the same as STD and STA which means no delay occurred.

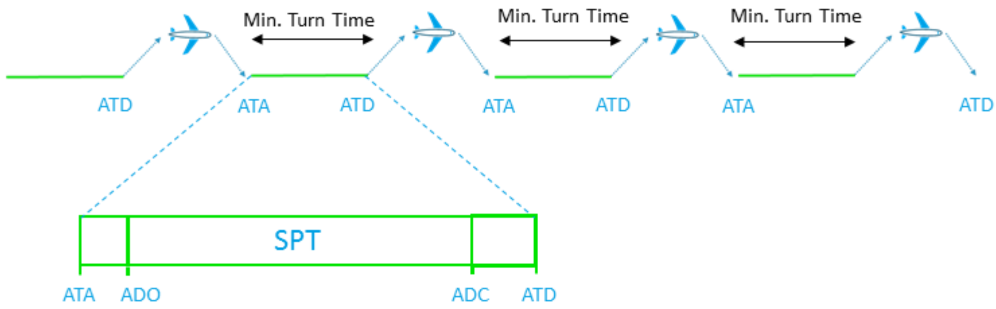


Figure 3.2: Conceptualization of flight process

When the aircraft has arrived at his destination, it is assigned to a gate. At the gate, the unloading of this arriving flight starts and the preparations of the next flight begin. This procedure is called the turnaround of an aircraft, which is the time between ATA of the arriving flight and ATD of the next flight. The ground processes start when any door has been opened at the gate, so either a passenger or baggage door. This is around three minutes later than ATA. This moment is called Any Door Open (ADO) and then the Standard Processing Time (SPT) starts. SPT is the time in which the aircraft is prepared for the next flight until all the doors are closed (ADC). According to the norm, the aircraft should leave soon after later after ADC; this moment is called ATD. The ATD of a flight is determined by air traffic control (ATC) of the origin station and this could be delayed due to, for example, other flights that have priority. This is defined as ATC delay and is the delay between ADC and ATD. When an aircraft gets permission to leave the gate, the breaks are released, the aircraft starts taxiing and takes off to the next destination. An brief overview of this flight procedure is shown in Figure 3.2.

3.1.2 Schedule Design

The flight schedule for a season is created by the network department. This department determines the destinations, important connecting flights, frequency of the flights and the aircraft type, and therewith the profitability of the schedule. They are also responsible for creating an operationally feasible schedule by taking into account constraints such as slots and crew (Barnhart & Cohn, 2004).

3.2 EVALUATION OF THE BASE CASE SCHEDULE

This section contains a detailed evaluation of the flight schedule of Summer 2019 for European flights. The period of the summer schedule of KLM is during daylight saving time in the Netherlands. The focus of this evaluation is on the months July and August. These months are the busiest of the entire year. Therefore, it is the most challenging and interesting to find a balance between delay minutes and connectivity value. Next to that, the months May and June are not suitable for analysis since there was one fewer runway than in a normal situation due to maintenance, which led to outliers in the data.

In this research, one week in July and one week in August is evaluated namely the week from July 22, 2019 to 28 July, 2019 and the week from August 12, 2019 to August 18, 2019. These weeks are chosen within July and August to test the novel optimization model. This is the base case schedule for the research. Note that the base case schedule already includes some buffer, thus the base case does not mean that there is no buffer between rotations. Currently, only one aircraft type is analyzed namely Boeing 737-900.

From the analysis, it can be there is no clear indication of the value of reliability due to the stochastic variance. On the connectivity side, there is no confidence interval for the percentage of missed connection by passengers. An explanation is that there is a cluster of connections that are close to the minimum transfer time or the maximum acceptable transfer time.

The performance of the base case schedule for July 22, 2019 to July 28, 2019 and August 12, 2019 to August 18, 2019 is used as a benchmark for evaluating new schedules with various ways of buffer scheduling.

4 FORMULATION OF MULTI-OBJECTIVE OPTIMIZATION MODEL

This chapter introduces a novel multi-objective optimization model for designing a flight schedule by trading off between reliability and connectivity. First, the methodology and search algorithm of this optimization is described. Hereafter, the objectives are discussed. Lastly, the model formulation is presented.

4.1 MULTI-OBJECTIVE OPTIMIZATION ALGORITHM

The problem of this research has a multi-objective nature because reliability needs to be minimized while connectivity needs to be maximized. Therefore, a multi-objective optimization methodology is applied. Another approach to solve this problem is to scalarize the multiple objectives to a single-objective problem (Deb, 2014). The goal of this approach is to find a single optimum solution. There are two disadvantages in solving the problem of this research as a single-objective problem namely, (i) accurately converting the two objectives to the same unit is currently impossible and (ii) the novel optimization model supports decision-making by giving insight in the trade-off between the objectives, this is not possible when obtaining only one optimal solution. The single-objective approach is not suitable for this research and thus, the multi-objective optimization methodology is adopted.

The main advantages of the multi-objective approach are that the objectives are optimized simultaneously, instead of sequentially, and the trade-offs between the objectives are identified with the help of the Pareto optimal front (Burke et al., 2010; Emmerich & Deutz, 2018; Kollat & Reed, 2007). The Pareto front is the set of non-dominated Pareto solutions. Non-dominated Pareto solutions are solutions that cannot improve one objective without deteriorating the performance of another objective (Emmerich & Deutz, 2018). In the context of this research, this means that, for a Pareto solution, buffer scheduling cannot be changed in such a manner that it improves reliability without decreasing connectivity revenue and vice versa.

A widely used approach for optimizing a multi-objective problem is the multi-objective evolutionary algorithm (MOEA), i.e. a population-based search algorithm (Vikhar, 2016). In this class, genetic algorithms (GA) are qualified to generate high-quality solutions for optimization problems based on the concept of natural selection in Darwin's theory of evolution (Mitchell, 1996). Over the years, numerous MOEAs are introduced by research such as NSGA-II, NSGA-III, IBEA, SPEA2, MOED/D, ε-NSGAII, ε-MOEA, BORG.

The most classic and popular approach to generate the Pareto front for a multi-objective problem is Non dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002; Reed et al., 2013). NSGA-II initially generates a random parent population P_0 which is sorted based on non-domination. Then it creates an offspring population Q_0 with binary tournament selection, recombination and mutation operators. Hereafter, a step-by-step procedure is followed: first, a combined population, R_t, is created from the random parent population, P_t, and offspring , Q_t. This combined population is sorted on non-domination and divided in sets $F_{1,\dots,l}$. F_1 is the best non-dominated set and will definitely be chosen for the next parent population, P_{t+1}. The remaining solutions for the new population of size N will be chosen based on the ranking and crowding distance sorting. The new population is used for creating new offspring Q_{t+1} with selection, crossover and mutation. The NSGA-II procedure is visualized in Figure 4.1 and is described in more detail by Deb et al. (2002).

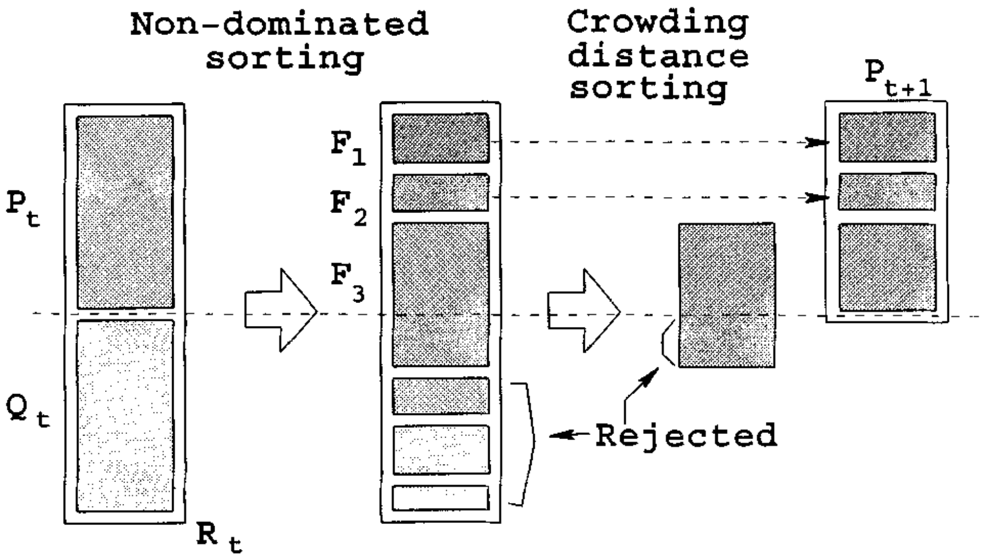


Figure 4.1: NSGA-II procedure (Deb et al., 2002, pp.186)

One of the main disadvantages of NSGA-II is the fixed-sized population, N. The algorithm replaces offspring with existing members in the population, which could lead to offspring being replaced with inferior solutions. This deteriorates the quality of the solutions (Reed et al., 2013). Therefore, ε-NSGAII with an adaptive population size and time continuation is introduced. This search algorithm merges ε-dominance archive with NSGA-II for more efficient, reliable and ease-to-use MOEA (Kollat & Reed, 2006, 2007). With ε-dominance, the user is able to set the search precision for each objective (Kollat & Reed, 2006). The adaptive population size improves the search quality by adapting the population based on the Pareto approximate set size (Kollat & Reed, 2006; Reed et al., 2013). Time continuation refers to the series of connected runs where small populations are exploited to pre-condition search and population size is adapted accordingly (Kollat & Reed, 2005, 2006).

Kollat & Reed (2006) describes the concept of ε-dominance, used in ε-NSGAII, with a three steps approach as shown in Figure 4.2. In step 1, the decision maker determines the ε search precision for each objective (f1 and f2) which results in a ε-grid. Thus, the search space is divided in grids with ε-precision. The smaller the ε, the finer the grid and the more solutions in the objective space. For grid blocks with multiple solutions and assuming minimization of the objectives, the solution closest to the lower left-hand corner of the block is kept. Step 2 conducts non-dominated sorting based on the grid block. For example, the solution in the leftmost column on row four from the bottom dominates the shaded grid blocks above and to the right in terms of required precision. Step 3 then eliminates the redundant solutions and presents a more even search of the objective space. For a more detailed description of ε-dominance, see Deb et al. (2002) and Laumanns et al. (2002).

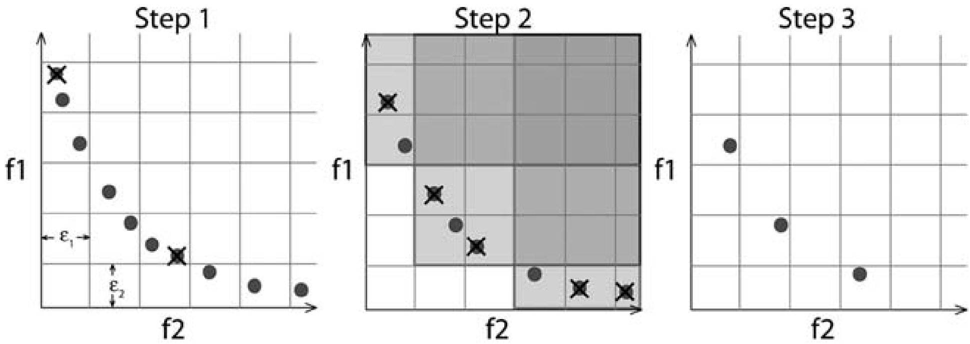


Figure 4.2: Illustration of the ε-dominance concept (Kollat & Reed, 2006, pp.797)

However, ε-NSGAII is performing worse when the number of decision variables increases (larger than 100) or when the complexity of the problem increases. Ward et al. (2015) shows that BORG would perform better than ε-NSGAII when the population size and number of decision variables increases. This is tested with the performance measure hypervolume. Hypervolume describes the ratio of dominated solution compared to the population in the objective space (Reed et al., 2013). Their paper states that BORG is expected to have a better performance, even the best, with many decision variables and for a complex problem.

BORG is an optimization algorithm that combines ε-domaninance, adaptive population size and time continuation with an adaptive operator selection (Hadka & Reed, 2013; Reed et al., 2013). It is an extension of the ε-NSGAII with adaptive operator selection. This means that the algorithm keeps track of the performance of the operators and adapts to the most appropriate operator. Operators of BORG are binary crossover, differential evolution, parent-centric recombination, unimodal normal distribution crossover, simplex crossover, polynominal mutation and uniform mutation. Next to that, BORG also keep track of ε-progression. Hereby, it can easily detect search stagnation. Another feature is that BORG is a steady-state algorithm. BORG is mostly used for many-objectives (more than four objectives), multimodal problems.

Due to the high complexity of the problem in this research, BORG is applied as algorithm for the optimization. Since this algorithm has a significantly better performance for large problems than other algorithms, this is the most suitable choice.

4.2 OBJECTIVES

A schedule consists of $u = 1, \ldots, U$ fleetlines and these fleetlines consist of $n = 1, \ldots, N$ rotations. Let $X_{u,n}$ be the time in which the arrival and departure of rotation n shifts in fleetline u and τ the unit of buffer time[1]. This is the primary decision variable of this research since this determines the buffer time of rotation n. One rotation consists of two or more flight legs; within the Europe schedule, one rotation generally consist of two flight legs namely a flight from the hub to an outstation as outbound flight, j, and a flight from an outstation to the hub as inbound flight, i.

Reliability Objective
The reliability objective is the average arrival delay minutes of all the rotations in the flight schedule. For rotation $n \in N$, the arrival reliability of rotation n is the arrival punctuality of the last flight leg arriving at the hub, the inbound flight. Thus flight i is equal to the outstation-to-hub flight of the rotation n. This gives:

$$R_a(i) = \frac{\sum_{f \in i} max(0, ata_f - [sta_f + X_n\tau])}{i} \tag{4.1}$$

The reliability objective for a flight schedule ξ with a set of fleetlines U is formulated as:

$$R_a(\xi) = \overline{\sum_{u \in U} \sum_{n \in N} R_a(X_{u,n}\tau)} \tag{4.2}$$

Connectivity Objective
The connectivity objective is total connectivity in euros of all the rotations N in the flight schedule. The connectivity of rotation n is divided into inbound connectivity and outbound connectivity of a rotation. Inbound connectivity is determined by the connections between inbound flights on hub station and the first flight leg of the rotation from hub station to outstation. Outbound connectivity is determined by the connections between the last flight leg of the rotation arriving at hub station from outstation and the related outbound connection from hub station. Thus, the outbound flight j of rotation n determines the inbound connectivity, C_{in}, and the inbound flight i of rotation n determines the outbound connectivity, C_{out}.

Inbound connectivity for flight j of rotation n can be defined as

$$C_{in} = \sum_i \begin{cases} p_{tp,k}(tt_{i,j})E[tp]_{i,j}\, r_{i,j}, & \text{if } MCT_k \leq tt_{i,j} \leq MACT_k \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

with $tt_{i,j} = [std_j + X_n\tau] - sta_i$

1 Most airlines make schedule adjustment with a unit of 5 minutes to ensure that too accurate scheduled departure and arrival times do not occur such as a scheduled arrival at 10:12 a.m. Therefore, this research takes into account the unit of the buffer time, τ.

Outbound connectivity for flight i of rotation n can be defined as

$$C_{out} = \sum_j \begin{cases} p_{tp,k}(tt_{i,j})E[tp]_{i,j}\, r_{i,j}, & \text{if } MCT_k \leq tt_{i,j} \leq MACT_k \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

with $tt_{i,j} = std_j - [sta_i + X_n\tau]$

The connectivity objective for a flight schedule ξ with a set of fleetlines U is defined as follows:

$$C(\xi) = \sum_{u\in U}\sum_{n\in N} C_{in}(X_{u,n}\tau) + C_{out}(X_{u,n}\tau) \tag{4.5}$$

4.3 MODEL FORMULATION

To find the optimal combination of $X_{u,n}$ in a chain of flights, the entire fleetline is optimized simultaneously. The interdependencies between the arrival and departure times are taken into account for this. There are two main objectives for this optimization namely the average delay minutes of the schedule and the sum of inbound and outbound connectivity of the schedule. Let U be the number of fleetlines in the schedule and N be the set of rotations within a schedule. This gives the following optimization problem:

$$\text{minimize} \quad \overline{\sum_{u\in U}\sum_{n\in N} R_a(X_{u,n}\tau)}, \quad \text{maximize} \sum_{u\in U}\sum_{n\in N} C_{in}(X_{u,n}\tau) + C_{out}(X_{u,n}\tau) \tag{4.6}$$

$$\text{subject to} \quad \sum_{n\in N}(X_{u,n}\tau) \leq T_{u,N} \qquad \forall u \tag{4.7}$$

$$MCT_k \leq tt_{i,j} \leq MACT_k \qquad \forall i,j \tag{4.8}$$

$$[std_{j,n} + (X_{u,n}\tau)] - sta_{i,n-1} \geq MTT_{ac} \qquad \forall u,n \tag{4.9}$$

$$std_{j,n+1} - [sta_{i,n} + (X_{u,n}\tau)] \geq MTT_{ac} \qquad \forall u,n \tag{4.10}$$

Constraint (3.5) ensures that no extra time could be added to a day. The total buffer time of the day per fleetline, T_N, can only be re-allocated. This means that the sum of the buffer time of all rotations of one fleetline should be less or equal to the total amount of buffer time that could be divided for a fleetline. This account for all fleetlines in the schedule. Constraint (3.6) ensures that the transfer time of a passenger between inbound flight i and outbound flight j is not smaller than MCT and not larger than MACT. Constraints (3.7) and (3.8) ensure that the turnaround time between two rotations in a fleetline is equal or more than the minimum turnaround time of the aircraft type. The sequence of the rotations is fixed since $n-1$ is the previous rotation of rotation n and $n+1$ is the following rotation of rotation n. Table 4.1 provides an overview of the mathematical notations.

Table 4.1: Table of Notation

Objectives	
C_{in}	Inbound connectivity in euros
C_{out}	Outbound connectivity in euros
C	Connectivity, sum of inbound and outbound connectivity, in euros
R_a	Arrival reliability in minutes
Decision Variable	
$X_{u,n}$	Buffer time for rotation n in fleetline u
Parameters	
ac	Aircraft type subscript
ata	Actual time of arrival
f	Flight repetition subscript
i	Inbound flight subscript
j	Outbound flight subscript
k	Connection type subscript
n	Rotation subscript ($n = 1, \ldots, N$)
$p_{tp,k}$	Probability that a passenger with connection type k will actually be boarded on the transferring flight
$r_{i,j}$	Revenue of one passenger connecting from flight i to flight j
sta	Scheduled time of arrival
std	Scheduled time of departure
$tt_{i,j}$	Transfer time between flight i and j
u	Fleetline subscript ($u = 1, \ldots, U$)
$E[tp]_{i,j}$	Expected number of passengers connecting from flight i to flight j
$MACT_k$	Maximum acceptable connecting time for connection type k
MCT_k	Minimum connecting time for connection type k
MTT_{ac}	Minimum turnaround time for aircraft type ac
T_N	Total buffer time for a fleetline with rotation set N
ξ	Flight schedule
τ	Unit of buffer time

5 CONFIGURATION OF OPTIMIZATION MODEL

This section describes the configuration of the optimization model as formulated in Chapter 4. First, the process of the simulation model and the assumptions are presented. Next, the optimization model is described in more depth with its' assumptions. Following, the interaction between the simulation model and optimization model is described, i.e. how these models intertwine. Lastly, the Key Performance Indicators (KPI) are defined.

5.1 SIMULATION MODEL

PRINCIPLE The simulation model Voyager is designed and developed by KLM. Voyager is a simulation tool of the departure process, propagated delay processes and collaborative decision making (CDM), i.e. joint decisions between ATC, Schiphol and airlines on the day of operation to improve the departure process. It reconstructs the departure queue and calculates the expected range of departure delay and propagated delay for a given KLM timetable. Voyager contains a detailed simulation of the CDM processes which have a direct effect on when the aircraft can actually depart. It gives insights into the relative operational performance of a flight schedule in terms of departure punctuality. Voyager only runs for one day of operation.

MAIN ASSUMPTIONS Modifications of the simulation model for this research led to additional relevant assumptions. The main assumptions of the simulation model are as follows:

- **Historical data of CDM is from 2017, 2018 and 2019.**
- **Historical rotation performance data is based on data from March 1, 2017 to October 29, 2019.** The period begins at the start of Summer 2017 and ends at the last day of Summer 2019. For consistency with CDM historical data, this time period is chosen for the rotation performance. Also this research focuses on the summer season and therefore, the summer period is chosen as starting and ending point.
- **Rotation delays, either positive or negative, of more than 60 minutes are not taken into account.**
- **Buffer time at the outstations is included in the arrival delay.** When buffer is planned at the outstation, some of the rotation delay can be captured by this buffer. Thus, the departure delay plus the sampled

rotation delay minus the buffer at the outstation accounts for the final arrival delay.

- **The number of replications is 10.** The number of required replications depends on the desired statistical accuracy of the research (Carson, 2004). For the inherent variability factors in the simulation, i.e. collaborative decision making moments drawn from a distribution and the rotation performance, 10 replications for one buffer allocation solution are necessary.

5.2 OPTIMIZATION MODEL

PRINCIPLE The principle of buffer scheduling in the optimization model works as follows. Buffer values per rotation are determined by the optimization algorithm of Section 4.1.

For one fleetline, the rotations can shift from the first arrival of the day and the last departure of the day, taken into account the MTT of that aircraft type. The relation within the fleetline determines the spread of buffer per rotation. For example, if rotation 1 shifts to 10 minutes earlier and there is 5 minutes buffer time initially scheduled between rotation 1 and 2, than rotation 2 could shift to 15 minutes earlier. This applies to each fleetline individually. Figure 5.1 shows the relation within fleetlines and how the (total) buffer times are determined.

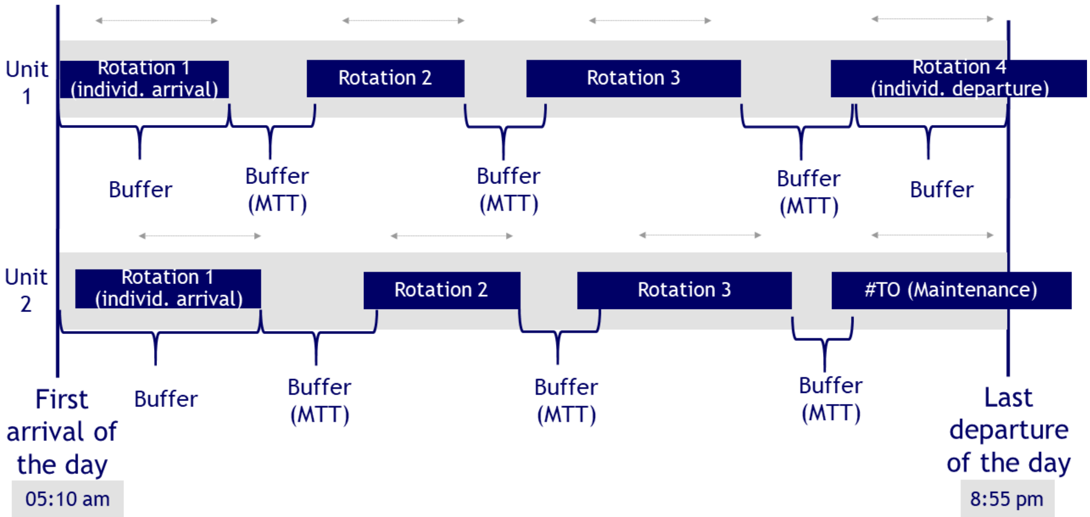


Figure 5.1: Relation within fleetlines

The relation between the fleetlines is based on connecting passengers. When a passenger is transferring from an inbound flight of a fleetline to an outbound flight from another fleetline, the shift of these flights affects the transfer time. The transfer time determines whether a connection can be offered and if the passengers actually make the transfer. Therefore, the relation between fleetlines is important for buffer scheduling in terms of connectivity. Figure 5.2 presents an example of this; there are three transferring passengers from rotation 1 of unit 1 to rotation 2 of unit 2. As an example, the initially scheduled transfer time is 50 minutes. Rotation 1 of unit 1 is shifted 10 minutes later and rotation 2 of unit 2 is shifted 5 minutes earlier; this is all possible according to the buffer time within the fleetline. However, the transfer time for these three transferring passenger is now 35 minutes which is lower than MCT. Thus, this connection could not be offered anymore. To determine the effect of buffer scheduling on connectivity, this relation between the fleetlines is included.

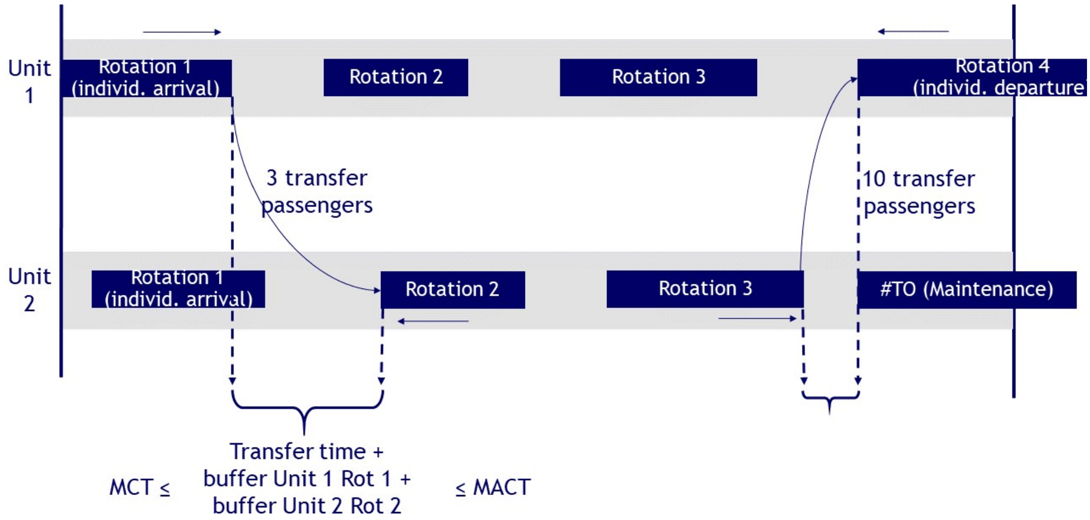


Figure 5.2: Relation between fleetlines

The principle of buffer scheduling relates to the creation of buffer values and the calculation of the objectives in the optimization model. The optimization itself follows the procedure of Section 5.3 and the steps of BORG algorithm as defined in Section 4.1.

MAIN ASSUMPTIONS The optimization model has the following main assumptions:

- **The model uses 10.000 function evaluations to optimize the problem.**
- **Time is defined as Coordinated Universal Time (UTC).**
- **The model optimizes for one day and one aircraft type.** The model only optimizes for one day at a time. Also, only the fleetlines of one aircraft type can be optimized at once.
- **There is a fixed fleet assignment.** It is assumed that a flight has to be performed with the assigned aircraft type.
- **The model assumes that the block time is always 100% accurate.** It is assumed that the scheduled block time is 100% accurate. This means that time between departing and arriving is the actual time that an aircraft needs to perform the flight, with 100% accuracy.
- **The model does not use early arrivals to compensate delays.** Compensating delays with early arrivals, is not be taken into account when allocating buffers. An early arrival is seen as "on-time", so no delay, in the model.

- **The model includes on-time maintenance.** Maintenance trips are included in the flight schedule to ensure they are executed. It is assumed that maintenance is always finished on time.
- **Probability of a passenger actually making their transfer based on the transfer time is determined per connection type from KLM data.** By this, the Schengen versus Non-Schengen connection types are incorporated.

More detailed description and other assumptions can be found in Appendix A.

5.3 SIMULATION-OPTIMIZATION MODEL

A simulation-optimization model is created to find optimal combinations of buffer allocation in a schedule. The model is developed based on the formulation of Section 4.3 and follows the optimization algorithm as defined in Section 4.1. The multi-objective optimization has two objectives namely, reliability, i.e. the average arrival delay minutes of a schedule, and connectivity, i.e. the sum of the revenue on inbound and outbound connections in the schedule. A simulation model is used to determine the reliability objective. This gives insight in the value of the average arrival delay, given a certain buffer allocation and sources of uncertainties. Next to this, the simulation model helps to evaluate and compare the optimal flight schedules with buffer scheduling.

The simulation-optimization model works as following; first, the user gives the desired flight date and aircraft type as input. With this part of the schedule, the optimization starts. The algorithm generates buffer values for all rotations in every fleetline in the schedule. Combining these buffer values with the initial schedule, a new schedule is created to analyze. Next, the objective values are calculated. Reliability is calculated by means of a simulation model with multiple replications to ensure statistical accuracy. With the results of the simulation model, the average arrival delay minutes of the new schedule is calculated as Equation 4.2. Simultaneously. the connectivity value is calculated by means of formulas as defined in Equation 4.5. Connectivity, as quantified in Section 2.2.2, has no source of uncertainty in the design of the flight schedule since this only involves the offered connections that are fixed before the actual day of operation. After the two objectives for the new way of buffer scheduling are computed, the constraints of Section 4.3 are applied. Following, it is checked whether this solution is non-dominated and saves it in the archive. New buffer values are generated following the optimization algorithm and the process repeats. The optimization continues until the user-defined number of function evaluations are completed and this gives the Pareto optimal solutions. This results in new flight schedules with optimal buffer scheduling. Figure 5.3 gives an overview of the simulation-optimization model and how they relate.

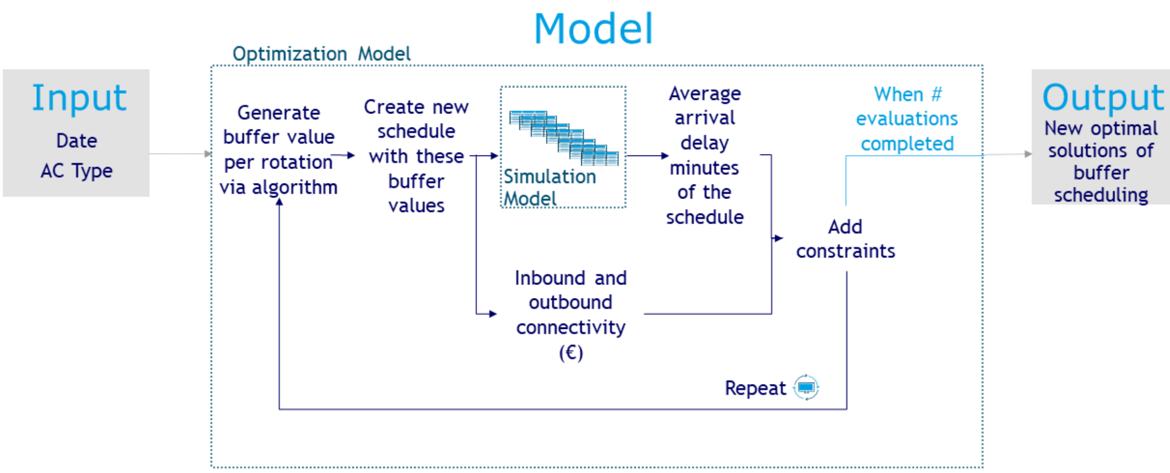


Figure 5.3: Visualization of the Simulation-Optimization Model

5.4 KEY PERFORMANCE INDICATORS

The performance of the schedules resulting from the simulation-optimization model is tested and compared to the current schedule. In order to test whether the Pareto solutions would perform better than the current schedule, several measurements are defined called KPIs. The current schedule is compared with the Pareto optimal schedules by five performance indicators. These indicators are set in consultation with schedule design experts. Table 5.1 gives an overview of the KPIs for evaluating the flight schedules with various ways of buffer allocations.

Table 5.1: Key Performance Indicators

Reliability	Connectivity
Average delay minutes of the schedule (min)	Connectivity of the schedule (€)
Total delay minutes of the schedule (min)	Percentage of missed connections (%)
	Average transfer time of a passenger at AMS (min)

As shown in Table 5.1, two categories are distinguished for evaluation namely, the two objectives reliability and connectivity. The values of the objectives, average arrival delay minutes of the schedule and connectivity in euros, are KPIs. Next to this, the arrival total delay minutes of the schedule is also an valuable measure since this gives an indication of the overall disruption. The connectivity in euros defines the offered connections whereas the percentage of missed connections defines the realized transfer with departure and arrival delay. The percentages of missed connections, i.e. the realized transfers with the buffer re-allocation, is also an important measure for connectivity purposes. This indicator depends on the arrival and departure delay of the connection flights. For example, the initial transfer time is 50 minutes; the inbound flight has an arrival delay of 10 minutes and the connecting outbound flight has a departure delay of 10 minutes. When only taken into account the arrival delay, the recalculated transfer time becomes 40 minutes and passengers would miss their connection. However, since the connecting flight has a departure delay of 10 minutes, the transfer time stays the same and passengers make their connections. Moreover, the average transfer time of a passenger at the hub (AMS) is a measure to indicate the customer experience of the flight schedule. A longer transfer time for a passenger results in a lower satisfaction level of this customers since the waiting time becomes longer.

For KPIs with a source of uncertainty, a 95% confidence interval is calculated in order to ensure statistical accuracy. This means that 95 of the 100 times the average of that value lies within this confidence interval. Only the measure connectivity of the schedule (€) has no confidence interval since this is a fixed sum of the price per connection.

6 RESULTS

This chapter presents the results of the optimization model. First, the results of the simulation-optimization model are discussed. This led to improvements in the optimization model which are explained hereafter. Note that no exact values of the results are presented due to confidentiality reasons.

6.1 RESULTS OF SIMULATION-OPTIMIZATION MODEL

The simulation-optimization model, as described in Chapter 5, is applied to the schedule of two days in July 2019 namely Tuesday 23 July 2019 and Thursday 25 July 2019. Optimization on the weeks from 22 July to 28 July and from 12 August to 18 August is not executed completely due to time limitations. The results of the simulation-optimization model are described with the help of the KPIs.

23 JULY 2019 Optimizing the flight schedule of 23 July 2019 resulted in two Pareto optimal solutions. Both solutions are better than the base case schedule in terms of reliability - less delay minutes - and connectivity. Figure 6.1 shows the Pareto front including the objectives of the base case schedule. The arrows on the axes point out the direction of desirability. Note that the connectivity axes are reversed: this means that the point closest to the lower-left corner is the most optimal.

There is a trade-off visible between the two Pareto optimal solution namely the solution with on average three arrival delay minutes has a lower connectivity than the solution with on average four arrival delay minutes.

Table 6.1 presents the buffer minutes added to the rotations of aircraft type Boeing 737-900 on 23 July 2019 compared to the base case. The Pareto optimal solutions are quite similar in terms of buffer scheduling; only five flights have a different buffer assignment namely KL1109 to ARN, KL1289 to EDI, KL1083 to MAN, KL1823 to TXL and KL0461 to TLV. The difference between the buffer assignment of these two optimal schedules is not more than 10 minutes per rotation. This means that a slight shift in buffer could already cause a different reliability value which suggests that the model is sensitivity to small buffer changes. In both optimal solutions, the most buffer time is added to the first flight or last flight of the day.

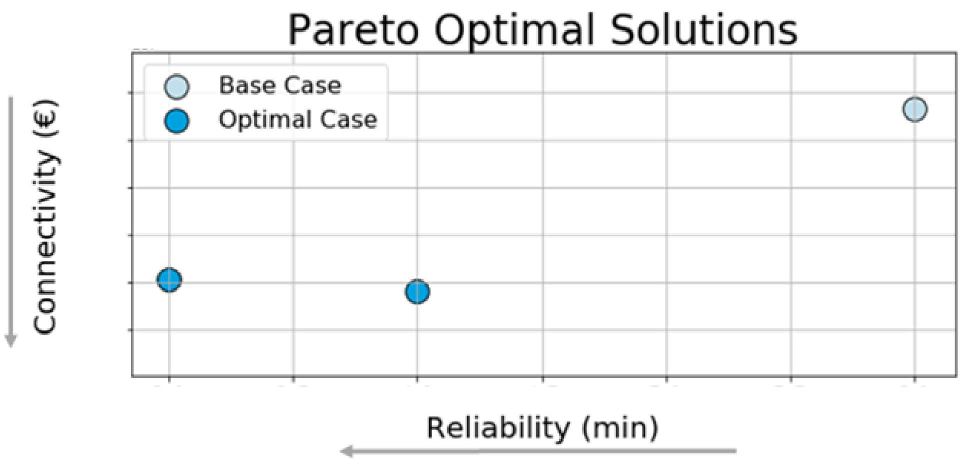


Figure 6.1: Simulation result: Pareto front for 23 July 2019

To ensure that the network planners make a well-informed choice for buffer scheduling, the KPIs for the Pareto optimal solutions are compared with the base case schedule to evaluate the performance.

Both Pareto optimal solutions are expected to have less arrival delay minutes on average than the base case. However, the wide of the 95% confidence interval is for all cases extremely large. This means that there is no clear indication of the actual value of arrival delay minutes. It could be explained by the high stochastic variance of the flights in the simulation model. Due to this wide confidence interval, it is not feasible to compare these values correctly. The same applies to the total arrival delay minutes of the schedule.

For the connectivity performance of the schedule, one of the KPIs is the connectivity objective. The comparison of the connectivity value between the base case and solutions is shown in Figure 6.2. This visualization shows that both Pareto optimal solutions have a larger connectivity value than the base case schedule. This means that more or more valuable connections can be offered with the Pareto optimal schedules. The solution with on average three arrival delay minutes has a slightly lower connectivity value than the other Pareto optimal solution. It is the decision of the network department if this difference is considered as significant or not.

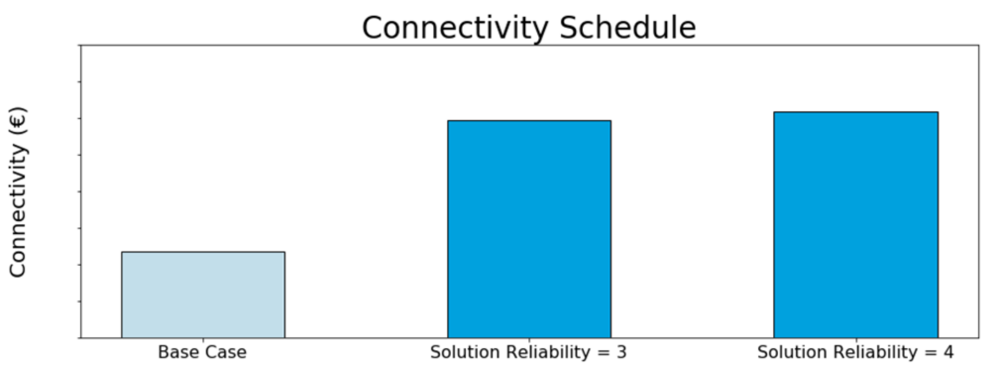


Figure 6.2: Simulation result: Connectivity value of the schedule on 23 July 2019

Table 6.1: Buffer values of the Pareto optimal solutions for 23 July 2019 in minutes

Rotation Number	Destination	Base Case	Solution Reliability = 3	Solution Reliability = 4
KL1070	MAN	0	10	10
KL1373	OTP	0	5	5
KL1705	MAD	0	5	5
KL1164	HEL	0	0	0
KL1109	ARN	0	-10	0
KL1289	EDI	0	15	10
KL1083	MAN	0	15	20
KL1372	OTP	0	0	0
KL1823	TXL	0	0	-5
KL1115	ARN	0	-5	-5
KL1379	OTP	0	0	0
KL1597	FCO	0	-25	-25
KL1169	HEL	0	5	5
KL1171	HEL	0	5	5
KL0462	TLV	0	-15	-15
KL1009	LHR	0	-5	-5
KL1189	BGO	0	5	5
KL0461	TLV	0	50	55

Figure 6.3 visualizes the percentage of missed connections by passengers in the schedule of 23 July 2019. Once more, the Pareto optimal solutions show a better performance in terms of percentage of missed connections. This is in line with the expectation that a more reliable schedule results in less missed connections. However, this does not apply for comparing the two optimal solutions. The solution with on average four delay minutes shows a lower percentage of missed connection by passenger than the solution with on average three delay minutes. This solution has a higher connectivity value, thus more (valuable) connections are offered. It could be that the solution offers more connections with a longer connecting time; passengers would always successfully make their transfer. So although the solution is less reliable, the transfer time of the offered connections plays an important role in the percentage of missed connection by passengers.

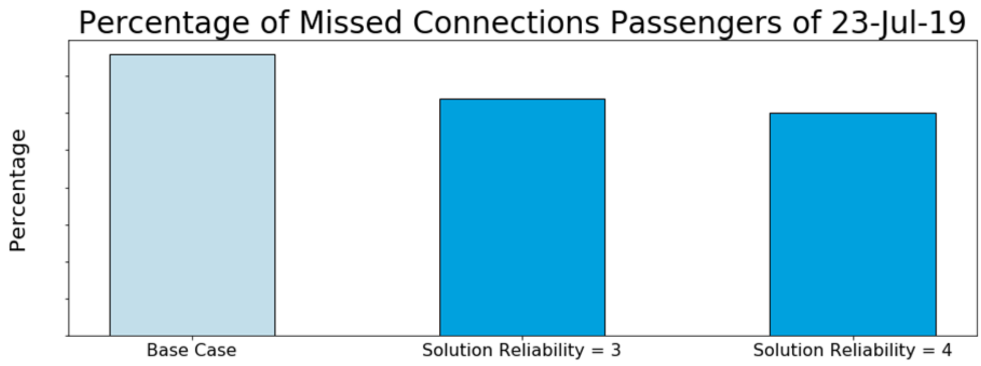


Figure 6.3: Simulation result: Percentage of missed connections by passengers in the schedule of 23 July 2019

When analyzing the average transfer time of the schedule, Figure 6.4 shows that the Pareto optimal solutions have a higher transfer time than the base case. Reasoning could be that the model tries to maximize the number of connections; the higher the average transfer time, the more connections could be offered. For example, a connection with a transfer time of 35 minutes cannot be offered but it is possible to sell a connection with a transfer time of 50 minutes. The difference of the average transfer time between the two Pareto optimal solutions is not significant.

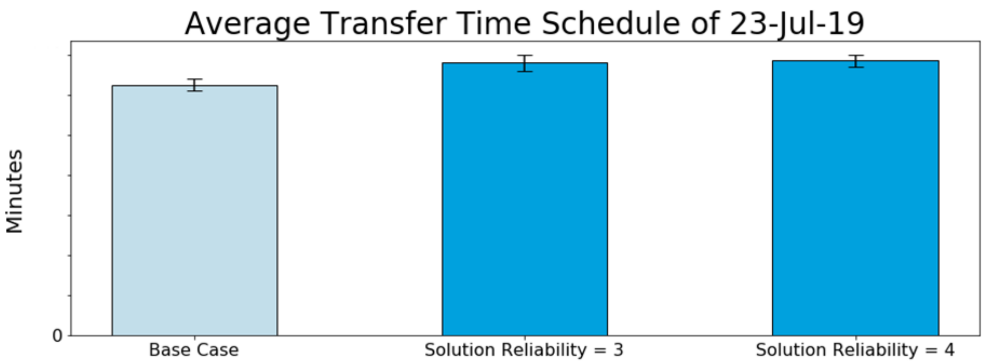


Figure 6.4: Simulation result: Average transfer time of the schedule on 23 July 2019

In terms of connectivity performance, it can be concluded that the Pareto optimal solutions perform better than the base case. It is recommended to reflect on the effect of increasing the transfer time for the customers satisfaction. The choice on which Pareto optimal solution is better depends on the decision whether (i) the difference of connectivity value is significant or (ii) the percentage of missed connections by passenger presents a significant difference.

On the reliability side, it is not possible to draw a correct conclusion as to which schedule is better. The wide of the confidence intervals of the KPIs of this performance element is too large to provide a meaningful comparison. This is caused by the high stochastic variance of the simulation model.

25 JULY 2019 The simulation-optimization model provides four Pareto optimal solutions for the flight schedule of Thursday 25 July 2019. Figure 6.5 shows the Pareto front including the objective values of the base case schedule. The arrows on the axes point out the direction of desirability. Note that the connectivity axes are reversed: this means that the point closest to the lower-left corner is the most optimal. All the solutions perform better in terms of reliability and connectivity than the base case. This means that any schedule resulting from the simulation-optimization model would be a better choice than the base case schedule.

Between the Pareto optimal solutions, there is a trade-off between reliability and connectivity visible, i.e. the less delay minutes, the higher the connectivity. However, the differences between the connectivity value of the Pareto optimal solutions are relatively small. It is in the hands of the decision maker, in this case network planners, whether this difference is significant or not.

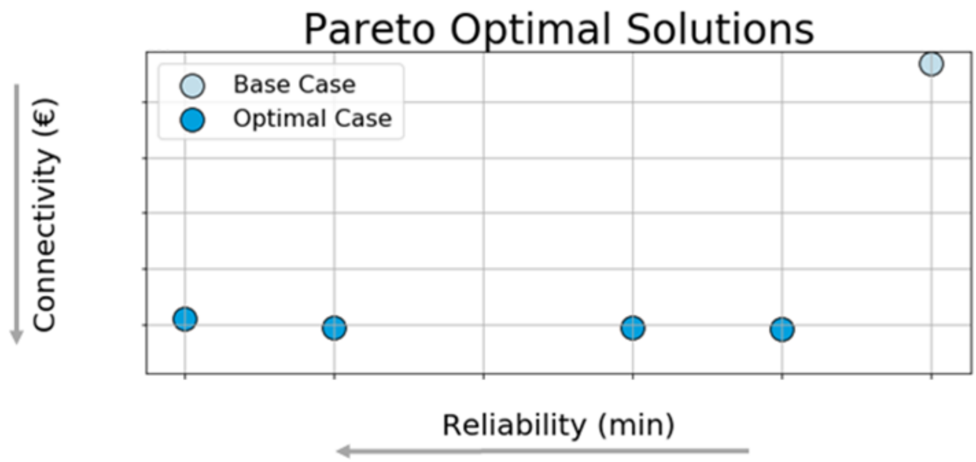


Figure 6.5: Simulation result: Pareto front for 25 July 2019

Table 6.2 presents the buffer minutes added to the rotations of aircraft type Boeing 737-900 on 25 July 2019 compared to the base case. From the 18 scheduled rotations, the Pareto optimal solutions differ in the buffer scheduling for only 8 rotations. This difference ranges between 5 minutes to 35 minutes. Similar to Tuesday 23 July, the most buffer time is added to the first flight or last flight of the day.

Table 6.2: Buffer values of the Pareto optimal solutions for 25 July 2019 in minutes

Rotation Number	Dest	Base Case	Solution Reliability = 3	Solution Reliability = 4	Solution Reliability = 6	Solution Reliability = 7
KL1164	HEL	0	20	50	25	50
KL1109	ARN	0	5	5	5	5
KL1289	EDI	0	10	10	10	10
KL1083	MAN	0	15	10	10	10
KL1597	FCO	0	-25	-25	-25	-25
KL1169	HEL	0	0	0	0	0
KL1171	HEL	0	25	30	15	15
KL1372	OTP	0	0	0	0	0
KL1009	LHR	0	0	-5	0	-5
KL1189	BGO	0	5	5	5	5
KL0461	TLV	0	50	55	55	55
KL1070	MAN	0	0	5	10	15
KL1373	OTP	0	5	5	5	5
KL1705	MAD	0	5	5	5	5
KL0462	TLV	0	-10	-10	-15	-15
KL1823	TXL	0	0	-5	-5	-5
KL1115	ARN	0	0	-5	-5	-5
KL1379	OTP	0	5	5	40	0

Based on the objectives, the Pareto optimal solutions perform better than the base case. However, there are other KPIs that are also important for assessing the performance of the schedule. Therefore, the base case is compared with the Pareto optimal solutions based on these KPIs.

Similar to the results of 23 July 2019, the 95% confidence interval on reliability is extremely large for all cases. This means that there is no clear indication of the actual value of average arrival delay minutes (see Figure 6.6) and total arrival delay minutes (see Figure 6.7). It is not possible to compare these values correctly due to this wide confidence interval.

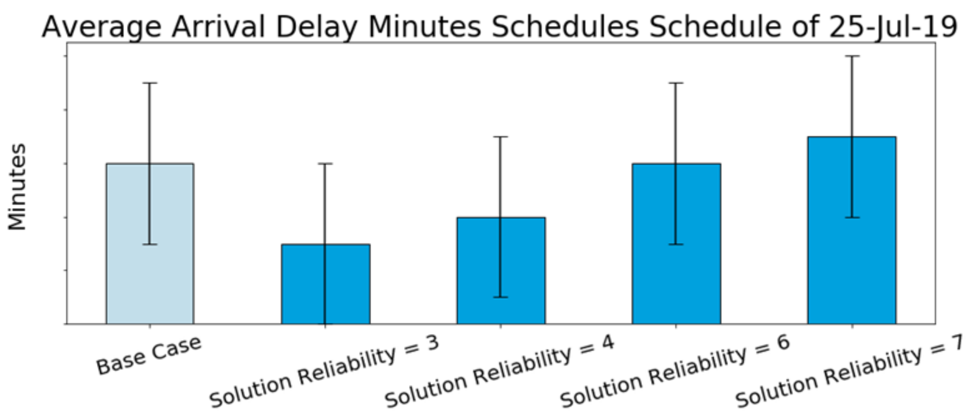


Figure 6.6: Simulation result: Average arrival delay minutes of the schedule on 25 July 2019

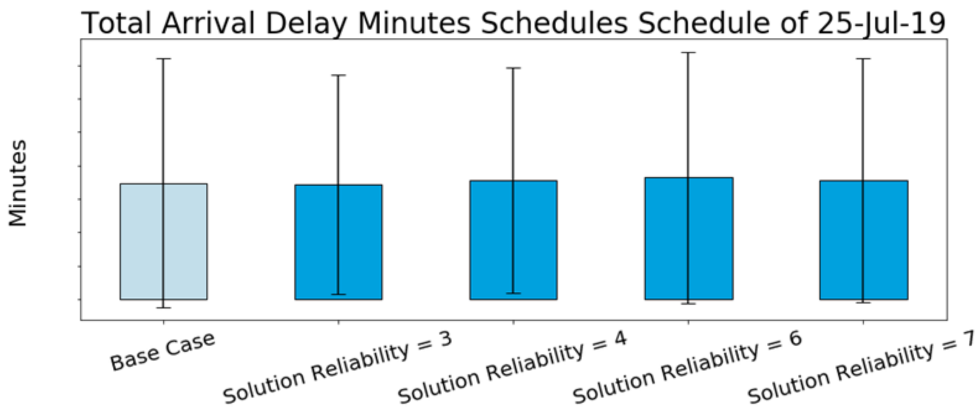


Figure 6.7: Simulation result: Total arrival delay minutes of the schedule on 25 July 2019

Complementary to the Pareto front, Figure 6.8 shows the connectivity value of the base case schedule and the Pareto optimal schedule of 25 July 2019. The difference between the base case and the Pareto optimal solutions is relatively large, thus the simulation-optimization model gives better solutions in terms of connectivity. Between the Pareto optimal solutions, the difference is relatively small. It is the decision of the network department if this is considered significant or not.

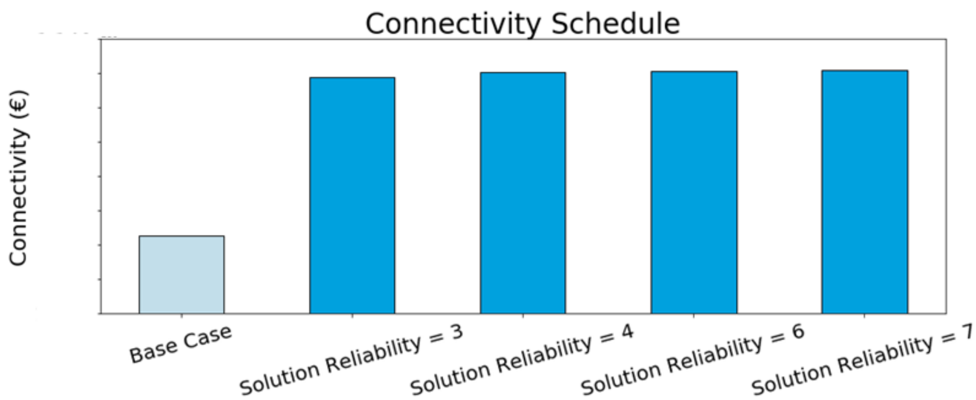


Figure 6.8: Simulation result: Connectivity value of the schedule on 25 July 2019

Figure 6.9 presents the percentage of missed connections by passengers of 25 July 2019. In most cases, a more reliable schedule results in a lower percentage of missed connections. Only there is one outlier for this rule namely the solution with on average four delay minutes. This solution results in the higher percentage of missed connections as the base case although the reliability has improved. Reasoning could be that the way of buffer scheduling in this solution improved in terms of arrival delay but not in departure delay. The percentage of missed connections depends on arrival and departure delay. It could be that this solution offers many connections with a short connection time, similar to the base case, and the shift of rotation by buffer scheduling was not enough to compensate for the departure delay. Thus, although the solution is more reliable, the transfer time of the offered connections plays a crucial role in the percentage of missed connections by passengers.

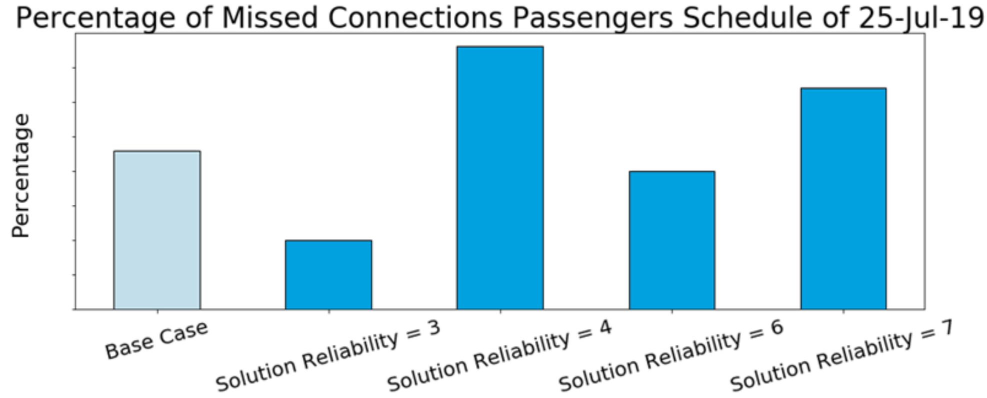


Figure 6.9: Simulation result: Percentage of missed connections by passengers in the schedule of 25 July 2019

In terms of average transfer time of the schedule, the Pareto optimal solutions result in nearly similar average transfer times as the base case (see Figure 6.10).

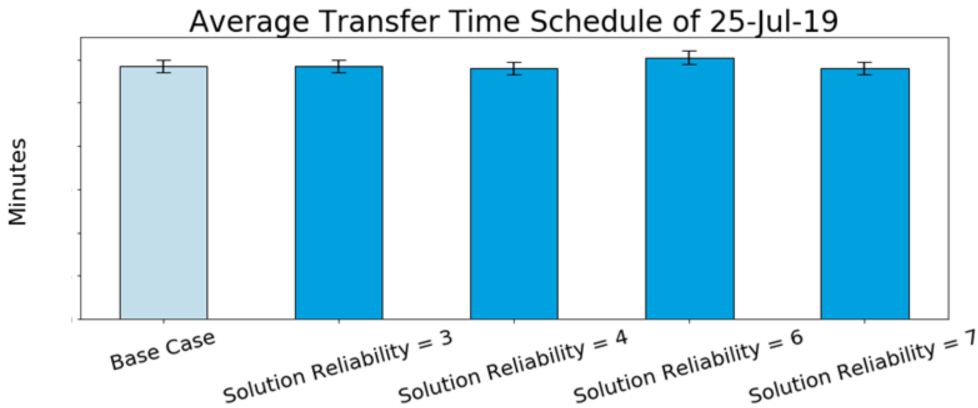


Figure 6.10: Simulation result: Average transfer time of the schedule on 25 July 2019

Concluding, it is not possible to draw a correct conclusion on the performance of the schedules on the reliability KPIs. The wide of the confidence intervals is too large to provide a meaningful comparison.

For the connectivity performance, it can be concluded that the Pareto optimal solutions perform better than or similar to the base case on the connectivity and the average transfer time. Choosing between the Pareto optimal solutions depends on (i) if the difference of the connectivity value is significant or not and (ii) how to rank the value of percentage of missed connections by passengers.

CONCLUSION From analyzing the results of the simulation-optimization model on Tuesday 23 July 2019 and Thursday 25 July 2019, the following conclusion can be drawn:

- It is not possible to compare and interpret the reliability performance due to the extremely wide confidence interval. This is the results of the high stochastic variance of the simulation model.
- The Pareto optimal solutions perform significantly better than the base case on connectivity. The decision on which Pareto optimal solution performs better is in the hands of the decision maker, in this case network planners. This depends on which difference in connectivity value makes a significant impact.
- The transfer time of the offered connections plays a crucial role in the percentage of missed connections by passengers.
- It is recommended to reflect on the effect of increasing or decreasing the transfer time for the customers satisfaction.

6.2 IMPROVEMENTS OF OPTIMIZATION MODEL

In response to the inability to interpret and compare the reliability performance of the simulation-optimization model, the simulation model is replaced by a deterministic model. As the simulation is responsible for calculating the reliability in the optimization model, replacing this with the deterministic model eliminates the stochastic element in the optimization. This deterministic manner of calculating reliability is included to test whether it would improve the performance of the optimization model.

First, the deterministic model is introduced. Hereafter, the results of the optimization model with a deterministic reliability calculation is evaluated. Multiple levels of stochastic variance are added to the deterministic model to test the effect of noise on the optimization model. These results are discussed last.

6.2.1 Deterministic Model

The deterministic model is designed such that it closely corresponds to the principle of the simulation model without stochastic variance.

The reliability objective is calculated by means of Equation 4.2. The average arrival delay minutes per rotation is calculated by means of Equation 4.1. The collection of flights i is determined based on historical data of Summer 2019. Adding more buffer time to a rotation follows the assumption that more buffer time leads to a better on-time performance and thus, less delay minutes. This means that shifting the rotation forward, i.e. creating more buffer, leads to less arrival delay minutes and vice versa.

Other assumptions of the deterministic model are:

- **Early arrivals are considered as on-time.** Thus, when historical data on a rotation shows a negative delay, the amount of delay minutes is set to zero. This operation is executed before the average arrival delay minutes of a rotation is calculated.
- **Buffer time at the outstations is included in the arrival delay.** Similar to the simulation model, the buffer time at the outstations is included in the arrival delay. This means that when a buffer is planned at the outstation, some of the calculated average arrival delay is captured by this buffer.

Due to the lack of a source of uncertainty in the model, no replication are required to ensure statistical accuracy. Therefore, this model calculates the reliability objective relatively fast (in computational time) compared to the simulation model.

6.2.2 Evaluate Pareto Optimal Solutions

The Pareto optimal solutions, resulting from the deterministic optimization model, for 23 July 2019 and 25 July 2019 are compared to the base case by means of the KPIs. Note that due to the elimination of stochastic variance in the calculation of reliability, there is no confidence interval present for the indicators.

23 JULY 2019 Optimizing the flight schedule of 23 July 2019 with the deterministic way of calculating reliability results in five Pareto optimal solutions. Every solution has a higher connectivity value than the base case schedule. Figure 6.11 shows the Pareto front of 23 July 2019; the graph is presented in the same manner as the previous section. This graph shows that almost all solutions (excluding one) are expected to perform better on both objectives than the base case schedule.

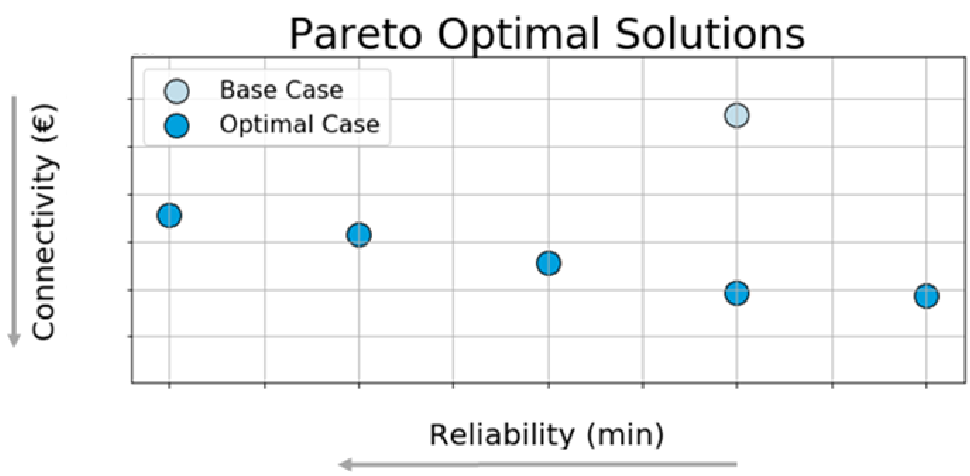


Figure 6.11: Deterministic result: Pareto front for 23 July 2019

Figure 6.12 presents the average arrival delay minutes of the schedule, which is similar to the reliability objective in the Pareto front. There is one optimal solution that is expected to lead to more delay minutes than the base case. Since the connectivity value of this solution is the highest of all Pareto optimal solutions (see Figure 6.11), it is still interesting to include this solution.

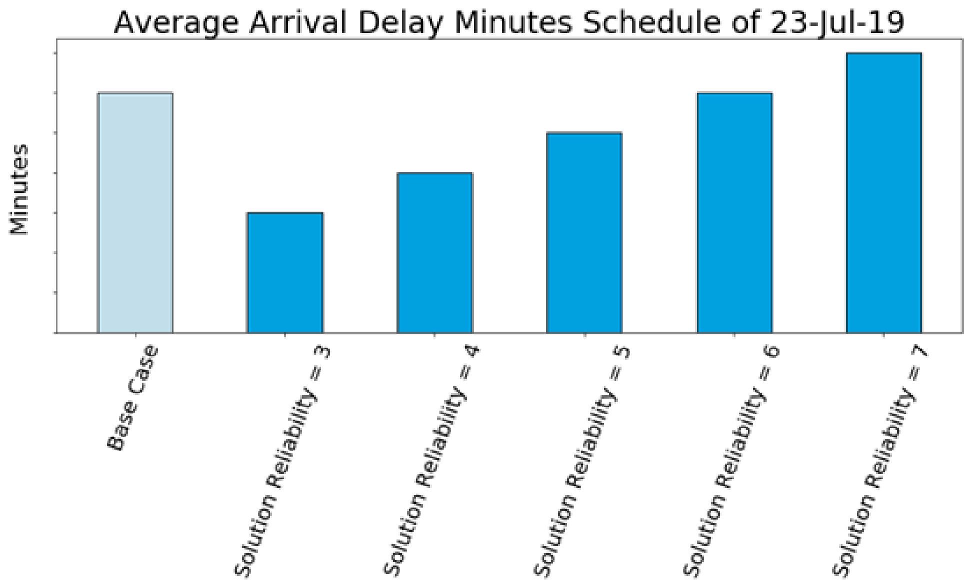


Figure 6.12: Deterministic result: Average arrival delay minutes of the schedule on 23 July 2019

In terms of total arrival delay minutes of the schedule, there are two Pareto optimal solutions that result in more total arrival delay minutes than the base case namely the solutions with on average 6 delay minutes and on average 7 delay minutes (see Figure 6.13). The base case has the same reliability value as the solution with on average 6 delay minutes, but the base case has less total arrival delay minutes. When comparing the base case and this particular solution, the buffer assignment of the base case is better in terms of reliability performance. However, the connectivity value of the Pareto optimal solution with on average 6 delay minutes is higher than the base case and thus, it is still valuable to include this solution. The total arrival delay minutes of the solution with on average 5 delay minutes is almost similar to the base case. Since this solution has less average arrival delay minutes, it can be noted that this solution would be superior to the base case.

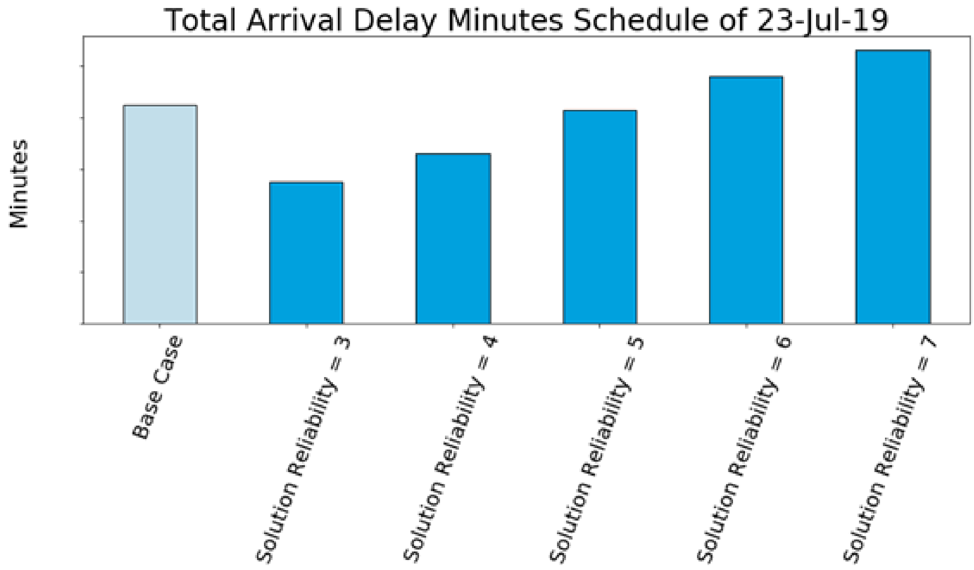


Figure 6.13: Deterministic result: Total arrival delay minutes of the schedule on 23 July 2019

On the connectivity side, Figure 6.14 shows that all solutions have a greater connectivity value than the base case. This means that, in terms of connectivity, it could be better to choose a solution created by the optimization model than the base schedule.

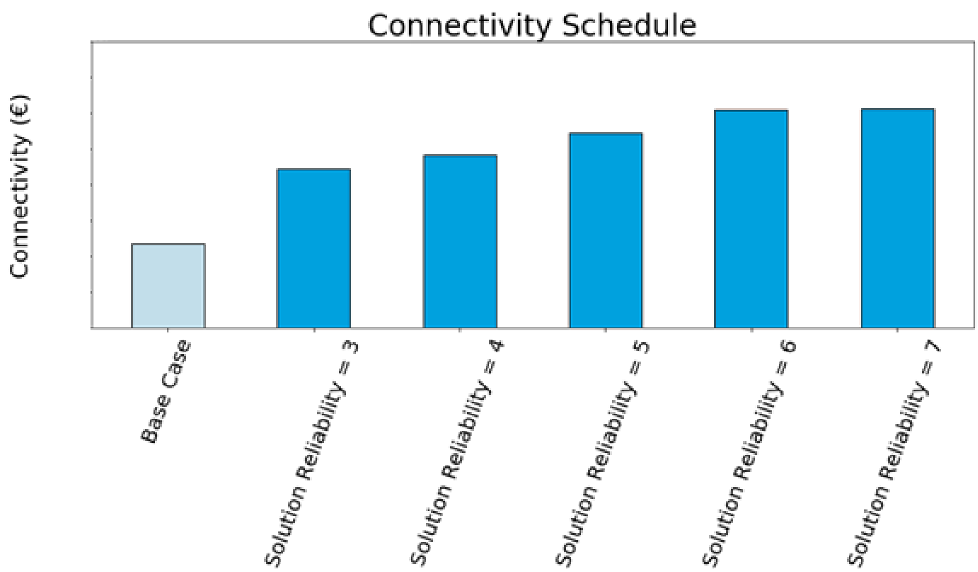


Figure 6.14: Deterministic result: Connectivity value of the schedule on 23 July 2019

Figure 6.15 presents the percentage of missed connections by passenger in the schedule. This shows whether the optimal schedule would actually positively contribute to decreasing the number of missed connections. Only the Pareto optimal solutions with on average 3 minutes and 6 minutes delay minutes are expected to have a lower percentage of missed connections by passengers. The solution with on average 4 delay minutes has the same percentage of missed connections as the base case; so although the reliability has improved, it does not necessarily contributes to a lower percentage of missed connections. The solution with reliability value of 5 has a slightly higher percentage of missed connections whereas it performed better on the reliability performance measures as the base case. The reason could be that the arrival delay has improved by the buffer allocation, but it did not compensate the departure delay. Next to that, the buffer assignment of the solution with reliability value of 6 shows a lower percentage of missed connections by passengers than the base case. It thus contributes to less missed connections although the reliability performance is worse. Again, the could be result of the departure delay; the buffer assignment could lead to less departure delay and this impacts the percentage of missed connections. Lastly, the optimal solution with the highest average arrival delay minutes performs the poorest on this KPI. With these results, it can be noted that there is a significant impact of departure delay on the percentage of missed connections by passengers.

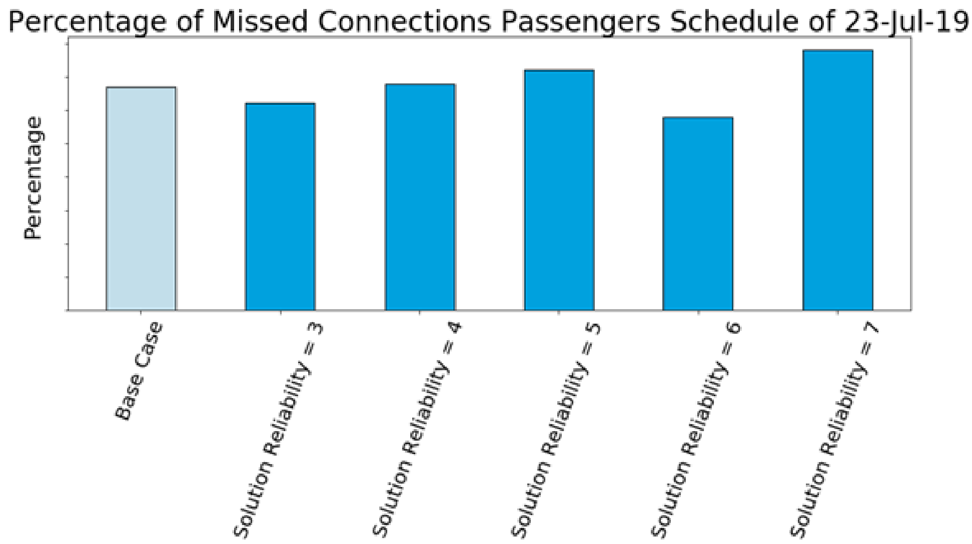


Figure 6.15: Deterministic result: Percentage of missed connections by passengers in the schedule of 23 July 2019

Figure 6.16 presents the results of the average transfer time of the schedule for the base case and the Pareto optimal solutions. The difference between the base case and the Pareto optimal solutions is quite significant but within the Pareto optimal solutions not. Since the optimization model tries to maximize connectivity, it could be that the model includes more connections with a long - close to MACT - transfer time. This increases the connectivity value and increases the average transfer time of the schedule.

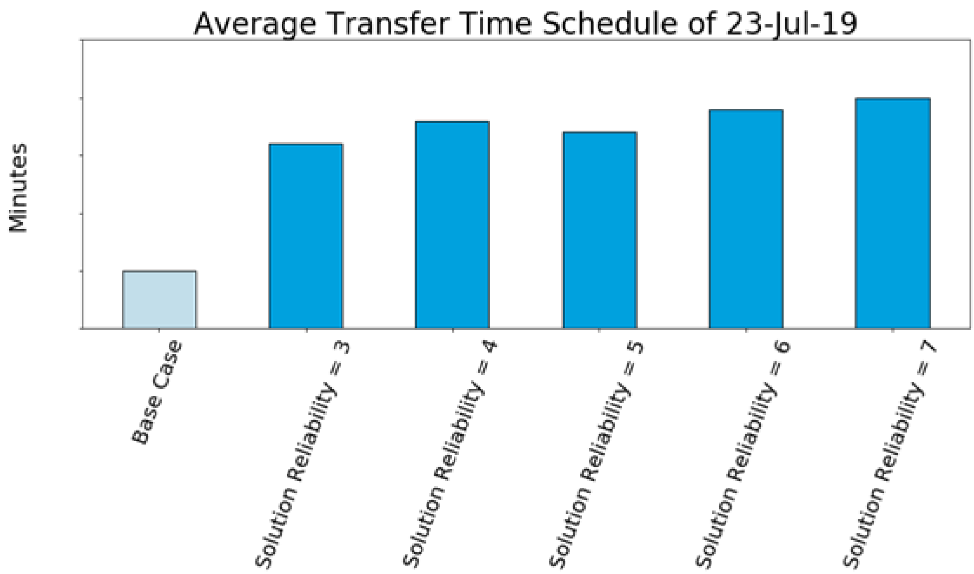


Figure 6.16: Deterministic result: Average transfer time of the schedule on 23 July 2019

Table 6.3 gives an overview of the performance of the Pareto optimal solutions compared to the base case. The Pareto optimal solution with on average 7 arrival delay minutes performs only better on connectivity. For the other solutions, it is up to the decision maker to determine which solution performs best. This overview indicates if the solution performs better or worse than the base case; however it does not indicate value and importance of the better or worse performance. The decision maker, in this case network planners, need to determine rating of the KPIs when choosing.

Table 6.3: Overview of performance Pareto optimal solutions compared to the base case of 23 July 2019

Key Performance Indicators / Solutions	Average arrival delay minutes	Total arrival delay minutes	Connectivity	Percentage of missed connections by passengers	Average transfer time
Solution Reliability = 3	+	+	+	+	+
Solution Reliability = 4	+	+	+	+/-	-
Solution Reliability = 5	+	+/-	+	-	-
Solution Reliability = 6	+/-	-	+	+	-
Solution Reliability = 7	-	-	+	-	-

+ = performance is better than base case; +/- = performance is similar to base case;
- = performance is worse than base case.

25 July 2019 Optimizing the flight schedule of 25 July 2019 with the deterministic way of calculating reliability results in seven Pareto optimal solutions. Every solution has a higher connectivity value than the base case schedule. Figure 6.17 shows the Pareto front of 25 July 2019; the graph is presented in the same manner as the previous section. A clear trade-off is visible between reliability and connectivity. The graph shows that almost all solutions (excluding one) are expected to perform better on both objectives than the base case schedule.

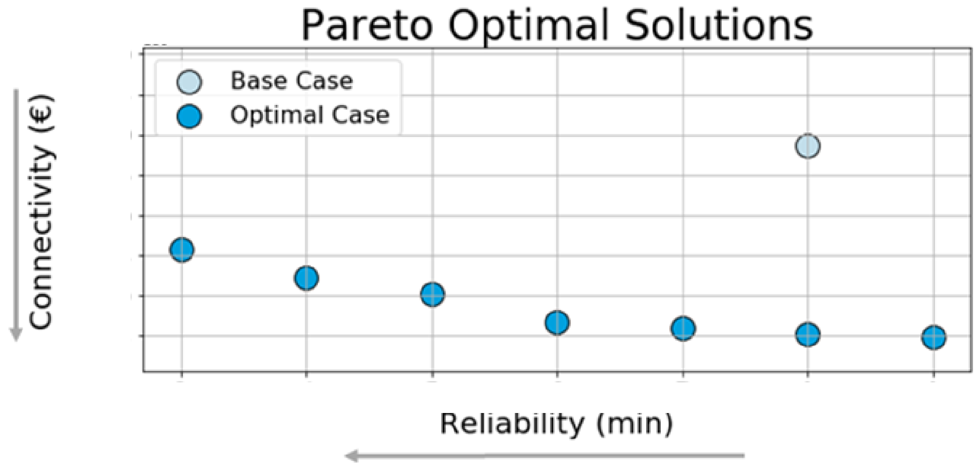


Figure 6.17: Deterministic result: Pareto front for 25 July 2019

For the average arrival delay minutes of the schedule (see Figure 6.18), the same applies here as to the results of 23 July 2019. Although one solution has a lower reliability than the base case, the connectivity value is interesting enough to include this solution in the analysis.

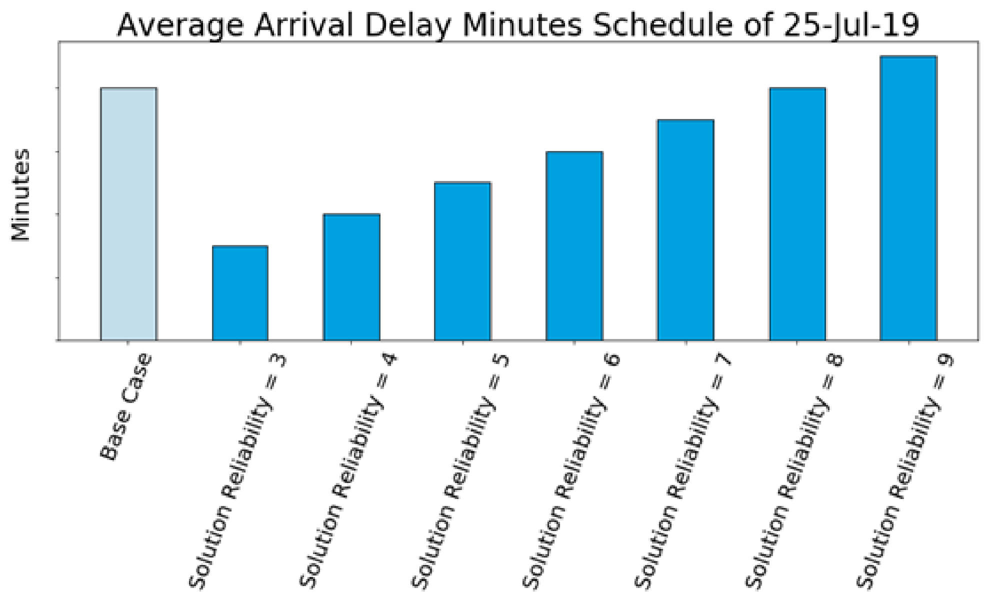


Figure 6.18: Deterministic result: Average arrival delay minutes of the schedule on 25 July 2019

The total arrival delay minutes has similar proportions between the solutions as the average arrival delay minutes of the schedule on 25 July 2019; the total arrival delay minutes increases proportional to the reliability. Concluding, only one solution performs worse on the reliability KPIs.

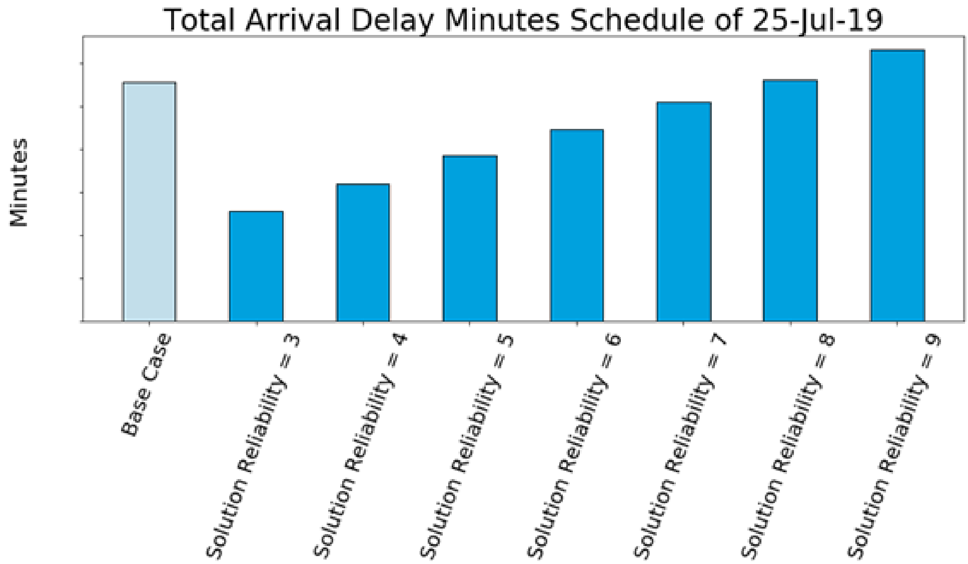


Figure 6.19: Deterministic result: Total arrival delay minutes of the schedule on 25 July 2019

In line with Figure 6.17, all Pareto optimal solutions have a higher connectivity value than the base case. This means that more connectivity revenue can be retrieved with these optimal solutions. Complementary to the trade-off between reliability and connectivity, Figure 6.20 shows that the less arrival delay minutes (higher reliability) leads to less connectivity. The graph flattens out when the number of average arrival delay minutes increases.

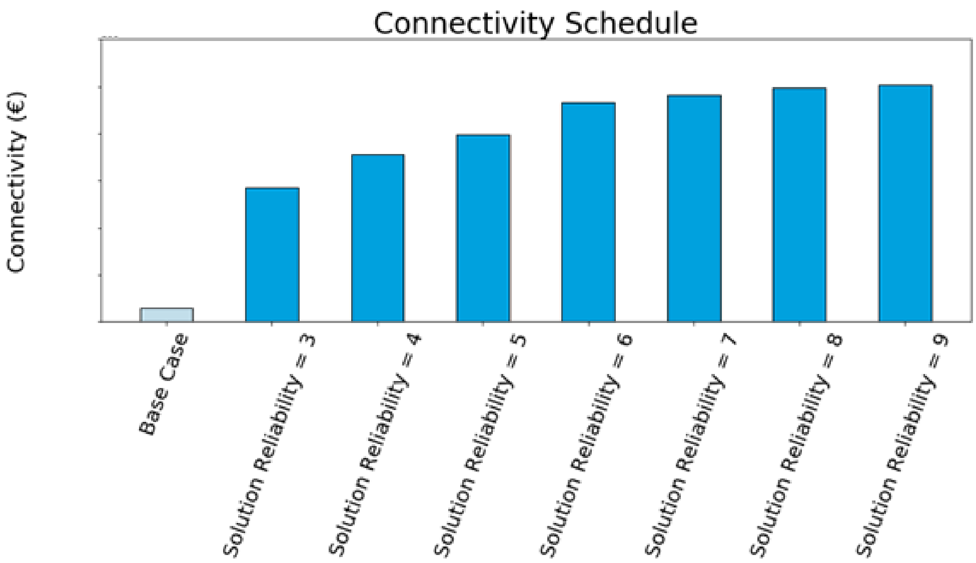


Figure 6.20: Deterministic result: Connectivity value of the schedule on 25 July 2019

Figure 6.21 displays the percentage of missed connections by passengers for the base case and the Pareto optimal solutions. The solution with the highest number of arrival delay minutes results in the same percentage of missed connections as the base case. The other solutions positively contribute to ensuring that less passengers miss their transfer. This graph shows that the higher the reliability, thus the lower the amount of arrival delay minutes, the less passengers miss their connection.

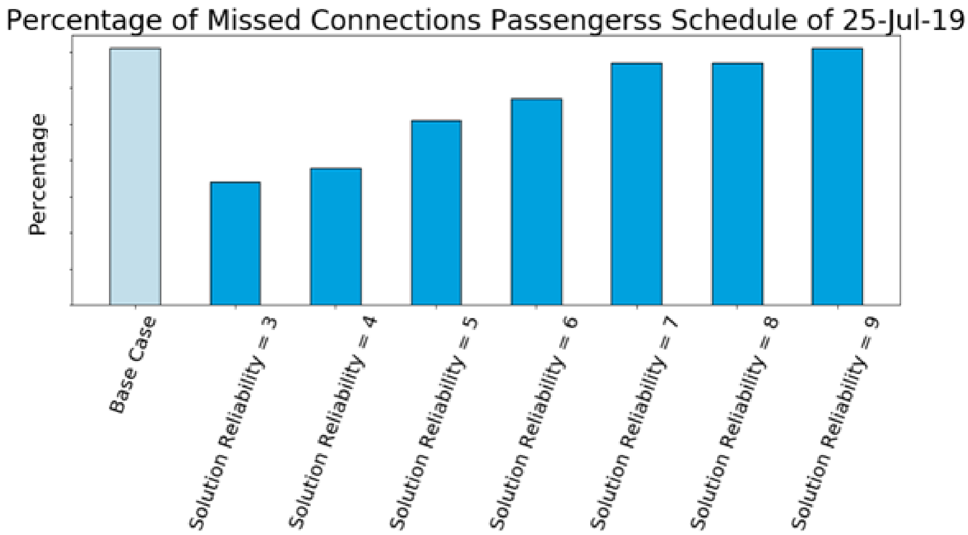


Figure 6.21: Deterministic result: Percentage of missed connections by passengers in the schedule of 25 July 2019

The average transfer time of the schedule varies (see Figure 6.22). Reasoning could be that more long connections are newly offered or short connections are cut-off by the optimization model. It would be valuable for KLM to evaluate the effect on the increase of transfer time on customer satisfaction. This could help with making a well-founded decision on the increase of average transfer time.

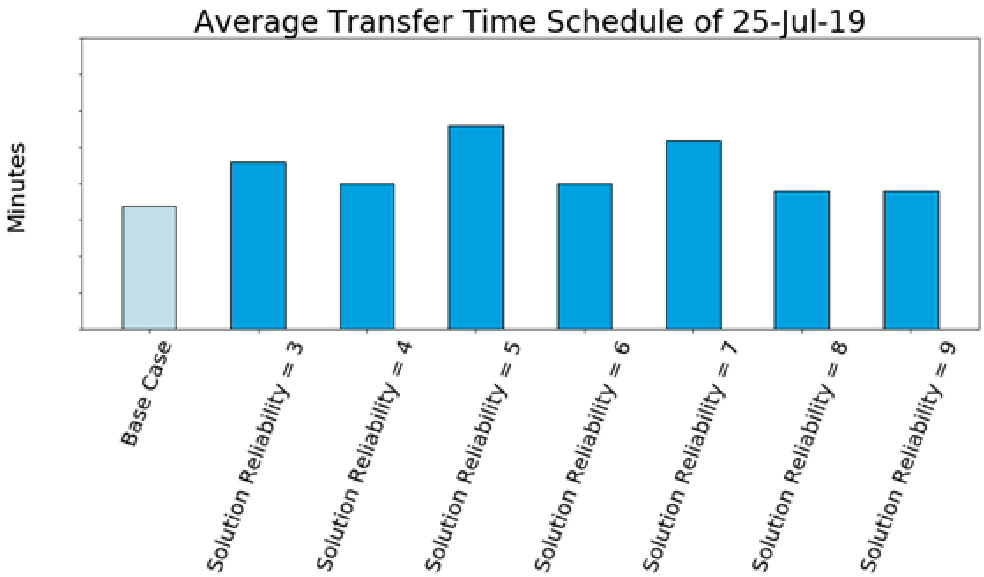


Figure 6.22: Deterministic result: Average transfer time of the schedule on 25 July 2019

Table 6.4 gives an overview of the performance of the Pareto optimal solutions compared to the base case. The solutions with reliability value of 3 to 7 perform similar with respect to the base case. The performance of the other two solutions (with higher average arrival delay) is more similar to the base case. The main contribution of these solutions is the increase in connectivity.

Note that this overview gives an indication whether the solution performs better or worse than the base case, not how much better or worse. The value of the performance needs to be determined by the decision maker. The graphs of the KPIs offer support for this. Moreover, for the KPI average transfer time, it is recommended to investigate the impact of increasing the transfer time on customer satisfaction.

Table 6.4: Overview of performance Pareto optimal solutions compared to the base case of 25 July 2019

Key Performance Indicators / Solutions	Average arrival delay minutes	Total arrival delay minutes	Connectivity	Percentage of missed connections by passengers	Average transfer time
Solution Reliability = 3	+	+	+	+	-
Solution Reliability = 4	+	+	+	+	-
Solution Reliability = 5	+	+	+	+	-
Solution Reliability = 6	+	+	+	+	-
Solution Reliability = 7	+	+	+	+	-
Solution Reliability = 8	+/-	+/-	+	+	-
Solution Reliability = 9	-	-	+	+/-	-

+ = performance is better than base case; +/- = performance is similar to base case;
- = performance is worse than base case.

6.2.3 Effect of Stochastic Variance

From the previous sections, there is a clear difference visible between the results of the optimization model with very stochastic reliability calculation and the model with a deterministic calculation. The deterministic optimization model did not only make the solutions more comparable but also increased the number of Pareto optimal solutions. This last improvement can be related to the performance of the search quality and progress of the optimization model, i.e. how fast the optimization convergences to optimal solutions. It appears, from this improvement, that stochastic variance has an effect on the performance of the optimization model. This section investigates this effect more in-depth.

In order to test the effect of stochastic variance on the performance of the optimization model, experiments with various noise levels are performed. Noise is added to the deterministic model on the reliability objective, i.e. the average arrival delay minutes per schedule. The simulation model is also included in the analysis; this model has the highest stochastic variance since it adds noise to every individual flight instead of solely to the reliability objective.

The performance of the optimization model is evaluated by (1) convergence, i.e. convergence of the solution set to the Pareto optimal front, and (2) diversity, i.e. diverse set of solutions in the objective space. (Goel & Stander, 2010; Reed et al., 2013). This research applies two performance measures to evaluate the convergence and diversity of the model namely ε-progress and hypervolume.

ε-progress describes the cumulative number of improvements between the ε-grids. It measures whether the optimization models found a substantially better solution based on the user-defined search precision ε. When the number of improvements stabilizes, the search of the optimization model has fully converged (Kwakkel et al., 2016). ε-progress thus measures the convergence of the model. Hypervolume describes the size of the objective space that is covered by the Pareto front. The user defines the minimum and maximum values of the objective space, and then the ratio of non-dominated outcomes to all possible outcomes in the space is calculated. It measures the convergence and diversity of the optimization model (Reed et al., 2013; Zitzler et al., 2007). A hypervolume of 0 means that there are no solutions are found in the objective space; a hypervolume of 1 means that the objective space merely contains Pareto optimal solutions (Garner & Keller, 2018). Similar to the ε-progress measure, the optimization model has completely converged to Pareto optimal solutions when the hypervolume stablizes. The value of the hypervolume determines the diversity of the optimization model; the higher the hypervolume, the more optimal solutions are located in the solution space (Sato et al., 2007; Trautmann et al., 2008).

Both performance measure are computed for each generation. In total, the number of function evaluation (NFE) for this experiment is 10.000. Due to an adaptive population size of BORG, the population differs per generation. The initial population is set to 100.

Figure 6.23 and Figure 6.24 visualize the ε-progress of two optimization cases namely Tuesday 23 July 2019 and Thursday 25 July 2019. In both cases, the optimization without stochastic variance (noise = 0%) has the highest number of cumulative improvements and the simulation model, with the most stochastic variance, has the least number of cumulative improvements. This means that the optimization model without any stochastic variance converges faster to the Pareto optimal front than the optimization model with much stochastic variance.

The ε-progress for 23 July 2019 for the other noise levels has formed a group in the midpoint of the deterministic model and the simulation model. No clear pattern is shaped yet between the various noise levels. An explanation could be that the ε-progress is not stabilized yet, and since the number of improvements for the different noise levels lie very close to each other, the order of ε-progress of the various noise levels can easily change.

Figure 6.24 convincingly shows that more noise leads to less ε-progress, and thus a slower convergence of the optimization model. Although the optimization model has not stabilized yet, the ε-progress is relatively spread and no change in sequence of the noise levels is expected.

From the ε-progress measure, it can be concluded that a completely deterministic optimization model, i.e. no stochastic variance, results in a faster convergence than the simulation-optimization model, i.e. a very high stochastic variance.

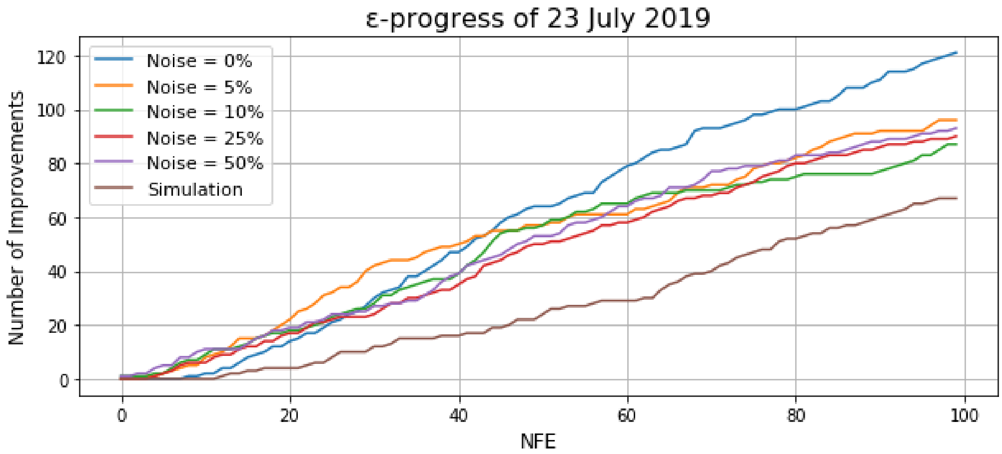


Figure 6.23: ε-progress with various noise levels of 23 July 2019

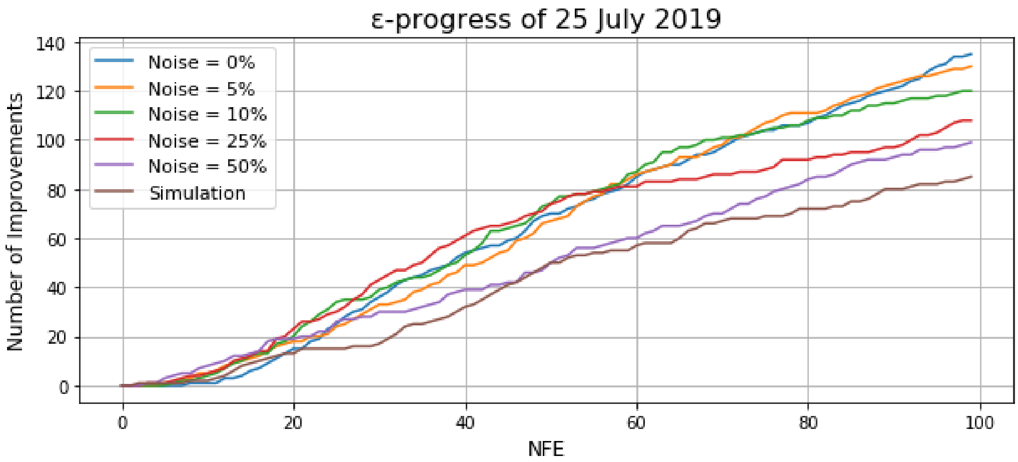


Figure 6.24: ε-progress with various noise levels of 25 July 2019

Figure 6.25 and Figure 6.26 show the comparison of the hypervolume measure. For 23 July 2019, the hypervolume remains fairly stable throughout each noise level after approximately 45 NFE. The value of the hypervolume indicates that for most noise levels, the Pareto front covers a large part of the objective space. The simulation-optimization model stabilizes relatively fast (after 22 NFE); the model convergences fast towards a more optimal solution while the other noise levels are still exploring the solution space (and hereby increase in hypervolume). For 25 July, the hypervolume of the simulation model and a noise level with 50% remains the most stable over multiple generations. This means that the optimization model with a high stochastic variance is not likely to find more optimal solutions than currently found. The hypervolume of the models with a lower noise level is still increasing and thus, can increase even more. This means that although the models with less stochastic variance take longer to find optimal solutions, the solutions will be more diverse and have a higher coverage of the objective space.

Thus, the hypervolume shows that the simulation-optimization model convergence relatively fast towards a solution while other noise levels are still exploring the solution space and hereby, maximizing their hypervolume.

Reasoning could be that the simulation-optimization model does not evaluate infeasible solutions; this is modelled to decrease the computational time. This could improve convergence towards a solution but not necessarily towards many dominated solutions that cover a significant part of the solution space.

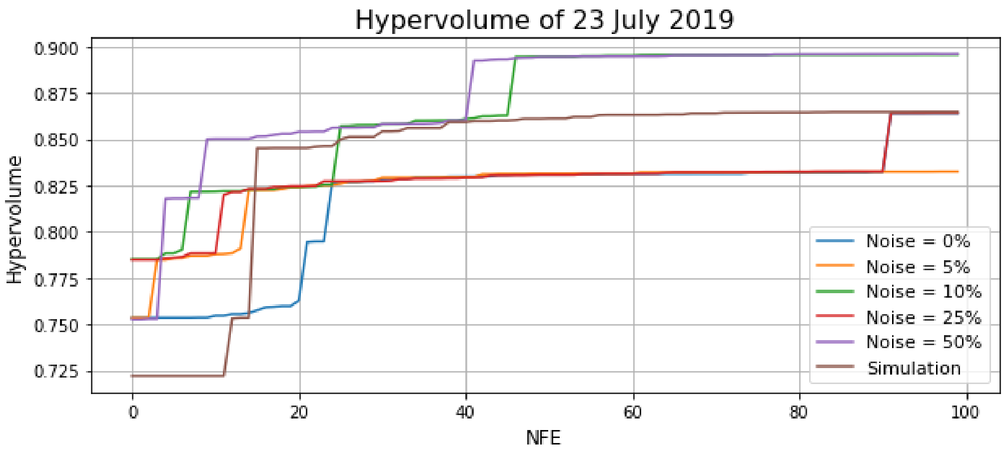


Figure 6.25: Hypervolume with various noise levels of 23 July 2019

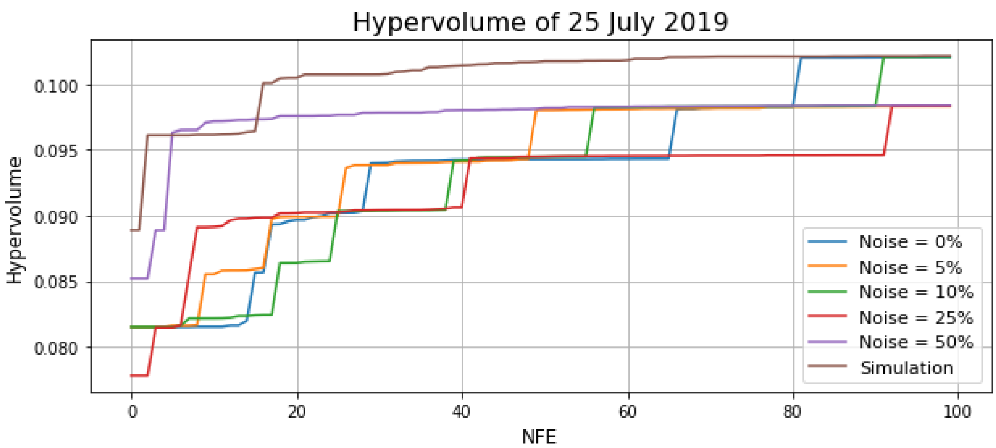


Figure 6.26: Hypervolume with various noise levels of 25 July 2019

Concluding, ε-progress shows that an optimization model without any stochastic variance leads to a faster convergence than the simulation-optimization model. When the optimization model is stabilized in terms of improvements, it is also expected to see that more stochastic variance leads to less improvements, and thus a slower convergence. This is already shown for one of the cases (25 July) but not yet for both. Moreover, the hypervolume measure shows that the simulation-optimization model convergences relatively fast towards a solution due to modelling techniques, while other noise level are still exploring the solution space. However, the simulation-optimization model will likely not present a diverse Pareto front according to the hypervolume measure. In combination with the ε-progress results, it can be stated that the simulation-optimization model convergence relatively fast towards a solution but most likely not towards the Pareto optimal solutions.

7 DISCUSSION ON RESULTS

This chapters discusses the results of the optimization model. Challenges for the use of the model and the results in academia, business are addressed sequentially. Lastly, the computational challenges are presented.

7.1 METHODOLOGICAL CHALLENGES

This section discusses the challenges related to the methodological part of the research. The focus lies on the process of designing the optimization model, the use of the algorithm and the effect on the results.

During the design and experiments of the optimization model, several challenges occurred on the methodological side. This has a great effect on the quality of the solutions. Challenges, applied solutions and the remaining difficulties encountered in this research are as follows:

USE OF GENETIC ALGORITHM: BORG BORG is chosen as optimization algorithm for this research since it is currently the best performing GA optimization algorithm for complex problems (Ward et al., 2015). Results from this study show that the optimization model has not reached the most optimal results yet for all cases. This means that the optimization model is still under performing. This could either be the cause of (i) too few function evaluations as stopping criteria, thus NFE should be increased or (ii) BORG is not suitable for handling this extremely complex multi-objective optimization problem. When increasing NFE to a reasonable limit does not improve the performance of the GA, then the problem lies in the algorithm itself. This has not been investigated in this research due to time limitations. It is difficult to find another better performing GA since BORG is currently the best performing GAs for multimodal and complex problems. The conclusion would then be that the problem cannot be sufficiently handled by current state-of-art GAs. Novel GAs need to be developed that can deal with the required degree of complexity for obtaining the most optimal results of this problem.

CONVERGENCE In the course of this study, convergence challenges are encountered. At first, the simulation-optimization model did not deliver any feasible solution (i.e. solution where the constraints are met). One of the causes was the quality of the convergence of the optimization model. The main problems of this poor convergence are (1) the completely randomly sampled initial population and (2) stochastic variance of the simulation model.

The first problem is the completely random sampling of the initial population. The standard algorithm, as defined in the Python library, samples a complete randomly initial population, i.e. the start point of the optimization model. When the start of the optimization model is too widespread, it is difficult and time consuming for the optimization to convergence in the right direction. In this case, the right direction is the solution space where the feasible solutions lie. The solution for this challenge is to change the sampling of the initial population. In this research, there is a base case where the Pareto optimal solutions are compared with. The base case meets all the constraints and is, thus, a feasible solution. To ensure convergence to the feasible solution, the base case and variations on the base case are set to fill 20% of the population. To guarantee diversity in the solutions, the other 80% of the initial population is sampled randomly. Figure 7.1 shows an example of the initial population that is injected in the optimization model. This way of sampling resulted in more feasible solutions.

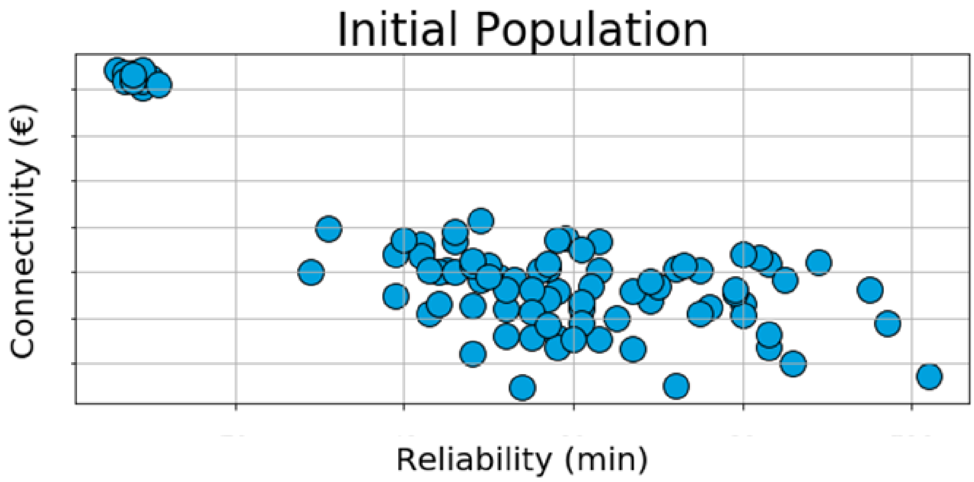


Figure 7.1: Initial Population with base case, 20% variations of base case and 80% random

The other problem is the effect of stochastic variance on convergence. As concluded from the results, the simulation-optimization model has the poorest convergence. This could be explained by the high stochastic variance of the simulation model and the extremely wide confidence intervals. As a consequence, the optimization algorithm is not consistent in determining whether this solution performs better or worse than other outcomes. So if a solution is evaluated twice by the model, it could give two completely different values for reliability. This makes it nearly impossible for the algorithm to convergence to feasible and optimal solutions. As a solution, this research eliminates the stochastic variance by using a deterministic model. Another solution would be to simplify the simulation model to get less stochastic variance and hereby, a faster convergence of the model.

SIZE OF THE SOLUTION SPACE Although the number of decision variables is quite small (approximately 18 rotations per optimization), the variable bounds are very large. The variables bounds depend on the total buffer time available in one fleetline and one decision variable could have 80 possible solutions. If all 18 rotations have around 80 possible buffer values, the number of possible solutions becomes extremely large. In this research, the number of function evaluations has been increased from 1000 to 10.000. The results show that the model still did not find the most optimal solutions. A remaining solution to obtain the most optimal results is to increase the number of function evaluations. In this way, the optimization model is able to evaluate more solutions of this extremely large solution space. However, this also increases the computational time tenfold. Another solution is to limit the decision variable bounds, however this is not desirable in terms of model validity.

CONSTRAINTS The optimization problem involves only four constraints, however these constraints are formulated quite strictly. There are many solutions that violate these constraints and this makes it difficult to get feasible solutions within a reasonable time. To reduce computational time, the simulation-optimization model does not evaluate a solution when it does not meet the constraints. However, this also limits the search quality of the optimization model by not evaluating near-feasible results which are close to more optimal results. In the end, the optimization did obtain feasible solutions and therefore, it was not necessary to take action. For further improvement of the optimization model, it would be worthwhile to handle the constraint challenge and hereby, obtain a higher search quality.

SEARCH PRECISION The search precision (ε) for each objective in the optimization model is user-defined. When setting the search precision either too small or too large, the optimization model does not return the right results. In this research, the initial search precision was set too large for reliability objective and too small for the connectivity objective. Whereas the expectation was that the optimization would give multiple Pareto optimal solutions, it only gave one solution. Patterns of the results could not properly be explained and the expected trade-off between the objectives was not visible. When changing the search precision to the right value, these problems were solved immediately. Thus, the search precision is a crucial element for the quality of the results. A suggestion for other optimizations is to normalize the objective values and the epsilons such that it is harder to make a mistake in this essential user-defined parameter.

The five methodological challenges encountered during this research are the use of a genetic algorithm, poor convergence of the model, large size of the solution space, tight constraint and incorrect search precision. These challenges are fundamental for the quality of the results.

7.2 BUSINESS CHALLENGES

This section discusses the challenges related to the business implementation of the model. For airlines, it is interesting to know whether it is feasible to use the novel optimization model for creating a flight schedule. This research defines various challenges for using the optimization model in practice. Based on the results, the following challenges and recommendations for the business are identified:

DECISION-MAKING The optimization model is a basis for trading off reliability and connectivity in the schedule design. Due to the novelty of the model and time limitations, this model does not include important constraints such as slots assignment and crew legislation. Therefore, this model is not suitable yet for stand-alone decision-making in practice. Human interference is needed for the actual decision-making. Nonetheless, this model offers help and clarity in the decision-making process and could already be used as a support for decision-making.

ADDING MORE COMPLEXITY TO THE MODEL The optimization model defined in this research is very simplistic compared to the real-life schedule design. This model only includes four simple constraints of numerous constraints that are applied in practice. Even with these two relatively simple constraints, the optimization models already has difficulties with finding the optimal solutions. This gives an idea of the high level of complexity of the problem. Therefore, when adding more constraints, it is going to be even harder to find the optimal solutions. It could be valuable to relax the constraints to get infeasible but good performing schedules which might actually work better than optimal solutions. Another commonly used solution is to apply a penalty to the objective function when a constraint is violated instead of setting it as a hard constraint. Thus, when adding complexity to the optimization model, a suggestion for the airlines is to be aware of the tightness of the constraints and to relax them or work with penalties instead.

EFFECT OF CHANGING THE TRANSFER TIME ON CUSTOMER SATISFACTION AND ON TICKET PRICES The effect of changing the transfer time on customer satisfaction and on ticket prices is not taken into account in the model. The customer satisfaction increases when the transfer time becomes slightly longer than the minimum transfer time; passengers do not have to run to catch their flight. When the transfer time increases significantly, customer satisfaction decreases; if there is no meaningful use of the waiting time, passengers tend to experience it as unpleasant. For them, it is mostly a trade-off between waiting time at the hub and leisure time at the origin. (Rietveld & Brons, 2001). As a consequence, this lower desirability of trips with longer transfer time leads to lower ticket prices. The average transfer time of a passenger is taken as a KPI in this research. However, the effect of increasing or decreasing the transfer time has not actively been implemented in the model due to time limitations. Therefore, the model tends towards maximizing transfer times since this would increase connectivity. For further development, it would be valuable to add this element by making the average transfer time an extra objective or by adding a penalty to the ticket price when the transfer time increases. However, it is worthwhile to note that the constraint that ensures no extra time could be added to a day minimizes the outgrow of transfer time, so transfer time does not get unrealistically large. Also, the outbound connectivity tends to function as a counterweight to keep the average transfer time of a rotation from becoming extremely large.

no future connectivity data In this research, connectivity is defined by historical connectivity data. Due to lack of data on new possible connections, the model does not add new connections and thus, the positive impact on revenues is not displayed. Therefore, with the current data, the optimal solutions are expected to stay relatively close to the base case in terms of connectivity. Shifting the rotations forward or backward would mostly lead to removing connections, and thus less revenue of tickets. In practice, shifting rotations could also lead to connections which have never been performed historically. Data on possible new connections of KLM was not available for research. Therefore, insights on these new connections still needs to be acquired by the commercial department of an airline.

stochastic variance of the simulation A recommendation for airlines is to limit the stochastic variance when including a simulation model in the optimization model. A way to limit the stochastic variance of a simulation model is to exclude extreme cases of delay. For example, only delays of less than 30 minutes are taken into account in the new simulation model.

computational speed of the (simulation) optimization model The computational speed of the simulation-optimization model is slow. It takes 3.5 days to get the Pareto front and the results show that the model has not even completely converged yet within this time. It is therefore not desirable to use the simulation-optimization model for fast decision-making. However, it is still suitable as support for decision-making during the schedule design process, since it is not real-time decision-making. Introducing the deterministic way of calculating reliability reduces the computational speed to 5 hours per optimization problem. This is more reasonable considering the speed of decision-making with respect to the schedule design. Concluding, the implementation of the optimization model in the business still faces some challenges. The current model is a basis for optimizing reliability and connectivity in the schedule design. Although human interference is still required, the optimization model is already able to support decision-making by giving a clear overview of the trade-off between reliability and connectivity. The main challenge is to extend the model for full optimization of the schedule.

7.3 COMPUTATIONAL CHALLENGES

This section discusses the challenges related to the computational time. As briefly mentioned in the other sections, there are some issues regarding computational speed. The general computational challenges are the following:

SPEED OF OPTIMIZATION MODEL The computational time of the simulation-optimization model is 3.5 days and the computation time of the deterministic optimization model is 5 hours. The experiments are run on a computer with 32GB Memory with parallel processing over 8 cores. The deterministic optimization model has a more reasonable computational time for decision-making than the simulation. However, it would be desirable to speed up the model for implementation purposes. There are many tricks to speed up GA but these are not applied in this research as this was not the focus. When implementing this model, it is recommended to have Computer Engineers speed up the optimization algorithm for faster performance.

NUMBER OF FUNCTION EVALUATIONS The results show that the model has not fully converged yet, i.e. ε has not stabilized yet. This means that the number of function evaluations needs to increase. This way, more solutions in the extremely large solution space are evaluated and the model has the ability to converge further. However, increasing the number of function evaluations results in an increase in run time. Therefore, it is crucial to take the increase in computational time into account when increasing the number of function evaluations.

ADDING CONSTRAINTS TO THE MODEL In order to use this model to truly optimize the schedule design, more constraints need to be added. Constraints such as slot assignments, crew legislation, and swapping flights then need to be included. However, including these constraints would it make it even harder to find feasible solutions within a reasonable time. Therefore, it is important to keep track of the increase of computational time when adding constraints. When the computational time increases tremendously, no results can be obtained within a reasonable time and the model becomes unusable for decision-making.

Concluding, the computational challenges are how to speed up to model, the increase of number of function evaluation and adding constraints to the model. It is important to be aware of the increase in computational time when adding more complexity to the model. When the computational time becomes too large, the model cannot give results in a reasonable time and becomes unusable.

8 ANALYSIS ON ENVIRONMENTAL SUSTAINABILITY

8.1 OPERATIONALIZATION OF ENVIRONMENTAL SUSTAINABILITY

INTRODUCTION Sustainability is a broad concept; there are three main types of sustainability namely social, economic and environmental. Social sustainability defines the well-being of the society. Economic sustainability means the stabilization of the economic capital. Environmental sustainability refers to improving humans welfare by protecting raw materials used for humans needs and humans waste is limited (Goodland, 1995; Morelli, 2011).

The environmental importance of aviation has grown in the last years since the impact of the airline industry on climate change has become more critical (H.-C. Wu et al., 2018). Airlines can perform many actions to reduce their environmental impact as defined in CORSIA, such as less food waste or less fuel use. For this research, the focus lies on the effect of reducing delays by buffer scheduling on fuel use.

Fuel use can be directly related to the CO_2 emissions of a flight; 1 kilogram (kg) of fuel for airplanes (Jet Fuel) is equal to 71.5 CO_2 kg per gigajoules (GJ) (Vreuls & Zijlema, 2009). Fuel use during a flight depends on many factors such as:

- speed in the air,
- route to destination,
- load weight of the passengers and baggage, and
- weight of extra fuel.

Fuel use is affected by delay with respect to the speed in the air. In particular, departure delay and airborne delay are correlated with fuel use. The pilot could decide to perform high-speed flying (HSF) to catch up with departure delay. The speed in the air increases so the fuel use increases as well. Airborne delay could cause rerouting the aircraft or increasing speed during the flight, which also increases the fuel use (Ryerson et al., 2014). Since this research does not change the airborne time but only the buffer on ground, sustainability is operationalized with departure delay.

Departure delay could cause HSF which leads to an increase in fuel use. Reverse this, less departure delay could lead to less fuel use and a reduction in CO_2 emissions.

KLM designed a sustainability model which calculates marginal fuel usage in kilograms of a time gain during a flight. It, for example, calculates how much additional kilograms of fuel is used when flying 5 minutes faster. This is used to calculate the fuel consumption.

8.2 ASSUMPTIONS

To determine the effect of the optimization model on environmental sustainability, the following assumptions are made:

- **An one-to-one relationship exists between departure delay minutes and HSF.** This means that if the departure delay is 5 minutes, than the pilot tries to catch up 5 minutes with HSF.
- **There are no other reasons to perform HSF.** It is assumed that the only reason for performing HSF is to recover departure delay. Other reasons such as being on-time for a slot or a specific connection, are not taken into account.
- **There is always the possibility to perform HSF.** This means that there is always enough fuel available.

8.3 RESULTS

The environmental impact of the Pareto optimal solutions is compared with the base case schedule. The environmental sustainability is evaluated by means of the average fuel use per flight of a schedule. Fuel use is defined as the additional fuel use per minute for HSF multiplied by the total flight time (including the time gained by HSF).

Similar to Chapter 6, the results of Tuesday 23 July 2019 and Thursday 25 July 2019 are presented. For this analysis, the departure delay resulting from the deterministic optimization model is used since this is part of the reliability calculation. The simulation model is not suitable due to its high stochastic variance. Note that no exact values of the results are presented due to confidentiality reasons.

23 JULY Figure 8.1 presents the average fuel use per flight in the base case schedule and the Pareto optimal solutions. It can be concluded that all optimal solutions use less fuel kg than the base case; this means that these solutions are more sustainable than the base case. It can also be observed that the more reliable the schedule is, thus the less delay minutes, the less average fuel per flight is used. It seems as if the average fuel use of the schedule is rather in line with the reliability.

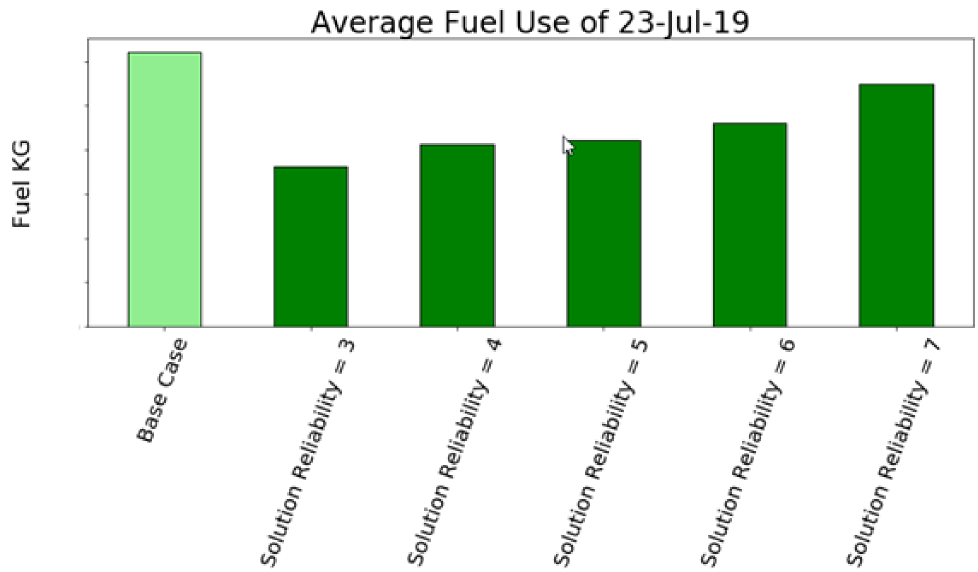


Figure 8.1: Average fuel use per flight in the schedule of 23 July 2019

To visualize whether the sustainability pattern is similar to the reliability pattern, a new Pareto front is created with sustainability as objective instead of reliability (see Figure 8.2). This research tries to minimize the fuel use. The arrows on the axes point out the direction of desirability. Note that the connectivity axes are reversed; this means that the point closest to the lower-left corner is the most optimal. Compared to the Pareto front of connectivity and reliability (Figure 6.11), the position of the solutions is quite similar expect for one solution. Excluding this one solution, it can be noted that less connectivity leads to more sustainable solutions and vice versa. Overall, it can be stated that any optimal solution could be more sustainable than the base case.

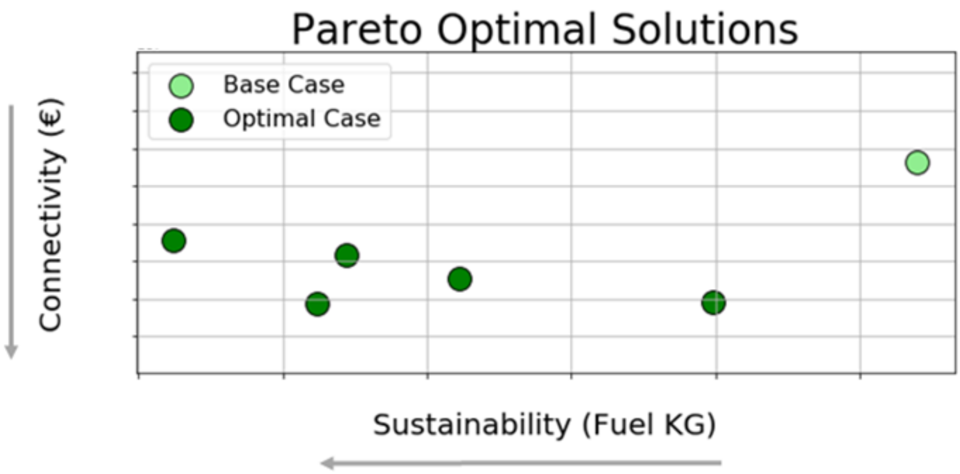


Figure 8.2: Pareto front of connectivity and sustainability of the schedule of 23 July 2019

25 JULY Similar to 23 July, Figure 8.3 shows that the more reliable, the less average fuel per flight is used and thus, the more sustainable the schedule is. Again, the Pareto optimal solutions provide more sustainable schedules than the base case.

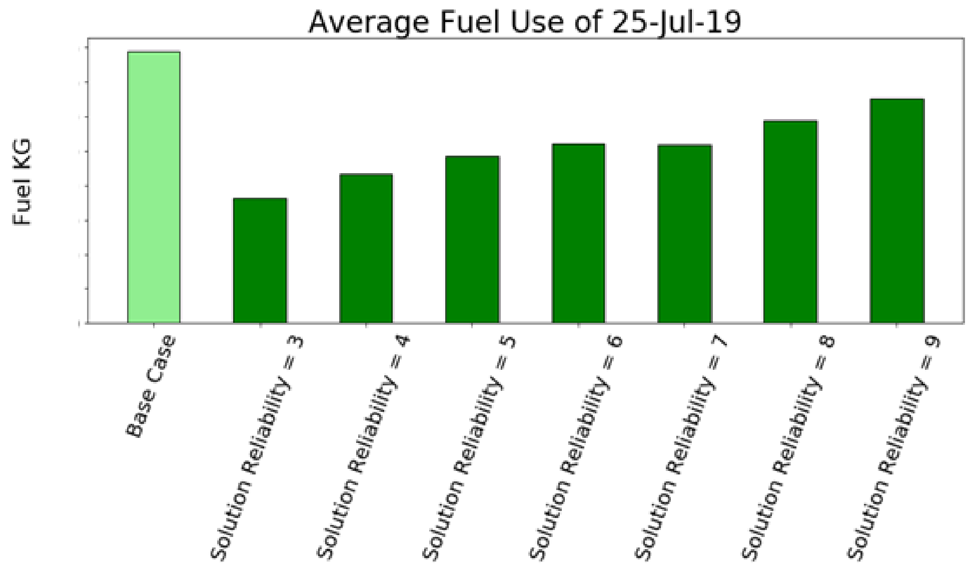


Figure 8.3: Average fuel use per flight in the schedule of 25 July 2019

In contrast to 23 July, the sustainability pattern is not comparable to the reliability in the Pareto front. Figure 8.4 visualizes the Pareto front in the same way as the previous paragraph. Compared to the Pareto front of reliability and connectivity (Figure 6.17), the solution in this Pareto front are randomly scattered. No clear trade-off between sustainability and connectivity is visible and thus, sustainability cannot be directly linked to reliability. It is again shown that the base case solution could perform worse in terms of sustainability and connectivity.

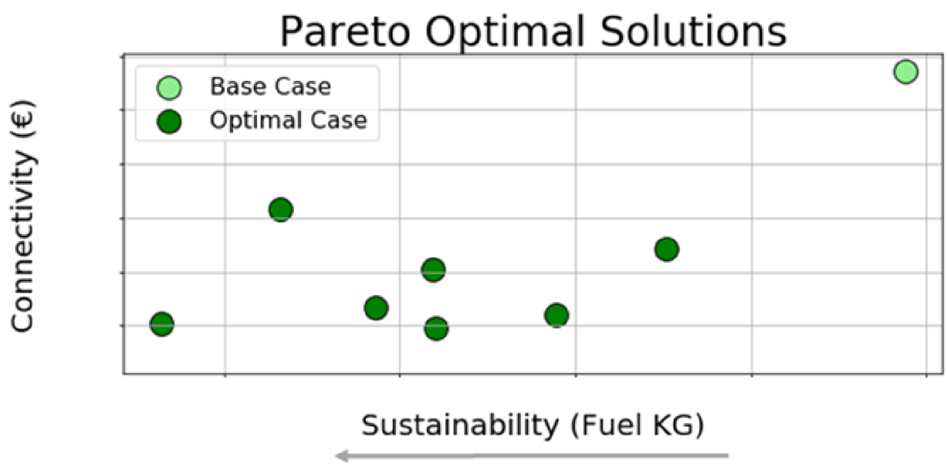


Figure 8.4: Pareto front of connectivity and sustainability of the schedule of 25 July 2019

CONCLUSION It can be concluded that the optimization model generates schedules that could perform better than the base case schedule in terms of sustainability. This means that the optimization model positively contributes to reduction of CO_2 emissions. Also the results show that the more reliable, the less average fuel per flight is used and thus, the more sustainable the schedule is. A more reliable schedule could save fuel and therefore, improve CO_2. Although it could be expected that sustainability follows the trend of reliability in this situation, there is no clear evidence for this.

8.4 DISCUSSION

The sustainability analysis in this research is focused on one element which affects the environmental sustainability of a schedule namely the departure delay. It is assumed that there exist an one-to-one relation between departure delay minutes and HSF, and HSF directly impacts the fuel use. This explains why the average fuel use per flight increases when the reliability decreases.

Moreover, it is concluded that the optimization model positively contributes to the environmental sustainability goals of airlines. The degree of contribution highly depends on the assumption that departure delay directly affects the fuel use. When adding more elements that influence the fuel use, the magnitude of this positive contribution of the optimization model could change. Due to the significant difference between the base case and the Pareto optimal solutions, it can be stated that the optimization model reduces the fuel use of a schedule.

When concluding on the greater picture of environmental sustainability without being aware of the limitations, it could lead to incorrect information for decision-making. Therefore, the results of the sustainability analysis should be interpreted with the limitations in mind.

9 CONCLUSION

The environmental impact of aviation has been an emerging topic of discussion worldwide. There is a need for reducing the CO_2 footprint of airlines while they keep contributing to globalization, i.e. more travelling passengers. Creating a more robust and predictable flight planning would reduce fuel consumption and thus, greenhouse gas emissions, because it results in less high-speed flying since pilots do not have to catch up with delays. However, a more reliable flight schedule often requires more buffer time and this takes a toll on the amount of connecting flights that can be offered. The goal of this research is to design a novel optimization model for balancing reliability and connectivity by buffer scheduling in the flight schedule, and to evaluate the impact on environmental sustainability. The following main research question is thereby formulated:

"How can airlines use buffer scheduling to ensure the optimal balance between the reliability of the flight schedule and the value of passenger connections, and what is the impact on environmental sustainability?"

The main research question is answered by the following sub questions:

1. How are reliability and connectivity operationalized in aviation?
2. What is the current state of reliability and connectivity of the short haul flight schedule of KLM Royal Dutch Airlines?
3. Which optimization model is suitable for creating a reliable schedule by buffer scheduling while trading off passengers connections?
4. How can reliability and connectivity of the short haul flight schedule of KLM Royal Dutch Airlines be improved by buffer scheduling with the optimization model from sub question 3?
5. What is the impact of this improvement of the short haul flight schedule of KLM Royal Dutch Airlines on environmental sustainability?

For the first sub question, the reliability and connectivity for airline industry are operationalized as follows:

* Reliability is defined as the weighted average arrival delay minutes for flight i during time window T.
* Connectivity is defined as the revenue from passengers transferring based on (i) the probability of passengers actually making the transfer

given the transfer time, (ii) the number of expected passengers per transfer and (iii) the income per passenger.

Secondly, the current state of reliability an connectivity of KLM defines a benchmark for evaluating the optimization model.

As answer to the third sub question, the optimization model that is most suitable for creating a reliable schedule by buffer scheduling, while trading off passenger connections, is a multi-objective optimization model. The constraints that are necessary to include in this optimization model are (1) the transfer time between two flights should be larger than the minimum connecting time for that connection type, (2) the turnaround time between connecting flights of the aircraft should be larger than the minimum turnaround time and (3) the total amount of slack within one fleet line can be divided only once. A good performance of the optimization model, i.e. a high search quality and progress to find the most optimal solutions, is guaranteed when (i) the stochastic variance in the objectives calculation is limited or none, (ii) including (variations of) the base case in the initial population and (iii) perform at least 10.000 function evaluations. The performance of this model and the quality of the results could be improved by relaxing or penalizing the constraints and increasing the number of function evaluations.

Fourthly, the reliability and connectivity of the short haul flight schedule of KLM Royal Dutch Airlines can be improved by presenting the Pareto optimum that results from the optimization model in a clear overview in such a way that the network planners can make an explicit trade-off between them. The results of the optimization model perform better than the base case, so the ways of buffer scheduling resulting from the optimization model could improve either or both the reliability and connectivity of the short haul flight schedule, depending on the choice of the network planners.

In response to sub question 5, a more reliable schedule, resulting from the optimization model, positively impacts the environmental sustainability of the schedule assuming that the departure delay solely and directly affects the fuel use. A more reliable schedule could save fuel and therefore, improve CO_2.

Concluding, the novel optimization model is a suitable way to use buffer scheduling for ensuring the optimal balance between the reliability of the flight schedule and the value of passenger connections. Considering that the model does not include all the constraints required for the schedule design and has a run time of 5 hours, the model supports decision-making by creating insight in the trade-off between reliability and connectivity, and by providing more optimal and sustainable solutions than the base case. Integrating it with a high level of stochastic variance, makes it harder to obtain Pareto optimal solutions due to the low search quality and progression. Therefore, it is recommended to use a deterministic optimization for supporting decision-making when balancing reliability and connectivity in the flight schedule.

Moreover, the optimization model positively impacts environmental sustainability, assuming that the departure delay is the only factor that directly influences fuel use. Adding more factors to the equation could change the degree of this positive contribution.

10 FURTHER RESEARCH

This research presents a novel optimization model for trading off the reliability and connectivity in the schedule design by buffer reallocation. Due to system boundaries and time limitations, elements are simplified and could be expanded in the model. The following elements could be expanded in further work:

- **Entire float of all aircraft types.** The current model optimizes the float for one aircraft type. For further work, it would be valuable to optimize the schedule for all aircraft types at once. Hereby, the connections between aircraft types are also included and this makes the connectivity measure even more realistic.
- **Time period.** Currently, the model optimizes for one day. It would be beneficial to expand this to an entire season to include the dependencies between the days. To expand the model to international flights, this would be necessary since the schedule of international flights often do not include overnight flights to cut the schedule into days, as the European flight schedule does.
- **Propagated delay.** The reliability measure is currently the average arrival delay minutes per flight of a schedule. It would be valuable to also define the propagated delay as measure for reliability since this is one of the main causes of delay in a hub-and-spoke network.
- **Value of transfer time for ticket prices.** The value of transfer time for the connectivity measure would bring the model closer to reality. As mentioned, the transfer times could increase or decrease the ticket prices and therefore, the connectivity value.
- **Value of time of the day for ticket prices.** The model currently takes historical ticket prices as revenue of a connection. However, when changing the time of a flight by buffer scheduling, the ticket prices could change based on the time of the day. For example, it is more profitable to fly on Monday morning at 8 a.m. to London than on 10 a.m. to facilitate business meetings. Therefore, adding this time element to the monetization of the connections would be interesting for further work.

There are also elements that are necessary for decision making and have not been included at all. To expand the model for further research, it would be interesting to incorporate these elements and get a fully integrated optimization model for the schedule design. The following elements could be included:

- **Slot assignment of the hub and outstations.** Slot assignment is a complex element to add; it involves many dependencies and politics. However, it is one of the main (and most difficult) constraint for creating a schedule. Schedule planners manually ensure that the number of slots available in a time window aligns with the created schedule. The importance and complexity of this constraint makes it very useful to add to the optimization model.
- **Crew legalisation.** Crew legalisation such as maximum working hours, are restrictive for creating a flight schedule. It could determine the flight time or how the flights are scheduled. Therefore, this constraint would be valuable to add. Ideally, the entire crew scheduling is aligned with the flight scheduling.
- **Resource assignment.** Resources such as fuel tanks, push-back truck, ground personnel, are limited for a day of operation. The schedule should also be aligned with the number of available resources to ensure that the schedule can be performed as intended. For example, ten aircraft need to have their baggage unloaded at the same time since their STA is the same but there are only nine baggage unloading crews available. This leads to a delay for one aircraft which could propagated over the day.
- **Fleet assignment.** Fleet assignment, i.e. which aircraft type a flight uses, is already fixed in the current optimization. In real life, schedule planners sometimes choose to change the aircraft type to create a more fitted schedule. It is also possible that a specific aircraft type can only fly on one airport,which makes a change in aircraft type impossible. Adding the possibility of changing the aircraft type would be interesting for further work.
- **Delay of a passenger.** The current model focuses on the delay of a flight. For customer satisfaction, the delay of a passenger is a more suitable measure to incorporate. Also non-performance cost are determined on passengers level. Therefore, it would be valuable to add this to the optimization model.
- **Block time as decision variable.** Block time, i.e. the time between STD and STA, is also interesting to take as a decision variable. The model could then choose whether it is better to add buffer to the ground time or to the block time.
- **Flight speed as decision variable.** Next to changing the block time with buffer scheduling, flight time could also be increased or decreased by the flight speed. This would also be an interesting element to add as decision variable, especially for the sustainability aspect of this research.

In addition to expanding the model, this research also gives insight in the effect of stochastic variance on the performance of the optimization model. In particular, it gives insight in the performance of the search quality and progress of the genetic algorithm BORG. More in-depth research on the effect of stochastic variance on the optimization model would be valuable.

It is also interesting to perform more research on the suitability of BORG for this highly complex optimization problem. As mentioned, it could be that BORG is not completely suitable for this level of complexity. Since BORG is one of the best performing GAs, it means that this complexity level cannot be sufficiently handled by the current state-of-art GA. Novel GAs need to be developed that include more complexity for obtaining the Pareto optimal front. Before this conclusion can be drawn, more research is required in the performance and suitability of BORG for extremely complex problem.

Lastly, in terms of environmental sustainability, other factors that influence fuel should be further investigated and incorporated in the sustainability measure. As environmental sustainability is grown in importance, airlines want to mark their environmental footprint. Therefore, it would be valuable to create an extensive formula that could describe the actual environmental impact of delays and different manner of buffer scheduling.

BIBLIOGRAPHY

Achenbach, A., & Spinler, S. (2018). Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights. *Operations Research Perspectives*, *5*, 265–279.

Ageeva, Y. (2000). *Approaches to incorporating robustness into airline scheduling* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Ahmadbeygi, S., Cohn, A., & Lapp, M. (2010). Decreasing airline delay propagation by re-allocating scheduled slack. *IIE transactions*, *42*(7), 478–489.

Air France-KLM. (2020). *Profile.* Retrieved 2020-03-05, from https://www.airfranceklm.com/en/group/profileL

Air Transport Action Group. (2020). *Facts & figures.* Retrieved 2020-03-05, from https://www.atag.org/facts-figures.html

Air Transport Netherlands. (2019). *Smart and sustainable.* Retrieved 2019-09-26, from https://www.klmtakescare.com/sites/default/files/Smart%20and%20Sustainable_final.pdf

Banks, J. (1998). Principles of simulation. In *Handbook of simulation: Principles, methodology, advances, applications, and practice* (pp. 3–30). New York: John Wiley & Sons, Inc.

Barnhart, C., & Cohn, A. (2004). Airline schedule planning: Accomplishments and opportunities. *Manufacturing & service operations management*, *6*(1), 3–22.

Baumgarten, P., Malina, R., & Lange, A. (2014). The impact of hubbing concentration on flight delays within airline networks: An empirical analysis of the us domestic market. *Transportation Research Part E: Logistics and Transportation Review*, *66*, 103–114.

Burghouwt, G., & de Wit, J. (2005). Temporal configurations of european airline networks. *Journal of Air Transport Management*, *11*(3), 185–198.

Burke, E. K., De Causmaecker, P., De Maere, G., Mulder, J., Paelinck, M., & Berghe, G. V. (2010). A multi-objective approach for robust airline scheduling. *Computers & Operations Research*, *37*(5), 822–832.

Button, K. (2008). The impacts of globalisation on international air transport activity, global forum on transport and environment in a globalising world. In *Global Forum on Transport and Environment in a Globalising World, Guadalajara, Mexico* (pp. 1–40).

Carlier, S., De Lépinay, I., Hustache, J.-C., & Jelinek, F. (2007). Environmental impact of air traffic flow management delays. In *7th USA/Europe air traffic management research and development seminar (ATM2007)* (Vol. 2, p. 16).

Carson, J. S. (2004). Introduction to modeling and simulation. In *Proceedings of the 2004 Winter Simulation Conference, 2004.* (Vol. 1, p. 16). Washington, DC, USA.

Chen, C., Skabardonis, A., & Varaiya, P. (2003). Travel-time reliability as a measure of service. *Transportation Research Record*, *1855*(1), 74–79.

Christidis, P., Rivas, J. N. I., et al. (2012). Measuring road congestion. *Institute for Prospective Technological Studies (IPTS), European Commission Joint Research Centre. Retrieved from http://ipts. jrc. ec. euro pa. eu/publications/pub. cfm*.

Churchill, A. M., Lovell, D. J., & Ball, M. O. (2010). Flight delay propagation impact on strategic air traffic flow management. *Transportation Research Record*, *2177*(1), 105–113.

Clausen, J., Larsen, A., Larsen, J., & Rezanova, N. J. (2010). Disruption management in the airline industry—concepts, models and methods. *Computers & Operations Research*, *37*(5), 809–821.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Danesi, A. (2006, 01). Measuring airline hub timetable co-ordination and connectivity: definition of a new index and application to a sample of european hubs. *European Transport Trasporti Europei*, *34*, 54-74.

Deb, K. (2014). Multi-objective optimization. In *Search methodologies* (pp. 403–449). Springer.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, *6*(2), 182–197.

Dessens, O., Köhler, M. O., Rogers, H. L., Jones, R. L., & Pyle, J. A. (2014). Aviation and climate change. *Transport Policy*, *34*, 14–20.

Dunbar, M., Froyland, G., & Wu, C.-L. (2012). Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework. *Transportation Science*, *46*(2), 204–216.

Emmerich, M. T., & Deutz, A. H. (2018). A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural computing*, *17*(3), 585–609.

Environmental and Energy Study Institute. (2019). *Fact sheet: The growth in greenhouse gas emissions from commercial aviation.* Retrieved 2020-03-05, from https://www.eesi.org/papers/view/fact-sheet-the-growth-in-greenhouse-gas-emissions-from-commercial-aviation

Etschmaier, M. M., & Mathaisel, D. F. (1985). Airline scheduling: An overview. *Transportation Science*, *19*(2), 127–138.

European Commission. (2019). *Reducing emissions from aviation.* Retrieved 2019-09-11, from https://ec.europa.eu/clima/policies/transport/aviation_en

European Commission, E. E. A. . E., European Aviation Safety Agency. (2019). *European aviation environmental report 2019.* Retrieved 2019-09-09, from https://ec.europa.eu/transport/sites/transport/files/2019-aviation-environmental-report.pdf

Fricke, H., & Schultz, M. (2009). Delay impacts onto turnaround performance. optimal time buffering for minimizing delay propagation. In *ATM Seminar.*

Garner, G. G., & Keller, K. (2018). Using direct policy search to identify robust strategies in adapting to uncertain sea-level rise and storm surge. *Environmental modelling & software*, *107*, 96–104.

Gill, T. (2015). *Aviation and climate change; impact and initiatives.* Retrieved 2019-09-11, from https://to70.com/aviation-and-climate-change-impact-and-initiatives/

Goel, T., & Stander, N. (2010). A study of the convergence characteristics of multiobjective evolutionary algorithms. In *13th aiaa/issmo multidisciplinary analysis optimization conference* (p. 9233).

Goodland, R. (1995). The concept of environmental sustainability. *Annual review of ecology and systematics*, *26*(1), 1–24.

Hadka, D., & Reed, P. (2013). Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary computation*, *21*(2), 231–259.

Hansen, M. M., Gillen, D., & Djafarian-Tehrani, R. (2001). Aviation infrastructure performance and airline cost: a statistical cost estimation approach. *Transportation Research Part E: Logistics and Transportation Review*, *37*(1), 1–23.

International Air Transport Association. (2018). *Iata forecast predicts 8.2 billion air travelers in 2037.* Retrieved 2019-09-11, from https://www.iata.org/pressroom/pr/Pages/2018-10-24-02.aspx

Jacquillat, A., & Vaze, V. (2017). Balancing reliability, efficiency and equity in airport scheduling interventions. In *Twelfth USA/Europe Air Traffic Management R&D Seminar.*

Jarrah, A. I., Yu, G., Krishnamurthy, N., & Rakshit, A. (1993). A decision support framework for airline flight cancellations and delays. *Transportation Science*, *27*(3), 266–280.

Kafle, N., & Zou, B. (2016). Modeling flight delay propagation: A new analytical-econometric approach. *Transportation Research Part B: Methodological*, *93*, 520–542.

Kim, J. Y., & Park, Y. (2012). Connectivity analysis of transshipments at a cargo hub airport. *Journal of Air Transport Management*, *18*(1), 12–15.

KLM. (2019). *Network and alliances.* Retrieved 2019-10-09, from https://www.klm.com/travel/nl_en/corporate/network_alliances.htm

Kollat, J. B., & Reed, P. M. (2005). The value of online adaptive search: a performance comparison of nsgaii, ε-nsgaii and εmoea. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 386–398).

Kollat, J. B., & Reed, P. M. (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources*, *29*(6), 792–807.

Kollat, J. B., & Reed, P. M. (2007). A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, *30*(3), 408–419.

Kondo, A. (2011). Impacts of delay propagation on airline operations: Network vs. point-to-point carriers. In *2011 Integrated Communications, Navigation, and Surveillance Conference Proceedings* (pp. L4–1–L4–7).

Kwakkel, J. H., Haasnoot, M., & Walker, W. E. (2016). Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty. *Environmental Modelling & Software*, *86*, 168–183.

Laumanns, M., Thiele, L., Deb, K., & Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary computation*, *10*(3), 263–282.

Law, A. M. (2008). How to build valid and credible simulation models. In *2008 Winter Simulation Conference* (pp. 39–47).

Lederer, P. J., & Nambimadom, R. S. (1998). Airline network design. *Operations Research*, *46*(6), 785–804.

Lee, L. H., Lee, C. U., & Tan, Y. P. (2007). A multi-objective genetic algorithm for robust flight scheduling using simulation. *European Journal of Operational Research*, *177*(3), 1948–1968.

Lee, S. Y., Yoo, K. E., & Park, Y. (2014). A continuous connectivity model for evaluation of hub-and-spoke operations. *Transportmetrica A: Transport Science*, *10*(10), 894–916.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, Massachusetts. London, England: MIT press.

Montlaur, A., & Delgado, L. (2017). Flight and passenger delay assignment optimization strategies. *Transportation Research Part C: Emerging Technologies*, *81*, 99–117.

Morelli, J. (2011). Environmental sustainability: A definition for environmental professionals. *Journal of environmental sustainability*, *1*(1), 2.

Olsson, N. O., & Haugland, H. (2004). Influencing factors on train punctuality—results from some norwegian studies. *Transport policy*, *11*(4), 387–397.

O'Connell, J. F., & Bueno, O. E. (2018). A study into the hub performance emirates, etihad airways and qatar airways and their competitive position against the major european hubbing airlines. *Journal of Air Transport Management*, *69*, 257–268.

Pels, E. (2008). The environmental impacts of increased international air transport past trends and future perspectives. In *Global Forum on Transport and Environment in a Globalising World, Guadalajara, Mexico* (pp. 10–12).

Peterson, E. B., Neels, K., Barczi, N., & Graham, T. (2013). The economic cost of airline flight delay. *Journal of Transport Economics and Policy (JTEP)*, *47*(1), 107–121.

Price, T., & Probert, D. (1995). Environmental impacts of air traffic. *Applied Energy*, *50*(2), 133–162.

Qin, S., Mou, J., Chen, S., & Lu, X. (2019). Modeling and optimizing the delay propagation in chinese aviation networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *29*(8), 081101.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in water resources*, *51*, 438–456.

Reynolds-Feighan, A. (2001). Traffic distribution in low-cost and full-service carrier networks in the us air transportation market. *Journal of Air Transport Management*, *7*(5), 265–275.

Rietveld, P., & Brons, M. (2001). Quality of hub-and-spoke networks; the effects of timetable co-ordination on waiting time and rescheduling time. *Journal of Air Transport Management*, *7*(4), 241–249.

Robinson, S. (2010). Conceptual modelling: Who needs it. *SCS M&S Magazine*, *2*(7).

Rothengatter, W. (2010). Climate change and the contribution of transport: Basic facts and the role of aviation. *Transportation Research Part D: Transport and Environment*, *15*(1), 5–13.

Ryerson, M. S., Hansen, M., & Bonn, J. (2014). Time to burn: Flight delay, terminal efficiency, and fuel consumption in the national airspace system. *Transportation Research Part A: Policy and Practice*, *69*, 286–298.

Şafak, Ö., Gürel, S., & Aktürk, M. S. (2017). Integrated aircraft-path assignment and robust schedule design with cruise speed control. *Computers & Operations Research*, *84*, 127–145.

Santos, B. F., Wormer, M. M., Achola, T. A., & Curran, R. (2017). Airline delay management problem with airport capacity constraints and priority decisions. *Journal of Air Transport Management*, *63*, 34–44.

Sato, H., Aguirre, H. E., & Tanaka, K. (2007). Controlling dominance area of solutions and its impact on the performance of moeas. In *International conference on evolutionary multi-criterion optimization* (pp. 5–20).

Shannon, R. E. (1998). Introduction to the art and science of simulation. In *1998 Winter Simulation Conference. Proceedings* (Vol. 1, pp. 7–14).

SkyTeam. (2020). *Onze member airlines.* Retrieved 2020-03-05, from https://www.skyteam.com/NL

Sohoni, M., Lee, Y.-C., & Klabjan, D. (2011). Robust airline scheduling under block-time uncertainty. *Transportation Science*, *45*(4), 451–464.

Sternberg, A., Soares, J., Carvalho, D., & Ogasawara, E. (2017). A review on flight delay prediction. *CoRR*. Retrieved from http://arxiv.org/abs/1703.06118

Thengvall, B. G., Bard, J. F., & Yu, G. (2000). Balancing user preferences for aircraft schedule recovery during irregular operations. *Iie Transactions*, *32*(3), 181–193.

Trautmann, H., Ligges, U., Mehnen, J., & Preuss, M. (2008). A convergence criterion for multiobjective evolutionary algorithms based on systematic statistical testing. In *International Conference on Parallel Problem Solving from Nature* (pp. 825–836).

Veiseth, M., Olsson, N., & Saetermo, I. (2007). Infrastructure's influence on rail punctuality. *WIT Transactions on The Built Environment*, *96*, 481–490.

Vikhar, P. A. (2016). Evolutionary algorithms: A critical review and its future prospects. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)* (pp. 261–265).

Vreuls, H. H., & Zijlema, P. J. (2009). The netherlands: list of fuels and standard co2 emission factors. *SenterNovem report to the Ministry of VROM (Spatial Planning, Housing and the Environment, Utrecht, the Netherlands*.

Ward, V. L., Singh, R., Reed, P. M., & Keller, K. (2015). Confronting tipping points: Can multi-objective evolutionary algorithms discover pollution control tradeoffs given environmental thresholds? *Environmental Modelling & Software*, *73*, 27–43.

Wong, J.-T., & Tsai, S.-C. (2012). A survival model for flight delay propagation. *Journal of Air Transport Management*, *23*, 5–11.

Wu, C.-L. (2005). Inherent delays and operational reliability of airline schedules. *Journal of Air Transport Management*, *11*(4), 273–282.

Wu, C.-L. (2006). Improving airline network robustness and operational reliability by sequential optimisation algorithms. *Networks and Spatial Economics*, *6*(3-4), 235–251.

Wu, C.-L. (2008). Monitoring aircraft turnaround operations–framework development, application and implications for airline operations. *Transportation Planning and Technology*, *31*(2), 215–228.

Wu, H.-C., Cheng, C.-C., & Ai, C.-H. (2018). An empirical analysis of green switching intentions in the airline industry. *Journal of Environmental Planning and Management*, *61*(8), 1438–1468.

Zitzler, E., Brockhoff, D., & Thiele, L. (2007). The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 862–876).

A ASSUMPTIONS

This appendix describes two types of assumptions, namely structural and data assumptions. Structural assumptions are assumptions about the operation of the real-world system. Data assumptions are assumptions on the data used in this research (Banks, 1998).

A.1 STRUCTURAL ASSUMPTIONS

- **There is a focus on scheduled passenger connections.**
  There are two types of passenger connections namely actual passenger connections and scheduled passenger connections. *Actual passenger connections* are the number of passengers that can actually make the transfer they intended on the day of operation. This relates to passenger disruptions. *Scheduled passenger connections* are the number of possible connections that can be made according to the schedule. Here, the revenue of a connection is important. In this research, there is a focus on the scheduled passenger connections for the hub connectivity since this is influencing the flight schedule and not the day of operation. By increasing reliability, the actual passenger connections is expected to increase.
- **Crew scheduling is excluded.**
  The crew scheduling is not taken into account in this research.
- **Baggage connectivity is excluded.**
  Together with passengers, baggage also needs to be transferred. Different norms and processes for transferring baggage are applied. However, when a passenger was able to transfer but the baggage was not, it could lead to a lower customer satisfaction level. Baggage transfers could also delay the aircraft. For this research, this baggage flow has not been taken into account with respect to connectivity.
- **Block time is always 100% accurate.**
  It is assumed that the scheduled block time is 100% accurate. This means that time between departing and arriving is the actual time that an aircraft needs to perform a flight, with 100% accuracy. In real life, the scheduled block time is not always in line with the actual flight time.
- **On-time maintenance routing is included.**
  Maintenance trips are included in the flight schedule. Hereby, it is assumed that the maintenance for an aircraft has to be performed and

during this time, the aircraft cannot be used for a flight. It is assumed that an aircraft is always on time after maintenance for the next flight. Also delays on starting maintenance does not affect the time to finalize maintenance. This means that, although the maintenance starts late due to delays, it will always finish on time. Thus, there is no focus on the maintenance routing but the trips itself are included in the flight schedule.

- **Standard Processing Time (SPT) for ground services is always on-time**
  The processing time of an aircraft on the ground from any door open (ADO) until all doors closed (ADC) is assumed to be always on time.
- **On-time arrival of international flights (ICA) is assumed.**
  ICA connections are taken into account for the connectivity element of the schedule. This means that the connections of long haul flights with short haul flights are taken into account. It is assumed long haul flights always arrive on time.
- **There is a fixed fleet assignment.**
  It is assumed that a flight has to be performed with the assigned aircraft type.
- **There is a fixed amount of aircraft available.**
  For each aircraft type, it is assumed that there is a fixed amount of aircraft available.
- **Early arrivals are considered as on-time.**
  Early arrival are considered as on-time in the model, i.e. they have zero delay minutes. The effect of early arrivals, thus compensating some delay due to early arrivals, is not be taken into account when allocating buffers.
- **Alternative flights for a transfer are excluded.**
  For this research, the possibility of taking alternative flights to the final destination is not taken into account for measuring connectivity. Thus, when a transfer cannot be made anymore, it is not weighted whether this is less important since there is another flight to the same destination shortly afterwards.

A.2 DATA ASSUMPTIONS

A.2.1 General

- **Time is defined as Coordinated Universal Time (UTC).**
  All time elements in this research are defined in UTC and not local time. This is mostly used by the airline industry to ensure mutual understanding on time.

A.2.2 Optimization Values

- The model uses 10.000 function evaluations to optimize the problem.
- Optimization in this research is performed for aircraft type Boeing 737-800 and 737-900.
- Two weeks in Summer 2019 are intended to be optimized and compared with the current schedule namely the week of July 22, 2019 to July 28, 2019 and the week of August 12, 2019 to August 18, 2019. Due to time limitations, only Tuesday 23 July 2019 and Thursday 25 July 2019 are optimized.
- The model only optimizes for one day at a time.
- The model optimizes only the fleetlines of one aircraft type at once.

A.2.3 Simulation Model

- **Historical data of CDM is from 2017, 2018 and 2019.**
- **Historical rotation performance data is based on data from March 1, 2017 to October 29, 2019.** The period begins at the start of Summer 2017 and ends at the last day of Summer 2019. For consistency with CDM historical data, this time period is chosen for the rotation performance. Also this research focuses on the summer season and therefore, the summer period is chosen as starting and ending point.
- **Rotation delays, either positive or negative, of more than 60 minutes are not taken into account.**
- **Buffer time at the outstations is included in the arrival delay.** When buffer is planned at the outstation, some of the rotation delay can be captured by this buffer. Thus, the departure delay plus the sampled rotation delay minus the buffer at the outstation accounts for the final arrival delay.
- **The number of replications is 10.** The number of required replications depends on the desired statistical accuracy of the research (Carson, 2004). There is no rule of thumb for this. For the inherent variability factors in the simulation, i.e. collaborative decision making moments drawn from a distribution and the rotation performance, 10 replications for one buffer allocation solution are necessary.

A.2.4 Connectivity

- **Connections with a connection time longer than six hours are excluded from the initial schedule.**
  Long connections of more than six hours are not desirable and thus, not offered. This assumption ensures that a rotation will not shift that much (e.g. more than three hours) such that these connections become interesting. Also the computational time of the model decreases when applying this limit.

- **Number of connecting passengers and price of the tickets is calculated per day of the week in this research.**
  The number of connecting passenger and the price of the tickets is grouped by the day of the week (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday).

- **Probability of a passenger actually making their transfer based on the transfer time is determined per connection type.**
  By this, the Schengen versus Non-Schengen connection types are incorporated.

- **Data on new possible connections is not included.**
  This research only has data on historical connections and not on future connections based on market share. The model cannot find opportunities to create new connections. Therefore, the optimal schedule is expected to be quite similar to the base case in terms of the type of connections made. Thus, the variability of the connectivity value could be limited due to this data pitfall.