D I S S E R T A T I O N

Robust Efficiency of Airline Resource Schedules

submitted to the
School of Business and Economics
of Freie Universität Berlin

in Candidacy for the Degree of
doctor rerum politicarum (Dr. rer. pol.)

by
Dipl.-Wirt.-Inf. Lucian Ionescu

Berlin, 2018

Referees

Univ.-Prof. Dr. Natalia Kliewer, Freie Universität Berlin
Univ.-Prof. Dr. Ralf Borndörfer, Freie Universität Berlin

Date of Disputation

February 12$^{\text{th}}$, 2018

Acknowledgement

This research was supported by a grant from the German Research Foundation (DFG, Grant No. KL2152/3-1).

Abstract

During operations, airlines frequently have to deal with disruptions, causing a gap between expected planned costs and real operational costs for the resource deployment. This issue is addressed by the concept of *robust efficiency*: While maintaining its planned cost efficiency as well as possible, the robustness of a schedule against the impact of unforeseeable delays is improved. There are two eligible schedule properties that affect the robustness, namely stability and flexibility. While stability refers to a schedules' ability to maintain feasibility in spite of emerging delays, flexibility addresses a schedules' capability to be easily adapted to changing environments. In this context, the objective of this work is to examine influential factors and effect mechanisms in the scope of robust efficiency. In particular, we refer to the airline crew scheduling along with the closely related aircraft scheduling.

Inevitably, the trade-off between robustness and planned cost efficiency depends on the accuracy of information on delay occurrence probabilities. Therefore, we investigate the potential of data-driven delay prediction in a first step. Meeting the requirements of long-term robust resource scheduling, interpretable decision rules for delay occurrence patterns are derived. Subsequently, their prediction accuracy is assessed in a subsequent statistical modeling step, following the concept of the analysis of covariance (ANCOVA). The actual influence of delay prediction on the trade-off between cost-efficiency and robustness of crew schedules is quantified in a scheduling and simulation study on real-world flight instances.

In a second step, we examine potential mutual impacts of stability and flexibility of crew schedules. For this purpose, we extend an existing prototypical scheduling and simulation framework by the following components: At first, we propose a rule-based recovery approach for crews and aircraft that aims at the evaluation of schedule flexibility. At second, we develop a stochastic optimization approach for increasing the flexibility of crew schedules. Obtained results show that the benefit of stability is generally dominant to flexibility in regular daily operations. In particular, schedules with an improved degree of stability necessitate substantially less recovery interventions. In contrast, the main advantage of an improved flexibility is that delay propagation chains on consecutively operated flights of the same crew can be prevented by improved recovery capabilities.

In summary, the results of this thesis show the benefit of understanding delay occurrence mechanisms and demonstrate the methodical potential of delay prediction for robust resource scheduling. Moreover, the identification of effect mechanisms between stability and flexibility offer valuable insights into the trade-off between planned cost efficiency and robustness.

Zusammenfassung

Während der operativen Phase sind die Ressourcen-Einsatzpläne von Fluggesellschaften oft unvermeidlichen Störungen im Betriebsablauf ausgesetzt. Durch die kostenintensive Notwendigkeit zur Planwiederherstellung führt dies mitunter zu deutlichen Unterschieden zwischen erwarteten Plankosten und den real anfallenden operativen Kosten für den Ressourceneinsatz. Diese Thematik wird von dem Konzept der *Robusten Effizienz* aufgegriffen. Es zielt darauf ab, die Robustheit von Plänen gegen unvorhersehbare Verspätungen zu erhöhen, während die Plankosteneffizienz zu größtmöglichen Teilen erhalten wird. Bezüglich des Einflusses auf die Robustheit sind die Planeigenschaften der Stabilität und Flexibilität wesentlich. Die Stabilität zielt darauf ab, die Ausführbarkeit des ursprünglichen Plans auch beim Auftreten von Verspätungen aufrechtzuerhalten. Im Gegensatz dazu bezeichnet die Flexibilität die Möglichkeit, einen Plan mit einfachen und überschaubaren Maßnahmen an gegebene Umstände anzupassen.

In diesem Zusammenhang behandelt diese Arbeit das Ziel, Einflussfaktoren und Wirkmechanismen der Robusten Effizienz zu identifizieren. Unsere Untersuchungen behandeln dabei schwerpunktmäßig die Dienstplanung für Cockpit-Besatzungen und die damit eng verbundene Flugzeug-Einsatzplanung. Die Austauschbeziehung zwischen Plankosteneffizienz und Robustheit hängt in hohem Maße von der Informationsgenauigkeit bezüglich des Auftretens von Verspätungen ab. Diesbezüglich untersuchen wir in einem ersten Schritt das Potenzial zur datengetriebenen Verspätungsvorhersage. Unter Berücksichtigung der langen Vorlaufzeit der Ressourceneinsatzplanung werden interpretierbare Entscheidungsregeln für das Eintreten von Verspätungen abgeleitet. Anschließend werden diese durch statistische Modellierung auf ihre Vorhersagegüte untersucht. Die Untersuchung basiert auf dem Konzept der Kovarianzanalyse (ANCOVA). Der konkrete Einfluss von Verspätungsvorhersagemodellen auf die Austauschbeziehung zwischen Plankosteneffizienz und Robustheit von Dienstplänen wird in einer Planungs- und Simulationsstudie basierend auf realen Flugplänen evaluiert.

In einem zweiten Schritt untersuchen wir mögliche wechselseitige Auswirkungen der Stabilität und Flexibilität von Dienstplänen. Zu diesem Zweck erfolgt die Erweiterung einer vorhandenen prototypischen Planungs- und Simulationsumgebung durch folgende Komponenten: Zunächst entwickeln wir ein regelbasiertes Wieder-

herstellungsverfahren für Dienst- und Flugzeugeinsatzpläne, welches die Evaluierung derer Flexibilität ermöglicht. Des Weiteren wird ein stochastisches Optimierungsverfahren vorgestellt, welches auf die Erhöhung der Dienstplan-Flexibilität abzielt. Ergebnisse der auf diesen Komponenten basierenden kalkulatorischen Studie zeigen eine Dominanz der Stabilität gegenüber der Flexibilität. Stabile Dienstpläne erfordern substanziell weniger operative Eingriffe zur Planwiederherstellung. Im Gegensatz dazu liegt der Vorteil eines verbesserten Flexibilitätsgrades darin, dass durch verbesserte Wiederherstellungsmaßnahmen verkettete Mehrfach-Propagationen von Verspätungen auf Flugfolgen von Crews zunehmend verhindert werden können.

Zusammenfassend lässt sich festhalten, dass die Ergebnisse dieser Arbeit das Verständnis von Auftritts-Mechanismen von Verspätungen fördern und das methodische Potenzial der Verspätungsvorhersage für die robuste Planung aufzeigen. Darüber hinaus bewirkt die Herausarbeitung von Wirkungsmechanismen zwischen den Planeigenschaften Stabilität und Flexibilität einen wertvollen Einblick in die Austauschbeziehung zwischen Plankosteneffizienz und Robustheit.

Contents

List of Figures

List of Tables

List of Algorithms

Part I

Airline Scheduling – Efficient and Robust Operations

Chapter 1

Introduction

In times of advancing globalization, air transportation has become an indispensable part in the lives of many people. For both holiday and business travels, attractive, reliable and widespreading air transportation plays an important role. Although the number of continental flights in Europe decreased from 10 million in 2008 to 9.5 million in 2012, EUROCONTROL expects an increase to 11.2 million flights in 2019[1]. A more recent study expects a growth of 2.4% just in 2016 with up to 11.5 million flights in 2022, see (EUROCONTROL, 2016a, p. 19).

Airlines respond to the growing demand with increased frequencies on existing flight routes and new destination offerings. It inevitably goes along with additional complexity of their flight networks. More crews, aircraft but also airport ground equipment and scarse airspace has to be efficiently scheduled in order to allow regular flight operations. Moreover, low-fare airlines sharpen an already tough competition and reinforce the need to operate flights as cost-efficient as possible[2]. To name a few examples, an airlines' flight network must be planned in a way that allows efficient transportation of passengers to their destination with few transfers. Reasonable positioning of spare resources and maintenance facilities avoids ferry flights without revenue. Fuel saving flight routes must be determined with respect to wind and geographical conditions while still complying with legal rules of air traffic control authorities. Itineraries of crews and aircraft must be scheduled in a way that reduces unnecessary idle times and eventually reduce the number of resources needed for operation.

[1]See EUROCONTROL (2013) for further details.

[2]Recent figures for 2015 indicate a market share of 53.7% for traditionally scheduled and 28% for low-fare flights. See (EUROCONTROL, 2016a, p. 9) for further details.

Accordingly, state-of-the-art information technology and decision support tools have become an inseparable part of airline scheduling and operations in the last decades. Commercial software suites for airlines are available from a large number of specialized vendors, e.g. Jeppesen, Lufthansa Systems (Netline Suite) or Sabre Airline Solutions, to only name a few. They cover a wide range of airline-specific routines from strategic decision making to resource scheduling and actual operations. In particular, Operations Research (OR) methods have been successfully applied to solve many kinds of airline scheduling problems. In this regard, the explicit focus of this thesis is on the *Airline Crew Scheduling* along with the closely related *Aircraft Scheduling*. Especially the crew scheduling has been widely discussed in the OR literature because of two aspects. Firstly, crew costs are usually second highest for airlines right after fuel costs, see (Anbil et al., 1991, p. 62) and (Klabjan et al., 2001, p. 75). Secondly, complex work rules and labor regulations result in a high computational complexity, necessitating the application of sophisticated optimization techniques. It is therefore one of the most relevant topics discussed in airline-related scientific literature.

Our research strives to combine the requirements imposed by contradictory objectives during scheduling and operations. Traditional approaches on crew and aircraft scheduling have primarily dealt with the reduction of deployment costs, aiming at high utilization levels of resources. On the downside, resulting tight schedules may offer too little slack times. These, however, are important when schedules are exposed to changing environments at the day of operations. Technical breakdowns of equipment, severe weather conditions, late passengers or airspace congestion, to only name a few, may disrupt regular operations and eventually lead to infeasible schedules. As a consequence, the airline operations control center (AOCC) has to either postpone subsequent flight departures or has to reallocate resources so that affected flights can be operated on-time. Otherwise, passengers can be rerouted or rebooked to other flights when possible. However, if all else fails and flights must be canceled, compensation and hotel accommodation costs for overnight stays may arise. Inevitably, real costs for an airline can substantially differ from originally planned costs: Referring to (Cook and Tanner, 2011, p. 8), the amount of delay-related reactionary costs were estimated to exceed 1.25 billion Euros just for the year 2010 in Europe which is around 81 Euros per minute of delay.

In this context, we introduce the concept of *robust efficiency* which aims at the minimization of real costs by already taking into account reactionary costs during the scheduling stage. This can be achieved by improving the robustness of schedules in two ways:

- Improving the schedules' ability to withstand the impact of disruptions and delays in order to maintain feasibility and execution as originally planned. We refer to this as the *operational stability* of a schedule.
- Improving the schedules' capability to be easily adapted to changing environments in order to maintain timely execution of tasks. This is referred to as a schedules' *operational flexibility*.

As a result, a stronger coupling of the objectives between scheduling and operations can be obtained. At best, the incorporation of scheduled slack times reduces the amount of emerging delays in case of unforeseen disruptions. Moreover, the necessity and severity of interventions by the AOCC can be reduced. Inevitably, the benefit of related robust scheduling approaches for crews and aircraft inevitably depends on information on delay occurrence probabilities: More accurate delay prediction promises an increase of schedule robustness at a lower level of planned cost increase.

In this regard, the objective of this work is to examine influential factors and effect mechanisms in the context of the robust efficiency of crew and aircraft schedules. Technically, our work bases on the prototypical scheduling and simulation framework created in a preceding project, see Dück (2010). We develop necessary extensions of the existing mathematical optimization approach and the stochastic simulation procedure in order to meet the demands of our research. In particular, we address the following interrelated topics: At first, we examine the **impact of refined delay prediction** on the robust efficiency of schedules. Therefore, findings on the potential of data-driven delay prediction are obtained in an analysis of historical delay data from a major European carrier. The actual impact on the robust efficiency is investigated in a simulation study under varying assumptions on delay occurrences. At second, we identify the influence of schedule characteristics on the robust efficiency. For this purpose, a **simulation-based evaluation technique for the robust efficiency** is developed that takes into account both stability and flexibility aspects. It forms the groundwork for a subsequent study on potential **mutual impacts between**

stability and flexibility in order to obtain a holistic view on effect mechanisms related to robust efficiency.

This thesis connects methods of data analysis, stochastic simulation and mathematical optimization. According to (Domschke et al., 2015, pp. 2-3), this corresponds to a recent shift of the focus from traditional OR – dealing with mathematical modeling and optimization – to an advancing integration with the methods of Analytics. For more information on the differentiation and interaction of OR and Analytics, we refer to Liberatore and Luo (2010) and Fink et al. (2014).

Organization of the Thesis

This thesis is split up into four main parts. In the remainder of the current Part I, we provide an overview of the airline scheduling and flight operation processes along with a literature survey on recent advances in the field of robust resource scheduling. Part II provides an analysis of historical delay data in order to examine the potential of data-driven delay prediction. Part III addresses the simulation of crew and aircraft schedule operations for the evaluation of robust efficiency. In Part IV, the focus is on scheduling strategies, dealing in particular with the identification of potential mutual impacts between stability and flexibility.

Part I (Airline Scheduling – Efficient and Robust Operations): The subsequent Chapter 2 introduces the airline planning process, focusing on crews and aircraft scheduling. We describe strategic, tactical and operational planning stages, their chronological order and interrelations between related decisions. Moreover, recent approaches on the (partial) integration of planning stages are outlined. In Chapter 3, the focus is on the implementation of generated resource schedules and operating flights. Based on a detailed description of activities during flight operations, we concentrate on the topic of inevitably emerging disruptions and delays, followed by outlining necessary responsive recovery actions of an airline. Subsequently, the concept of robust efficiency is introduced, aiming at the adaption of operational objectives already during the scheduling phase. Chapter 4 provides an overview of fundamental approaches and recent advances in robust crew and aircraft scheduling. In particular, we address optimization techniques as well as the usage of historical delay data for delay prediction in recent stochastic scheduling approaches.

Based on these considerations, we derive four research objectives that are tackled in the scope of this thesis in Chapter 5. In addition, necessary extensions of the base prototypical scheduling and simulation framework of Dück (2010) are introduced.

Part II (Study on Exogenous Delays in Airline Networks): Part II comprises topics that relate to the analysis of historical delay data. The data set is introduced in Chapter 6. After an initial description and contextualization of attributes, a time course analysis is provided for the examination of general trends and patterns in delay occurrences. In Chapter 7, the potential of primary delay prediction for robust resource scheduling is examined. Decision rules for delay occurrences are obtained in an exploratory analysis and subsequently assessed regarding their prediction accuracy by statistical modeling.

Part III (Evaluating the Robust Efficiency of Crew and Aircraft Schedules): Part III addresses the evaluation of the robust efficiency by simulation. Chapter 8 introduces the event-driven stochastic discrete simulation procedure for the evaluation of crew and aircraft schedules. Besides, a formalization of delay propagation mechanisms is provided and the prediction accuracy of a theoretical delay propagation model is assessed on real-world data. In Chapter 9, the influence of primary delay prediction on the robust efficiency of crew schedules is examined. We assess statistical models obtained in Chapter 7 and subsequently perform a sensitivity analysis on primary delay occurrence assumptions. Chapter 10 deals with the conception and implementation of a rule-based recovery procedure for crews and aircraft for the purpose of schedule flexibility evaluation. The impact of the approach on the robustness outcome of evaluated schedules is assessed in computational experiments.

Part IV (Mutual Impacts of Operational Stability and Flexibility): The final Part IV deals with the examination of schedule characteristics regarding their effect on the robust efficiency. In particular, the mutual impact between stability and flexibility is investigated. Therefore, a mathematical formulation and a solution approach for increasing the flexibility of airline crew schedules is described in Chapter 11. Afterwards, mutual impacts are elaborated in Chapter 12.

Finally, we summarize the findings of this thesis and propose directions for future research in Chapter 13.

Chapter 2

The Airline Planning Process

In this chapter, we introduce the airline planning process. It includes several stages that are characterized by a high degree of complexity and interdependencies resulting from close interconnections: In the long-term, strategic supply and capacity planning have to be performed. Based on this, the flight network has to be determined and the aircraft fleet has to be acquired. The main objectives are to maintain and expand the airlines' market shares and to safeguard and strengthen the competitiveness. In the medium-term tactical planning, actual flight schedules have to be constructed by setting up service frequencies for every route in the flight network. In addition, the capacity planning has to be specified accordingly by determining which flight is served by which aircraft type. The main objective in tactical planning is the maximization of expected revenue. In the subsequent operational planning, resources such as crews and aircraft have to be scheduled for operating the flight schedule. The objective is to schedule available resources as efficiently as possible in order to minimize operating costs.

Traditionally, the whole airline planning process is carried out sequentially. Each stage is considered separately with respect to decisions taken in previous stages. This approach is often a result of an organizational split of departments that are each responsible for one particular planning task, e.g. the scheduling of (cockpit) crews. The split is also reflected in decision support systems which usually relate to one individual stage due to their high complexity. In consequence, decisions taken early in the planning process may lead to sub-optimal solutions in later stages. This issue can partially be resolved by an iterative reconsideration of previous stages. Moreover, the progressive development of decision support systems and computer hardware allow advances in the (partial) integration of planning stages on the technical side.

The remainder of this chapter is organized as follows: Section 2.1 introduces determinants for the strategic orientation of an airline and resulting planning problems. Tactical planning stages are described in Section 2.2. A detailed presentation of operational planning with a focus on crew and aircraft scheduling is provided in Section 2.3. We refer to the (partial) integration of planning stages in Section 2.4. Eventually, a summary and implications for subsequent studies of this thesis are provided in Section 2.5.

2.1 Strategic Airline Planning

Fundamental decisions on the strategic orientation of an airline form the basis for the entire planning process. In the beginning of this section, we outline the effects of the market deregulation and the arising necessities for airlines to maintain competitiveness. Subsequently, we outline strategic decisions that have to be made in the long-term.

For a long time, airlines have been in possession of state governments. Examples for these so called former *flag carriers* are Deutsche Lufthansa (Germany), KLM (Netherlands), British Airways (United Kingdom) or Air France (France). The privatization of airlines, that has been initialized in the 1980 and 1990s by a market deregulation, is almost completed in Europe[1]. As an example, the privatization of Deutsche Lufthansa as the largest German airline was finalized in 1997[2].

In the aftermath of the U.S. Airline Deregulation Act, which came into force in 1987, long-term effects of deregulation can be observed: Ticket prices have dropped, the utilization of aircraft has been increased and a more appropriate use of different aircraft types for short- and long-haul operations could be observed[3]. In addition, low-fare airlines were encouraged for market entry, offering inexpensive flights by reducing extra services. In consequence, traditional airlines are exposed to an increasing competition, resulting in the necessity to consider their strategic orientation more

[1] For an updated list of government-owned and privatized airlines, see `https://www.icao.int/sustainability/Documents/PrivatizedAirlines.pdf`, last access: January 8$^{\text{th}}$, 2018.

[2] For details, we refer to `https://www.lufthansagroup.com/en/company/history.html`, last access: January 9$^{\text{th}}$, 2018.

[3] Report of the Government Accountability Office (GAO) on the effects on Airline Deregulation, `https://www.gao.gov/archive/1997/rc97004.pdf`, last accessed: January 8$^{\text{th}}$, 2018.

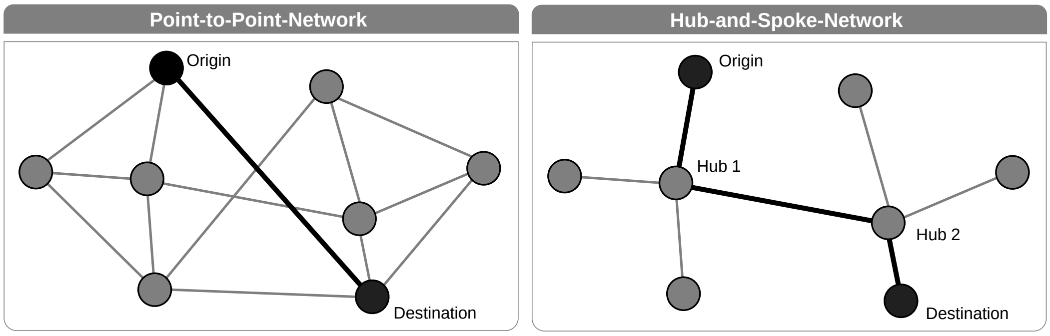

Figure 2.1: Two fundamental network types for airline operations

thoroughly. Among other things, this includes the consideration of whether and to what extent specific market segments are served. In close connection, it has to be decided which customer target group shall be addressed. Moreover, related branding and marketing strategies have to be developed. The strategic orientation of an airline may be revised occasionally, e.g. when entering new markets and offering new routes. Hence, strategic planning is subject to continuous iterative adjustments, always taking into account current demand situations and competing carriers. Prior to any decision related to airline operations it is therefore necessary to evaluate prospects of market success and demand forecasts.

In the following, we present two strategic planning problems that have a relevant impact on our research, namely the *Network Design* and *Fleet Planning*. In the Network Design, an airline selects connections between airports, resulting in a set of origin-and-destinations (so called *O&Ds*). O&Ds may sometimes consist of more than one flight, depending on the network structure. If two cities and their respective airports are not connected by a direct flight, it is still possible to reach the destination by one or even more stopovers and transfers. Thus, a direct association between O&Ds and flights is not mandatory. Flights offered by an airline form its flight network. Commonly, there are two fundamental theoretical network types, illustrated in Figure 2.1.

On the one hand, a network may contain connections between all airports, which is called a point-to-point network. Considering n airports in a theoretical complete point-to-point network, the connection count is $(n-1)+(n-2)+\cdots+1 = n\cdot(n-1)/2$. However, not all airports are interconnected in practice, but the focus is on O&Ds

that offer a certain level of profitability. Such network types are mostly implemented by small regional and low-cost carriers[4].

On the other hand, there is the hub-and-spoke network structure, primarily used by large carriers[5]. The nodes of the network represent airports and are split up in two types of airports – hubs and spokes. Commonly, hubs are major airports where nearly all important processes are bundled. This especially includes the location of crew bases and maintenance facilities but also the placement of spare resources and reserve crew deployment. A hub often is a large airport that can sufficiently handle all needs regarding quantity and complexity of large carrier operations. Examples for large hub airports in Europe are London/Heathrow (British Airways), Frankfurt/Main (Lufthansa) or Paris/Charles-de-Gaulle (Air France), see EUROCONTROL (2016b). Due to efficiency reasons, wavelike departures and arrivals are used at hubs, commonly referred to as banks, see (Rexing et al., 2000, p. 6). In this way, passengers are transferred to the hub by many flight arrivals in a specific time window and afterwards distributed on outgoing flights, bringing them to their destination airports. Banks are especially helpful for long-haul flights with high capacity aircraft whose passengers must be collected from several origins first. In contrast to hubs, spoke airports are connected by flights from or to the hub only. Thus, a pure hub-and-spoke network provides connections between two spokes only via transfer at a hub. This network design allows significantly more connection possibilities with less flights. Considering one hub airport, and n spokes, a complete connectivity is derived by n connections. This advantage is the reason why characteristics of hub-and-spoke networks have increasingly been established as a consequence of the market deregulation, see Burghouwt et al. (2003) and Adler and Golany (2001) for studies on the development of flight networks in the European market. However, pure hub-and-spoke structures rarely exist in practice. This is because direct connections are often favorable for domestic short-distance O&Ds where airlines often compete with railway and individual traffic, see Meffert et al. (2005) and Jiménez and Betancor (2012)).

[4] For a detailed list of airlines that operate in a point-to-point-related network structure, we refer to (Conrady et al., 2013, p. 237).

[5] An exemplary list of airlines operating in hub-and-spoke network structures is provided by (Conrady et al., 2013, p. 227).

Also based on the strategic orientation, the composition of the airlines' aircraft fleet is specified in the Fleet Planning step. Since aircraft acquisition or leasing is a long-term decision, operational efficiency of aircraft is of high importance. According to Bazargan (2010), operating and maintenance costs are the major cost drivers to be considered at Fleet Planning. Since each aircraft type is related to a specific passenger capacity, the Fleet Planning depends on strategic demand forecasts and strategic considerations. Aircraft types can be classified by their family, e.g. Airbus A320 or Boeing 737 for short-/medium-haul and Airbus A340/350 or Boeing 757 for long-haul flights. Usually, an aircraft family contains a set of aircraft with identical technical configurations and crew requirements, primarily differing in fuselage length, seating configuration and passenger capacity. A major airline's fleet usually consists of several aircraft families, e.g. Airbus A320 and Boeing 737 for short haul and the Airbus 330/340 or Boeing 767/777 for long haul operations. The fleet of the major German carrier Lufthansa comprises 15 aircraft types for various fields of application[6]. In contrast, low cost carrier often concentrate on one or at least a small number of different aircraft families. The main reason for this are advantages in maintenance and crew requirement specifications. In this regard, the low-cost carrier Ryan Air operates its continental flight network with over 300 aircraft of the same Boeing 737-800 type[7].

2.2 Tactical Planning

In the tactical *Schedule Design*, frequencies for each O&D of the flight network are specified, determining how often and when an O&D is served. According to Lohatepanont and Barnhart (2004), the Schedule Design consists of the two consecutive steps frequency planning and timetable development. When setting up frequencies, deviations in the demand structure by time of day, weekday and season must be taken into account. The result is a *Service Plan*, containing desired destinations and their frequencies that promise high revenue. Based on the service plan, a flight schedule is elaborated, including specific departure and arrival airports and times. It must

[6] `https://www.lufthansagroup.com/en/company/fleet/lufthansa-and-regional-partners.html`, last access: April 21st, 2017.

[7] `https://www.ryanair.com/gb/en/useful-info/about-ryanair/fleet`, last acces:s April 21st, 2017.

comply with operational constraints such as the availability of airport departure and arrival slots and available resources such as aircraft and staff, see (Klabjan et al., 2001, p. 4) for further details.

The parts of the Schedule Design are usually formulated and solved as mathematical problems, making it reasonable to tackle them with methods of Operations Research. Essential scientific studies on this topic can be found in Etschmaier and Mathaisel (1985), Teodorović and Krcmar-Nozić (1989) and, more recently, Erdmann et al. (2001). Due to the possibility of shifting departure times within certain time windows, Schedule Design decisions are often integrated into subsequent resource scheduling stages. By doing so, the degree of freedom for resource scheduling decisions can be increased, allowing considerable cost reductions. An exemplary examination of the benefits is provided by Klabjan et al. (2002). For further discussion on aspects of integration, we refer to Section 2.4.

In the *Fleet Assignment*, an aircraft type is assigned to each flight of the flight schedule. The objective is to meet demand forecasts and aircraft seating capacities as well as possible. Demand potentials shall be exploited and low load factors of aircraft shall be prevented. Mismatches between aircraft capacities and passenger demands lead to inefficiency or lost revenue. Selecting an aircraft that is too small to fully exploit the passenger demand of a flight leads to so called *spill costs*, while an aircraft being too large necessitates higher operational costs for operations without additional benefit.

The Fleet Assignment Problem has been well addressed in OR literature. A basic Fleet Assignment Model can be formulated as a multi commodity network flow problem, see Hane et al. (1995). Hard constraints are demand restrictions, the fleet structure including aircraft range and capacity as well as limits in ground operations like the number of available gates at airports. Also flow constraints are considered, i.e. departures and arrivals has to be balanced at each airport.

Although major carriers' fleets consist of a variety of different aircraft types, the problem is easy to solve by standard IP solvers, as stated by Klabjan (2005). Solution approaches can be found in Rushmeier and Kontogiorgis (1997), Barnhart et al. (2001), Barnhart et al. (2002) and Sherali et al. (2006). Potential solution improvements by new optimization techniques have almost been tapped. More recent approaches therefore address the stochastic components in a more dedicated way, e.g. by capturing spill mechanisms and uncertainty in passenger demand (Barnhart et al. (2009)). Sherali and Zhu (2008) propose a two-stage model capturing demand uncertainty. The model is solved several months prior to the day of operations and then being resolved when the real passenger demand is known. Duquesne (2013) propose a market-driven Fleet Assignment Model considering the stochastic nature of passenger demand forecasting. As in the Fleet Assignment crucial decisions for all upcoming scheduling stages are taken, the problem is often used for integrating planning stages, see Section 2.4.

In the context of this thesis, decisions taken in strategic and tactical planning stages are assumed to be fixed and irreversible. They form the basis for the subsequent operational planning stage and, in particular, the crew and aircraft scheduling.

2.3 Operational Planning of Crews and Aircraft

Decisions made in previous tactical planning are intended to maximize profit by meeting the demand as well as possible. In the following scheduling stages, the objective is to minimize the costs for operating the flight schedules. This can be achieved by high utilization levels of available resource capacities and low idle times. In the following, we present the individual planning stage with respect to scheduling decisions and resulting interdependencies. For crew and aircraft scheduling problems, mathematical formulations and solution approaches are provided.

As a result of the Fleet Assignment, an aircraft type is assigned to each flight. By the usage of flow constraints, potential aircraft routes and crew itineraries are already anticipated and pre-selected. However, it is not determined yet which flights are operated in consecutive order by the same aircraft or crew. This issue is now addressed in the *Aircraft Scheduling* and *Crew Scheduling*. Note that by crews we generally refer to pilots and, more generally, cockpit crews in our context. As cockpit crews may hold a license for only one aircraft type at a time, these two scheduling steps can be solved separately for sub-network operated by the same aircraft type. The very similar problem of flight attendant scheduling is far more complex since they often are allowed to serve more than one aircraft type. Due to the similar problem structure similar solution approaches may be applied, as stated in (Vance et al., 1997, p. 188).

2.3.1 Aircraft Scheduling

The Aircraft Scheduling deals with the assignment of aircraft to scheduled flights. It is generally decomposed into *Aircraft Routing* and *Tail Assignment*. In the first step, so called *Aircraft Rotations* are built, i.e. sequences of flights operated by any aircraft of the same type. Rotations have to fulfill certain rules concerning maintenance intervals and minimum connection times for ground handling between consecutive flights. The set of possible connections between flights is predetermined by the fleet assignment solution. In Aircraft Routing it is preferable to prevent very long ground times known as *drips*, see Barnhart and Smith (2012). In the subsequent Tail Assignment, individual aircraft differentiated by their distinct tail number are assigned to the previously constructed rotations. This process allows continuous changes until a few days prior to departure, see (Lapp and Cohn, 2012, pp. 2052). In related literature, transitions between the two stages are often fluent. As an example, (Grönkvist, 2006, p. 2919) only differentiates between Fleet Assignment and Tail Assignment, with the latter also including maintenance planning. However, restrictions and the set of decisions to make are consistent throughout all literature approaches, independently from the question in which stage they are considered. Therefore, we adhere to the presented two-stage approach, allowing a distinctive description of all relevant issues.

Firstly, maintenance restrictions that have to be considered when constructing rotations, differing from time interval, duration and intensity. According to (Klabjan, 2005, p. 7), so called A-checks take 3 to 10 hours and are usually performed overnight to not affect daily operations. They mostly consist of fundamental visual inspections. More intense visual B-checks are required every few months. Barnhart et al. (1998) refer to routes that are flown by an aircraft between two checks as *strings*. For maintenance scheduling, often approximate solutions are generated, e.g. by maximizing the number of aircraft on the ground for overnight maintenance without explicitly schedule time windows for checks. The list of checks is completed by C- and D-checks are month-long extensive investigations every few years which require to partial disassembly of the aircraft. Due to their long lead time, they are not considered in tactical aircraft scheduling.

Secondly, aircraft rotations must comply with minimum time intervals for ground handling between flights. This time span is commonly referred to as *ground time* or *turn time*. Several ground operations have to be performed, including technical checks, refueling, catering, (un-)boarding of passengers, and (un-)loading baggage. The turnaround process is described more detailed in Section 3.

From OR perspective, the Aircraft Routing problem can be formulated as a *Set-Partitioning-Problem (SPP)*, a combinatorical decision problem for integer solutions, aiming at the coverage of all flights in F by a subset of all possible rotations R. The formulation is given in Model 2.1.

$$\min \sum_{j \in R} c_j x_j \tag{2.1a}$$

$$s.t. \sum_{j \in R} a_{ij} x_j = 1 \qquad \forall i \in F \tag{2.1b}$$

$$x_j \in \{0, 1\} \qquad \forall j \in R \tag{2.1c}$$

c_j are the costs assigned to each rotation $j \in R$. The following applies to all elements $a_{ij} \in A$:

$$a_{ij} = \begin{cases} 1, & \text{if flight } i \text{ is part of rotation } j, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

Constraints 2.1b ensure that all flights are covered exactly once. Moreover, only integer solutions are accepted (2.1c). Additional constraints may ensure that the maximum number of aircraft is not exceeded, see (Maher, 2013, p. 10) for a respective enhancement of the formulation.

The costs for operating a flight by an aircraft of the same type can be assumed to be almost identical. Given the maximum number of aircraft, the problem turns into a feasibility problem (Lan et al., 2006, p. 17). Nonetheless, consecutively operating certain flights by the same aircraft may be beneficial, e.g. it is desirable for connecting passengers to remain on the same aircraft, see (Weide, 2009, p. 69). This is especially the case for important O&Ds with high passenger loads where no direct connections are offered. Connections between flights that shall be flown by the same aircraft, are called *Through Connections*. In the objective function, corresponding *Through Values* can then be considered in order to prefer rotations that contain certain connections, see Clarke et al. (1997). According to (Lan et al., 2006, p. 17) and Sarmadi (2004), this may only result in a minor increase of actual revenue since Through Connections may be broken by recovery actions during operations.

The Aircraft Routing SPP can be solved by column generation, basing on the Dantzig-Wolfe decomposition[8]. Early solution approaches can be found in Desaulniers et al. (1997) and Barnhart et al. (1998). In contrast, (Barnhart and Talluri, 1997, pp. 445) and Gopalan and Talluri (1998) describe a different solution approach using graph theory applications. Clarke et al. (1997) use a Lagrangian relaxation with subgradient optimization, leaning on the asymmetric traveling salesman problem. The more recent approach of Grönkvist (2006) has gained attention since the proposed constraint propagation extension for column generation provides significantly faster solutions.

A few weeks or even days prior to departure, particular aircraft are assigned to the previously constructed rotations in the *Tail Assignment*. The term Tail Assignment is also often used for the Aircraft Routing Problem as presented before, see Grönkvist (2006) and Borndörfer et al. (2010).

2.3.2 Crew Scheduling

The Crew Scheduling includes both the *Crew Pairing* step where anonymous crew itineraries are generated[9] and the *Crew Rostering* step which aims at assigning individual staff members to these itineraries. This two-stage approach shows strong analogy with the Aircraft Routing and Tail Assignment.

Crews generally operate from their destined crew base where their service starts and ends. A sequence of flights, that is operated on one single day is called a *crew duty*. Break times between two flights within a duty are called *sit time*, whereas long breaks between two duties are called *rest time*, generally associated with overnight stays. The alternating succession of duties and rest periods defines a *crew pairing*.

For crews, there exists a set of complex labor and flying time restrictions, comprising minimum rest and sit times, maximum flying times per day, month and year as well as specific duty extension rules. According to the *European Regulations on*

[8]The column generation approach is used in this work and therefore is presented in further detail in Appendix A.

[9]Examples for specialized commercial software for the Crew Pairing are *Jeppesen Crew Pairing*, *Netline/Crew Pairing xOPT optimizer suite* (Lufthansa Systems) or *Sabre AirCentre Crew Manager*.

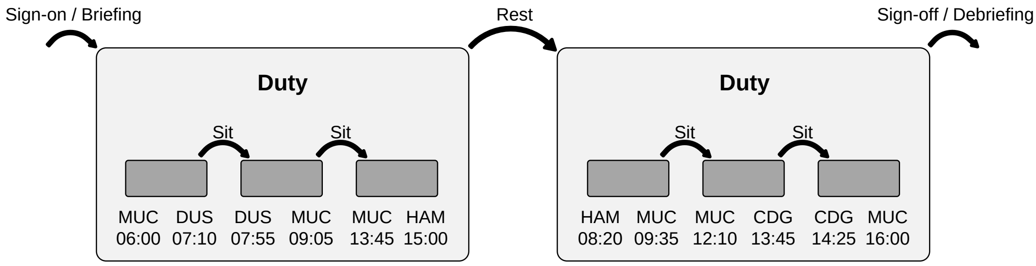

Figure 2.2: An exemplary Crew Pairing, illustration based on (Dück, 2010, p. 12)

Crew Work-Time[10], 60 hours of duty are allowed within seven consecutive days and 190 duty hours in 28 consecutive days. Concerning flying hours, at most 100 hours of flying time are allowed in 28 days and 900 hours per calendar year. A crew pairing may last up to 96 hours which is equivalent to 4 full days. Between two duties, a rest period of at least 10 hours must be ensured. The maximum flying time within one duty is 8 hours. Including sit times, a duty must be conducted in 10 hours. However, it can be extended by to 12 or 14 hours, depending on the duration of the subsequent rest period. For a detailed description of the set of rules that are applied in all following studies, we refer to (Dück, 2010, pp. 11-12).

In Figure 2.2, we illustrate an exemplary crew pairing containing two duties and six flights for a cockpit crew. Departing from the crew base in MUC and after completing the 60-minutes sign-in and briefing, the crew's first scheduled flight is to DUS at 6 p.m. After a sit time of 45 minutes, they head back to MUC, followed by a comparatively long sit time of 4 hours and 40 minutes. So called out-and-back flights are a common instrument in hub-and-spoke networks, referred to as cycles (see Rosenberger et al. (2004)). The last flight of the day is scheduled for HAM. The duty ends here for an overnight stay lasting 17 hours and 20 minutes until the second duty starts at 8:20 a.m., back to MUC. After a sit time of 2 hours 35 minutes, a continental flight to CDG. After a short sit time, the final flight of both the second duty and the pairing leaves for MUC. Thus, after three cycles to different

[10]Regulation (EC) No 1899/2006 of the European Parliament and the of the Council Subpart Q, retrieved from `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006R1899`, last access: April 24th, 2017.

destinations, the pairing ends at the crew base after sign-off and debriefing within 30 minutes[11].

Regularities and legal requirements for crews are very different from aircraft maintenance and flying time restrictions. This may lead to crews changing the aircraft several times in a duty. If a crew changes its aircraft during a sit connection, they have to leave the aircraft ready for another crew taking over, proceed to the new aircraft and perform additional checks on it before the next departure. In consequence, the minimum sit time must be longer when changing aircraft rather than staying on the current aircraft. Considered the other way around, crews staying on the aircraft are an inevitable instrument for obtaining high utilization levels and keeping sit time duration to a minimum. According to Sandhu and Klabjan (2007), this is the reason why the solution of the previous Aircraft Routing is an input to Crew Pairing in the sequential scheduling process.

In this thesis, the Crew Pairing Problem is paramount for all scheduling-related examinations. The objective of the Crew Pairing Problem is to find a set of crew pairings to operate all scheduled flights at a minimum of deployment costs. This can be achieved by obtaining high utilization levels and minimizing the number of necessary crews. Barnhart et al. (2003) state that Airline Crew Pairing is one of the hardest crew scheduling problems. However, crew costs are one of the main cost drivers for an airline that however is controllable to a large part, see (Anbil et al., 1991, p. 62). The Crew Pairing Problem has therefore gained high attention in scientific studies and is tackled by a variety of OR methods in order to provide cost-efficient schedules. It can be formulated as a mathematical problem similar to the Aircraft Routing Problem. It deals with the coverage of every flight by exactly one crew while ensuring compliance with the complex labor and flying time restrictions for crews. A corresponding set-partitioning formulation is given in Model 2.3.

$$\min \sum_{j \in P} (c_j) x_j \tag{2.3a}$$

$$s.t. \sum_{j \in P} a_{ij} x_j = 1 \qquad \forall i \in F \tag{2.3b}$$

$$x_j \in \{0, 1\} \qquad \forall j \in P \tag{2.3c}$$

[11] A deeper insight into the tasks and obligations of a crew during a crew pairing is provided at the beginning of Chapter 3.

Analogously to Model 2.1, F is the set of flights given by the flight schedule. P is the set of all pairings that meet the crew regulations. Again, a_{ij} has value 1 if pairing $j \in P$ includes flight i, otherwise 0. Constraints 2.3b ensure that each flight is covered exactly once. In case an airline allows *overcovers*, the equality sign can be replaced by $\geq$, potentially resulting in deadhead crews. Constraints 2.3c guarantee integer solutions.

While the selection of pairings can be formulated as an IP, the enumeration of all possible pairings is hardly practicable for large instances. In addition, the generation of feasible pairings itself is not a linear task. Therefore, a common approach is the application of the Dantzig-Wolfe decomposition that is solved by column generation, see Appendix A for details. For a general overview, we refer to the literature surveys of Barnhart et al. (2003), Gopalakrishnan and Johnson (2005) and Kasirzadeh et al. (2015). In more detail, fundamental solution approaches using column generation are described by Desaulniers et al. (1997) and Anbil et al. (1998). For the determination of pairings, Vance et al. (1997) illustrates a two-stage approach that firstly generates feasible duties and afterwards connects them to full pairings. Vance et al. (1997) and Klabjan et al. (2001) both present problem-related branching strategies depend on the question whether a flight i is succeeded by flight i' or not. Such strategies are referred to as follow-on branching. They adapt the fundamental idea of Ryan and Foster (1981) where an integer programming approach and problem-related branching strategies for crew scheduling in public transport is proposed. Rasmussen et al. (2011) introduce the concept of *subsequences* that restrict the set of flights that are considered for the continuation of a duty. More recently, Dück et al. (2011) propose a branch-and-price-and-cut framework that is used as the technical groundwork for scheduling-related tasks in this thesis.

Subsequent to the Crew Pairing, in *Crew Rostering* an assignment of actual staff members to the anonymous pairings is realized. All connections between flights are determined in the crew pairing phase already. Thus, it is an assignment problem, where individual demands and requests of the employees can be taken into account. The objective of cost minimization is no longer the major focus. Mathematical modeling and solution techniques for the crew rostering can be found in Ryan (1992), Lučic and Teodorovic (1999) and Kohl and Karisch (2004). In the context of this work, costs and robustness of a crew schedule are fully determined by decisions made in Crew Pairing. The realization of crew rosters does not influence them and therefore they are not dealt with in further detail at this point.

2.4 Interdependencies between Planning Stages

The preceding presentation of the individual scheduling stages affecting crew and aircraft schedules gives an insight on relevant decisions as well as their time horizon and complexity. In this section, we provide further details on the topic of dependencies between the stages and review approaches for partial integration.

Referring to the traditional way the airline scheduling process is carried out, (Grönkvist, 2006, p. 2919) states that "[...] steps are often done sequentially, without feedback, which makes it crucial to solve each step as well as possible." In addition, presumed optimal decisions in early planning stages may narrow down the degree of freedom in subsequent stages. This leads to suboptimal resource utilization levels at higher planned costs than theoretically possible. First of all, the Schedule Design determines connection possibilities between flights for aircraft and crews as both resources require certain minimum turn times. In Fleet Assignment, the network is split up into sub-networks for which aircraft and crews are scheduled separately. Unless there are flow constraints that guarantee feasible solutions, a different partitioning potentially leads to a larger decrease of operating costs than the expected increase in revenue which is the main objective in Fleet Assignment. Crews require longer minimum sit times between flights when not staying at the same aircraft, therefore the aircraft routing restricts the solution possibilities for the Crew Pairing Problem.

In order to resolve issues like that, recent studies has been carried out that use fully or partial integration of objectives and restrictions from two or more scheduling stages. In the following presentation of these approaches, we focus on the elaborating paths of integration paths for scheduling decisions rather than technical aspects.

2.4.1 Fleet-Assignment-based Integration

First studies on integrative aspects consider the early determination of resource network partitioning and flows in Fleet Assignment. Since in Fleet Assignment many decisions on Aircraft Scheduling must necessarily be anticipated, an integrated consideration appears self-evident. In this regard, Clarke et al. (1996) propose to include aircraft maintenance restrictions and general crew requirements in the fleet assignment approach of Hane et al. (1995). The number of general aircraft maintenance opportunities for checks with a maximum duration of 24 hours are considered rather than individual aircraft checks. Concerning cockpit crews, possible *lonely overnight rests* are balanced with the fleeting costs and revenue. Desaulniers et al. (1997) construct aircraft routes during fleet assignment for daily schedules. Departures are not fixed but have to take place within predefined time windows. The problem is formulated as set partitioning as well as an equivalent multi-commodity-flow problem, solved by column generation and Dantzig-Wolfe decomposition, respectively. Barnhart et al. (1998) propose the consideration of so called flight-strings in fleet assignment. Aircraft routes and maintenance restrictions are considered simultaneously with fleeting, an extension also enables the possibility to balance utilization of operating aircraft. The resulting models are solved by a branch-and-price algorithm.

Rexing et al. (2000) presents an approach for fleet assignment with discrete time windows, enabling possibilities to shift departure times for obtaining additional flight connections. For every flight several arcs are generated in the connection network, representing varying departure times. Additional constraints ensure that only one of the arcs is selected. In the same context, Lohatepanont and Barnhart (2004) discuss integrated and approximate modeling and solution approaches for simultaneous schedule design and fleet assignment.

2.4.2 Integration of Crew and Aircraft Scheduling

Certainly, the most complex integration path is to simultaneously determine scheduling decisions for aircraft and crews. Cordeau et al. (2001) has presented a first fully integrated connection-based model of crew pairing and aircraft routing in a point-to-point flight network. It is solved by a combination of branch-and-price and Benders' decomposition. Klabjan et al. (2002) solve the crew pairing problem in the first place and then construct suitable aircraft routes. Therefore, plane-count constraints are part of the crew pairing model in order to guarantee feasibility of the aircraft routing. As an additional degree of freedom, departure and arrival times can be shifted within certain time windows. In a subsequent publication, Sandhu and Klabjan (2007) extent the model by integrating plane count constraints and crew pairing decisions in fleeting. The model is solved by both column generation and Benders' decomposition.

Cohn and Barnhart (2003) include solutions of the aircraft routing as additional variables in crew pairing and solve the resulting model by column generation. In a comparative study, Mercier et al. (2005) propose a Benders' decomposition based solution approach that promises better solutions. Mercier and Soumis (2007) extend the model by departure time shifts, leading to an additional decrease of costs while still being able to cope with increased complexity. Papadakos (2009) provides a full integration of crew and aircraft scheduling while simplifying crew regularities and cost structures. In a study that primarily focuses on the robustness of crew and aircraft schedules, Weide et al. (2010) iteratively solve the Crew Pairing and Aircraft Scheduling Problem. Convergence is artificially achieved by parameter settings. The approach is adapted by Dunbar et al. (2012) and Dück et al. (2012).

More recently, a new branch-and-price approach for aircraft routing, tail assignment and crew pairing is provided by Ruther et al. (2013). The inevitable increase in complexity is dealt with by obtaining specific solution strategies that reduce the number of pricing problems to be resolved during branch-and-price. This is achieved by either only selecting a subset of pricing problems for solving or using superimposed pricing problems to aggregate sets of original pricing problems that accrue during branch-and-price.

2.5 Summary and Implications

In this chapter, we have provided an overview of the airline planning process. In the long-term, strategic decisions have to be made concerning served O&Ds and the aircraft fleet composition. With market situation analyses and demand forecasts serving as the main input factors, the objective is to consolidate or expand the market position. In tactical planning, flight schedules are constructed in the Schedule Design. Subsequently, the Fleet Assignment deals with assigning an aircraft type to each scheduled flight in a way that the potential revenue is maximized.

In the subsequent operational planning, the focus is on the efficient usage of available resources in order to minimize operating costs for scheduled flights. In the Aircraft Routing, aircraft rotations are constructed separately for every sub-fleet so that maintenance requirements are met. Due to the consideration of flow constraints already in Fleet Assignment, the existence of feasible solutions is ensured. The Crew Scheduling is split up in the construction of anonymous crew itineraries (Pairing Optimization) and the subsequent assignment of specific staff members to these itineraries (Crew Rostering).

The presented planning stages are traditionally tackled by OR-related methods, including mathematical modeling and optimization approaches. In general, the airline planning process is characterized by a high level of computer-aided decision support. A wide range of sophisticated commercial optimization suites and IT solutions has been developed in recent decades from specialized providers such as Amadeus, EDS, Jeppesen, Sabre Airline Solutions, SITA or Lufthansa Systems, see (Bazargan, 2010, p. 208) for details. Commercial decision support systems are characterized by a strong relation to scientific literature. Cooperatively developed approaches are often directly incorporated into commercial software suites, see for example Desaulniers et al. (1997), (Grönkvist, 2005, p. 4) and Borndörfer et al. (2006). In addition, periodic meetings for researchers, airline representatives and software vendors are organized by the AGIFORS[12].

State of the art solution techniques for the individual planning stages are widely exploited, providing solutions of high quality and making further development less beneficial. While the integration of planning decisions has been an issue for a long time already, its influence becomes more and more important, going along with steady developments in computing power. Possible paths of integration typically arise from the dependencies resulting from sequential scheduling. A natural implication is given for Fleet Assignment and Aircraft Routing since the former already must at least rudimentarily take into account actual aircraft rotations. Crew Pairing and Aircraft Scheduling pose similar optimization problems with a different set of rules for crew pairings and aircraft rotations. However, their integration is not straightforward and does not offer substantial simplifications in decision making. Promising advances are achieved by Benders' Decomposition (Mercier et al. (2005)) for comparably smaller instances or iterative approaches as provided by Weide et al. (2010). Schedule Design decisions can basically be incorporated into all subsequent stages, however, at the costs of significant increase of the complexity. This is mostly

[12] Airline Group of the International Federation of Operational Research Society (AGIFORS), `http://www.agifors.org/`, last access: January 9th, 2018.

because deviations in departure and arrival times technically imply additional arcs a flight network, see Mercier and Soumis (2007).

Overall, it is evident that even partial integration promises an increased degree of freedom for scheduling decisions. Inevitably, this is always at the expense of additional model complexity which can mostly be responded by heuristic solution strategies (Ruther et al. (2013)) or relaxations of problem assumptions (Papadakos (2009)). Although integration provides improved results in terms of cost-minimization, systematical differences to the sequential approach do not become apparent.

In this work, we concentrate on the Crew Pairing Problem for all scheduling-related studies, using the optimization framework of Dück et al. (2012). A detailed description of the framework and its extensions is provided in Chapter 5. Further development of optimization techniques and the integration of scheduling stages form the basis of potential future research as addressed in Chapter 13 (Summary & Outlook).

Chapter 3

Operating Flights – Disruptions, Delays and Schedule Recovery

Once the scheduling phase is completed, resource schedules are implemented and submitted to the Airline Operations Control Center (AOCC) and air traffic control authorities, e.g. EUROCONTROL[1] and DFS[2]. They are now responsible for the safe and on-time execution of flights. This is a demanding task, consisting of a complex set of interactions between air traffic authorities, ground operations providers and the actual airline. Therefore, the implementation process starts with a lead time of about two weeks prior to the actual departure, allowing for short-term modifications in response to emerging infeasibility due to resource unavailability or unforeseen external circumstances. During the actual operation, the AOCC has to perform a variety of closely interconnected tasks, comprising aircraft ground handling and maintenance issues, passenger and baggage handling, allocation of airport facilities such as gates and runways, and actually carrying out the flights. The close interconnection of these tasks necessitates coordination and collaboration of employees and technical equipment.

For the investigation of the robust efficiency of resource schedules, the understanding of operational tasks and potential disruptions that may lead to flight delays is essential. In the following, we therefore provide an overview of crew- and aircraft-related procedures in flight and ground operations based on Midkiff et al. (2015)

[1]European Organisation for the Safety of Air Navigation (EUROCONTROL) is an international organization for coordination and control of air traffic in Europe. The headquarters are situated in Brussels. For more information, see `https://www.eurocontrol.int`, last access: January 8$^{\text{th}}$, 2018.

[2]Deutsche Flugsicherung GmbH (DFS), public organization responsible for the control of air traffic in Germany. For more information, see `https://www.dfs.de`, last access: January 8$^{\text{th}}$, 2018.

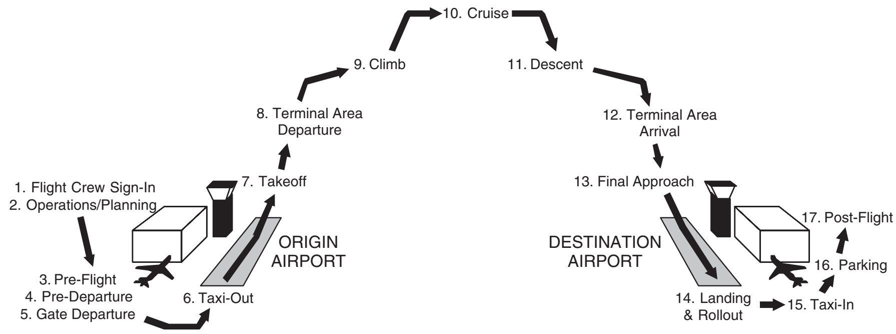

Figure 3.1: 17 phases of a flight (Midkiff et al., 2015, p. 234)

and Dück (2010). It forms the basis for the description of disruptions and delays, possible recovery actions of the AOCC in the subsequent sections.

Midkiff et al. (2015) distinguishes seventeen typical phases of a flight with special regard to the pilot's perspective, see Figure 3.1. In this illustration, Phases 1 to 4 are related to organizational issues. Crews have to sign-in about one hour prior to the first departure of a day (Phase 1). The lead time may differ depending on the aircraft and the flight type (continental or intercontinental). At a first crew meeting, security- and flight-related issues are discussed. Since cockpit and cabin crews are often newly composed for a flight, it is necessary to gain knowledge and agree on collective behavior in special situations. It includes the selection of alternate airports for landing and general rules of conduct, e.g. in case of rejected take-off or landing. In Phase 2 (Operations/Planning), the AOCC ensures awareness of the flight and creates an optimal flight route concerning duration, fuel burn, weather conditions and availability of airspace. The resulting flight plan is communicated to the cockpit crew and submitted to the aircraft automation. Prior to the flight, the crew must check the ability of the aircraft to operate the flight securely (Phase 3). This includes both visual interior and exterior inspections as well as cockpit setup and system checks. The Pre-Departure Phase includes continuously checking and updating fuel consumption, the flight plan and automation parameters since environmental conditions may have changed in the meantime. Relevant actors of the turnaround process (captain, gate agent, ground crew chief) are responsible to meet all requirements for departure. The air traffic control (ATC) gives clearance for the flight route, potentially leading to adaptions and reprogramming of the flight plan. Also, the departure time can be adjusted in case of air traffic congestion. In cases of significant delays in departure, the boarding of passengers may be postponed. In terms of on-time performance, it is desirable to complete the boarding as early as possible so that security-related additional processes, e.g. unloading bags of passengers not turning up, can be performed.

After closing the gate and aircraft doors, external ground processes must soon be finalized. The aircraft is disconnected from external power supplies, gates, boarding bridges and afterwards wheel chocks are removed. After all tasks are finished and the aircraft is safe for movement, the *push-back* from the gate is initiated, generally performed by a tug (Phase 5). Only in rare cases, reverse throttle is used for leaving the gate, a process known as *powerback*. Using a tug for push-back, the aircraft is allowed to use its own engine for movement after receiving clearance from the ground crew. Any ground movement during taxi-out (Phase 6) must be acknowledged by the airport control. Weather or congestion effects may force the usage of a different runway, implying delays due to unexpected taxi duration. At some time, the captain commonly welcomes the flight guests and gives explanations on the today's flight route, the place of arrival and environmental conditions of interest. In addition, taxi and pre-takeoff checks have to be completed and the aircraft enqueues for take-off. Prevailing weather conditions may necessitate de-icing of the aircraft which is up to the captain to decide. De-icing mostly happens close to the runway in order to prevent the aircraft from recurring icing prior to departure in case of taxiway congestion.

In order to guarantee an efficient use of runway times, the aircraft can be positioned at the starting point of the runway even if predecessor or crossing aircraft do not allow take-off yet. Taking off subsequently to a larger aircraft requires some lead time because air turbulences may destabilize the aircraft. Final checks are performed, considering cross winds and constitution of the runway. Close to the take-off, the cockpit crew also informs the cabin crew to be securely seated. For the take-off (Phase 7), a V-speed threshold is determined in advance. If the speed of the aircraft exceeds this threshold, take-off will not be aborted because the remaining runway length is too short to stop the aircraft securely. Closely after the aircraft leaves ground, gear and flaps are retracted. Unless specific ATC restrictions are issued, there are standard procedures for leaving the terminal area (Phase 8) for the coordination of departing and arriving aircraft. The aircraft climbs (Phase 9) and adjusts the flight direction in order to reach its flight route. Air traffic authorities control departing and arriving aircraft in order to ensure secure operations whilst providing efficient flow rates.

After the determined altitude is reached, the cruise begins (Phase 10). Even if the autopilot is activated, the cockpit crew must continuously check fuel consumption level and weather conditions with reference to the expected arrival time. Alternate airports for emergency landings must be updated depending on the current position. The actual flying time can differ from its scheduled equivalent in case of headwinds. In some cases, the flight route and altitude can be recomputed if it has positive effects in terms of winds and resulting air resistance. Crossing boarders often goes along with changing ATC authorities requiring different communication equipment and procedures.

Approaching the destination airport, the descent may start at up to 100 miles before the runway, depending on the aircraft type and flight distance (Phase 11). The local ATC is responsible to assign a secure and fuel-efficient route for the approach in the terminal area (Phase 12). The crew has to process real-time information on weather and runway conditions for the approach. The cabin crew meanwhile ensures security of the passengers who are also informed on potential arrival delays and resulting effects on connecting flights. The final approach (Phase 13) is supported by a variety of navigation and control systems, e.g. the Instrument Landing System (ILS) and must be cleared by the ATC in order to keep sufficient distances between consecutive aircraft. After touch down on the runway, brake flaps and wheel brakes are used to decelerate the aircraft, often supported by reverse throttle (Phase 14). Depending on the airport situation, it is desired to leave the runway as early as possible. Afterwards, the aircraft is guided to the arrival gate on taxiways (Phase 15). If the gate is still occupied and no alternative gates are available, the aircraft must wait at a temporary parking position for clearance. Once the gate is reached and brake pads are positioned (Phase 16), external power supply is provided and boarding bridges are connected for passengers to disembark.

The crew's post-flight debriefing consists of reports on possible non-standard situations associated to e.g. emergencies, mechanical failures or ATC violations (Phase

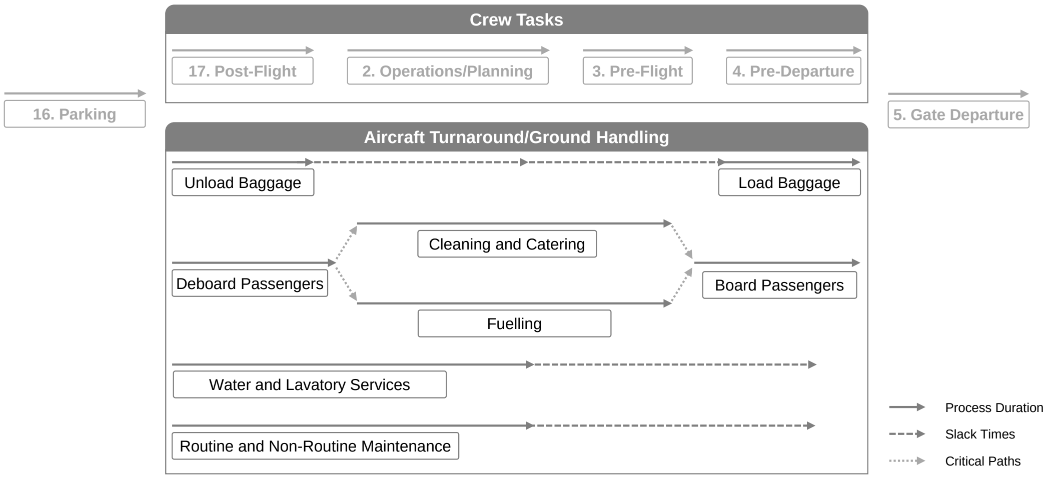

Figure 3.2: The aircraft turnaround process, based on (Dück, 2010, p. 17)

17). In case of an aircraft change, it is preferred that the inbound crew meets the outbound crew before proceeding to the new gate for taking over another aircraft. If a daily crew duty ends, transportation to a crew hotel must be ensured if they stay in a city different to the crew base.

If the next flight of the aircraft starts soon, the aircraft needs to be *turned around* in a short period of time. This is also referred to as ground handling. The set of interdependent sequential and parallel tasks of a generalized turnaround is illustrated in Figure 3.2. The depicted crew tasks refer to the case that the crew does not start a new duty and therefore does not need to sign-in. Moreover, the case is depicted when the aircraft turnaround must be performed in the least time possible, i.e. the connection does not contain explicit buffer times.

Soon after the aircraft arrives at the gate, loader vehicles approach the aircraft for unloading baggage. The subsequent transportation is carried out by dollies or trucks. At the same time, boarding bridges or stairways are connected to the doors so that passenger deboarding can begin. In the meantime, water and lavatory service trucks empty and refill tanks of used and unused water. Fueling is generally performed between boarding and deboarding. With passengers on board, fueling is only permitted in specific cases and under additional security precautions. The same holds for catering and cleaning which is also performed between deboarding and boarding. According to Fricke and Schultz (2009), the chronological order of these processes result in the two main critical paths during the turnaround: Delays during deboarding, catering/cleaning or fueling may affect the timely completion of following tasks and therefore have a greater impact on the overall turnaround duration. In contrast, there are implicit slack times for baggage handling, water and lavatory services that allow certain deviations from the planned time schedule without affecting the subsequent flight departure. If the connection time between the flights is larger than the minimum time for the turnaround, critical paths are segregated due to additional buffer times.

During the whole turnaround process, routine and non-routine maintenance can be performed, especially if damages have emerged during the last flight. Depending on the severeness of damage, the departure of the upcoming flight may be delayed or even the aircraft is disallowed to continue operations. In this case, the AOCC is responsible to recover schedules, e.g. by rerouting or checking availability of spare aircraft. Although checks on the aircraft are generally performed before departure, exterior visual post-flight checks may apply if the aircraft is not turned around immediately. This ensures that fixing obvious damages can be scheduled as early as possible, especially at overnight stays.

The preceding presentation offers an insight into the many sub-processes of flight and ground operations, leading to inevitable unforeseen deviations from original schedules. Therefore, we discuss the topic of disruptions and delays in operations in the following Section 3.1. Operational response actions are carved out in Section 3.2, focusing on the topic of schedule recovery. Based on these considerations, the final Section 3.3 deals with the introduction of the concept of robust efficiency, aiming at the adaption of operational objectives in the scheduling phase.

3.1 Disruptions and Delays in Operations

Besides the inevitable complexity, the AOCC has to cope with unpredictable environmental influences to which all stages of a flight are exposed. These may be bad weather conditions, airport and airspace congestion, equipment failure and technical breakdowns, or missing passengers, to name only a few of them. The International Air Transport Association (IATA), the world's largest umbrella organization of air-

Codes	Category
00-09	Internal
11-19	Passenger/Baggage
21-29	Cargo/Mail
31-39	Handling
41-48	Technical
51-58	Damage/Failure
61-69	Operation
71-77	Weather
81-89	Air Traffic Control
91-99	Reactionary/Miscellaneous

Table 3.1: Categories of the IATA Delay Codes

lines, has created a set of IATA Delay Codes[3] in order to standardize data recording and reporting. Delays are split into nine categories, based on their cause. These categories are presented in Table 3.1.

In our context of crew and aircraft scheduling, an important distinction is drawn between *primary* and *secondary* delays. The former are caused by exogenous disruptions and cannot be prevented by scheduling decisions. They are therefore often also called *exogenous* delays. Typical examples are weather conditions (codes 71-77), aircraft defects (41) or air traffic capacity issues (81). In contrast, the latter is induced by delayed arrival of previous tasks using at least one common resource and therefore can be influenced by scheduling decisions. Secondary delays are also referred to as *reactionary* or *propagated* delays. Note that the presented differentiation is highly context-sensitive and we adhere to a definition directly related to crew and aircraft scheduling. An eminent example is the postponed departure of a flight due to the late arrival of the aircraft (code 93). In this case, the buffer time between the two flights is not sufficient to absorb the delay. The same holds for cabin crews (code 94) and cockpit crews (95) if a late arrival impedes a timely transfer to the next aircraft. These cases lead to the so called effect of *delay propagation*. In particular, cost-optimized schedules are characterized by a high utilization factor and following tight connections and numerous interactions between resources such as crews and aircraft. Consequently, even slight disruptions may lead to cascading delay effects between different resource layer networks.

[3] See the full list in Appendix B.

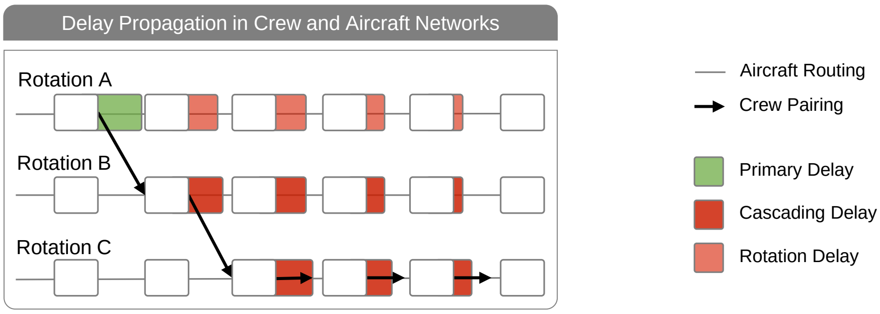

Figure 3.3: Cascading delay propagation in multi-layer resource networks

In this regard, Figure 3.3 exemplifies the way how a single primary delay (green mark) may cause multiple secondary delays by affecting various aircraft routings and crew pairings as a result of insufficient buffer times before it can be fully absorbed. On the one hand, subsequent flights Rotation A are affected due to the usage of the same aircraft (light red marks). It is therefore referred to as *rotation delay*. Additionally, the aircraft changes of the crew (black arrow) lead to further cascading propagation to Rotation B and subsequently to Rotation C (red marks).

According to the Central Office for Delay Analysis (CODA) of EUROCONTROL, "[secondary] delays have an increasing influence on the financial and operational results of airlines in Europe" (CODA, 2011, p. 5). Referring to this, one minute of primary delay induced 0.54 minutes (32 seconds) of secondary delay in 2005. This value increased significantly to 0.83 (50 seconds) in 2008, see (CODA, 2011, pp.5) for details. The most recent CODA report for the year 2016 reports an annual average flight departure delay of 11.3 minutes, ranging from a low of 8 minutes in February to 16 Minutes in July, see (CODA, 2017, p. 1). The annual average includes 5.1 minutes of secondary delay, resulting in an average proportion of 0.82 (50 seconds). Taking a closer look on a daily basis, the share of secondary delays constantly increases over the course of the day (CODA, 2017, p. 16). Secondary delays also drive the on-time performance (OTP), one main indicator for the reliability and customer satisfaction of an airline. It describes the fraction of flights arriving on-time. The inclusion of an additional delay tolerance threshold, e.g. 0, 5 and 15 minutes is possible. For airline operations the threshold of 15 minutes is the most important one, see Burke et al. (2010) or Shebalov and Klabjan (2006) for details. According to (CODA, 2017, p.

10), 79.9% of all flights departed within the 15 minutes threshold in 2016, down from 81.8 in 2015.

For airlines, there exists a range of various instruments to tackle delays. In this regard, Sodi (2011) distinguishes three approaches that are transferred to the context of this thesis:

- **The reactive method** comprises actions in response to events that already happened. In these cases, the AOCC reacts to disruptions in real time using a set of available recovery actions.
- **The proactive method** aims at the active prevention of potential delay risks by constantly monitoring and analyzing operational activities. Events that are likely to disrupt operations are identified based on past events or trends. At best, emerging disruptions and delays can be resolved in advance.
- **The predictive method** reviews and identifies the potential for incidents that may affect future operations. Findings from the analysis of historical data and operational expertise are used to anticipate potential risks and make schedules more resilient against the impact of disruptions.

In our context, actions attributed to the reactive and proactive methods are mainly carried out by the AOCC in response to or expectation of schedule disruptions. Essential details on recovery-related issues are presented in the following Section 3.2. In contrast, the predictive method is related to long-term activities that aim at constantly improving the adaption of schedules to potential disruption risks. It is therefore situated in the scheduling phase and addresses the *schedule robustness*. The main objective of this work is related to the predictive method. However, since scheduling and operations are inevitably and closely interconnected, it is desirable that predictive approaches already involve the anticipation of recovery procedures.

3.2 Recovery of Airline Resource Schedules

In case disruptions lead to infeasibility of the originally planned resource schedules, the AOCC is responsible for initiating actions to return to regular operations. The process of finding new schedules is called *recovery* or *disruption management* and comprises a variety of resources, e.g. airport gate assignments, departure and arrival slots, ATC clearance, passenger connections as well as aircraft and crew schedules. Traditionally, there are individual responsibilities of staff members for specific resources that must negotiate and coordinate their activities. In addition, decisions must be communicated and coordinated with airport and ATC authorities. Altogether, these aspects make the recovery a complex and time-critical task that usually has to be carried out in real-time.

In accordance with the scope of this work, we focus on the recovery of crew and aircraft schedules. Several instruments are at the AOCC's disposal for responding to disruptions and emerging delays. The following list is based on the survey of Kohl et al. (2007):

Delay subsequent flights The straightforward way to recover infeasible schedules is to postpone subsequent flights until all necessary resources are available. As a result, reactionary delays are induced. Rosenberger et al. (2002) refer to this strategy as *push-back*.

Increased cruise speed In case of late departure, the cruise speed can be increased. However, the benefits has to be set into relation with the associated higher fuel consumption. A recent calculation model is provided in Marla et al. (2016).

Swaps Especially in hub-and-spoke networks it is likely that several aircraft of the same type are situated at the same airport and at the same time. This allows the reallocation of buffer times by swapping the assignment of aircraft or crews. Crew regulations and maintenance restrictions must be complied when performing a swap. For more details on the application of swaps, we refer to Rosenberger et al. (2002).

Reserve crews A common way to react to severe disruptions is the usage of reserve resources. These are usually located at a hub airport and guarantee a short response time. Rosenberger et al. (2000) and Bratu and Barnhart (2006) address the usage of reserve crews for recovery.

Cancellations Flights may be canceled as a result of severe weather conditions, equipment failure, extensive delays or industrial actions, to name only a few. If a flight has to be canceled, subsequent flights are inevitably affected. In this matter, Rosenberger et al. (2004) define cancellation cycles as a set of flights that has to be canceled in order to continue operations from the point where resources are available. Cancellations are usually the last option for an airline since they entail costly passenger rerouting and potential repositioning of crews and aircraft. According to CODA (2017), operational cancellations have increased by 0.1% to 1.6% in 2016.

Ferry Flights Ferrying describe the repositioning of empty aircraft to a specific location where it continues its service. Since fuel costs are not compensated by any revenue, ferry flights are only used in rare cases, e.g. after industrial actions, environmental impacts or unscheduled maintenance.

While monitoring operations as well as the selection and implementation of recovery instruments have been carried out manually for decades, decision support tools increasingly gain in importance. Analogously to the airline planning process, a close interconnection between scientific literature and the airline practice can be observed: The availability of real-time information, their visualization via Gantt diagrams as well as the usage of data warehouses for all relevant data have become a crucial and indispensable part of everyday work, see (Kohl et al., 2007, p. 152). In related scientific literature, the representation of real-world recovery policies is an elementary part for the evaluation of resource schedules. They are mostly incorporated into frameworks for the stochastic simulation of airline operations: For aircraft schedule simulation, Burke et al. (2010) use a "[. . .] heuristic approach that is based on the current practices at KLM's operations control department. The recovery strategies include swapping aircraft, canceling flights and accepting delays. The crew schedule is not explicitly taken into account in the solution model." The *SimAir* framework makes use of delay propagation, swapping, cancellation, passenger rerouting, see Rosenberger et al. (2000). Different recovery policies can be implemented in order to reflect real-world processes of the AOCC. According to (Schaefer et al., 2005, p. 343), SimAir has been used by major US carriers and airline software firms. Among others, it is applied by Rosenberger et al. (2004), Schaefer et al. (2005) and Shebalov and Klabjan (2006) for the evaluation of scheduling approaches with scientific background. Concerning this matter, the discussion of stochastic simulation of airline operations in the context of robust scheduling is continued in Chapter 4.

Besides the decision support for adapted manual recovery, the provision of sophisticated recovery solutions based on mathematical optimization has evolved in the last two decades. Initial optimization-based approaches in Lettovský et al. (2000), Yu et al. (2003) and Abdelghany et al. (2004) for crew recovery are highly related to techniques used in the crew scheduling stage. For this reason, they are often referred to as *rescheduling* approaches. The limitation of the scheduling horizon to a specific recovery time window as well as the incorporation of actual crew positions and availabilities are essential adaptions. Rosenberger et al. (2003) provides an heuristic approach for aircraft recovery that minimizes rerouting and cancellation costs. Since recovery solutions for aircraft do not necessarily maintain crew and passenger connections, a revised model is presented that additionally takes them into account for the assessment of obtained solutions.

In this regard, the focus has increasingly moved towards integrative decision making aiming at the simultaneous provision of recovery solutions for multiple resources. Eggenberg et al. (2010a) present a general recovery network that allows the usage of operations-specific constraints in a column generation framework. The benefit is demonstrated for aircraft recovery with special attention on the feasibility of maintenance events. Petersen et al. (2012) propose an optimization approach consisting of flight retiming as well as aircraft, crew and passenger recovery. For a deeper insight on optimization-based rescheduling, we refer to the comprehensive study in Clausen et al. (2010).

In the scope of this work, recovery-related actions are a necessary instrument for the assessment of different robust scheduling strategies for regular daily operations in terms of measures for potential reactionary costs. Highly competitive rescheduling approaches may promise best-possible results in terms of on-time performance or reliability because schedules can be adapted to current circumstances by changing them to a large extent. However, as consequence they are no appropriate instrument for the assessment of scheduling strategies. In comparison, adapted manual recovery actions that make use of certain schedule characteristics appear to be more suitable. Based on these preliminary considerations, the selection of specific recovery instruments for the evaluation of robust scheduling strategies is further discussed in Chapter 5 (Required Work).

3.3 The Concept of Robust Efficiency

In Chapter 2, cost minimization has turned out to be the main objective in traditional airline resource scheduling. It preferably results in schedules that are characterized by the efficient usage of available capacities at preferably high utilization levels. Since idle times of crews and aircraft are costly, the ground times between flights tend towards a minimum in cost-optimized schedules. We refer to the costs of a schedule that can be operated as planned as *nominal costs*.

In contrast, the major challenge during operations is to maintain punctuality and reliability of schedules as well as passenger connectivity. Buffer times for delay absorption and possibilities to easily adapt schedules to current circumstances are important for the achievement of these goals. For this reason, pure cost-optimized schedules with minimum idle times imply higher delay propagation risks and tend to operate worse. According to (Tam et al., 2011, p. 52), "[...] schedules happen to become 'de-optimised' in actual operation, as they are easily disrupted and chain impacts are usually found as a result." In consequence, unforeseen disruptions may lead to additional, so called *reactionary costs* for resource reassignment, schedule recovery or passenger rerouting. The *real costs* actually incurring for an airline to operate its schedules can therefore substantially differ from nominal costs. The apparent trade-off between nominal and reactionary costs is commonly known in the field of crew and aircraft scheduling, see for example Ehrgott and Ryan (2002), Schaefer et al. (2005) and Yen and Birge (2006).

Based on these considerations, Figure 3.4 provides a posteriori view on the composition of real costs for hypothetical schedules with different utilization levels, i.e. delay occurrences are completely known. Note that in the context of crew and aircraft scheduling, reactionary costs resulting from exogenous delays must be assumed to be unavoidable by definition. They can therefore be assumed as fixed costs and are not considered for assessment of schedules. In contrast, propagated delays are subject to the influence of scheduling decisions. Referring to this, the far left of the diagram represents schedules with lowest utilization levels. Reactionary costs emerging from propagated delays are theoretically non-existent. Schedules with increasing utilization tend to induce a growth in reactionary costs. Finally, the far right of the diagram represent cost-optimized schedules that are assumed to be highly susceptible to disruptions, leading to potentially high reactionary costs emerging from propa-

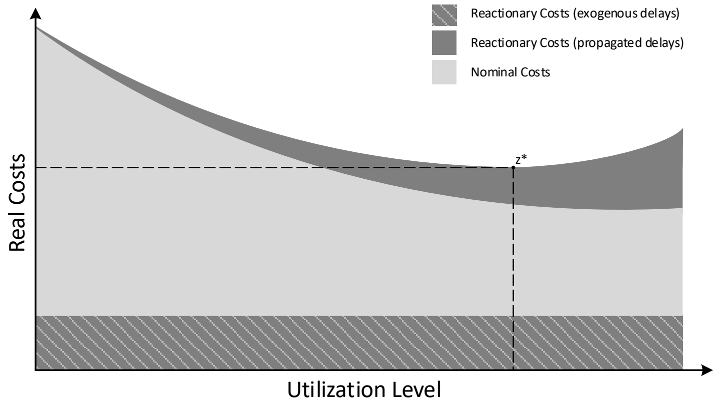

Figure 3.4: Real costs as the sum of nominal and reactionary costs

gated delays. The sole usage of nominal costs for assessing a schedule's performance underestimates the real costs. However, the main issue in this regard has been well stated by Weide (2009):

> "It is easy to measure the performance of the schedule – once it has been operated [. . .]."

In the depicted scenario, all delay occurrences are known and the optimal trade-off between nominal and reactionary costs can easily be determined, indicated by z^* in Figure 3.4. Since the scheduling phase starts months before actual operations, this is not possible in reality. In contrast, stochastic assumptions on reactionary costs have to be factored in the assessment of schedules. This issue is addressed by the concept of *robust efficiency*, aiming at the minimization of real costs by already taking into account probable reactionary costs during the scheduling stage. Besides cost-efficiency, the additional objective of robustness is considered. The term *robustness* has its origin in the Latin word *robustus* which originally means *made of oak* and has gained the meaning of hard, strong and solid over time. According to Scholl (2001), the abstract concept of robustness describes the potential of a system to maintain its function independently from exogenous conditions.

There are two eligible properties of resource schedules that influence the robustness, namely *stability* and *flexibility*. Stability describes the ability of a system to work properly without changes and adjustments in case of disruptions. It therefore has to remain feasible under changing operational environments. In the airline scheduling context, stability is the capability of schedules to absorb delays so that no or less delay propagation occurs. The main instrument for increasing stability is the incorporation or reallocation of buffer times between flights that use the same resource[4]. In contrast, flexibility is the ability of a schedule to be adapted to changing environments by manageable and cost-neutral recovery actions. For schedules with a high degree of flexibility the feasibility can be restored by local and restricted interventions during operations, e.g. by swaps of resources. Both stability and flexibility aim at improving the operational performance in regular daily operations rather than immense disruptions such as temporary airport closures or industrial actions. In these cases, schedules are highly distorted or even suspended and their initial properties have no effect.

Note that in this work, we focus on the amount of *expected propagated delay* and certain derivative measures as main indicators for reactionary costs rather than dealing with specific monetary conversions. There are many calculation models for translating propagated delays to reactionary costs with varying assumption, see Ball et al. (2010), Cook and Tanner (2011) or Ferguson et al. (2013). However, in the scope of this work it is hard to determine incontestable models for reactionary costs. The usage of propagated delay as the main indicator deliberately prevents additional bias resulting from the adaption of specific assumptions whose suitability is hard to determine in a theoretical context. Based on profound contextual knowledge, specific calculation models can straightforwardly be applied to our findings if a monetary assessment is desired.

The trade-off between nominal costs and expected propagated delay results in Pareto optimal solutions. Figure 3.5 illustrates the difference between pure cost-optimization and multi-objective optimization. Line (I) illustrates the single objective optimization targeting at nominal cost optimality. It results in one explicit solution that is referred to as the *cost-efficient solution* in the scope of our studies. The 100% marks on both axes represents the standardized nominal costs and amount of expected delay propagation for this solution. For multi-objective optimization, a

[4]Note that often the terms robustness and stability are used synonymously, see Liebchen et al. (2009), Eggenberg et al. (2010b) or Froyland et al. (2014). In contrast, Burke et al. (2010) refer to our definition of stability as reliability.

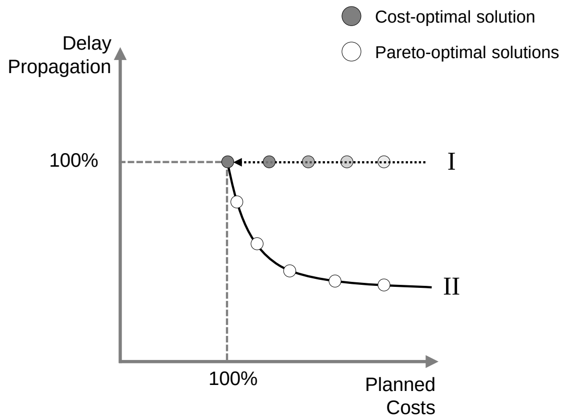

Figure 3.5: Pareto-optimality for cost-efficient and delay tolerant schedules

Pareto front (II) can be constructed from non-dominated solutions, i.e. no other solution provides better results in both nominal costs and the amount of delay propagation at the same time[5]. The final choice of a solution depends on the risk aversion level of the decision maker.

3.4 Summary and Implications

In this chapter, we have presented responsibilities and emerging challenges in airline operations with a focus on crew- and aircraft-related tasks. Based on a generalized description of ground and airborne processes, a context-sensitive distinction between (exogenous) primary and (propagated) secondary delays has been introduced. The distinction is important for the assessment of schedule robustness since only secondary delays can be influenced by scheduling decisions while primary delays must be regarded as exogenous input factors.

For tackling delays, there are three main concepts, namely the reactive, proactive and predictive approach. The former two are usually carried out by the AOCC, continuously monitoring operations and making interventions if disruptions lead to

[5]In a related research, Amberg (2017) examine the effects of sequential, partially integrated and integrated vehicle and crew scheduling for urban public bus transportation. Especially by the latter, the robustness can be significantly improved without increasing the nominal costs of schedules.

infeasible schedules. The AOCC can make use of a number of instruments for schedule recovery like delay propagation, resource swaps, reserve resources, cancellations or ferry flights. Traditional manual recovery evolved in recent decades as a result of computational decision support. In this context, an exemplary joint work of researchers and airline practitioners has led to the development of the airline operation simulation framework SimAir (Rosenberger et al. (2002)). Besides real-time information presentation and processing of recovery possibilities for decision support, there are highly competitive rescheduling approaches that increasingly consider integrated decisions for multiple resources and provide sophisticated solutions for complex recovery tasks. In particular, these approaches appear beneficial in case of severe disruptions such has partial or full hub closures, see Petersen et al. (2012). However, (Clausen et al., 2010, p. 820) state that "[...] the airlines' requirements for recovery decision support systems are still substantially different from the services offered by commercial tools and from most of the prototype tools proposed in the literature". The main reason is that recovery approaches are often derived from scheduling approaches with particular alterations and therefore do not sufficiently reflect the operational reality.

In contrast, the predictive approach is equivalent to the consideration of schedule robustness during the scheduling phase. By doing so, it is possible to minimize the complexity of recovery in regular daily operations. It is consistent with (Clausen et al., 2010, p. 820) who also state that "[r]obustness can be seen as the [...] counterpart of recovery, and we believe that these two concepts will be central in the process of minimizing the effect of disruptions on the daily operation of airline companies." In this regard, the concept of robust efficiency aims at the a more realistic assessment of the real schedule costs by taking into account nominal and reactionary costs. Concerning crew and aircraft schedules, the main driver for reactionary costs are propagated delays. They can be influenced by improving the stability or flexibility of schedules. While stability aims at the execution of schedules 'as planned', an increased degree of flexibility offers additional opportunities to react to disruptions in case of unforeseen disruptions. Concerning the latter, we focus on the incorporation of swaps into schedules as cost-effective, elementary and simple recovery options. The state-of-the-art in related scientific literature on robust scheduling is presented and discussed in the following chapter.

Chapter 4

State of the Art in Airline Resource Scheduling and Operations under Disruptions – Fundamental Techniques and Recent Advances

In the previous chapter, the chronological order of scheduling process has been presented along with emerging decision problems for every planning stage as well as mutual influences between them. The emerging need to join the contrary objectives of airline scheduling and operations has led the concept of robust efficiency, dealing with the consideration of operational requirements already during scheduling. In this chapter, we provide a review of state-of-the-art approaches from scientific literature that aim at improving the robustness of crew and aircraft schedules.

Mathematical modeling and optimization approaches that simultaneously consider the objectives of cost-efficiency and robustness are presented in Section 4.1. Since delay prediction is an important part of recent stochastic approaches, Section 4.2 is dedicated to this issue. Special attention is drawn to the usage of historical delay data for primary delay prediction modeling. For both topics, a summary and implications are provided at the end of respective sections.

4.1 Incorporating Robustness in Airline Resource Schedules

In order to provide a holistic view on the topic, scheduling techniques and robustness indicators are carved out for every planning stage in which crew and aircraft schedules are affected.

4.1.1 Time Windows in Schedule Design

Concerning the scheduling policies of cost-efficiency and robustness, a high degree of freedom for resource scheduling decisions is a crucial condition for good solutions. Since flight schedules are commonly generated prior to crew and aircraft scheduling, many connection possibilities are impossible although they are preferable in terms of cost-efficiency. However, the degree of freedom can be increased by varying departure and arrival times of a predefined flight schedule within certain time windows. Fundamental techniques and restrictions related to the Crew Pairing and Aircraft Routing are described by Klabjan et al. (2002).

With the objective of minimizing passenger delays and disruptions, both a sequential and an integrated model for robust aircraft routing are proposed by Sarmadi (2004). Aircraft routings for one fleet are constructed under consideration of the ability to shift flight departure times. For a previously generated set of feasible aircraft routings the total amount propagated delay is computed for each of routing. Model parameters are taken from historical delay data – for each flight the departure and arrival delay is calculated as the average over all days the flight is operated.

Wu (2006) proposes the reallocation of buffer times for existing aircraft routings. The goal is to achieve a minimum schedule on-time performance. Based on real-world schedule and delay data, scheduled flights are shifted based on historical delay occurrences while maintaining the order of flights in each aircraft routing.

AhmadBeygi et al. (2008) deal with potential delay propagation by passenger itineraries in the flight network. They introduce the concept of propagation trees, containing all resource connections that may propagate delays to subsequent flights. Based on this, Ahmadbeygi et al. (2010) re-allocate slack times in crew and aircraft schedules by retiming flights in order to minimize delay propagation. Historical delay data of a twelve month time span is used to generate empirical distributions, differentiated by the respective departure airport.

Besides the usage of time-windows in subsequent scheduling stages, Sohoni et al. (2011) propose a stochastic optimization model for the determination of block times, incorporating uncertainties of their duration. They aim at maximizing flight schedule profits while maintaining the on-time performance at a specific level. Improved actual block time predictions can implicitly reduce propagation of en-route delays.

4.1.2 Fleet Assignment

Decisions made in fleet assignment not only determine the possibilities to minimize planned costs, but also affect robustness potentials of crew and aircraft schedules. Delay propagation and recovery possibilities cannot be addressed directly but both aspects are predetermined by the subnetworks resulting from the Fleet Assignment. In this context, Rosenberger et al. (2004) present an approach in which hubs are isolated regarding aircraft rotations and crew itineraries. Commonly, hubs are large airports where many interacting processes are carried out, implying higher delay propagation risks. If a network contains several hubs, it is preferable that different hubs operate as separately as possible. Therefore, the *hub connectivity* by aircraft, crews or passengers shall be kept to a minimum. In this way, hubs are less or not affected by disruptions at other hubs. However, hub isolation leads to increased planned costs for aircraft schedules. In addition, so called *short cycles* used for crews and aircraft. They only operate a small number of flights until returning to the hub, closely related to the *out-and-back* principle. It leads to a reduced number of cancellations in case of disruptions.

Smith and Johnson (2006) introduce the concept of *station purity* by restricting the number of fleet types that serve each airport. They follow the idea that an increased frequency of single fleets at an airport increases the degree of freedom and therefore obtain the possibility for more robust solutions for subsequent crew and aircraft scheduling. However, the resulting models are hard to solve, making it necessary to use a station decomposition solution approach tailored towards the given flight network structure.

4.1.3 Aircraft Routing

Concerning the robustness, the Aircraft Routing has to deal with improving the stability or flexibility of aircraft schedules. For the former, sufficient buffer times between flights that are successively operated by the same aircraft have to be determined. The latter deals with explicitly incorporating recovery possibilities into schedules for the event of disruptions in operations.

Lan et al. (2006) improves the aircraft schedule stability by minimizing the expected total propagated delay for a given number of aircraft. Real-world delay data is used to estimate delay propagation between two consecutive flights. Subsequent propagation on following flights is not taken into account. Delays for flight connections that do not exist in historical data are estimated by the total arrival delay and the ground time before the aircraft departs again.

Borndörfer et al. (2010) use a column generation approach to construct daily aircraft tail assignments with reduced delay propagation risk. The authors distinguish between block (taxi and en-route) and gate (ground operations) times that can be delayed independently from each other. The resulting stochastic optimization model is based on Grönkvist (2005). The objective is to find a set of rotations with minimal delay propagation probabilities. Based on real world delay data of an European short haul carrier, distributions for length and occurrence probability of primary delays for both gate and bock times are fitted. Gate delay length is assumed to be independent from spatio-temporal factors as airport and time of day. Gate delay occurrence probabilities are modeled as log-normal distributions. Block delays are related to scheduled block length and are modeled by a log-logistic distribution.

Lapp and Cohn (2012) consider the robustness of maintenance checks in daily aircraft schedules, aiming at increased feasibility of maintenance schedules. For this purpose, they develop a maintenance reachability index that is used as an objective in an optimization model.

Froyland et al. (2014) transfer the concept of *recoverable robustness* (see Liebchen et al. (2009)) to the robust tail assignment problem. The objective is to construct aircraft routes that can be recovered with limited effort. They address the concept of *proxy robustness*. The schedule process and structures are evaluated concerning aspects that may improve the delay tolerance during operations. The problem of proxy robust solutions is the low connection between scheduling and operations. Furthermore, *feedback robustness* describes the incorporation of operation-based measuring of delay tolerance. A one day aircraft tail assignment problem addresses costs for delays, cancellations and also swapping opportunities. The objective is to provide a solution promising a minimum of recovery. It is solved by a Benders' decomposition. The approach is compared to an exemplary proxy robust approach which is an extension of Grönkvist (2005) that penalizes flight connections regarding their actual ground time. Maher et al. (2014) additionally consider maintenance restrictions for large-scale instances.

Ageeva (2000) propose an aircraft routing model including flexibility measures. Based on the flight string model of Barnhart et al. (1998) aircraft rotations that meet as often as possible are generated. An important restriction is that an aircraft must be swapped back in order to return to the original route after a swap is performed. In the corresponding solution approach several solutions are generated consecutively. Afterwards the solution is selected where flight-strings meet each other as often as possible in order to provide a maximum number of swap opportunities.

Burke et al. (2010) construct a multi-objective optimization model with time windows to both incorporate stability and flexibility into aircraft schedules while maintaining the given fleet assignment. In contrast to Ageeva (2000), they consider *single point swaps* without back swaps. Also, no interdependencies between crew and aircraft are considered. The authors derive the probability of delayed departures by convolution from historical delay data of a major European carrier: Flying and departure handling times are modeled as Gamma distributions, arrival handling times by deterministic distributions, differentiated by attributes concerning time of day, time of the year, fleet type, and O&Ds. Network effects of delay propagation are not taken into account. Based on a simulation study on generated schedules, it is constituted that stability is dominant to flexibility concerning its influence on the schedule robustness.

4.1.4 Crew Pairing

The Crew Pairing problem is the most discussed airline scheduling problem which also holds in terms of robustness. In this section, we give a survey on techniques on how to incorporate robustness into crew schedules.

Ehrgott and Ryan (2002) propose a bi-criteria approach for solving the robust crew pairing problem by considering stability aspects of crews changing aircraft. They use a deterministic local indicator-based approach to check aircraft changes of crews for possible delay propagation between the two successive flights within a particular aircraft change. However, the propagation is measured locally; no further propagation in the affected pairing is considered.

Mercier et al. (2005) introduce a robustness indicator for the integrated crew and aircraft scheduling problem that is based on ground times for crews changing aircraft between two flights. These connections are called *restricted* if the ground time is lower than a certain threshold. Schedules with less restricted aircraft changes are assumed to be more robust. However, no network effects, especially further delay propagation due to crew pairings and aircraft rotations, are taken into account. Weide et al. (2010) use the same local non-robustness indicator for iterative crew scheduling and aircraft routing.

Schaefer et al. (2005) present an approach to incorporate stability into crew schedules by using expected cost rather than planned cost for crew pairings. The expected costs are computed by simulation of standalone pairings. Thus, no delay propagation between pairings, e.g. by sharing the same aircraft consecutively, are considered.

Yen and Birge (2006) formulate the crew pairing problem with a non-linear stochastic recourse to measure probable delay propagation between both crews and aircraft. The heuristic solution approach is computational expensive, thus only results for small instances with up to 79 flights are presented. Tam et al. (2011) compare the multi-objective approach of Ehrgott and Ryan (2002) and the stochastic programming approach of Yen and Birge (2006) in terms of schedule characteristics and robustness of the solutions. It turns out that the stochastic approach leads to better solutions when robustness is the major objective. In contrast, the deterministic robustness indicator of Ehrgott and Ryan (2002) allows to cope with delays at low planned cost increase.

Gao et al. (2009) discuss the influence of decisions taken in fleet assignment on crew scheduling. The authors extend the *station purity* concept of Smith and Johnson (2006) to crew bases, i.e. airports are served by a restricted number of fleets and crew bases. Note that only crew connections are considered rather than explicit crew pairings.

Weide et al. (2010) propose an iterative scheduling approach for robust crew and aircraft scheduling. For the evaluation of robustness, the concept of restricted aircraft changes of Mercier et al. (2005) is used. Again, the evaluation of robustness is only performed locally for aircraft changes.

Dück et al. (2012) propose a stochastic recourse model for delay propagation estimation in simultaneous crew and aircraft scheduling. The recourse model bases on delay scenarios that are generated by stochastic distributions derived from historical delay data of a major European carrier. Due to non-linear constraints, the model is not tractable for large real-world instances. Therefore a decomposed model is derived, enabling the application of an iterative solution approach, see Weide et al. (2010). In detail, during the generation of crew pairings propagation effects are measured with regard to the rotation schedule, and vice versa. The results indicate that the stochastic approach provides solutions comparable to Mercier et al. (2005). However, the advantage of the stochastic approach is the ability of self-calibration by the underlying delay scenarios.

In a comparable study, Dunbar et al. (2012) propose a refined delay propagation estimation based on the work of Wu (2006) and Lan et al. (2006). Initial primary delays are taken from historical data depending on flight departure times. The concept is tested within an iterative crew and aircraft scheduling framework. The work of Dunbar et al. (2014) additionally discusses the incorporation of delay scenarios and the influence of flight re-timing abilities.

In the scope of public bus transportation, Amberg et al. (2018) investigate the effect of interactions between vehicle and crew schedules on the robust efficiency. It turns out that integrated vehicle and crew scheduling is superior to sequential and partial integrated scheduling concerning planned cost efficiency and robustness. In particular, the integrated approach is capable to improve the robustness to a significant extent without affecting the planned costs. As a key result of the related work of Amberg (2017), the density of crew schedules is determined as the main influential factor for robustness. In this context, the density is defined as the level of task aggregation in crew duties. It has a superior influence compared to effects resulting from interconnections between by crew and vehicle schedules referred to as crew changeovers. In addition, the authors prove that even pure cost-efficient schedules already provide a substantial number of swap opportunities for recovery.

Concerning the flexibility of airline crew schedules, Shebalov and Klabjan (2006) introduce a robustness indicator called move-up crews. A move-up crew provides a swap opportunity for another crew if severe disruptions occur at the day of operations. To maintain pairing feasibility, move-up crews are only considered between pairings from the same crew base that end almost at the same time. Varying delay risks for different swap opportunities are not considered. Thus, the benefit of all move-up crews is assumed to be the same, regardless of time and airport. For the evaluation of the approach, they consider temporary hub airport closures at the beginning of a day for the daily scheduling problem. These disruption scenarios are resolved by the SimAir simulation framework using crew swapping, reserve crews, deadheading and flight cancellations for recovery. Results provide figures on the number of reserve crews, deadheads, cancellations and certain monetary values. However, potential improvement in terms of delay propagation is not discussed.

4.1.5 Summary and Implications

Uniformly, robustness is defined as the ability of a schedule to provide a subjectively sufficient level of fulfillment in operations. The wide variety of approaches deal with directly assessing and preventing the amount of delay propagation emerging from crew and aircraft connections. Besides, maintenance robustness for aircraft (Lapp and Cohn (2012)) and crew rule compliance (Dück et al. (2012)) are to mention as supplementary objectives. In addition, the preservation of passenger itineraries is often used for assessing the schedules' performance (Lan et al. (2006); Rosenberger et al. (2004)).

The robustness can be indirectly improved by increasing the degree of freedom for crew and aircraft scheduling by either adapting optimization policies in preceding scheduling stages or using (partial) integrated approaches. Examples are shifting departure and arrival times and thus allow either the re-allocation of slack time or provide additional beneficiary connections. The same holds for the Fleet Assignment in which connection possibilities are preselected for subsequent crew and aircraft scheduling. Analogously to the findings concerning pure cost-efficient scheduling in Section 2.4, partial or fully integrated solutions promise improved results. Changes in systematical behavior or robustness indicators different to the one's from sequential scheduling cannot be observed.

For improving the stability of crew and aircraft schedules, the main instrument is the (re-)allocation of buffer time within the scheduled ground time between consecutive flights. Furthermore, the concept of crew-follows-aircraft is stated as one of the most important factors since cascading delay propagation can be prevented. A general distinction can be drawn between approaches based on key performance indicators (KPI) on the one hand and stochastic approaches on the other hand. In KPI-based approaches, identical delay risks are implicitly assumed for all flights. Short connection times are generally tried to be avoided, especially when crews change aircraft. In contrast, stochastic approaches on the other side use historical delay information to adapt varying delay risks. Connections are prevented or sufficient buffer times are incorporated only if delay propagation is likely. Since it plays an important role in the scope of this thesis, we devote the separate Section 4.2 to the topic of modeling of primary delays from historical data.

Concerning flexibility, the concepts of station purity, hub isolation and short cycles can be attributed to flexibility since they allow simpler recovery actions during operations. Directly influencing the flexibility of crew or aircraft schedules can be achieved by improving swapping capabilities. Approaches of Ageeva (2000) and Shebalov and Klabjan (2006) are KPI-based, Burke et al. (2010) use probability values derived from historical data for the evaluation of delay risks per flight. No approach considers network effects in terms of further propagation on subsequent flights. To the best of our knowledge, Burke et al. (2010) is the only study so far that deals with the simultaneous consideration of stability and flexibility. It is stated that stability has a major impact on the schedule performance. However, potential mutual influences between the robustness objectives are not yet discussed. The related question is if an increased degree of stability affects the degree of flexibility or vice versa.

Concerning the evaluation of robustness, a distinction can be drawn between indicator-based and simulation-based assessment of the schedule performance. The former deals with the identification of schedule characteristics that are assumed to have an impact on the robustness, see Weide et al. (2010) or Ageeva (2000). According to (Dück, 2010, p. 41), the major drawback of such indicators is that they are difficult to interpret by decisions makers since they do not necessarily provide information on actual delay propagation. Conversely, there is the possibility of subsequently simulating the operation of generated schedules. We see the advantage that sensitivity analyses can be performed, e.g. schedules are tested in environments with differing assumptions on delay occurrences or propagation mechanisms. Furthermore, the schedule performance can be measured by adapting real-world recovery actions in order to obtain results closer to real operations.

4.2 Primary Delay Prediction for Robust Resource Scheduling

Deterministic approaches consider the same risk for all flights and as a consequence implicitly allocate equally distributed buffer times. In contrast, the benefit of stochastic robust scheduling strategies depend on the assumptions being made concerning primary delay occurrences risks. It is desirable to predict primary delays as well as possible in order to hold the nominal cost increase low when improving the robustness. Figure 4.1 illustrates three ways of delay misestimation. For a given flight schedule, the left panel shows predicted delays in scheduling. Light green marks show correct predictions, gray and red marks show different ways of misestimation. The right panel shows delays that actually occur when the schedule is operated (dark green marks). Misestimation Cases depict examples when delays are

- **Case 1:** overestimated,
- **Case 2:** not recognized,
- **Case 3:** falsely predicted, or
- **Case 4:** underestimated.

In Cases (1) and (3), nominal costs are increased without an actual benefit for the robustness because delays are falsely predicted or overestimated. In contrast, Case (2) is assumed to have no effect on the schedule, i.e. the robustness is not increased because an emerging delay is not recognized. Moreover, the underestimation in Case (4) leads to insufficient robustness as delay risks are only partially considered.

Aiming at improving the trade-off between nominal costs and robustness, the prediction of primary delays based on the analysis of historical data has increasingly attracted the scheduler's attention. The usage of historical delay data can be traced back to the beginning of the commercial air traffic and Operations Research appliance. Examples are Benders (1962) and Dunlay and Horonjeff (1976), discussing

Figure 4.1: Misestimation of delays

preceding models of today's queuing theory and traffic control strategies. In recent decades, the data availability has even increased since techniques for automated data collection and recording has significantly improved. Governmental authorities such as the Research and Innovative Technology Administration (RITA) for the USA[1], EUROCONTROL for Europe and DFS for Germany collect data and provide annual status reports on general issues concerning the operational performance in airline traffic. Besides, airlines have strong interests in the assessment and steady improvement of their schedules for which data on their performance is recorded, see for example Schlegel (2010) and Ehrgott and Ryan (2002).

Two core issues are essential for historical delay data analysis. On the one hand the operational system behavior shall be understood in order to allow sensitivity analyses which are inevitable when constructing new schedules since operational performance can only be simulated at this stage. On the other hand, the generation of statistical models for delay prediction is essential for stochastic scheduling approaches. In the following, we address both topics and their related studies.

4.2.1 Operation-based Delay Studies

Shifting the focus away from robust scheduling for a moment, there is a variety of recent studies on the comprehension of delay occurrence mechanisms. Recent results include a large set of operational decision rules.

[1] `https://www.rita.dot.gov/bts/data_and_statistics/index.html`, last access: November 7th, 2017.

In a survey on recovery approaches for aircraft, crews and passengers, Ball et al. (2007) emphasize the usage of historical data. They are refined in a study on flight delays, in which the buffer management is demonstrated conceptually and empirically.

In a statistical modeling approach for arrival delay, Hsiao and Hansen (2006) consider queuing, weather and seasonal effects. The authors discover negative daytime trends for queuing effects, i.e. delay occurring in the morning has a greater impact on delay propagation than in the evening. The study extracts a large number of variables influencing delays which leads to a high explanatory power.

Xu et al. (2008) use regression models for estimating airport-related delays for the usage by operations control authorities. Again, besides the scheduled departure time and the scheduled turnaround time, especially short-term variables such as weather, operation demand in relation to airport capacity, ground holding and inbound delays are considered. The prediction error is estimated by applying the model to an unknown test set.

Tu et al. (2008) present a statistical analysis with predictive modeling for strategic decision support. The authors consider seasonal, propagation and random delay patterns while concentrating on a specific airport in order to predict congestion effects. Deshpande and Arikan (2012) analyze the impact of scheduled block-times on the on-time arrival probabilities. Arıkan et al. (2013) aim at examining the impact of airline network structures and schedules on the reliability of the air-travel infrastructure. Therefore, stochastic models for actual block times following a log-Laplace distribution are discussed. Secondly, they develop a model for measuring the delay propagation through the flight network based on aircraft rotations.

4.2.2 Delay Modeling in the Context of Robust Airline Resource Scheduling

Traditional airline resource scheduling deals with the minimization of planned costs. The usage of empirical delay information has become important to the field of robust resource scheduling. In this section we present recent advances with special regard to the usage of historical delay information. Furthermore, recent data mining approaches for large data sets are discussed.

Ageeva (2000) presents an approach to increase flexibility of schedules by incorporating swapping opportunities for aircraft. However, delay risk are not considered for incorporating swaps. The evaluation of the approach is based on an increased number of swap opportunities which is considered as an indicator for increased flexibility.

The scheduled crew ground time is used as a deterministic indicator for stability in (Ehrgott and Ryan, 2002, p. 141). Therefore, the difference between slack duration and expected duration of a departure delay, specified by flight routes, is used as a penalty factor for non-robustness. Weide et al. (2010) use a related measure for a heuristic iterative crew and aircraft scheduling approach. Schaefer et al. (2005) incorporate robustness by considering operational costs of crew pairing instead of planned costs. The operational costs are determined by separately simulated crew pairings in SimAir, a simulation framework that uses empirical delay distributions gained from historical data, see (Rosenberger et al., 2002, p. 373).

(Yen and Birge, 2006, p. 10) fit truncated gamma and log-normal distributions to real world data from Air New Zealand in order to generate disruption scenarios for a stochastic crew scheduling model. No information on the goodness-of-fit is given. (Lan et al., 2006, p. 19) improve the stability of schedules by considering the delay propagation on aircraft routes. For the estimation, the authors use historical data from the ASQP database. Gamma, log-normal and Weibull distributions are compared by means of classical goodness-of-fit tests. As a result, the log-normal distribution is found the best fit for 84% of all flight arrival delays. The approach is also used by Dunbar et al. (2012). Note that both Yen and Birge (2006) and Lan et al. (2006) do not separately examine possible impacts per attributes such as time and location attributes of a flight in delay models.

(Tam, 2011, pp. 89-121) also uses historical data for delay estimation. The flight delay is modeled by multiple-regression for every weekday. The regression terms consider the departure and arrival airport and the departure and arrival time. Note that no interactions between the variables are taken into account. The quality of the models is measured only for the intra-sample and the prediction error over an unknown data set is not assessed. (Dück et al., 2012, pp. 54-55) present a stochastic model for increasing the stability of crew and aircraft schedules. They use log-logistic and log-normal distributions per delay reason for the generation of primary delay scenarios. The expected delay of a flight is based on the convolution of different delay reasons. Delays due to weather, airspace and airport congestion are not considered in scheduling but in the subsequent simulation of schedule operations.

4.2.3 Summary and Implications

The usage of historical data has become a standard procedure for a priori flight delay estimation. In the preceding literature survey, it becomes apparent that a distinction has to be drawn between operations- and scheduling-related delay modeling. There are extensive studies on operational delays that are able to explicitly take into account cause-and-effect relations. On this microscopic level, explanatory variables comprise operational aspects such as airport characteristics, short-term congestion effects or current weather conditions. As a consequence high prediction accuracy levels can be reached.

However, the time horizon of scheduling as a long- and medium-term task does not allow the application of resulting statistical prediction models due to three main reasons: Firstly, short-term prediction variables with reasonable impact on the prediction accuracy are not available during the long- and medium-term resource scheduling process. Secondly, specific operational rules, e.g., for a single airport (Wesonga et al. (2012); Tu et al. (2008)), cannot be adequately adapted to entire resource networks. At last, the generalization of models with a large number of explanatory variables and resulting high prediction accuracy (e.g., Hsiao and Hansen (2006)) for complete networks cannot be performed easily as it leads to an unmanageable model complexity.

Due to these limitations, the usage of historical data is often reduced to an automated distribution-fitting, e.g. Lan et al. (2006), Tam (2011) or Dück et al. (2012). With increasing frequency, delay patterns for different spatio-temporal parameters of flights are taken into consideration. According to Burke et al. (2010), systematical patterns in operational delays can be implicitly described by attributes such as time of day, season or O&Ds. We refer to this approach as macroscopic delay prediction. However, delay estimation on a macro-level is still a black box with often nontransparent assumptions. Selected explanatory spatio-temporal variables are mainly selected based on their availability or practitioners' expertise rather than a data-driven evaluation of their impact. Prediction models comprising different sets of spatio-temporal explanatory variables are not evaluated or set into relation to each other. It must furthermore be noted that delay modeling often has a lower priority than the actual optimization approach in publications related to robust scheduling. This observation is in line with Mortenson et al. (2015) who state that "[...] the amount of academic research into analytics published in journals associated with the OR/MS discipline will be shown to be surprisingly low."

We therefore see the necessity of examining the potential of interpretable decision rules that can be used as groundwork for the generation of realistic primary delay prediction models for robust scheduling. Statistical models for primary delay prediction therefore have to capture interpretable delay occurrence mechanisms on a macroscopic level.

Chapter 5

Required Work

In Part I of this thesis, we have provided an overview of airline scheduling and operations, highlighting their contrary objectives. The concept of *robust efficiency* aims at bringing them together by dealing with the consideration of probable reactionary costs during the scheduling stage. Reactionary costs are generally difficult to determine and depend on both strategic and tactical decisions of an airline concerning their recovery instruments. In the context of crew and aircraft scheduling, the main influential factor is the amount of potential delay propagation. Besides, some approaches aim at minimizing of crew rule disruptions or improving the aircraft maintenance schedule feasibility. Related scheduling problems are usually tackled using OR-related modeling and optimization techniques. Both commercial schedulers and scientific prototypes mostly rely on the fundamental column generation approach, often embedded in a branch-and-price framework. The degree of freedom for both cost-efficient and robust scheduling can be increased by the integration of planning stages, e.g. by applying Benders' Decomposition for simultaneous crew and aircraft scheduling. However, resulting integrated formulations suffer from the high level of complexity and are therefore difficult to solve.

Concerning the assessment of schedule robustness, the stochastic nature of primary delay occurrences is more and more taken into consideration by recent approaches. Understanding and modeling realistic delay occurrence and propagation mechanisms are essential for an assessment of the robust efficiency close to operational reality. However, in Section 4.2 it has become apparent that the potential of long-term delay prediction for fulfilling the requirements of robust scheduling has not been sufficiently examined yet.

The evaluation of the schedule robustness is commonly performed by a stochastic simulation of airline operations. Regarding the evaluation of stability, the postponement of flights is usually used as the only recovery instrument that is widely used. In contrast, the consideration of flexibility necessitates the application of suitable recovery instruments. An outstanding example for reflecting operational reality of recovery procedures is the SimAir simulation framework that is used in many publications focusing on the US airline market. For the purpose of our study, the main drawback is that is is not possible to trace changes in operational reliability explicitly back to either schedule robustness or sophisticated recovery actions.

From the current state-of-the-art in related scientific literature, we derive four important research objectives that are tackled in the subsequent studies:

R1 *Examining the potential of data-driven primary delay prediction,*

R2 *Assessing the influence of refined primary delay prediction on the robust efficiency of schedules,*

R3 *Developing recovery instruments that are suitable to evaluate the flexibility of crew and aircraft schedules,* and

R4 *Investigating mutual impacts between stability and flexibility as groundwork for a holistic view on the robust efficiency of schedules.*

Research Objectives R1-R4 concentrate on the examination of mechanisms and influential factors that directly affect the robust efficiency of schedules. Technical aspects such as improving optimization techniques and tackling the integration of scheduling stages for improving the degree of freedom for planning decisions play a subordinate role in our research. Nevertheless, they are subject to future work and will be addressed in Chapter 13 (Summary & Outlook).

The overall methodical approach of this thesis is associated with a long tradition of OR-related scientific publications on airline-specific resource scheduling as presented in Chapter 4. Analytics-related studies in Part II base on the conceptual groundwork of Hand et al. (2001) and Breiman (2001b). On the technical side, we refer to the machine learning approach of Hastie et al. (2009). A detailed discussion on this topic is provided at the beginning of Chapter 7.

In the remainder of this chapter, we firstly present the technical framework in which all related studies take place. Afterwards, each research objective is substantiated, specified and set into context.

The Prototypical Scheduling and Simulation Framework

All research objectives are investigated in a coherent prototypical scheduling and simulation framework that technically bases upon the work of Dück (2010). Figure 5.1 shows the base framework and the extensions developed in the context of this thesis (highlighted in darker color). For a given *Flight Schedule*, *Resource Schedules*

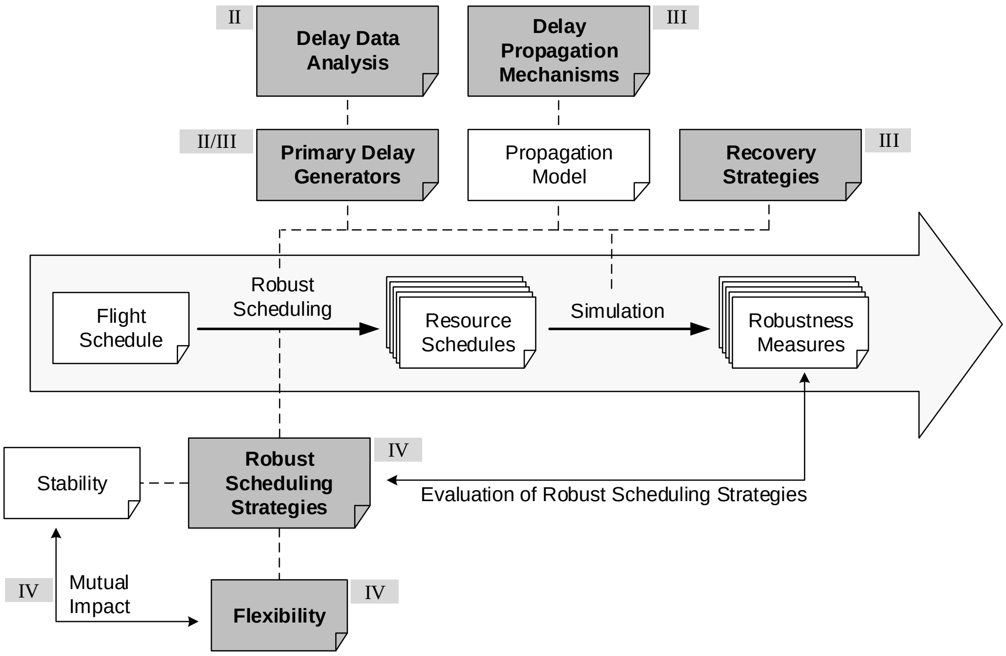

Figure 5.1: The base scheduling and simulation framework and the extensions developed in this thesis (highlighted in gray)

(e.g. for crews or aircraft) are generated following specified *Robust Scheduling Strategies* that take into account the robustness by penalizing potential delay propagation risks. We concentrate on the stochastic approach rather than deterministic robustness indicators by using primary delay scenarios for the assessment of potentially entailing propagation effects during the generation of schedules. Delay scenarios are constructed by a *Primary Delay Generator*, adapting findings of a preceding *Delay Data Analysis* that is performed in the following Part II of this thesis. Entailing delay propagation effects are then approximated based on a specified *Propagation Model*. The predictive quality of the propagation model can be assessed by comparing its estimations to *Delay Propagation Mechanisms* obtained from real-world data.

The assessment of the actual robustness of constructed schedules can then be performed by means of event-driven stochastic discrete simulation. Several simulation runs can be performed for varying primary delay scenarios. For sensitivity analyses on the potential of a schedule to maintain its robustness in changing environments, considered primary delay scenarios may deviate from the ones considered during scheduling. Applied *Recovery Strategies* include both postponement of subsequent flights and swapping of crews and aircraft, enabling the evaluation of both schedule stability and flexibility. Based on the simulation results, several *Robustness Measures* can be computed, e.g. the average amount of propagated delay over all simulation runs. In the context of this thesis, aspects related to the evaluation of robustness are encapsulated in Part III.

Eventually, conclusions on the benefit of the robust scheduling strategies can be drawn by comparing stability and flexibility indicators from scheduling with the robustness measures gained by simulation. According to (Dück, 2010, p. 89), a robust scheduling strategy is *predictable*, if its indicators correlate with the actual robustness measure outcome. Moreover, it is *efficient*, if a high level of robustness can be achieved for a relatively low increase of nominal costs compared to cost-optimal solutions. These definitions are important for the comparison of stability- and flexibility-related scheduling strategies in Part IV.

The highlighted extensions of the base framework arise from the necessity to tackle the research objectives that are described in the following. Roman numerals in Figure 5.1 indicate the Part of this thesis in which the extensions are elaborated in detail.

R1 *Examining the potential of data-driven primary delay prediction*
In the literature survey, the necessity for realistic prediction of exogenous primary delays in the context of robust resource scheduling has become apparent: In the recent decade, scheduling approaches increasingly focus on taking into account the stochastic nature of primary delay occurrences as exogenous input factors. Inevitably, the benefit of these approaches stand or fall with the accuracy of primary delay

occurrence prediction. It can be assumed that a better understanding of primary delay occurrence mechanisms leads to a better trade-off between cost-efficiency and robustness. However, this topic is still rather neglected in recent approaches which mostly deal with simple distribution fitting schemes and intra-sample estimation of the prediction accuracy. As a result, prediction models are often a black box for the user. In contrast, studies based on operational cause-and-effect relations are not applicable for robust scheduling in a straightforward manner due to their complexity and the short-term nature of explanatory variables.

The main question is therefore to what extent exogenous delays can be accurately predicted with respect to the demands of robust resource scheduling. Therefore, we want to examine the potential of interpretable decision rules that can be used as groundwork for the generation of realistic primary delay scenarios. For meeting the demands of robust resource scheduling, long-term predictor variables have to be taken into account rather than short-term cause-and-effect relations. In contrast to intra-sample-based evaluation schemes such as the R^2-based approach of Tam (2011), the prediction accuracy shall be assessed for unknown data sets.

R1 is addressed in Part II (Study on Exogenous Delays in Airline Networks). At first, the historical data form a major European carrier is introduced in Chapter 6. Flight network characteristics and fundamental patterns in delays are identified. The purpose is to initially examine influential parameters that are relevant for the description of delay occurrences. Subsequently, the potential of data-driven detection and statistical modeling of decision rules for primary delay occurrences based on spatio-temporal attributes is examined in Chapter 7. Obtained results shall give an insight into both the nature of primary delay occurrence and the methodical potential of delay prediction in the context of robust resource scheduling.

R2 *Assessing the influence of refined delay prediction on the robust efficiency of schedules*

This research objective results as a direct consequence of R1. Findings on the theoretical potential of primary delay prediction can be used for implementing primary delay generators for the usage in robust resource scheduling. However, improved primary delay prediction must not necessarily reflect in an identical improvement of the robust efficiency of schedules. This is because scheduling decisions may be restricted by other influential factors such as crew rules or the network structure.

A straightforward example is the out-and-back principle in hub-and-spoke networks, inevitably predetermining crew and aircraft connectivity independent from their potential delay propagation risk. In this regard, the general influence of primary delay prediction on the robust efficiency must be assessed. An important question is in how far the trade-off between cost-efficiency and robustness is affected. Besides the benefit of refined prediction models, effects resulting from prediction inaccuracies during scheduling has to be examined.

R2 is addressed in Part III. Initially, the process of the stochastic simulation of airline operations is introduced in Chapter 8. Besides, fundamental assumptions on the propagation model of Dück et al. (2012) are presented and evaluated regarding their correspondence with real-world delay propagation mechanisms. For the actual study on the influence of primary delay prediction in Chapter 9, we concentrate on crew schedules which are constructed based on existing aircraft rotations in a sequential scheduling approach. At first, we assess the influence of refined primary delay prediction on the robustness and the nominal costs of crew schedules, directly taking into account the findings from R1. In a subsequent sensitivity analysis, we estimate further effects resulting from over-, under- and also misestimation of primary delays during scheduling.

R3 *Developing recovery instruments that are suitable to evaluate the flexibility of crew and aircraft schedules*

The common recovery instrument for the evaluation of schedule stability is to postpone subsequent flights in case of delays until crews and aircraft are available. The stability can be measured by the resulting amount of propagated delay. In contrast, a holistic evaluation of the robust efficiency in simulation must also deal with recovery strategies that make use of the particular aspects of flexibility, corresponding with its definition from Section 3.3.

For this purpose, we present a rule-based recovery approach for crews and aircraft and evaluate its influence on the schedule robustness assessment in Chapter 3.2. Previous studies on this topic by Rosenberger et al. (2002), Shebalov and Klabjan (2006) and Burke et al. (2010) are taken into account. The approach is necessary groundwork for the subsequent holistic consideration of robust efficiency that is addressed in the following Research Objective R4.

R4 *Investigating mutual impacts between stability and flexibility as groundwork for a holistic view on the robust efficiency of schedules*
The concept of robust efficiency deals with reducing the gap between contrary objectives in scheduling and operations. In this regard, the consideration of both stability and flexibility during the scheduling stage is essential. Stability and flexibility are generally considered separately in related literature although a simultaneous consideration appears to be advantageous for a holistic assessment of the robust efficiency. However, the potential of such an approach may be limited by mutual impacts or even a trade-off between stability and flexibility. It is therefore essential to examine in how far an increased degree of stability affects the swapping capabilities. The other way around, a schedule with increased flexibility may implicitly tend towards more or less delay propagation. In this regard, the following research questions can be derived:

(A) Do stable schedules prevent delay propagation?

(B) Do flexible schedules prevent delay propagation?

(C) Do stable schedules provide swap opportunities?

(D) Do flexible schedules provide swap opportunities?

Questions (A) and (D) relate to the efficiency of scheduling strategies for stability or flexibility and related quantitative results serve as reference values. Questions (B) and (C) directly address potential mutual impacts. The findings can be used as groundwork for future work including a holistic assessment of the robust efficiency and the elaboration of scheduling strategies that incorporate both stability and flexibility aspects at the same time.

R4 is addressed in Part IV. Analogously to Research Objective R2, we concentrate on generating airline crew schedules with increased stability or flexibility in the scope of this study. Chapter 11 deals with the development of a stochastic optimization approach to improve the flexibility of airline crew schedules. For reasons of subsequent comparability, the approach has to be adaptable to the stochastic delay propagation evaluation of Dück et al. (2012). In terms of both methodical definition and technical implementation, swapping mechanisms are taken into account as elaborated in the context of R3. In the consecutive Chapter 12, research questions (A)-(D) are examined.

Assumptions and Restrictions for Subsequent Analyses

All subsequent analyses are performed with regard to the following restrictions and assumptions:

- We focus on delays in turnaround and ground operations determined by IATA Delay Codes. En-route delays are not modeled explicitly in the simulation. Since aircraft are kept on the ground if airspace or airport congestion limitations may be exceeded (so called *ground-holding*[1]), we assume that the impact on our results is negligible in this stage of research. Delay propagation is not affected by this restriction.
- Propagation effects by additional network layers such as passenger itineraries, technical equipment for ground operations, airport infrastructure or air traffic control are not considered.
- Corresponding with the assumptions of Ageeva (2000) and Shebalov and Klabjan (2006), swaps of crews and aircraft during operations are assumed to be cost-neutral. Although at least indirect costs for the implementational effort of swaps may incur, this assumption allows us an unfiltered view on the potential of increased flexibility.

[1] See Hirata et al. (2013) for details on controlling instruments of the Air Traffic Control.

Part II

Study on Exogenous Delays in Airline Networks

Chapter 6

Introducing the Data Set

In this chapter, we present the data set that forms the basis for all subsequent analyses. The first Section 6.1 deals with an initial preparation and cleaning of the data, followed by a domain-specific description and examination of data records and attributes. Subsequently, a time course analysis is performed in order to examine potential systematic changes over the years in Section 6.2. A summary of findings is provided in Section 6.3.

6.1 Preparation and Description of the Data

The following analysis is performed on a data set consisting of 2,545,248 flight delay records provided by a major European airline for the time period from March 2003 to February 2007. In the initial data cleaning and preparation, the data quality is examined and a subset of the data that is appropriate for our analysis is extracted. Afterwards, an overview of available data dimensions is given.

6.1.1 Data Extraction, Cleaning and Preparation

Since all subsequent studies refer to regular scheduled passage domestic and continental flights, the respective records are extracted in a first step. All flights declared as charter, cargo, mail, non-revenue, additional and ferry flights are removed. For 30 undeclared flights attributes such as O&D, aircraft type, rotation label or number of passengers are checked, with the unambiguous result that they are not incorrectly recorded scheduled passenger flights and therefore can be excluded. In summary, 108,542 records are removed.

Regarding traffic areas, flights that depart and arrive in Germany are classified as *domestic*. Flights either departing or arriving in another European country are labeled as *continental*. Flights into or from countries outside Europe are labeled as *intercontinental*. There are 752,989 domestic and 1,678,710 international flights. The latter consist of 1,504,610 continental and 174,100 intercontinental flights. There are 5,007 flight records without a traffic area code, although they unexceptionally contain a rotation code. Therefore, the traffic area has been resolved by considering fleet types and departure and arrival airports.

Especially in large data sets, invalid or (partially) missing data records are unavoidable. In the next step, the general data quality and consistency is checked. For 78 flights, there is no complete information on the scheduled and actual departure and arrival times. For 3 flights, departure and arrival times imply negative block times. 981 flights start and end at the same airport. As it is not possible to sufficiently resolve these issues, the records are removed. Subsequently, the consistency of delay records is checked. 27 flights have a delay reason record without a related duration and also without differences in scheduled and actual departure times. In addition, some flights contain delay reason and duration records without actual departure delays. Delay reason entries are removed in both cases. In relation to the number of flight records, the general data quality can be considered as very good.

In the following, we examine the characteristics of spatio-temporal and flight-dependent attributes.

Based on these attributes, further delimitation of the data set is performed in order to guarantee a suitable basis for the analyses of systematical effects. Conceivable issues are occasional flights that are operated in spite of night flying restrictions or O&Ds that are exceptionally served.

6.1.2 Airports and Network Structure

The network is based on a hub-and-spoke structure with two major hubs. 19.9% of all flights depart at hub 1, 18.8% at hub 2. The remaining 61,3% are distributed across 303 spoke airports that immensely differ in the number of departures and arrivals.

The network is illustrated by chord diagrams in Figure 6.1, separated by flights from and to Hub A (Panel A), Hub B (Panel B) and for inter-spoke connections

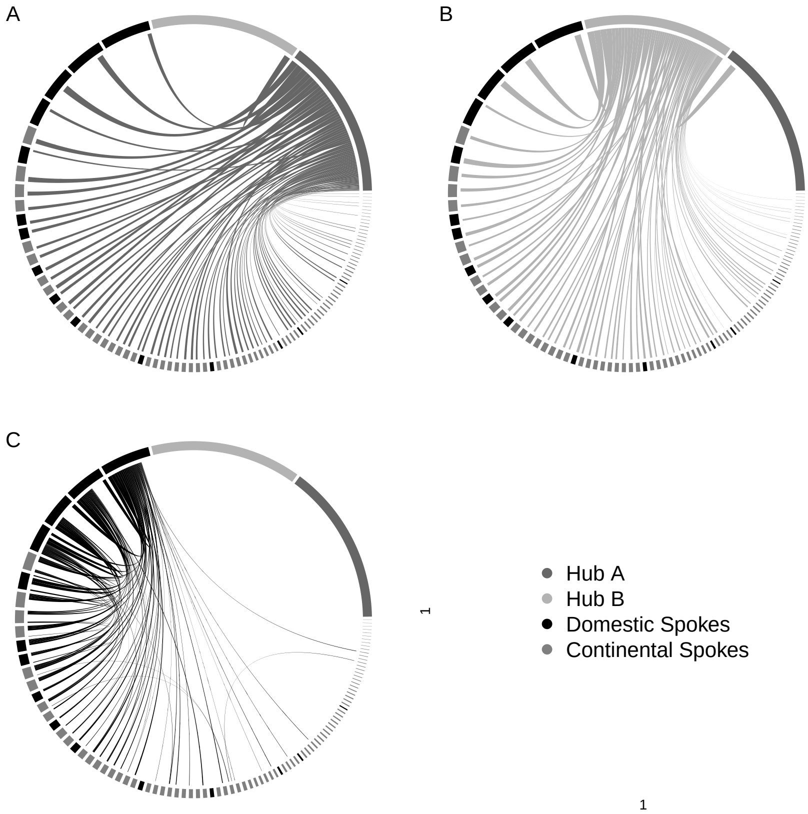

Figure 6.1: Routes in the flight network, separated by Hub 1 (A), Hub 2 (B) and spoke airports (C)

(Panel C). Routes are considered only if they are flown at least 1,000 times in total. The outer grid shows the relative size of the airports depending on the number of arrivals and departures. Lines indicate the respective traffic volume between airports. Panels A and B clearly show the general hub-based structure. Hub A is connected to 79.59%, Hub B to 75.51% of all spokes. 36.96% are hub-to-spoke, 36.89% spoke-to-hub, 1.70% hub-to-hub connections[1]. A ratio of 24.45% spoke-to-spoke connections indicate that there is a considerable amount of point-to-point flights. They mostly consist of flights between large domestic spokes (see Panel C). In contrast to the

[1] Hub-to-hub connections refer to the single route between Hubs A and B. Since analyses on specific routes may not be published, this category is not explicitly taken into account for further delay analyses.

hubs, the most spokes connect to less than 5% of all other spokes. Exceptions are the four most frequently served domestic airports with values of 35.71%, 20.41%, 17.35% and 10.2%, forming a point-to-point part of the network.

6.1.3 Time-based Attributes

Time stamps are available for departure and arrival times. Starting with seasonal and monthly patterns in flight movements, we also discuss differences by weekday and time of day. Note that an examination of trends over the full course of time is provided in Time Series Analysis in Section 6.2.

Seasonal Attributes

Regarding seasonal patterns, there are about 10% less flights compared during January and December, compared to the overall average, see Figure 6.2. The wavelike course of the year has two peaks in June and September, separated by a decrease in August. We assume that the number of flight movements straightforwardly reflects the customer's demand over the year. Differences between summer months may be explained by outward and return journeys at the beginning and end of summer holidays. The low values in January and December are primarily induced by Christmas and New Year's Eve. Other public holidays do not have an influence on the seasonal trend, however their impact on a daily basis can be observed as presented in the following section.

Weekdays and Holidays

Data splits by weekday show that the average number of flight movements is almost constant between values of 1,609 (Fridays) and 1,618 (Thursdays) at working days. Interestingly, initially assumed high values at the beginning and the end of the working week due to business outward and return journeys cannot be observed. The most flight movements rather take place on Tuesdays and Thursdays. The least flights depart on Saturdays (1,234) and Sundays (1,338).

Besides the regular weekly structure, 60 (public) holidays between March 2003 and February 2007 do also effect the flight schedule. On holidays at working days there are significantly less flights (1,432), the same holds for holidays at the weekend (1,160).

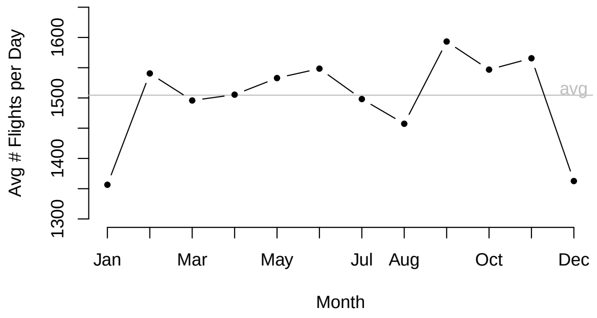

Figure 6.2: Average number of flights per month

Time of Day

The time of day is determined by scheduled and actual departure (STD, ATD) and arrival times (STA, ATA) in Coordinated Universal Time (UTC) stamps. Standardized time specifications are important for managing flight networks that stretch over multiple time zones. As an example, a scheduled flight departing at 08:00 UTC on a two hour flight presumably arrives at 10:00 UTC. If the destination airport is situated in a different time zone, local time stamps do not provide an appropriate schedule specification.

Regarding the context of this work, network effects can be made apparent due to absent local deviations by using standardized UTC times. Daylight saving times (DST) may also be taken into account for balancing seasonal deviations, enabling an unbiased view on itineraries through the network. However, there are certain areas that require local times in this work. Especially when it comes to capturing delay effects that may be closely related to local concerns and workflow procedures at departure and arrival airports. In particular, they include possible congestion effects during morning and evening hours.

The determination of local times requires the integration of information on time zone and daylight savings time. Suitable information on time zones and DST is

Offset	0	1	2	3	4	5	6
#Flights	39,036	867,357	1,242,082	29,545	18,161	0	1,225

Table 6.1: UTC departure time shifts due to time zone and daylight savings time

provided by the OpenFlights Airport Database[2]. Based on this data, we compute a *local UTC offset* for departure times. Table 6.1 provides details on the number of affected flights. Ranging from 0 to 6, the average local UTC offset is 01:36:07. Since 92.52% of all flights depart in the CET time zone and are affected by daylight savings time, the most frequent shifts are one or two hours.

With time attributes set up, we have a closer look at departures and arrivals in the course of the day. Histograms in Figure 6.3 illustrate departures (dark, in the front) and arrivals (light, in the back) in the course of the day. Dark gray areas indicate overlays of both metrics. We use bins of five minutes for discretization, following the principle that flights are scheduled within five minute intervals. While graphics for scheduled times show a quite angular picture, actual departure and arrival times appear smoother. This is a natural result of the distribution of flights within their scheduled slots of 5 minutes. It is also striking that departures and arrivals follow a wavelike shape with alternating peaks, indicating an important characteristic of hub-and-spoke networks called *banks*. By many arrivals within a short period of time, passengers are collected at a hub to be distributed on many departing flights afterwards.

Going into detail, Figure 6.4 depict the difference between departure and arrival counts for the two hubs as well as all spoke airports. The effect of banks is made more obvious here. One may assume that the first wave consists of incoming flights into the hubs. Interestingly, there are more departures from the hubs in early morning hours. A reason may be that hubs are the homebase for most crews and provide aircraft maintenance stations, aircraft spread out into the network first. Peaks are quite alternating between the hubs, too, leading to a more balanced trend at the spokes. In the late evening, a majority of flights heads back to the hubs again.

Speaking about the end of daily airline operations, the topic of night flying restrictions (referred to as *night curfew*) becomes imminent. Night curfew apply to

[2]OpenFlights Airports Database, retrieved from: `https://sourceforge.net/p/openflights/code/757/tree/openflights/data/airports.dat`, accessed November 12th, 2015.

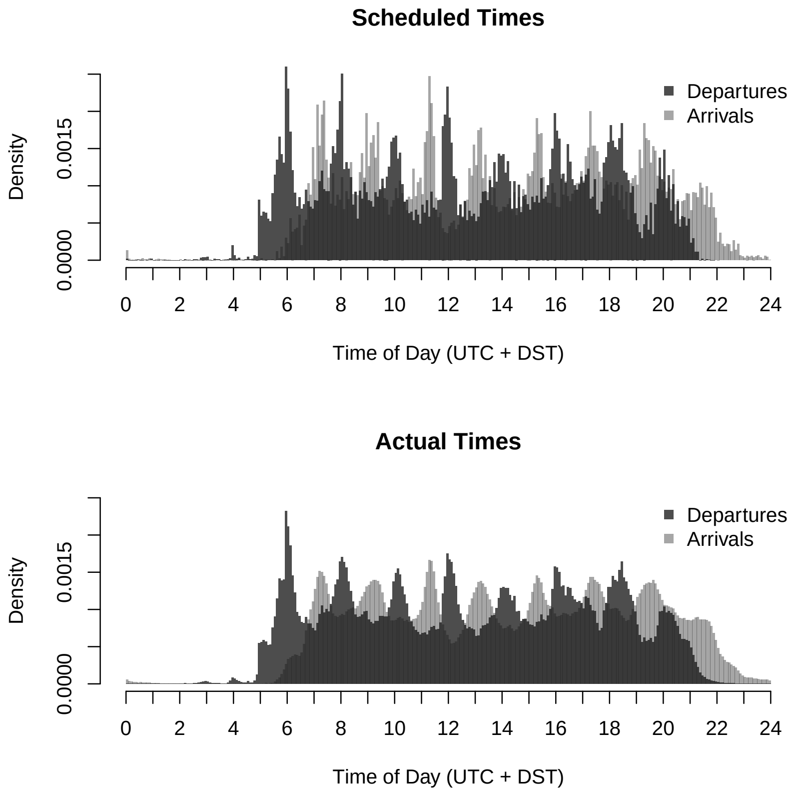

Figure 6.3: Departures and arrivals histograms

single airports and are imposed by (local) public administration in order to keep noise emissions low during the night hours. Rules are rather complex and precise for all kinds of flight movements at the airport[3]. Exceptions are provided in some cases, e.g. there are certain contingents for delayed arrivals that affect the whole airport or just particular runways. Night curfew restrictions do never apply in any case of security-related issues.

Although night flying restrictions relate to single airports, they affect the whole flight network. Looking again at the network-based Figure 6.3, scheduled daily

[3]For exemplary details on airport night flying regulations see for example `https://www.munich-airport.com/night-flight-264466`, last access: January 8$^{\text{th}}$, 2018.

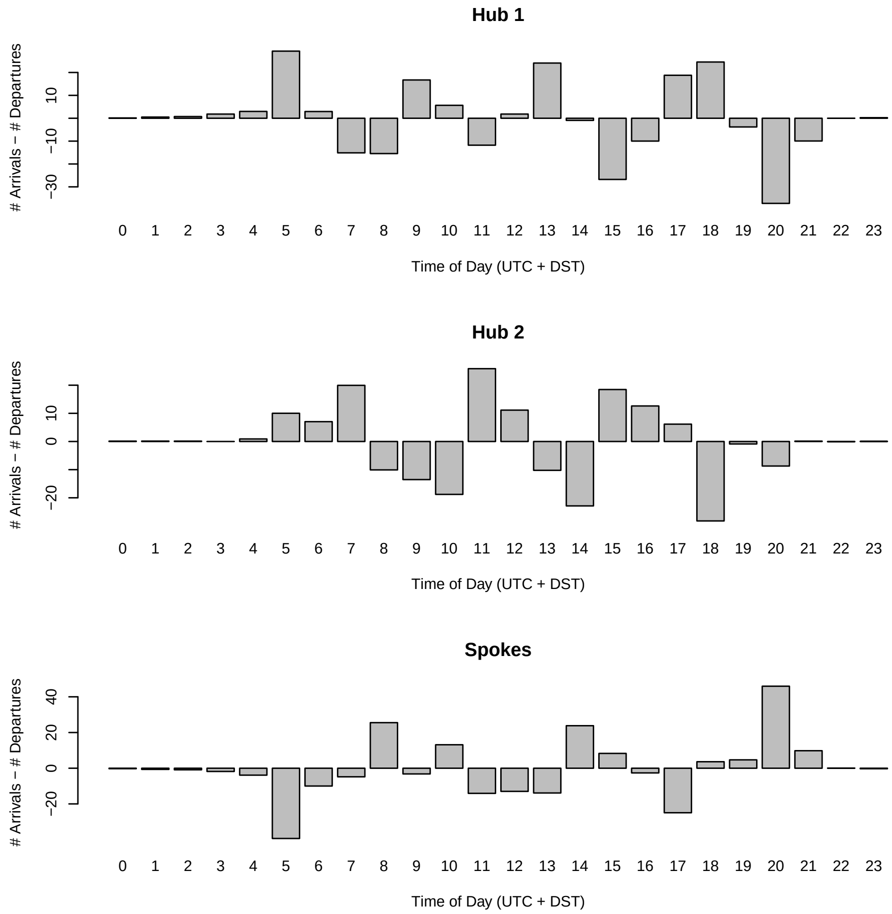

Figure 6.4: Difference between the number of arrivals and departures at Hub 1, Hub 2 and all spokes, aggregated per hour

operations start at 05:00 UTC+DST and end at 21:00 UTC+DST. When considering local airport times, 99.79% of all flights depart between 06:00 and 22:00 o'clock[4].

Over all four years, there are 19,208 flights departing between 22:00 at 6:00 local time. They mostly consist of distant destinations from Hub 1 to Turkey, Cyprus and the Caucasus and a few spoke-to-spoke connections. Nevertheless, flights between 22:00 and 23:00 local time are already partially restricted by night curfew. In addition to the small number of affected flights, the limitation on flights between 06:00

[4]For a graphical illustration based on local airport times, we refer to Figure C.1 in the Appendix.

and 22:00 local departure time ensures a target-oriented approximation in terms of ensuring simplicity while being compliant with our study design of analyzing regular daily operations.

6.1.4 Flight-specific Attributes

Concerning flight-specific attributes, rotation codes and fleet types allow us to retrace rotation connections between flights. Furthermore, we compute block and ground times from departure and arrival times since they are important metrics in terms of delay propagation. For the scheduled (SBT) and actual block times (ABT) of a flight f it holds:

$$\begin{aligned} SBT_f &= STA_f - STD_f, \\ ABT_f &= ATA_f - ATD_f. \end{aligned}$$

The general average block time is 87.74 minutes, with major differences between domestic (64.84) and continental routes (99.6). In addition, block times are 2.19% longer for the same routes in winter months.

Aircraft ground times are computed in the following way: For the scheduled ground time SGT^a for aircraft it holds $SGT_f^a = STD_f^a - STA_p^a$, where flight p is the predecessor of flight f in the rotation. Analogously, the actual ground time AGT^a of flight f is computed as $AGT_f^a = ATD_f^a - ATA_p^a$. Figure 6.5 represents empirical distributions for SGT and AGT values. Overnight stays of aircraft are visible as a second peak between 400 and 800 minutes. The average SGT value is 53.68 (not including overnight stays) for all rotation connections. As a first indicator for operational delays in ground processes, the average AGT of 57.57 minutes is 7.25% higher than the average SGT.

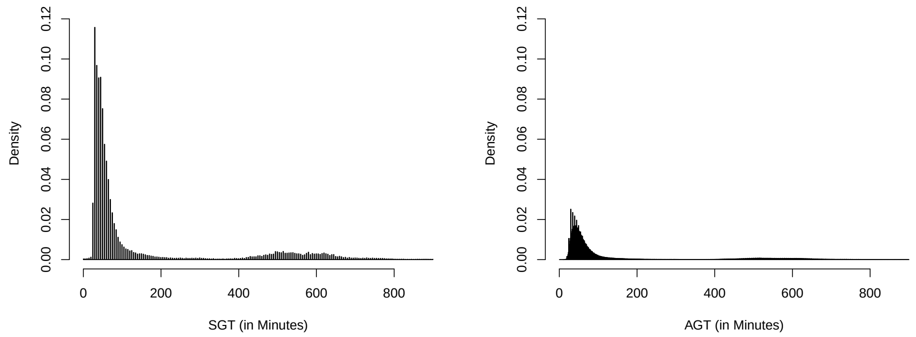

Figure 6.5: Scheduled and Actual Ground Times for all rotation connections

6.1.5 Delay-related Attributes

In our context, a delay is defined as the non-negative deviation between the scheduled and actual departure and arrival times of a flight. For the departure delay d_f^d and the arrival delay d_f^a it holds:

$$d_f^d = max(0, ATD_f - STD_f), and \tag{6.2a}$$

$$d_f^a = max(0, ATA_f - STA_f). \tag{6.2b}$$

Per flight, up to four different departure delay reasons and their duration are recorded. They are categorized by standardized IATA Delay Codes. Primary delays are denoted with codes from 1 to 89 for airline internal reasons, turnaround process disruptions, equipment failure, technical damages, or airport and airspace congestion, just to mention the main categories[5]. Reactionary delays are denoted with codes from 90 to 96, including waiting for passenger or load connections, the late arrival of aircraft or crew, and operations control. Inevitably, one may assume that some codes are in the gray area between primary and secondary delays. These include delays due to weather conditions and especially congestion effects at either the arrival or destination airport. In this point, we refer to the strict separation by whether delay propagation is directly induces by resource scheduling decisions for crews or

[5]For a complete overview and classification of all IATA Delay Codes, see Appendix B.

# Delay Records	0	1	2	3	4
Absolute Frequency	911,568	1,043,622	236,148	17,184	1,732
Relative Frequency	41.24%	47.22%	10.68%	0.78%	0.08%

Table 6.2: Occurrence frequencies of multiple delay records

aircraft (secondary delay) or not (primary delay). Indirect effects, e.g. by ATFM restrictions, are always referred to as exogenous primary delay.

In the data set, 47.6% of all flights are primarily and 19.75% secondarily delayed. 8.63% of the secondarily delayed flights also contain primary delays. Table 6.2 presents frequencies per number of departure delay records. In case of multiple records, secondary delay is almost always recorded first.

It also becomes apparent that delay records at spoke airports seem to be inaccurate due to rounding. Figure 6.6 illustrates differences in departure delay histograms at hub and spoke airports. While at Hub 1 it turns out to be a smooth curve shape, delay measures at Hub 2 seem to be rounded in many cases. This effect becomes even more obvious at spoke airports.

Another issue refers to the way of delay recording which may lead to an underestimation of delays for two reasons. Delays are recorded only if they actually lead to postponed departures. Delays absorbed by buffers in the turnaround process are not traceable in the data. Secondly, the four delay records per flight cumulatively form the total deviation between scheduled and actual departure. More precisely, for a flight with more than one delay reason, the second delay is not measured until the first delay has been resolved or ended. Thus, it is neither possible to determine whether delays overlap in time, nor the full length of potential second delays. As propagated delays are generally recorded first, the underestimation mainly affects primary delays.

Figure 6.7 provides an illustration of possible cases concerning multiple pre-departure delays. One delay can be completely covered, e.g. waiting for technical equipment for the turnaround process (Second Delay Reason) while the aircraft is not available yet (First Delay Reason). It also can be partially overlapped or start after the first reason has ended. The latter two cases cannot be distinguished retrospectively. Referring to Table 6.2, 11.54% of all records may be affected by underestimations. Besides, it is not evident how many *hidden* delays potentially exist for flights with

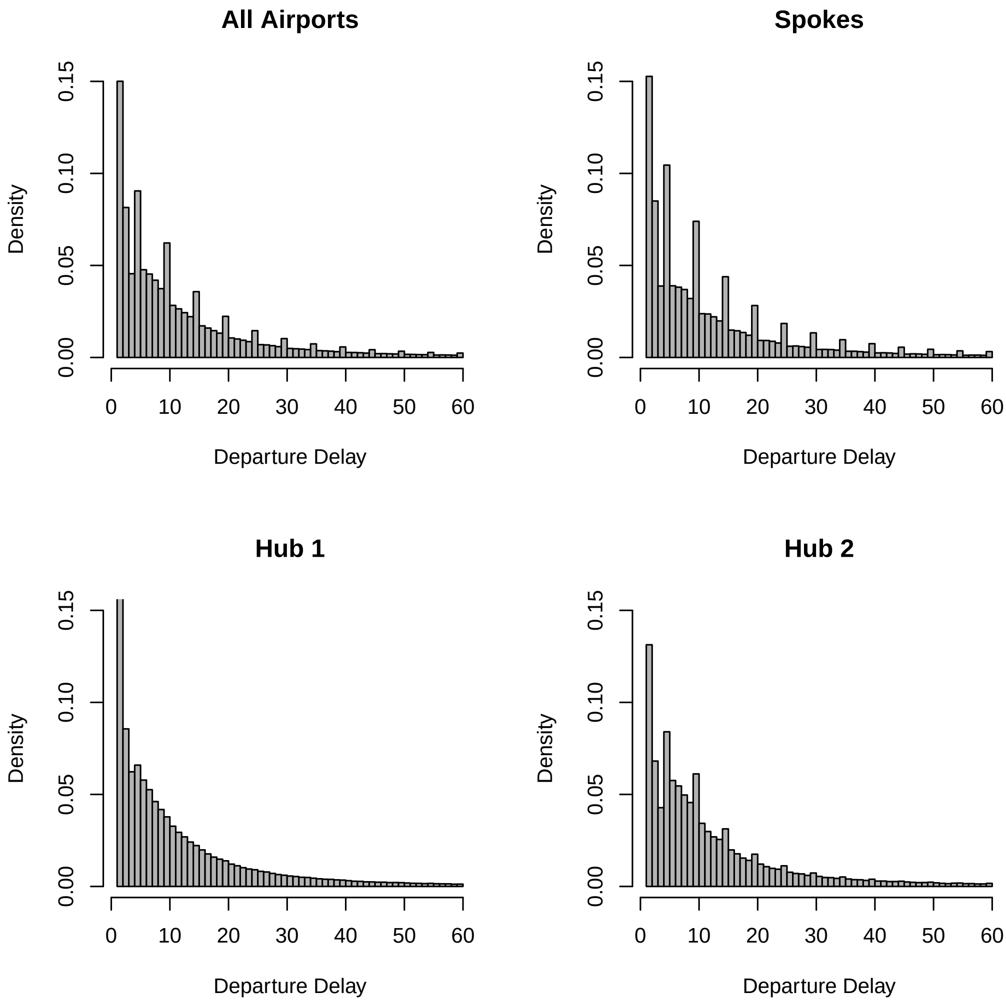

Figure 6.6: Delay record accuracy at hubs and spokes

only one delay record. Assuming that delay propagation shall be prevented by different crew pairing and aircraft routing decisions, it is not certain if the departure may be on-time due to these hidden delays. This case already illustrates that there is no straightforward solution for delay recording as these highly depend on the actual operated schedule.

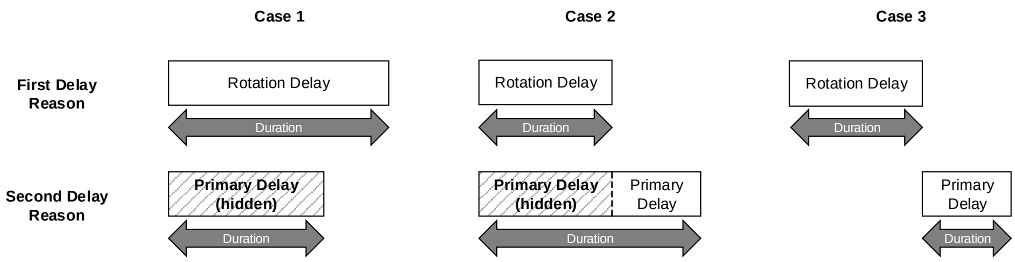

Figure 6.7: Delay reasons overlapping in time

6.2 Time Course Analysis – Examining General Trends and Seasonality

In this section, developments of the flight schedule and emerging delays in the course of time are evaluated. We discuss the schedule structure in terms of the number of offered flights, primary and secondary delays. Considering data in the time course, Hyndman and Athanasopoulos (2014) distinguish between trends, seasonal and cyclic patterns:

- **Trends** "exist when there is a long-term increase or decrease in the data."
- **Seasonality** "exists when a series is influenced by seasonal factors (e.g., quarter of the year, month, or day of the week). Seasonality is always of a fixed and known period."
- **Cyclic Patterns** exist "when data exhibit rises and falls that are not of fixed period in the long term (usually of at least 2 years)".

These aspects are discussed in the following, beginning with the examination of potential trends in the schedule structure and delay occurrences over the course of time in Section 6.2.1. Seasonality and cyclic patterns are investigated in Section 6.2.2 in close connection to the issue of potential autocorrelations. Based on the data examination in Section 6.1, the analyses are performed on a subset of 2,210,243 flight records, defined by the following restrictions:

- Since only regular daily operations are discussed, departure delays are only considered if they do not exceed the length of 180 minutes. The threshold is chosen based on experience from the descriptive analysis. 99.95% of all flight records meet this restriction.

- The analysis focuses on continental and domestic flights that are predominantly operated by Boeing 737 and Airbus 32x.
- Occasional flights between 10 p.m. and 6 a.m. are excluded since due to night curfews no systematical flight movements take place in this time span.
- In order to prevent zero-inflation, we consider delay frequency and delay length separately.

6.2.1 Trends over the Course of Time

The identification of potential trends over the course of time is an important issue when analyzing long period data. One main assumption for the application of ordinary least square (OLS) based models is the stationarity of the data. Data is called stationary if the observations do not depend on the time of record[6]. In a first step, the general flight schedule development as well as primary and secondary delay occurrences are examined. Subsequently, delay ratio and duration are addressed in more detail, followed by block times of flights and ground times for aircraft.

General Flight Schedule Trends

Figure 6.8 illustrates the total number of flights per day in the upper oscillating curve, as well as the number of primarily and secondarily delayed flights (middle and lower oscillating curves). The flight schedule itself reveals a clear recurring weekly structure that does not reflect in the graphical representations of delayed flights at first sight. We assume that the flights operated by just one airline are not a sufficiently coherent congestion indicator. Fully coping with congestion effects requires an analysis of the whole domestic or continental airspace network and airport structure, however, going beyond the scope of this work. Furthermore, at the end of every year significantly less flights are operated, obviously related to holidays at Christmas and the turn of the year. There is also a slight reduction during summer months. These general time trends and also seasonality is made visible by local estimations of a time trend (LOWESS[7]). This means that for a time point t, delays in its neighborhood are weighted in decreasing direction in order to determine its conditional mean. For

[6]For more information, see (Maindonald and Braun, 2010, p. 286).
[7]Local weighted scatterplot smoothing, see Cleveland (1981) for details.

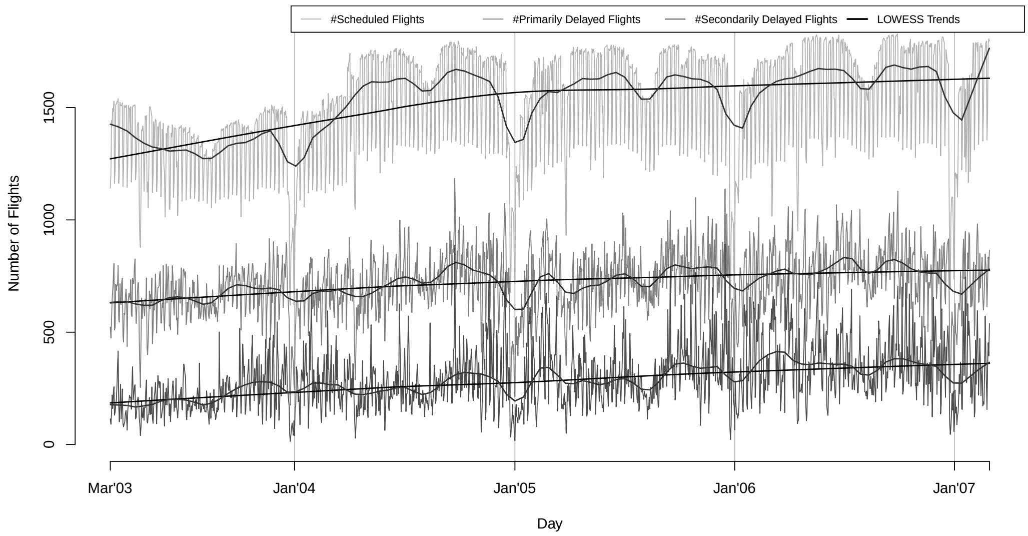

Figure 6.8: Scheduled and delayed flights over the course of time

each aspect in Figure 6.8 there are two time trends, each with a different smoothing parameter. General trends are indicated by the nearly straight lines, seasonal trends by a less smoothed lines.

The latter trend lines show a seasonal structure over the years. This important topic is discussed below. The former straight trend lines indicate an increase in offered flights over the years that goes along with an increase in the absolute number of delayed flights. On first sight, the unidirectional development of the trends indicates a quite constant delay ratio. Going into detail, Table 6.3 provides numerical representations on this topic. Abbreviations pd and sd stand for primary and secondary delays, respectively. As the available data starts in March 2003, a period is defined as the interval from March to February of the following year each for better comparability. The relative differences $\Delta(t_{i-1}, t_i$ are computed pairwise for succeeding periods. There is a considerable leap of 16.8% more flights from period 1 (Mar 2003 - Feb 2004) to period 2 (Mar 2004 - Feb 2005), presumably due to schedule adaptions as a persistent consequence of September 11th, 2001. The latter may also be an explanation for the considerable increase of primarily and secondarily delayed flights in the same in period 2. In summary, the number of scheduled flights increase

Period	1	2	3	4
Avg #Flights per Day	1319.85	1543.70	1557.04	1598.16
$\boldsymbol{\Delta(t_{i-1}, t_i)}$	-	16.98%	0.86%	2.75%
Avg #Flights with pd per Day	651.73	715.77	732.44	764.24
$\boldsymbol{\Delta(t_{i-1}, t_i)}$	-	9.53%	2.33%	4.34%
Avg #Flights with sd per Day	225.28	280.90	317.23	363.04
$\boldsymbol{\Delta(t_{i-1}, t_i)}$	-	24.35%	12.93%	14.44%

Table 6.3: Flight schedule development over the time course

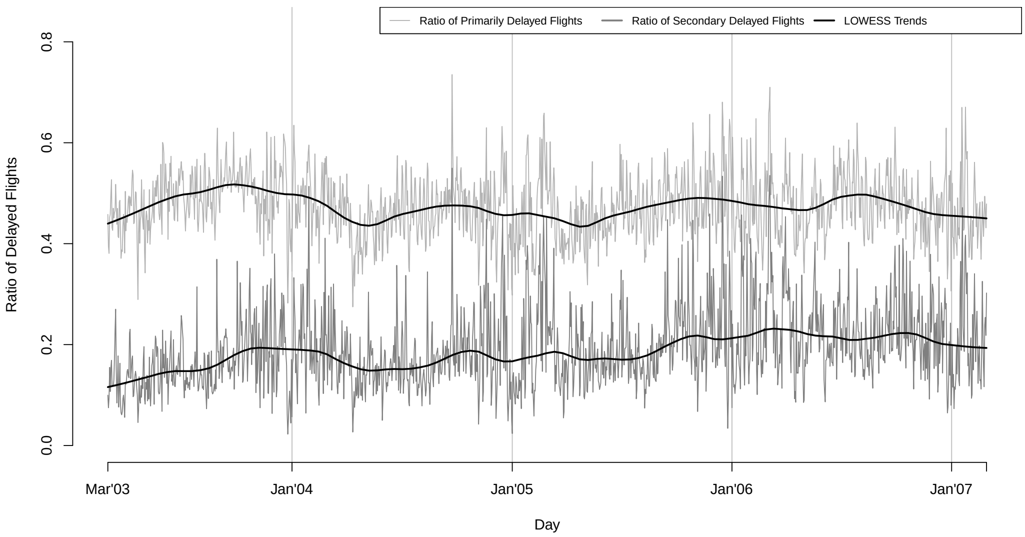

Figure 6.9: Ratio of primarily and secondarily delayed flights

over the time with a big leap between period 1 and 2. There is no obvious connection between the number of scheduled and delayed flights on a daily or weekly basis. Nevertheless, seasonal trends already become apparent.

Trends in Delay Ratio and Duration

Following the examination of the general flight schedule development in time, we examine trends in delay ratio and duration. Figure 6.9 illustrates the ratio of delayed flights per day. The primary delay ratio ranges from 27.48% to 73.51% with a

Period	**1**	**2**	**3**	**4**
Avg Primary Delay	10.75	10.55	10.49	10.83
$\boldsymbol{\Delta(t_{i-1}, t_i)}$	-	-1.81%	-0.06%	3.20%
Primary Delay Ratio	49.32%	46.30%	46.98%	47.77%
$\boldsymbol{d(t_{i-1}, t_i)}$	-	-6.12%	1.47%	1.69%

Table 6.4: Primary delay trends over the course of time

Period	**1**	**2**	**3**	**4**
Avg Secondary Delay	18.74	19.21	19.22	19.55
$\boldsymbol{\Delta(t_{i-1}, t_i)}$	-	2.49%	0.01%	1.75%
Secondary Delay Ratio	17.12%	18.25%	20.42%	22.75%
$\boldsymbol{d(t_{i-1}, t_i)}$	-	6.63%	11.89%	11.4%

Table 6.5: Secondary delay trends over the course of time

25% (75%) quantile of 43.35% (51.87%, respectively). Secondary delays occur less frequently, ranging from 2.26% to 54.96%, with 25%- (75%-) quantile of 13.75% (23.53%, respectively). In total, there are 0.396 secondarily delayed flights for every primarily delayed flight. Going into detail, Tables 6.4 and 6.5 provide numerical representations of primary and secondary delay per period. In primary delay there are no obvious patterns. This is interesting, because one may expect increasing (exogenous) delays due to the steady increase of flight offers and demand. It seems that either slack capacity is still available in the ground processes and the airspace system, or that the amount of resources grows with increasing demand. The positive trend in secondary delay becomes more obvious since it rises by 5.63% in total from period 1 to 4.

Not only the question whether a flight is delayed or not is important, but also and especially the resulting delay duration. In this regard, the average duration is illustrated both for primary and secondary delays in Figure 6.10. The delays seem to oscillate with a considerable amplitude during the entire span of time. Average primary delays range from 5.92 to 32.81 minutes, average secondary delays from 9.19 to 48.66 minutes. During winter months the average delay reaches its peak values more frequently. Again, the black lines are local estimations of a time trend (LOWESS) using different smoothing parameter. One can see that the line (1) is horizontally straight, indicating the absence of systematic changes of delay over time.

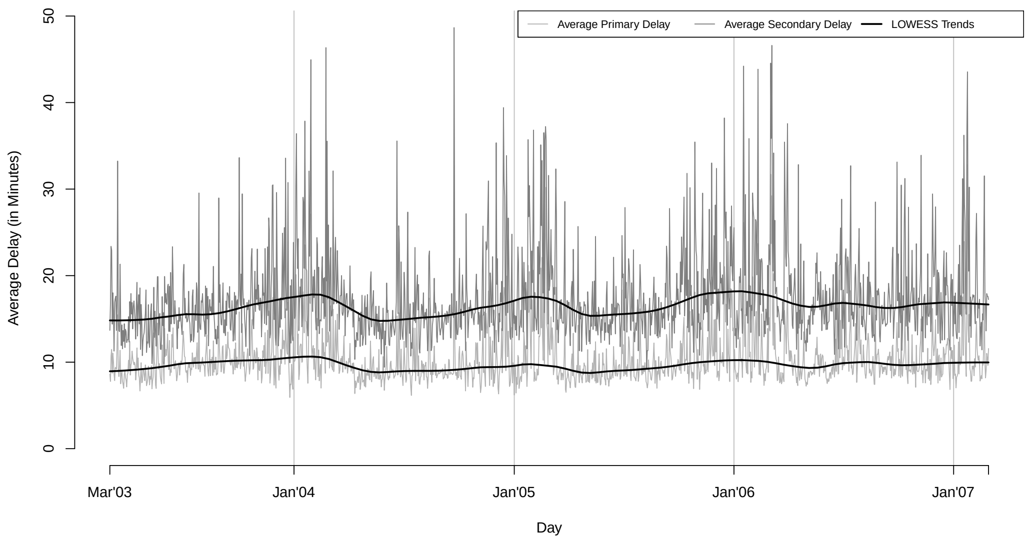

Figure 6.10: Average primary and secondary delay per day over the time course

See also Tables 6.4 and 6.5 for numerical representations per period. Line (2) with a lower smoothing parameter indicates seasonality as in winter months the trend reaches its peak values while in summer month the trend is constant at low levels. This behaviour repeats in every period. In addition, it becomes evident that the secondary follows the primary delay trend in attenuated form, clearly indicating propagation effects.

At last, we check the stationarity of of delay records over the course of time. Stationarity means that observation values are not related to the time of observation. If data is non-stationary it may cause spurious modeling parameters whose interpretation is difficult, see (Lindsey, 2004, p. 5), (Maindonald and Braun, 2010, p. 286), (Gwiggner, 2007, p. 90).

We use the Augmented Dickey Fuller (ADF) test whose null-hypothesis is that the data is non-stationary. The null-hypothesis is clearly rejected for all considered attributes. To confirm the results, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is applied which reverses the hypotheses. All data is confirmed to be stationary except for secondary delay ratio. Although the results are ambiguous, they confirm the visual indications of a slight trend in secondary delay ratio, potentially requiring differencing. One has to carefully exploit the benefit of trend adjustments and differencing of certain attributes in a complex natural system that contains more variables and interrelations than it is apparent from our point of view. It may not be applicable to perform differencing on a single (aggregated) attribute without adding unnecessary disturbance to this system. Subsequent analyses and modeling steps must take into account that interpretations of model parameters cannot be made straightforwardly. In accordance with (Gwiggner, 2007, p. 90), formal inferences about estimated parameters in stochastic models may not be performed on secondary delays.

Trends in Average Ground and Block Times

An important issue is the development of block and ground times since deviations in their duration directly affect delays. The average scheduled and actual block time per day is illustrated in Figure 6.11. Average actual block times are always shorter than scheduled ones. Seasonal patterns become apparent since in winter block times are generally longer, indicating additional seasonal buffer times. The general trend is almost constant with an average block time varying between 84.9 and 86.6 minutes per period.

Figure 6.12 depicts the average ground time between consecutive flights of single aircraft per day. In order to prevent bias, only ground times from 0 to 180 minutes are considered. Seasonality aspects are apparent, directly opposing the number of daily flights. This behavior is natural on first sight. Nevertheless, it can be assumed that days with less air traffic are not operated with significantly less resources. Referring to the general trend, the average ground time per day notably decreases by 7.5% during the course of time. Assuming that no technical speed up of the ground processes and no systematical changes in flight routes have emerged, the buffer time for ground operations decreases while airborne buffers remain quite constant. Along with the secondary delay ratio trend, it may be an indicator for ongoing optimization of aircraft rotations[8].

[8]For further examination, the development of the secondary delay ratio is plotted against average ground times in Figure C.2 in Appendix C.2. Although there is an observable contrary development in time, a straightforward conclusion on a direct cause-effect relationship cannot be justifiably drawn in the scope of this work due to the complexity of interrelated processes.

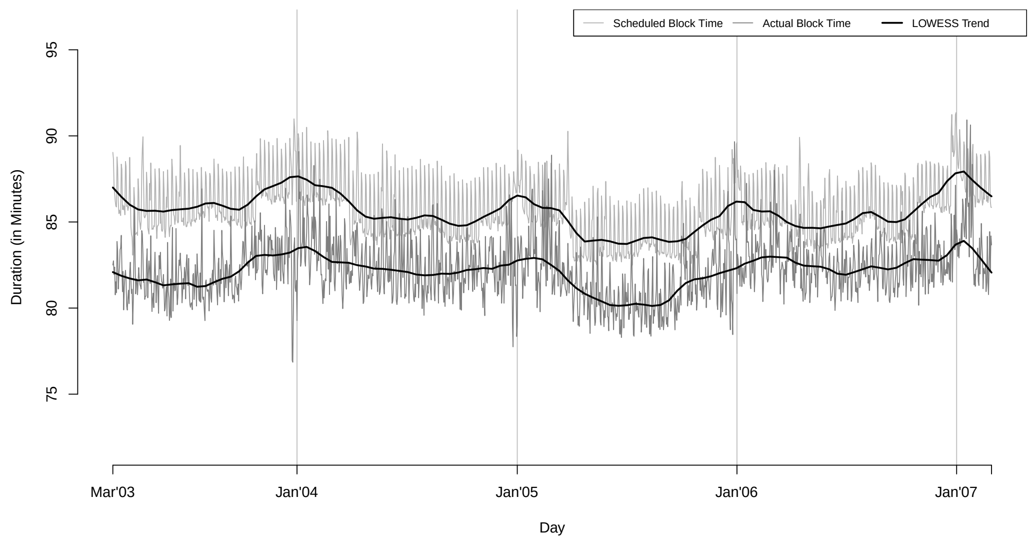

Figure 6.11: Block times of all flights over the course of time

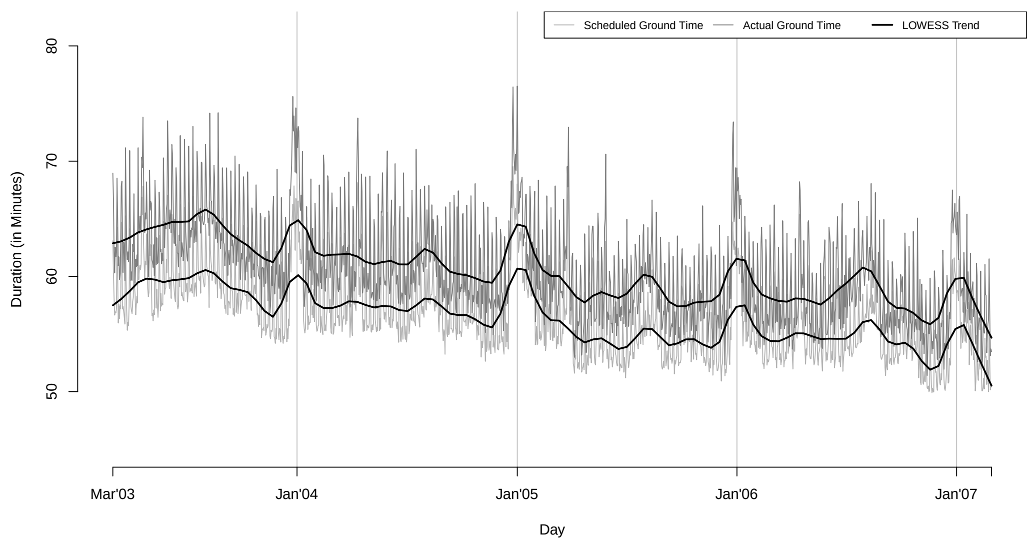

Figure 6.12: Ground times of rotations over the course of time

6.2.2 Seasonality and Cyclic Patterns

Cyclic behavior of delay frequency and length has become apparent in the time trend analysis already. In this section, we have a closer look at these patterns with special regard to autocorrelations. According to (Maindonald and Braun, 2010, pp. 284), autocorrelation means that there is a significant correlation between an observation x_t and its predecessor $x_{(t-1)}$. Moreover, an autocorrelation model of order p is defined as the regression of x_t against $x_{(t-1)}, x_{(t-2)}, ..., x_{(t-p)}$. Autocorrelations complicate the estimation of parameters such as standard errors because randomness is a key assumption of univariate statistical processes. However, autocorrelations can be used as an indicator for recurring seasonal patterns in our case.

First visual investigations can be found in lag plots. Dependencies concerning the average number of flights per day exist especially for lag 7 that is related to the weekly structure[9]. These dependencies do not show up in average primary and secondary delays per day, no meaningful patterns can be observed[10]. Further autocorrelation tests show that there are weak dependencies between Monday and Tuesday as well as between Thursday and Friday. We assume that this pattern can be explained by outward and return journeys at the beginning and the end of the working week.

Figure 6.13 illustrates (partial) autocorrelation functions for average primary and secondary delays per day. The dashed lines delimit the 95% confidence intervals for autocorrelation of a process with independent and identically distributed variables. Apparently, the data contains a strong seasonal component with recurring effects on primary and secondary delays in the long term.

Visualizing observed seasonal components on a monthly basis, Figure 6.14 shows the average primary and secondary delay per month. Each line corresponds to one year. Both primary and secondary delays reach their overall peak during winter months, but also in July there is a local peak. Figure 6.15 analogously illustrates the delay ratio per month. Note that intentionally the y-axes have different ranges in both panels since primary and secondary delay ratios appear in a significantly different scale. Effect comparable to average delay lengths cannot be observed. Although there is a certain fluctuation in primary and secondary delay ratios, a strong seasonal component is not apparent.

[9] The related lag plot is provided in Figure C.3 in Appendix C.2.

[10] See Figure C.4 for primary delays and Figure C.5 for secondary delays in Appendix C.2.

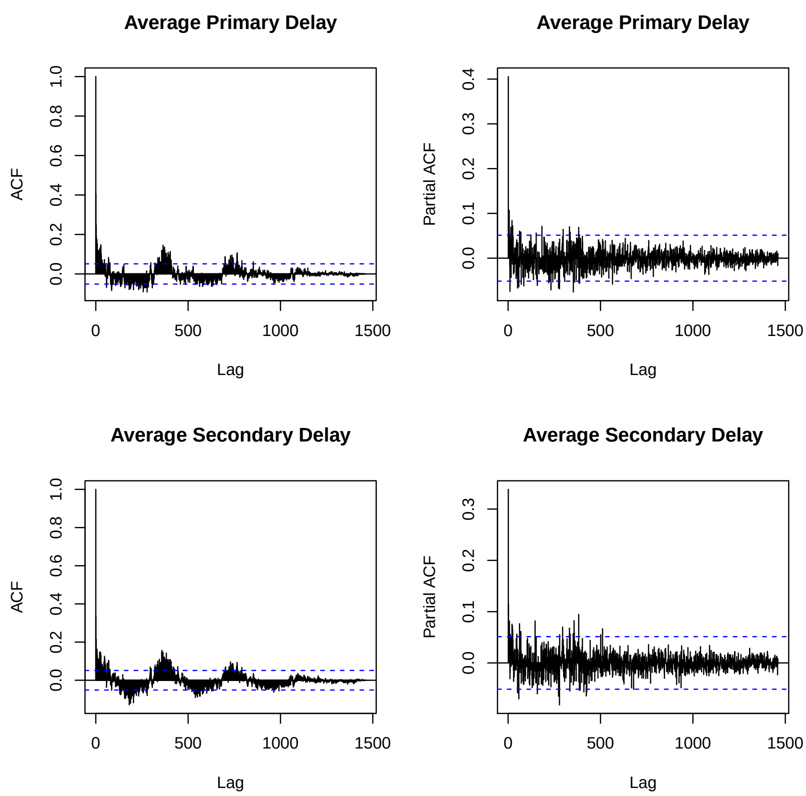

Figure 6.13: Lag plots for average primary and secondary delays per day

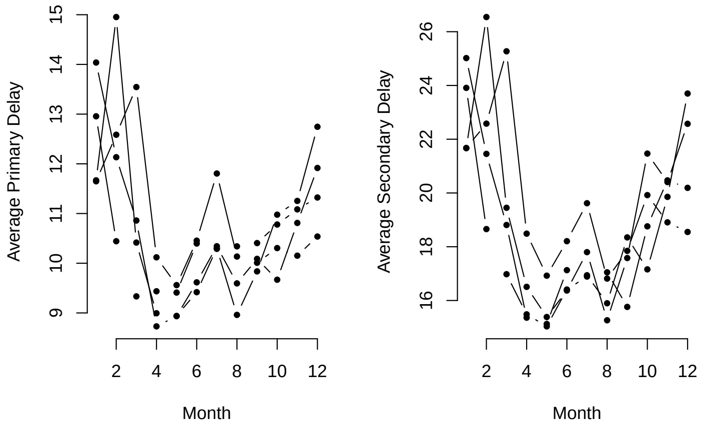

Figure 6.14: Month plots for average primary and secondary delays

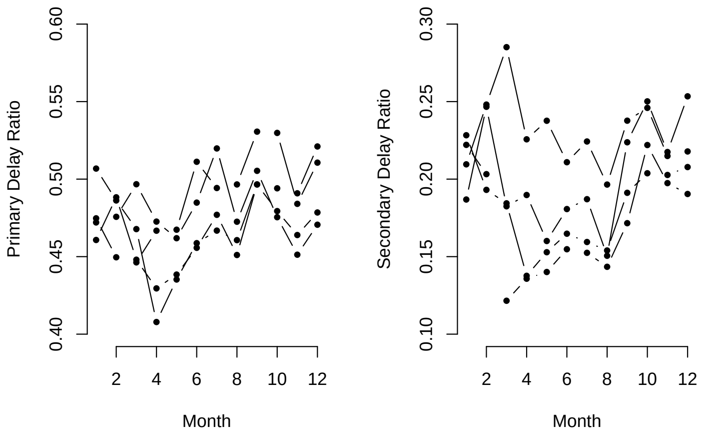

Figure 6.15: Month plots for the primary and secondary delay ratio

6.3 Summary and Implications

In this chapter, we have introduced the data set and provided details on properties of the underlying flight schedule. The flight network is a hub-and-spoke structure with two hubs with a substantial proportion of (mostly domestic) point-to-point connections. Departures and arrivals are organized in banks, accumulating inward and outward journeys at the same time from and to the hubs. The number of operated flights per day shows significant alterations per season, weekday and time of day. Regarding the way of delay recording, there is a lack of accuracy due to two reasons. Firstly, delays overlapping in time are not fully recorded. Secondly, only delays are recorded that actually lead to a departure delay. Both issues lead to partial underestimations of actual delays.

In the time course analysis, the following aspects have become apparent:

- There is a substantial increase of scheduled flights after March 2003.
- For the following years, the number of offered flights increases only slightly.
- The average delay varies within a constant amplitude over the course of time.
- There is an increasing development of the secondary delay ratio.
- Scheduled aircraft ground times between successive flights decrease over the course of time.
- Autocorrelations can be observed, implying seasonality and weekday patterns in delay length.
- Seasonality can be observed in average delays but not in the delay ratio.

Although the number of offered flights increases, primary and secondary delay duration do not change notably. Thus, a straightforward derivation of delay occurrences from the flight schedule development (e.g. in terms of congestion effects) cannot be made on a macroscopic level. Exogenous primary delay duration and ratio remain almost constant, speaking for a stable environmental system. It can be no surprise that flight and resource schedules are continuously optimized in the long run, potentially leading to the contrary development of secondary delay ratios and scheduled aircraft ground times. The related findings can be appropriately put into context by practitioners. For further analysis, the observed patterns simplify the situation, since we can concentrate on seasonal effects in absence of complicating time trends and autocorrelations. Corrections of effects related to stationarity and autocorrelations are not appropriate since indications are too weak to justify interventions on certain attributes that eventually distort the natural system's behavior. As a consequence, formal inferences about estimated model parameters will not be performed.

Chapter 7

Examining the Potential of Primary Delay Prediction

In the context of robust scheduling, primary delays are defined as an exogenous influential factor whose occurrence cannot be avoided by scheduling decisions. However, they are the initial source for any propagation effect in a resource network. If in resource scheduling, assumptions on their occurrence probabilities lack accuracy, schedule robustness may be severely overestimated because buffer times and swap opportunities are incorporated at the wrong positions in a schedule. As a consequence, a nominal cost increase leads to a smaller increase of the robustness than possible.

As observed in previous Sections 6.1 and 6.2, delay risks differ vary widely for different network- or time-related attributes such as O&Ds, season and time of day. The main goal of the following analysis is now to examine the potential of data-driven delay modeling for robust resource scheduling. Based on an initial descriptive analysis, we derive time of day patterns of primary delays from historical data and evaluate their prediction accuracy by statistical modeling. Since resource scheduling is a long- and medium-term process, the focus of interest is on spatio-temporal attributes.

The development of prediction models and decision rules takes place in a field of tension between prediction accuracy and interpretability. The *prediction accuracy* describes the relation between the model and the real data. High prediction accuracy means that there is a strong correlation between the predicted and the real value. In our context the usage of empirical distributions for delay predictions would have high prediction accuracy for short-term forecasts. However, this can lead to erroneous interpretations of the underlying mechanisms and result in wrong decision making. An alternative approach is to focus on the delay occurrence mechanisms, leading to the aspect of *interpretability*. The benefit in this approach is the understanding of underlying decision rules. The prediction accuracy might be lower as only the most important patterns are captured.

Both targets of prediction accuracy and interpretability are addressed in the following analysis. Based on an exploratory analysis on the influence of spatio-temporal patterns, we derive decision rules for the description of primary delay occurrences. For the analysis, we use the same data set as for the time course analysis of Section 6.2. The subsequent statistical model selection and assessment follows the idea of the Analysis of Covariance (ANCOVA). Interactions between attributes are explicitly taken into account, e.g. varying time of day trends for different seasons or weekdays. Final results are compared to random forests as a non-parametric, automated modeling approach. The derivation of decision rules and the generation of predictive models are closely related to the field of data mining. Data Mining is often defined as the extraction of unexpected patterns in large data sets, see Hand et al. (2001) and Hastie et al. (2009). It uses statistical and algorithmic methods for descriptive and predictive problems. Large data sets with thousands or millions of variables and observations pose challenges to formal statistical reasoning. For example, performing a large number of significance tests will reject by design a certain percentage of null hypotheses (e.g., Efron (2010)). Moreover, with large sample sizes, standard errors of estimators tend to become so small, that even *unimportant* differences between measured and true values are reported as significant. In predictive modeling, Big Data risks to favor complex models that *mimic* the sample and its statistical fluctuation, but do not necessarily extract its underlying mechanisms, see (Hand et al., 2001, ch. 4.6.2) and (Hastie et al., 2009, ch. 7). While for the purpose of short-term prediction some of these issues are resolved (for example by assessment of the bias-variance trade-off), the data mining methodology does currently not provide a sound basis for the automated extraction of interpretable patterns (Breiman (2001b); Cox (2006); Cox and Wermuth (1996)). As mentioned above, our strategy to avoid these pitfalls is to rely on descriptive methods, complemented by formal inferences whenever possible.[1]

[1]This paragraph bases on the publication Ionescu et al. (2016).

The remainder of this chapter is organized as follows[2]: In Section 7.1, we discuss statistical assumptions as necessary groundwork for the model generation. Observed correlations between the central moments for distributions (mean, standard deviation and skewness) are addressed and resulting implications are provided. In addition, distributions are analyzed regarding their goodness-of-fit. Section 7.2 deals with an exploratory data analysis. As a result, decision rules concerning daytime trends for primary delays are derived. In particular, we take into account the findings on seasonal and network-related patterns that have been carved out in Sections 6.1 and 6.2. Finally, the prediction accuracy of the rules is evaluated by a statistical model selection in Section 7.3. In Section 7.4, findings of the analysis are summarized and contextualized with regard to the entire Part II of this work. Derived stochastic models and decision rules can be used to refine generators for primary delay in robust resource scheduling and the simulation of delay propagation. The associated analysis is performed in Part III of this work.

7.1 Statistical Assumptions for Primary Delay Distributions

Prior to the exploratory analysis of patterns in primary delays, we provide details on statistical assumptions for primary delay data. In a first step, potential dependencies between central moments (mean, standard deviation and skewness) of empirical distributions for primary delays are analyzed. The examination of correlations between moments is an important groundwork for the categorization of primary delays by specific attributes in the following exploratory analysis. An exemplary category can contain all flights that depart at a hub airport in morning hours during winter months. If the standard deviation and the skewness depend on the conditional mean value of that category, the distinction of distributions according to the conditional mean will be more significant. In a second step, we examine the goodness-of-fit of problem-related density functions for the representation of primary delay distributions. Results are taken into account in the model selection in Section 7.3.

[2] Parts of this analysis (Sections 7.1.2, 7.2.3 7.3 and partially 7.4) have been published in: Ionescu, L., Gwiggner, C., Kliewer, N. (2016). *Data analysis of delays in airline networks.* Business & Information Systems Engineering 58(2), 119-133.

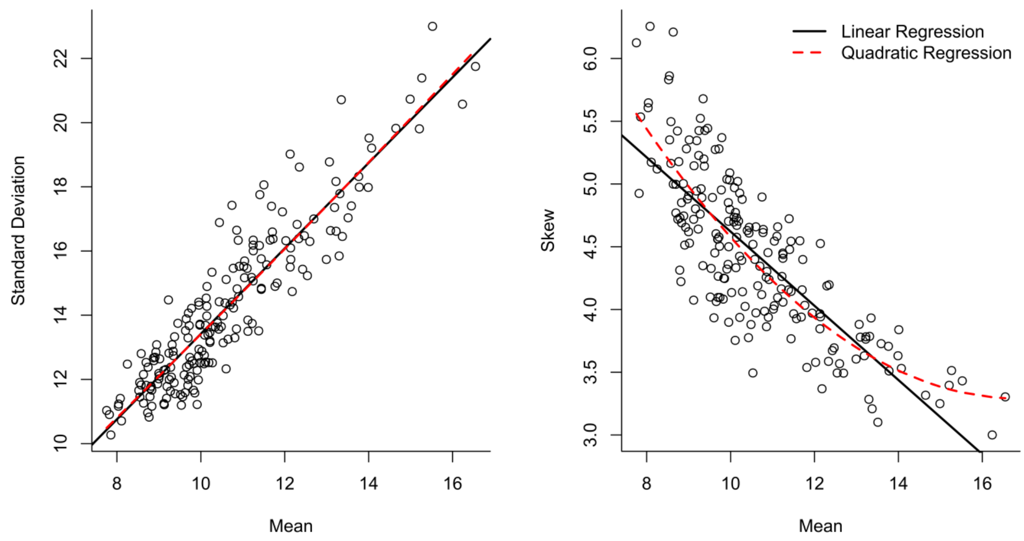

Figure 7.1: Linear and quadratic regression between central moments of distributions

7.1.1 Dependencies between Central Moments of Delay Distributions

Potential mutual impacts between central moments arise from the nature of the analyzed delay data. As the majority of delays is rather small, all distributions show a positive skewness. Due to the positive correlation of the number of small delays with the skewness and the decreasing effect on the mean value, a prediction seems reasonable. Furthermore an increasing variation of the delays can only be caused by a higher number of large delays, because the delay distribution is limited to values greater than zero. Therefore we expect a positive relationship between the standard deviation and the mean value. For the complete data set the average delay per flight is 10.70 minutes, the standard deviation is 14.48, and the skewness is 4.45. In Figure 7.1 the values of the central moments are based on distributions obtained by splitting the data by month and time of day. The 192 resulting data excerpts contain at least 1,487 flights and 3,965 flights on average. It can be observed that the correlation between the mean value and the standard deviation tends to be strong. The results of the OLS-regression in Table 7.1 underline the first impression, as 84%

Model	$\widehat{\textbf{Int.}}$	$\hat{\beta}_1(\bar{x})$	$\widehat{\beta_2}(\bar{x}^2)$	R^2	Std.Err.	Std.Err.	BIC
$\hat{\sigma} \sim \bar{x}$	0.104	1.332***	–	0.84	0.042	–	563
$\hat{\sigma} \sim \bar{x} + \bar{x}^2$	0.833	1.201**	0.006	0.84	0.417	0.018	568
$\hat{\upsilon} \sim \bar{x}$	7.577***	−0.296***	–	0.63	0.016	–	199
$\hat{\upsilon} \sim \bar{x} + \bar{x}^2$	11.077***	−0.924***	0.027***	0.66	0.155	0.007	188

Table 7.1: Regression model parameters

of the variance of the standard deviation can be explained by the mean value. As the quadratic regression does not provide more information of the relationship, the linear model seems to be more convenient. Despite the lower explanatory power of the regressions between the mean value and the skewness, the test results still prove a convincing relationship. In general, the standard errors are very small for all models. Due to its lower BIC-value, the quadratic regression can be preferred in this case. For a further description of the Bayesian information criterion (BIC) we refer to (Hastie et al., 2009, pp. 233).

The mean value determines the moments of distributions to a large extent. Alternative data splits do not show significant differences in the represented behavior. Due to this reason the subsequent descriptive analysis concentrates on the conditional mean value for describing delay occurrences.

7.1.2 Statistical Distributions for the Description of Primary Delay Data

In the absence of strong autocorrelations and time trends, we empirically identify density functions that describe the delay during the different seasons. A first visual indication for well-fitting distributions is obtained by quantile-quantile-plots with a family of event-related distributions, see (Lindsey, 2004, ch. 4). The log-normal, log-logistic and the Weibull turned out to be reasonable candidates.These distributions are fitted by Maximum Likelihood to the empirical data. The left panel of Figure 7.2 illustrates an exemplary fit for summer months (May, Jun, Jul and Aug). Log-normal and log-logistic seem to fit slightly better than Weibull. These results are consistent with (Lan et al., 2006, p. 19) who also consider these distributions as there are many small delays and only few very large delays. Taking the logarithm of the data leads to good fits with the normal and logistic distributions (right panel of Figure 7.2).

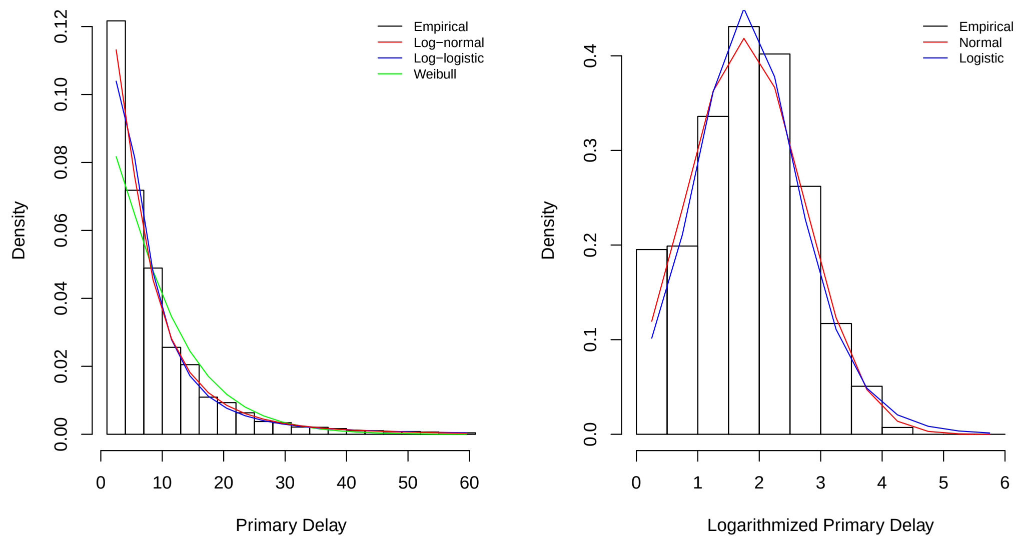

Figure 7.2: Distribution for primary delays in summer months

The logistic has a slightly better fit. However, there are crucial differences of both distributions for small values between 0 and 1. This can be explained by the data quality: only delay larger than one minute was considered, and the measurement unit is in minutes. Therefore, the logarithm for values smaller equal one will be distorted. The results can be reproduced for other data excerpts, i.e. not just for summer months.

Conventional χ^2-tests are not suitable for determining goodness-of-fit of theoretical distributions for large data as already small differences between observed and theoretical frequencies lead to a rejection of the null-hypothesis. For example, the null-hypothesis for the log-normal distribution is rejected on a 5% level at values larger than 47.40 with 33 degrees of freedom. Our sample statistics has a value of over 2000. This problem is already known from Berkson (1938).

7.2 Cyclic Patterns of Primary Delay Occurrences

In this section, we provide the findings from an extensive exploratory analysis for retrieving cyclic patterns in primary delays that shall eventually be categorized by spatio-temporal and network attributes of flights. Location parameters of a flight are related to its departure and arrival airports. Network-related attributes consist of O&Ds and flow directions in the hub-and-spoke network. Temporal attributes can be split up into multiple layers. Seasonal attributes are seasons, months and weeks. Weeks can be divided into weekdays or, more generally, into working days (Monday to Friday) and weekend days (Saturday and Sunday). The local time, measured at each departure airport, is used to determine daytime trends. Finally, the local departure time is used to determine fluctuations and trends over the time of day. Hourly bins are used for the departure times in the remainder of the analysis. Smaller intervals have resulted in comparable results.

The analysis intentionally focuses on the hub-and-spoke network structure rather than individual airports. In particular, the number of flights at individual spoke airports is so small that it is impossible to derive general patterns for them. The same holds for an entirely route-based evaluation. In order to prevent zero-inflation, we distinguish between delay occurrence ratio and delay length. In this regard, 47.59% of all flights are primarily delayed. The overall average primary delay is 10.67 minutes. Systematic dependencies between congestion indicators, such as the number of passengers and the primary delay length, cannot be observed. This is another indication that the airline has already eliminated predictable delay in its schedules.

7.2.1 Location- and Network-based Attributes

The average primary delay is 10.02 minutes at Hub 1 and 11.57 minutes at Hub 2. Figure 7.3 represents the average primary delay as well as the quantiles per airport. Especially the .95-quantile values vary more widely at very small (mostly regional) airports, often approached less than two times a day. We assume that in these cases bias is dominant to a potential systematic pattern. On this topic, see also the discussion on inaccurate delay recording in Section 6.1.5. Since the hub-and-spoke structure is a significant factor for all operations, we investigate the

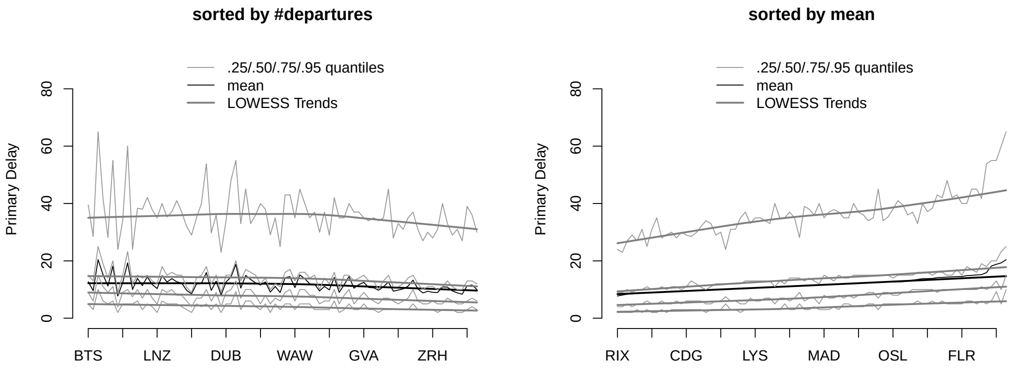

Figure 7.3: Mean and quartiles per airport

Direction	Avg Primary Delay
hub-to-spoke	10.75
spoke-to-hub	11.16
hub-to-hub	10.31
spoke-to-spoke	9.66

Table 7.2: Average primary delay per direction

ability to describe primary delays by network attributes. At Hub 1, the average primary delay is 10.02 minutes, at Hub 2 11.57 minutes and at spoke airports 10.60 minutes. Moreover, Table 7.2 offers average primary delays per direction. Note that hub-to-hub connections only cover the route between the two hubs.

Further disaggregation can be performed based on O&Ds. Deviations are apparent with an average primary delay ranging from 6.51 to 14.21 minutes for routes flown more than 5,000 times. The results are in the order of the ones obtained by the direction-based analysis and confirm its findings since the average primary delays per O&D are almost normally distributed around the mean value for each related direction type.

7.2.2 Time-based Attributes

Seasonal trends have already become obvious in the Time Series Analysis in Section 6.2. Primary delays are generally longer in winter months. Moreover, there is a

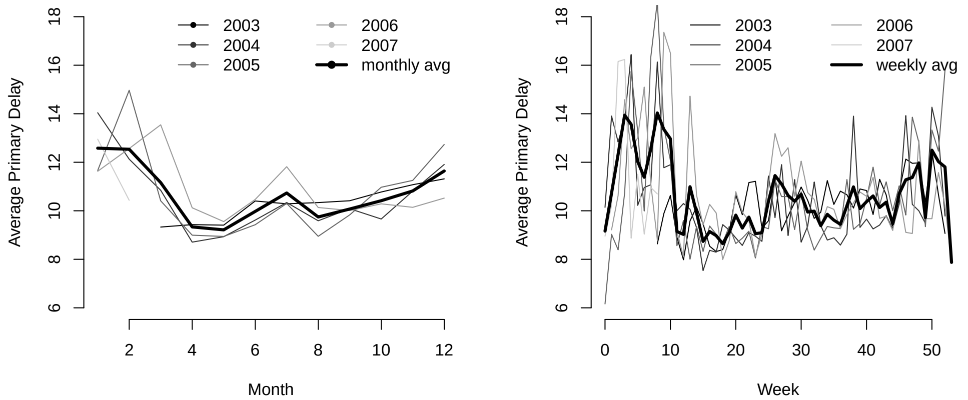

Figure 7.4: Seasonal patterns in primary delays

seasonal peak in July. Figure 7.4 depicts the average primary delay per month (left panel) and week (right panel). The same general pattern can be observed. The weekly structure offers a more detailed pattern, but also increased bias since yearly structures may be different, e.g. due to deviations in holiday dates. These cannot be exactly reflected by week numbers.

Considering data splits by weekdays, average primary delays have their peak values on Thursday (11.22 minutes) and Friday (11.16). The delay ratio follows this trend. While the delay length decreases at the weekend, the delay ratio does not significantly change in this time. However, this general weekly pattern does not show up every month. During winter months there is a larger decrease at the weekend. In February and June peak values are reached on Wednesday. Further monthly differences are very unique and do not follow any pattern.

7.2.3 Network- and Time-dependent Daytime Trends

The last step is the examination of trends depending on the time of day. In the overall course of day peak values of primary delays are reached in the morning and early evening hours. During the afternoon, the lowest values are measured. In the following, this pattern is further refined. Figure 7.6 exemplarily shows daytime trends in the different months (Jan to Dec from left to right) for spoke-to-hub flights

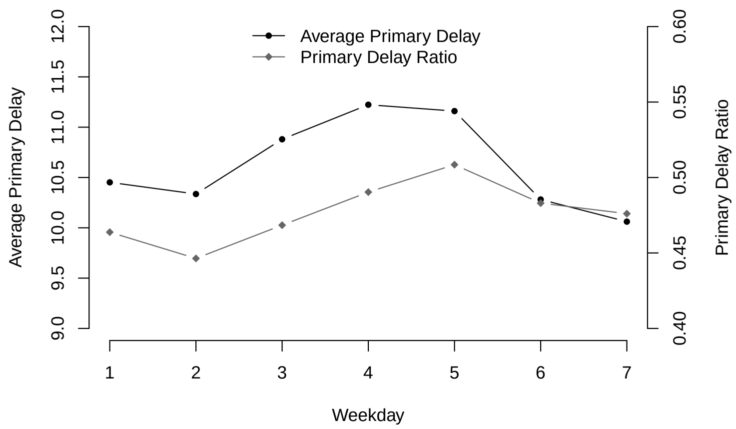

Figure 7.5: Weekday patters in primary delays

at working days. Thinner lines in the background indicate the conditional average delay. The overlapping bold lines are the result of linear regressions. Note that the vertical axis depicts the logarithm of the actual delay. In general, delay either grows or decreases during the day. Most months show a negative daytime trend as longer delays occur more often during morning hours. In contrast, in the summer months (May, Jun, Jul and Aug) a reverse daytime trend can be observed as evening hours display a higher average delay than morning hours. In conclusion, the daytime trend differs between months.

Systematical daytime trends can also be observed for different categories, i.e. for other flight directions and for the weekend. The daytime trends for these categories are similar with a slightly lower explanatory power. For the presentation of daytime patterns we use months as a seasonal attribute. The consideration of weeks instead of months is slightly more precise, especially for the location of transition points between winter and summer months. The summer cycle begins in week 17 (mostly end of April) and ends in week 36 (beginning of September). Without an exception, winter and summer weeks always follow their seasonal daytime trend.

Figure 7.7 illustrates spoke-to-spoke flights with the additional distinction between weekdays for the summer months and the rest of the year, respectively. The expected increasing daytime trend for summer is not valid for Monday and Saturday. All other

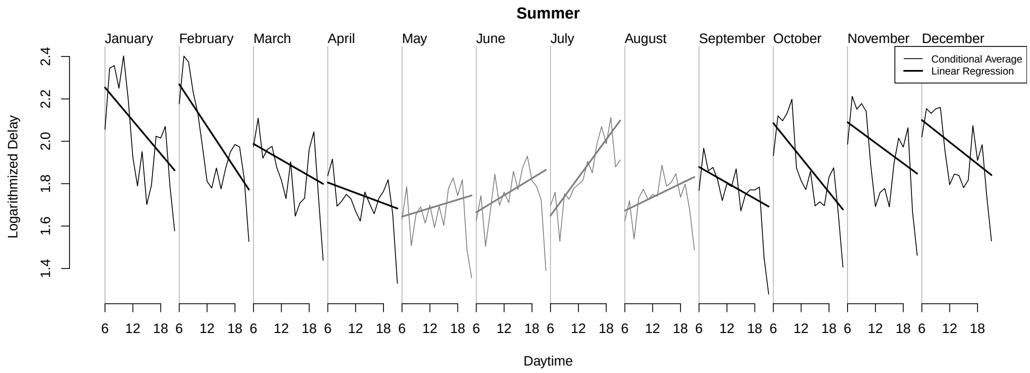

Figure 7.6: Daytime trends per month for flights into hubs

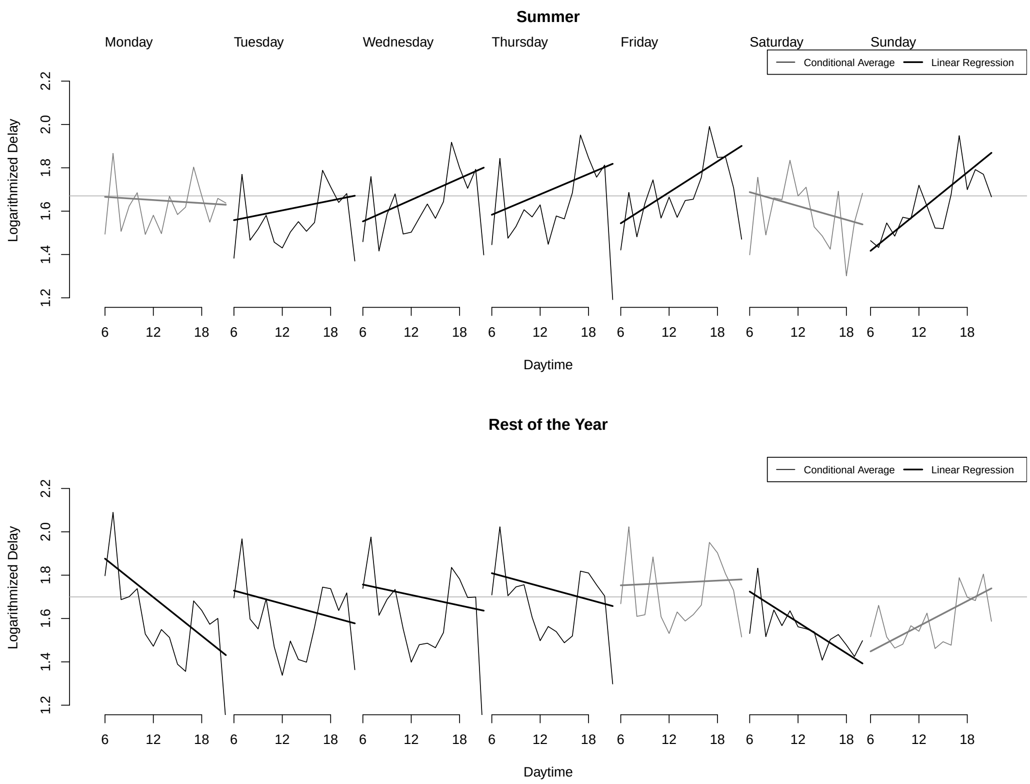

Figure 7.7: Daytime trends per weekday for spoke-to-spoke connections

weekdays, however, show the previously observed seasonal daytime trend. In the rest of the year, Friday and Sunday show a behavior that differs from the expected seasonal daytime trend. Additionally it has to be said that for hub-to-spoke flights the dependency of weekdays can be observed, too – in the summer months there are negative daytime trends for Monday and Friday. However, the daytime trend for hub-to-spoke flights during the rest of the year is still slightly negative for Friday and Saturday, though it almost flattens out. Spoke-to-hub flights do not show a dependency on weekdays as their daytime trends follow the seasonal trend both in winter and summer. We assume that there are peaks in the week structure that overlay the seasonal trends. On Monday morning and Friday evening there are peak values due to increased demands. By contrast, on Saturday evening very low demand levels are expected. An important fact is that spoke-to-hub flights do not show these effects as they monotonously follow the daytime trend of the current season for every weekday. We suppose that this difference is an implication of the fact that the hub is not the final destination for most passengers.

Summarizing the above, the following decision rules can be derived from our exploratory analysis:

1. Regarding the average delay per hour, there is a positive daytime trend during the summer months (May, Jun, Jul, Aug), except on Mondays and Saturdays in case the arrival airport is a spoke.

2. By contrast, a negative daytime trend can be observed during the rest of the year, except on Fridays and Sundays for spoke-to-spoke connections.

We validate the rules for every single day. The first rule is valid for 57.86% of all considered days, the second for 65.46%, respectively. The weighted average for all days is 62.90%. Interestingly, by taking into account seasons only, the error increases by just 2%. The results align with results obtained by a CART analysis where the error remains almost constant when forcing additional splits beyond the seasonal one. This first evaluation resulting in poor validity of the rules strongly demands for a modeling approach that is capable of capturing the complexity of these mechanisms.

Another important topic is that rough data splits with only a few classes already give hints on systematic behavior that can be refined by more detailed splits. Examples have been found in monthly and weekly as well as direction- and O&D-based

splits. This may be related to the bias-variance trade-off (Hastie et al. (2009)[pp. 219-221]), potentially emerging in the following model selection step. We also assume that existing patterns in primary delay occurrence are already taken into account by airlines in scheduling and systematical behavior can only be depicted to a lesser extent. Accordingly, as we will see in Section 7.3.3, the remaining signal-to-noise ratio is rather low in the data.

7.3 Model Selection and Assessment

In the previous sections we identified seasonal and monthly daytime trends that were positive in summer and negative during the rest of the year. We also discovered that on a daily level, these trends sometimes deviate from their seasonal component: Mondays and Saturdays during summer show a negative daytime trend, whereas Fridays and Sundays during winter show a positive trend. In this section we set up statistical models to quantify the predictive power of these findings. In our context, a statistical model is a model of the joint distribution of the observed data, along the common definitions such as Cox (2006) or Hastie et al. (2009).

More precisely, our problem is to model daytime trends in a number of spatio-temporal categories, such as flight directions, weekdays, and a seasonal component given by seasons, months or weeks. This is commonly referred to as ANCOVA (Analysis of Covariance). Two particularities of this approach are

- **Daytime trends instead of average values:** In an analysis of variance (ANOVA), the average value is estimated for each category. Categories with significantly different average values are identified by hypothesis tests. In our models, a daytime trend is fitted instead of the simple mean values. Such an analysis with a mix between categorical and continuous explanatory variables is called Analysis of Covariance (ANCOVA), see (Lindsey, 2004, p. 20).

- **Interactions:** Classical ANCOVA models introduce different intercepts for each unique category only. This gives us for example one daytime trend for summer and a shifted one for winter. As identified in the previous chapter, these trends may differ in slope across the categories; summer trends are positive and for winter months they are negative. Such patterns can be modeled by interactions between the continuous and categorical covariates. Then, the prediction for time t in the k-th category is

$$\mu(t) = \beta_{0k} + \beta_{1k} t, \tag{7.1}$$

corresponding to a linear model with intercept β_{0k} and slope β_{1k}. Instead of linear time-trends, we will later also fit cubic splines, corresponding to a basis expansion of the form

$$\mu(t) = \beta_{0k} + \sum_j \beta_{jk} f_j(t), \tag{7.2}$$

where f_j is the j-th transformation of t. Note that the interaction between the k-th category and the time-variate is captured by the parameter β_{1k}.

The main assumptions in these models are their additive structure and their stochastic behavior. This means that for every category, the response is considered to be a random variable with mean being a function of time and constant variance. Instead of Gaussian variables, we assume log-normal variables. Other distributions, especially those belonging to the exponential family, are natural extensions to our approach. Note that we do not perform significance tests on estimated parameters, but only assess prediction accuracy of our models. Thus, distributional assumptions, including independence in the residuals, are not required at this stage of research (see Weisberg (2005) for a discussion on stochastic assumptions in regression, and where they are needed). Auto-correlated data is traditionally modeled by time-series analysis. However, autocorrelation can sometimes already be explained by appropriate time-dependent covariates, see (Lindsey, 2004, p. 10). Based on the descriptive analysis, on our purpose of the models and on the fact that we perform a macroscopic analysis, we believe that neglecting possible autocorrelation can be justified.

The remainder of this section is organized as follows: In a first step the predictive power of categorical variables of the identified daytime trends is determined. In order to maintain relative comparability, single models for the whole data set are considered. Subsequently, a residual analysis is performed and the prediction error on unknown data is estimated. All model fitting is carried out in the R programming language (version 3.1.2) on an Intel i7 950 Quad-Core processor with 3.07GHz and

		Model	ΔAIC	ΔBIC	#Parameters
	L0	Mean Value	-	-	1
	L1	D	-462	-450	2
	L2	D°S	-2,038	-2,016	4
Linear	**L3**	D°M	-2,355	-2,126	24
Models	**L4**	D°W	-3,332	-2,371	108
	L5	D°S°DI°WE	-1,031	-1,305	84
	L6	D°M°DI°WE	-5,142	-338	504
	L7	D°W°DI°WE	-	-	2,268
	S1	D(3)°S°DI°WE	-4,614	-3,927	168
Spline	**S2**	D(5)°S°DI°WE	-2,779	-1,819	252
Models	**S3**	D(15)°S°DI°WE	-	-	672

Table 7.3: Structural Model Selection

24GB RAM. The most complex models can be estimated within several minutes; however, the bottleneck is the availability of memory.

7.3.1 Structural Model Selection

The first step deals with the selection of the model structure, and especially with the determination of the categorical variables. All models are fitted by maximum likelihood, more precisely by the iteratively reweighted least squares method. Table 7.3 illustrates the results. All models are based on delays on a log-scale. As a measure for the predictive quality of the models we penalize the likelihood by model complexity with the known information criteria AIC and BIC, see for example (Hastie et al., 2009, pp. 230). The AIC is used to give the relative quality of different models on a given set of data. The differences of the AIC values are given as the differences to the previous model each, except for model S1, referring to L5.

We use the nominal parameters S (season), M (month), W (week), WE (weekday) and DI (direction) to determine the category. A linear regression is then fitted for the interaction between the categorical variables and the continuous variable D (daytime). The simplest models are a single daytime trend for all levels and categories (L1), and one daytime trend per season (L2). AIC already improves by 462 and 2,038 units. Allowing for a daytime trend per month and week improves the AIC by 2,355 (L3) and additional 3,332 (L4) units. Our decision rule of the previous chapter, namely that daytime trends are permitted to differ among seasons, weekdays and directions, clearly improves the fit (L5, L6). Note that L6 already contains 504 regression parameters, and that model L7 cannot be computed anymore on the available system.

In order to further improve the prediction accuracy, we fit cubic splines (models S1 and S2) instead of linear trends (L5). Fitting regression splines is still linear in the parameters. What makes the difference, however, is a previous transformation of the continuous daytime variable according to a basis expansion, see (Hastie et al., 2009, ch. 5.2). An example can be seen in Figure 7.8 where we display the linear trend and two splines for a smaller data excerpt with the highest possible degree of freedom (15). The more degrees of freedom, the more splines are allowed. Numerical experiments showed us that the highest possible degree of freedom for model L5 is 5. The improvement in AIC and BIC is considerable, and higher values up to the maximum of 15 are in principle desirable. However, the resulting models can no longer be estimated or interpreted due to their complexity. Due to the same reason, the application of splines in models L6 and L7 is not possible.

In summary, the results confirm the observations of the exploratory analysis. All previously identified categorical variables lead to a considerable improvement of the prediction accuracy. Cubic splines improve the linear trends within the categories, although the most complex models can no longer be computed. However, this can be done in the following model assessment step. We concentrate on models represented by D(15)°X°DI°WE, where X defines the seasonal component (season, month or week).

7.3.2 Residual Analysis

The main assumptions for our regression models were that for every category and every time point t, the data can be described by a log-normal distribution with the mean value as a cubic spline function of the time t and constant variance across the time. If these assumptions are true, then the regression residuals, i.e. the differences between the predicted and observed values, have zero mean and follow a normal distribution with constant variance across time (normal because the logarithm of the delays was taken). A non-parametric bootstrap was performed to validate these assumptions. Figure 7.9 shows typical results for a category of the summer months, taking into account only flights into hubs. On the horizontal axis, the daytime is

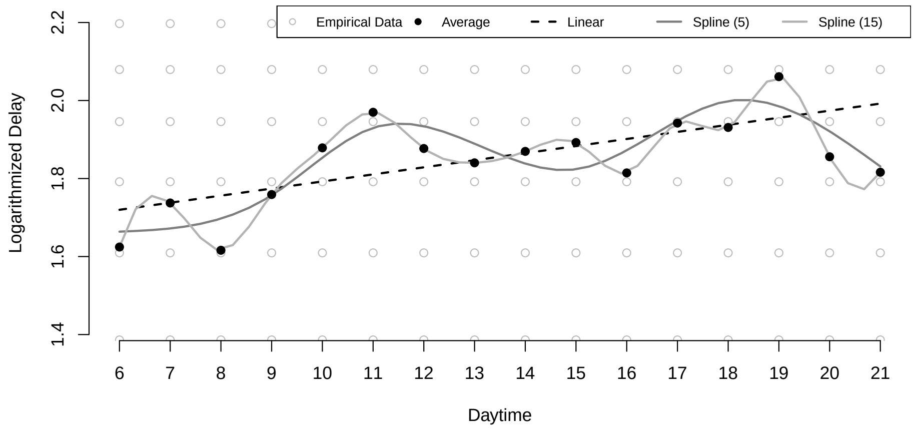

Figure 7.8: Exemplary Spline Models for Daytime Trends

displayed in hourly slots. On the vertical axis, bootstrapped statistics of the residuals are displayed.

The black line depicts the residual averages. They follow a straight line on the zero value, thus the spline model does indeed capture the conditional mean of the data. The blue lines are pointwise estimates of the residual standard deviation at time t. If the homoscedasticity assumption is true, then they should be constant overtime. This is reasonably the case, although care should be taken regarding a few time-points, e.g., at $t = 12$ or $t = 21$. Finally, the red lines depict 16% and 84% quantiles of the residuals. They were selected according to those values which match the standard deviation of a normal distribution. Therefore, if the blue lines correspond to the red lines, the normality assumption is reasonable, at least to the second order. The 84% quantiles (upper line) meet this condition very well. For the 16% quantiles (lower line), differences to the standard deviation in the order of 10^{-1} are the rule, not the exception. This means that the model does not accurately predict delays that are smaller than the average delay at time t. This finding also corresponds to the poor fit of the normal distribution on the left tail in Figure 4, although due to the regression function, no general relationship between marginal and residual distributions exist.

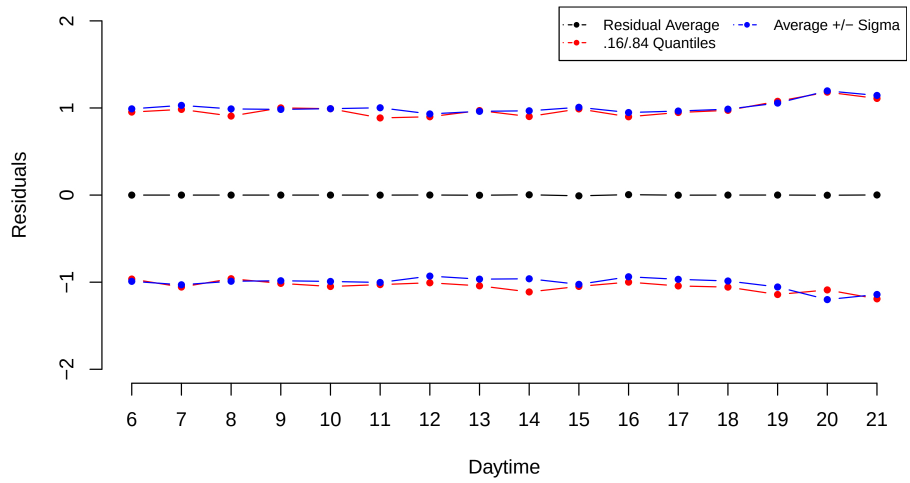

Figure 7.9: Typical behavior of the residuals per daytime

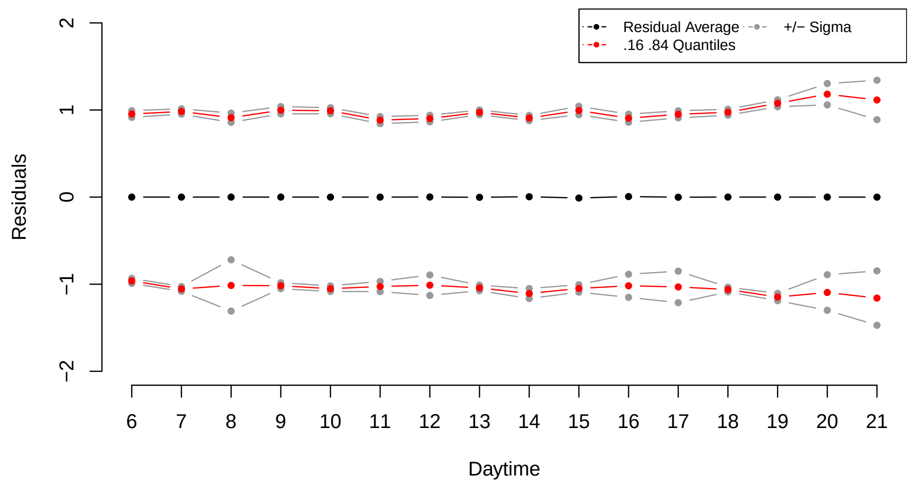

Figure 7.10: +/- Sigma for the .16 and .84 quantiles

The quality of the bootstrap estimates is also assessed. It turns out that the standard errors of these estimates are in the order of 10^{-2}, see Figure 7.10. We conclude that the model assumptions are reasonably met for large delay predictions and require care for small delay predictions. Particularly for small categories, e.g., when splitting up the data by season, flight direction and weekdays, the effects of the left tail become apparent. Note that formal statistical inference about model parameters are not the target of this analysis, thus independence assumptions of the residuals are not verified.

7.3.3 Prediction Accuracy

While the structural model selection above is based on the analytical information criteria AIC and BIC, this section deals with a resampling technique to validate the prediction accuracy of the best structural models. As in this step the relative comparability is not of primary concern, the models can be split up into smaller and therefore less complex categorical models. The parameter estimations remain the same.

We follow the approach to model assessment, as described in (Hastie et al., 2009, ch. 7). The target of this approach is the estimation of the expected extra-sample prediction error (EEPE), the prediction error that is independent of a given training data set. For comparison, the in-sample error (IE) for the training set is provided. In analogy with the idea of ANOVA, the EEPE of our model m is computed as the residual sum of squares $RSS_{m,\tau}$ for each category concerning a validation set τ. It is compared to the $RSS_{a,\tau}$ of a model a that predicts the average value of each category, respectively. Then, the improvement factor

$$imp_{m,\tau}^{EEPE} = 1 - \frac{RSS_{m,\tau}}{RSS_{a,\tau}} \tag{7.3}$$

gives the amount of the variance in a validation set τ that can be explained by our model and thus the improvement of the EEPE obtained by model m. Concerning the IE, the computation of imp_m^{IE} follows analogously for the training set.

For the analysis we repeatedly split the data into training and validation data (70/30) in a large number of runs and estimate the corresponding test errors. In the end we average them for all categories, weighted by the respective number of flights.

	X	#Cat.	∅#Flights	EEPE<0	IE	EEPE	Abs
E1	S	42	23,365	-	2.14%	**1.95%**	0.964
E2	M	252	3,894	1.50%	3.02%	1.93%	0.971
E3	W	1,134	865	7.1%	**5.73%**	1.23%	0.984

Table 7.4: Model Evaluation

It turns out that with 100 runs, the averages converge towards a stable number. Table 7.4 shows the results for the models D(15)°X°DI°WE, where X is one of the seasonal variables S, M and W. Other models from section 7.3.1 are dominated by these models in both IE and EPEE.

On average, the EEPE can be improved by 1.95% for model E1. In this model, 7.5% of all categories show an EPEE improvement of more than 3%. One can also observe the bias-variance-trade-off by means of IE and EEPE. While the best in-sample error can be achieved by choosing more categories (E3), these models perform quite poorly regarding the EEPE. Furthermore, some categories in E2 and E3 have a negative EEPE value, especially in categories containing a small number of flights. In particular, these are the categories for spoke-to-spoke flights and during the weekend. It turns out that this effect is merely associated with the small category sizes. For an illustration, see Figure 7.11 which exemplarily shows the relative improvement of the EPEE under the squared error for model E2. The horizontal-axis shows the number of flights within the category referred to. As we consider flights that are delayed at least by one minute, it is obvious that categories with hub-to-spoke flights are larger than those with spoke-to-hub flights, as the former are more likely to be delayed. Due to the hub-and-spoke structure there are less direct connections between spoke airports. Regarding the IE, the results are comparable with the ones obtained by non-parametric random forests that independently grow a large number of regression trees by repeated bootstrap sampling of the data, see Breiman (2001a). In our case, random forests are applied independently from our previous decision rules. However, they cannot be applied to the whole data set due to memory restrictions, and repeated sub-sampling does not provide reasonable results. The same holds for the seasonal model (E1). An application based on monthly (E2) and weekly (E3) basis results in IE values of 3.71% and 6.39%, respectively. It becomes apparent that these results are slightly better but still in the order of the ones obtained by our

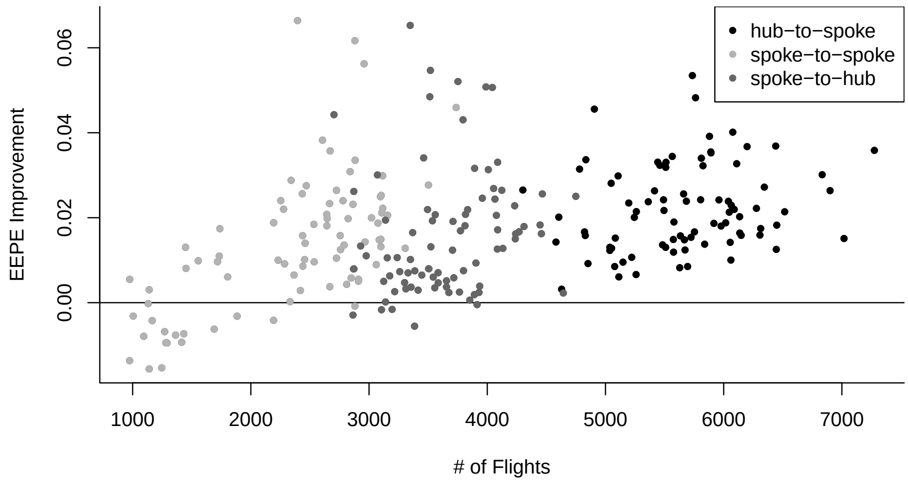

Figure 7.11: Improvement per category for model E2: D(15)°M°DI°WE

modeling approach. Prediction accuracy estimation for unknown data is performed internally in random forests by out-of-bag sampling, see (Hastie et al., 2009, pp. 592). For the prediction of x_i trees are used that do not contain the observation for x_i in their bootstrap sample used for growth. The prediction accuracy can be increased by 3.20% (E2) and 4.13% (E3), respectively. These values are expected since, in contrast to our decision rules, random forests implicitly provide an individual dynamic rule selection for all categories – leading to higher prediction accuracy in conjunction with a lack of interpretability. Finally, the absolute deviation between observed and predicted delay is a metric that is easy to interpret, since its unit is in minutes. While the average absolute deviation of the mean model L0 is 8.29 minutes, it can be reduced by nearly one minute to 7.32 by using the best ANCOVA-model E1. The best categories even show an improvement of the average absolute prediction error of 2.5 minutes, see Figure 7.12. These results are based on the presented daytime trends only. The absolute improvement can be used for estimating the benefit in real costs by linking them to specific airline's delay cost rates.

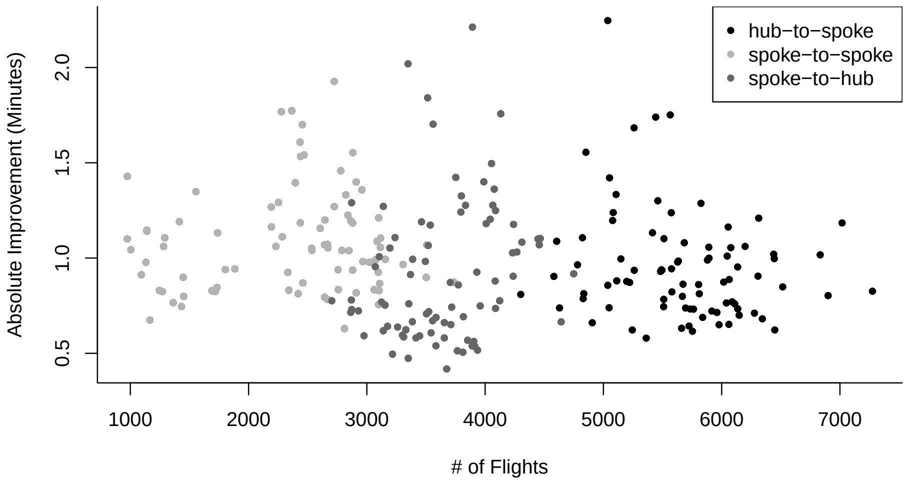

Figure 7.12: Absolute improvement (in minutes) per category for model E2: D(15)°M°DI°WE

7.4 Summary and Implications

During airline operations, exogenous disruptions often lead to delay that may result in infeasible resource schedules. The estimation of delays based on historical data has become an increasingly important topic in the context of robust airline resource scheduling. A better understanding of delay occurrence mechanisms may lead to a better trade-off between (nominal) cost-efficiency and robustness and is therefore the purpose of this study. We provide a regression modeling approach for daytime delay trends based on a data-driven detection of spatio-temporal patterns. The focus is on interpretable rules whose prediction accuracy is compared to random forests as a non-parametric, automated modeling approach.

Firstly, decision rules were derived that describe daytime delay trends in spatio-temporal categories. For example, there is a positive daytime trend during summer, except on Mondays and Saturdays when the arrival airport is a spoke. Thus, we can state that the daytime trend depends on the interaction of the considered attributes. In order to validate these rules, we performed a quantitative evaluation by means of statistical modeling. From a technical point of view, the nature of our problem is related to the analysis of covariance (ANCOVA). The highest prediction accuracy so far can be achieved by spline models for daytime trends, taking into account interactions between the categorical variables season, direction and weekday. Although the derived decision rules, taken as a whole, are valid for only 62.90% of all days, this leads to a reduction of the absolute prediction error by about one minute on average. In particular categories, our approach leads to an even higher improvement of the prediction accuracy. The overall prediction accuracy is comparable to non-parametric random forests that imply an individual categorization and rule selection but lack interpretability.

However, we can assume that in general, primary delays are inherently hard to predict in the long-term on a macroscopic level. In this context, one always has to take into account that delay recording underlies constraints that lead to underestimation, e.g., predictable delay may already be prevented by scheduling decisions of an airline. In close connection to this, it is desirable to check to which extent the findings may be generalized regarding other airline delay data.

A lesson learned of the analysis is the discovery of low signal-to-noise ratio of any trends in primary delay in the long-term. Time-related trends look promising on aggregated data, asking for further investigation and interpretation. During the statistical analysis, the variance of the delays around these time trends became apparent. Methodologically interesting is that, due to the large data sets, the standard errors of statistical estimators were so small that the resulting inferences were no longer conclusive. This is a general challenge of data-driven approaches that aim to argue by other means than predictive accuracy.

Future work shall therefore identify the conditions, under which accurate predictions of primary delay are feasible. Resulting statistical prediction models can be implemented into a scheduling and simulation framework in order to obtain a more realistic evaluation of schedule robustness. The emerging question is to what extent an improved delay prediction affects the buffer management in hub-and-spoke networks and whether it actually leads to significant improvements of the robustness of schedules. This topic is addressed in Chapter 9 in the subsequent Part III.

Part III

Evaluating the Robust Efficiency of Crew and Aircraft Schedules

Chapter 8

Modeling Propagation Mechanisms for a Stochastic Simulation of Schedule Operations

The robustness of crew and aircraft schedules can be evaluated based on the amount of propagated delay in an event-driven stochastic discrete simulation of schedule operations. Figure 8.1 shows the general simulation procedure. For given crew and/or aircraft schedules, primary delays are generated following a specific prediction model as discussed in Section 4.2. Resulting delay events are saved into an event queue which notifies an event processor at the time of their recognition. Delays are tackled then by the disruption management. Depending on the recovery policy, interventions may be carried out if possible. Otherwise, delays are propagated and resulting secondary delays are added to the event queue. If crews or aircraft are rerouted or flight departures are postponed, the underlying schedules are updated.

In the current Part III of this thesis, we address three key features that affect the evaluation of schedule robustness: At first, the propagation model is introduced and its prediction accuracy is examined in this chapter. Due to the lack of crew connection data, the analysis can only be performed for aircraft turnarounds. Subsequently, the influence of primary delay prediction modeling is examined in Chapter 9. Eventually, a rule-based recovery technique for crews and aircraft is developed in Chapter 10.

The remainder of this chapter is organized as follows[1]: We present a formalization of aircraft turnaround processes and the related propagation model in the subse-

[1] Parts of the following analysis have been accepted for publication in:
Ionescu, L., Kliewer, N. (2018). *Examining Delay Propagation Mechanisms for Aircraft Rotations.* Proceedings of MKWI 2018.

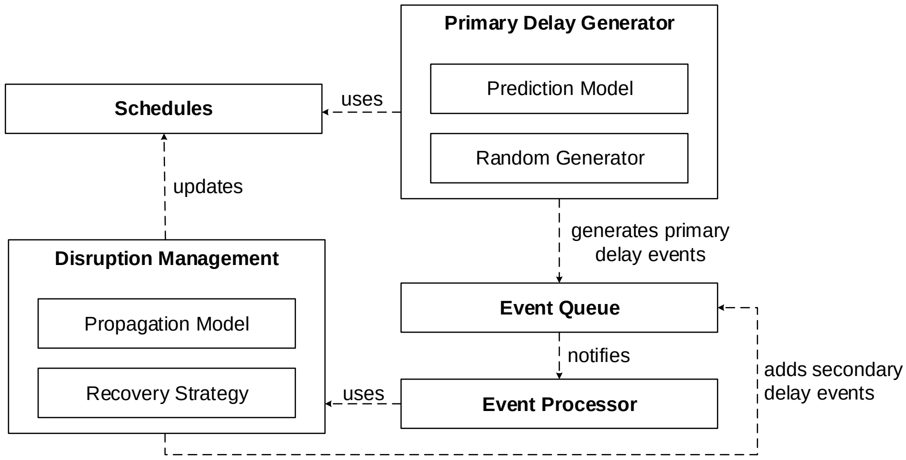

Figure 8.1: Event-driven stochastic discrete simulation of schedule operations

quent Section 8.1. One important input factor for correct estimation of propagated delays are minimum ground times which, however, are only available for the main hub. Therefore, Section 8.2 deals with the the derivation and estimation of minimum ground time values for all airports of the flight network. Based on this, propagation mechanisms are analyzed in Section 8.3 and the prediction accuracy of the propagation model is assessed. Since data is available only for aircraft rotations while crew itineraries are not retrievable, the analysis concentrates on rotation delays. Finally, conclusions and implications are provided in the final Section 8.4.

8.1 Fundamental Assumptions on the Turnaround Process and Propagation Effects

The generalized aircraft turnaround process between two consecutive flights within a rotation R is illustrated in Figure 8.2. Case (1) represents scheduled times in regular operations. Every flight $f \in R$ has a scheduled departure time STD_f and a scheduled arrival time STA_f. The turnaround process starts with the aircraft arriving at the gate. The scheduled ground time $sgt_{a(f)f}$ between flight f and its aircraft predecessor $a(f)$ consists of the minimum ground time $mgt^A_{a(f)f}$ and a potential buffer time $b^A_{a(f)f} \geq 0$.

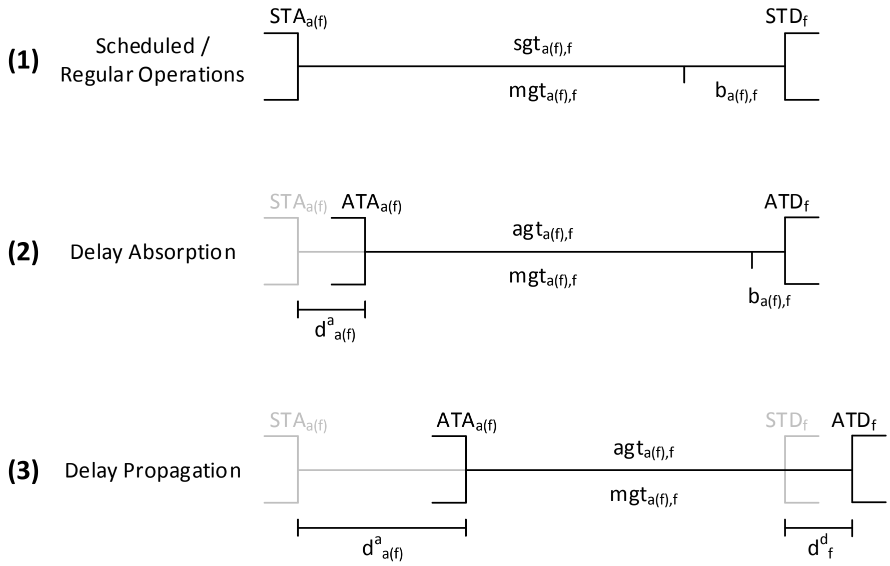

Figure 8.2: Ground Times of the Aircraft Turnaround Process

Cases (2) and (3) depict cases in which flight $a(f)$ arrives late. ATD_f and ATA_f describe the *actual* times of departure and arrival of flight f, respectively. d_f^D denotes the departure delay $ATD_f - STD_f$ and d_f^A the arrival delay $ATA_f - STA_f$. $agt_{a(f),f}$ represents the actual ground time $ATD_f - ATA_{a(f)}$. In Case (2), the delay can be absorbed by the buffer time and it holds

$$agt_{a(f),f} \geq mgt^A_{a(f),f}, \tag{8.1}$$

$$b^A_{a(f),f} = 0. \tag{8.2}$$

In contrast, Case (3) implies a delay propagation to flight f with

$$agt_{a(f),f} = mgt^A_{a(f),f}, \tag{8.3}$$

$$b^A_{a(f),f} = 0. \tag{8.4}$$

Delay propagation by crew itineraries are assumed to underlie the same general mechanisms with the respective minimum ground time $mgt^C_{c(f)f}$ between flight f and the crews' predecessor flight $c(f)$. Based on these assumptions, a basic propagation model for crews and aircraft is formulated in Equations 8.5 - 8.9, following Dück et al. (2012):

$$
\begin{aligned}
ATA_f &= \max\left\{STA_f, ATD_f + t_f\right\}, && \forall f \in F \quad (8.5)\\
ATD_f &= \max\left\{STD_f, \max\left\{\begin{array}{c} ATA_{a(f)} + mgt^A_{a(f),f}, \\ ATA_{c(f)} + mgt^C_{c(f),c} \end{array}\right\}\right\} + X_f, && \forall f \in F \quad (8.6)\\
d^D_f &= ATD_f - STD_f, && \forall f \in F \quad (8.7)\\
d^A_f &= ATA_f - STA_f, && \forall f \in F \quad (8.8)\\
s_f &= d^D_f - X_f, && \forall f \in F \quad (8.9)
\end{aligned}
$$

X_f represents a stochastic variable for primary delays in ground processes that can be modeled based on findings from Chapter 7. A block time t_f is associated to each flight f, determining the time from gate to gate, including taxi-out, flying time and taxi-in.

Equation 8.5 ensures the connection between actual departure time and actual arrival time for all flights $f \in F$. Early arrivals are not considered. Moreover, a flight f can depart only if both crew and aircraft are available (Equation 8.6). Equations 8.7 and 8.8 explicitly describe departure and arrival delays, respectively. A main assumption of the model is that negative delays do not propagate, i.e. an early arrival does not imply an early departures of subsequent flights. s_f depicts the amount of propagated departure delay for every flight f and therefore is the target value concerning the evaluation of schedule robustness (Equation 8.9). Since the duration of the minimum ground time plays a central role for the computation of s_f, its specification is discussed in the following.

8.2 Examining Minimum, Scheduled and Actual Ground Times

The minimum ground time *mgt* are determined prior to the construction of aircraft rotations. It is defined as the shortest time span in which the turnaround process can be performed. Several – partly interconnected – tasks have to be carried out, leading to a fairly complex system of interactive processes. Several studies on a high granular level have been carried out on this topic, see Fricke and Schultz (2009) and Schlegel (2010). According to (Schlegel, 2010, p. 81), the mgt for continental fleets (Airbus A320/A321, Boeing 737) is 45 minutes at the main hub. It can be reduced to 40 minutes when both the inbound and outbound flights are domestic. Since mgt values are only available for the main hub, a sufficient estimator for mgt values at all other airports of the flight network based on scheduled ground times (*sgt*) must be derived.

At first, possible influential factors that may affect the turnaround duration are checked in a CART analysis[2]. The results confirm the straightforward expectation that fleets are the most substantial influential factor for the sgt, followed by O&Ds. As an additional parameter we use a binary variable indicating whether the inbound and outbound flights are both domestic or not. Significant differences by seasonal attributes cannot be observed. We check the following candidates as mgt estimators:

- **Minimum Value** as a theoretically valid threshold if the sgt records are without outliers,
- **.01, .05, .075, .10, and .15 Quantiles** as estimators that are assumed to be robust against outliers,
- **Modal Value**, based on the assumption that most turnarounds are performed within the minimum time available

The results for the main hub are presented in Table 8.1. The second and third column depict the mean and standard deviation of $cgt - mgt$ where cgt is the minimum ground time computed using the respective estimator and mgt is the minimum ground time value given in the data set. The last two columns show the relative amount of flights for which the minimum ground time is derived exactly or within a tolerance range of ± 5. It turns out that the .10 quantile suits best for retrieving mgt values from sgt records, allowing an exact derivation for 86% of all flights. Within the tolerance range of ± 5 this value increases up to 95.9%.

For further illustration, Figure 8.3 shows the histogram for the .10 quantile in comparison to minimum and modal values, indicating the extreme cases of under- and overestimation[3]. One must be aware, of course, that the transfer of this estimation

[2]For details, see `https://cran.r-project.org/package=rpart`, last access: November 9$^{\text{th}}$, 2017
[3]Histograms for the remaining tested estimators can be found in the Appendix C.6.

	Deviation		Accuracy	
Model	**Mean**	**SD**	**Exact**	**±5min**
min	33.705	13.339	0.029	0.059
.01 Quantile	20.152	11.129	0.043	0.115
.05 Quantile	3.514	6.916	0.587	0.911
.075 Quantile	1.734	5.765	0.832	0.951
.10 Quantile	1.326	5.132	0.860	0.959
.15 Quantile	1.901	4.475	0.690	0.944
Mode	6.615	10.820	0.396	0.753
Mean	24.436	7.565	0.000	0.001

Table 8.1: Results of various models for minimum ground time prediction

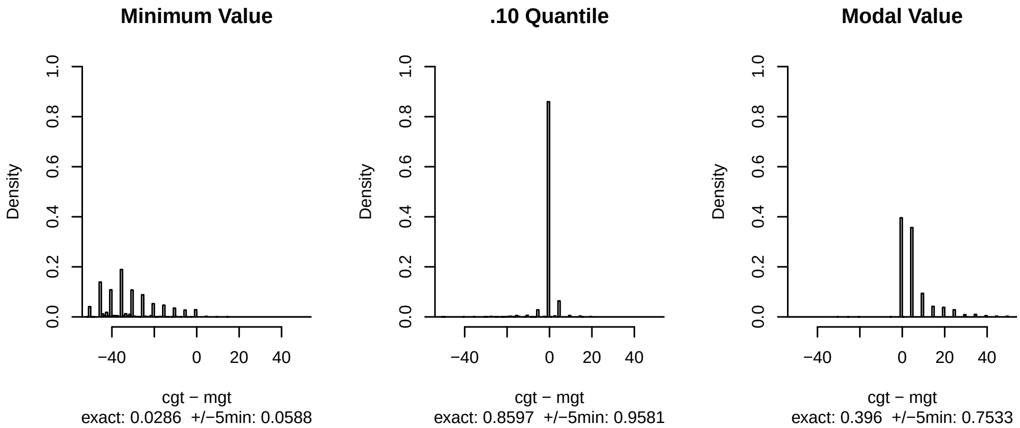

Figure 8.3: Best predictor for minimum (scheduled) ground time

scheme to turnaround stations other than the main hub may be slightly distorted. In combination with figures of Table 8.3, we can assume that the .10 quantile is the best available estimator for capturing station-specific mgt values in the scope of this analysis.

8.3 Examination of Propagation Mechanisms on Real-World Data

Prior to the assessment of the prediction accuracy of the propagation model, we give an introducing example of a daily rotation based at the main hub that is affected by

O&D	STD	ATD	STA	ATA	ΔGT	ΔBT	IN	ROT	PRIM
$S_1 - H_1$	06:00	07:09	07:10	08:42	–	23	14	0	69
$H_1 - S_2$	08:20	09:28	09:15	10:39	-24	16	92	68	0
$S_2 - H_1$	09:55	11:26	11:10	12:41	7	0	84	91	0
$H_1 - S_3$	15:25	15:25	16:35	16:26	-91	-9	91	0	0
$S_3 - H_1$	17:25	17:22	18:55	18:42	6	-10	-9	0	0
$H_1 - S_4$	19:55	21:17	20:40	22:01	95	-1	-13	0	0

Table 8.2: Exemplary one-day rotation of an Airbus A321

primary delays. Table 8.2 represents the rotation of an A321 at one representative day in summer. ΔGT indicates the difference between scheduled and actual ground time of flight f and its predecessor $a(f)$:

$$\Delta GT = (STD_f - STA_{a(f)}) - (ATD_f - ATA_{a(f)}). \tag{8.10}$$

Negative ΔGT values are induced by the usage of buffers for the absorption of incoming delays. In case of a delaye turnaround, the actual ground time exceeds the scheduled ground time, leading to positive values of ΔGT. Analogously, ΔBT describes the block time difference

$$\Delta BT = (ATA_f - ATD_f) - (STA_f - STD_f). \tag{8.11}$$

Concerning O&Ds, H_1 stands for the main hub while $S_1, \ldots, S_4$ depict the four spoke airports that are served by the out-and-back principle. Column IN depicts the (incoming) arrival delay of the predecessor flight, columns ROT and $PRIM$ show the actual rotation delay and primary delay. The first flight has a primary departure delay of 69 minutes due to weather conditions at the destination airport (IATA Delay Code 84). Additionally, the scheduled block time is exceeded by 23 minutes, resulting in an arrival delay of 92 minutes. 24 minutes can be absorbed during the following ground time. The turnaround is performed within 46 minutes, close to the minimum ground time of 45 minutes. Nevertheless, the second flight departs 68 minutes late and again experiences an increase of the scheduled block time duration (16 minutes).

The ground time between flight 2 and 3 does not contain any buffer time and takes 7 minutes more than scheduled. The initial delay can finally be absorbed prior to

	all	**all**$_d$	**all**$_c$	H_1^d	H_1^c	H_2^d	H_2^c	S^d	S^c
Avg	-0.54	-0.49	-0.57	-0.42	-0.20	-0.07	-0.15	-0.78	-0.91
SD	3.51	2.87	3.81	2.98	3.00	2.47	2.27	2.93	4.43
$\pm$**0**(%)	70.80	74.91	68.64	79.03	76.54	79.21	86.30	68.96	60.47
$\pm$**5**(%)	90.30	92.67	89.05	92.49	92.45	94.88	97.15	91.71	85.49
$r\Delta_{ta}$(%)	13.33	13.62	13.18	7.99	6.65	8.36	3.81	21.42	19.43
$a\Delta_{ta}$	-5.18	-4.58	-5.51	-4.65	-3.91	-3.69	-3.27	-4.73	-6.00

Table 8.3: Estimated rotation delay when the preceding flight of the aircraft arrives late

departure of flight 4. The aircraft spends 165 minutes on ground rather than the originally scheduled 255 minutes. The extended ground time prior to flight 5 is a result from the early arrival of flight 4. Eventually, flight 6 is severely delayed, awaiting a crew interchange (IATA Delay Code 95, *flight deck or entire crew*). Overall, a primary delay of 69 minutes in combination with additional 39 minutes of block time delay entails accumulated propagated delays of 159 minutes. In the following, we examine the prediction accuracy of the propagation model concerning its estimation of delay propagation and absorption. The key question is if operational propagation effects are sufficiently represented or if systematical over- or underestimation becomes apparent. The estimated rotation delay $\hat{d}_f^r$ of flight f is computed as the difference between the arrival delay of the preceding flight (denoted as IN) and the scheduled buffer time:

$$\hat{d}_f^r = max(max(IN, 0) - (sgt - mgt), 0). \tag{8.12}$$

We validate the accuracy of the rotation delay estimation for all 28,502 daily rotations of the A321 fleet that are available in the data set. Each rotation consists of 4.86 daily flights on average. For 58.09% of the underlying 138,662 flights, the predecessor flight in the rotation arrives late so that a rotation delay may emerge. An average of 7.30 minutes of incoming arrival delay leads to 3.30 minutes of outgoing rotation delay which is an absorption rate of 45.24%.

Table 8.3 shows figures concerning the accuracy of rotation delay estimations for flights when the predecessor flight of the aircraft arrives late. To compare, Table 8.4 shows the same values for flights whose aircraft predecessor flight arrives on-time. H_1 and H_2 stand for Hub 1 and 2, respectively, while S depicts spoke airports.

	all	**all**$_d$	**all**$_c$	$\boldsymbol{H_1^d}$	$\boldsymbol{H_1^c}$	$\boldsymbol{H_2^d}$	$\boldsymbol{H_2^c}$	$\boldsymbol{S^d}$	$\boldsymbol{S^c}$
Avg	0.02	0.02	0.02	0.02	0.01	0.06	0.04	0.00	0.02
SD	0.44	0.45	0.43	0.41	0.38	0.76	0.77	0.17	0.41
$\pm\mathbf{0}$(%)	99.72	99.68	99.75	99.78	99.81	99.14	99.50	99.91	99.72
$\pm\mathbf{5}$(%)	99.87	99.86	99.87	99.85	99.88	99.66	99.77	99.98	99.88
$\boldsymbol{r\Delta_{ta}}$(%)	1.02	1.41	0.78	0.35	0.24	0.61	0.14	2.78	1.39
$\boldsymbol{a\Delta_{ta}}$	-2.58	-2.42	-2.78	-2.07	-2.58	-2.16	-2.00	-2.49	-2.82

Table 8.4: Estimated rotation delay when the preceding flight of the aircraft arrives on-time

Furthermore, we differentiate between domestic (index d) and continental flights (index c). Note that differentiation on temporal attributes does not show observable patterns and is therefore neglected.

In a first step, we discuss the overall estimation accuracy. The first row depicts the average deviation between estimated rotation delay $\hat{d}_f^r$ and recorded rotation delay d_f^r. Throughout all categories, the value is negative, indicating a general overestimation of rotation delays by the model. For continental flights departing at spokes (column S^c), rotation delays are overestimated most, while values are closest to zero for H_2. The rotation delay is estimated correctly on the precise minute for 70.8% of all flights. Overestimation emerges in 19.87%, underestimation in 9.33% of all cases. Best results can be achieved for H_2, followed by H_1. Obtaining a ± 5 minute threshold, results are significantly better but still in the same order for all categories. This is a direct consequence of the fact that departure and arrival times are commonly scheduled in 5 minute intervals. In addition, the delay recording at many spoke airports is performed within 5 minute steps, see Section 6.1.5.

Table 8.4 shows respective values for all turnarounds without an incoming arrival delay of the preceding flight. It allows counter-checking if rotation delays occur even when they are theoretically impossible. Slightly positive values in the first row are implied by 160 flights containing rotation delay records without a preceding late aircraft arrival. Apart from obviously inconsistent or imprecise records, no systematic distortion effects become apparent.

Concerning the rotation delay overestimation of the propagation model in case of incoming arrival delays (Table 8.3), there are differences between the actual turnaround duration and generalized mgt values. Target mgt values do not necessarily reflect the minimum operational turnaround duration but may already contain slack times. Referring to this, the two bottom rows of Tables 8.3 and 8.4 depict:

- $\boldsymbol{r\Delta_{ta}}$ as the relative share of turnarounds that are performed in less than the minimum ground time ($agt < mgt$) if a late aircraft arrival is likely to postpone the following flight departure, and
- $\boldsymbol{a\Delta_{ta}}$ as the average of the absolute difference ($agt - mgt$) in minutes.

The actual ground time falls below the minimum ground time at spoke airports significantly more often (columns S^d and S^c). Assumingly, this is due to conservatively scheduled mgt values at spoke airports since the lack of strong presence makes it harder for an airline to perform recovery actions in case of unscheduled events. On the one hand, spare resources like reserve crews are mostly available at hubs rather than spokes, on the other hand less possibilities for aircraft swaps exist at spokes. The importance of this effect becomes apparent when comparing the results to corresponding values in Table 8.4. If an aircraft arrives late for the turnaround, it is thirteen times more likely that the turnaround is performed faster than scheduled.

Going into more detail, we check this mechanism by specifically considering data for H_1. Depending on the exact arrival delay of the aircraft predecessor flight, we examine if length and frequency of turnaround speed-ups change accordingly. Finally, we check the overall operational importance in terms of the total amount of savings in propagated delay.

For a start, Figure 8.4 shows the ratio of turnarounds with $agt < mgt$ depending on the arrival delay of the aircraft predecessor flight. We use bins of 5 minutes for arrival delays up to 60 minutes. There is a substantial and steady increase of faster turnarounds until an arrival delay of 25 minutes. For larger arrival delays, inconclusive fluctuations at high levels can be observed.

In addition, the left panel of Figure 8.5 shows the average absolute difference ($agt - mgt$) in minutes depending on the arrival delay. Values are in a constant interval, however, with considerable fluctuations. No relevant pattern can be observed. Finally, the right panel of Figure 8.5 is intended to indicate the operational relevance of turnaround speed-ups as the sum of absorbed arrival delay due to turnaround speed-ups. The y-axis shows the sum of absorbed arrival delay by turnaround speed-ups in minutes. Thus, values depend on both absolute turnaround speed-up and

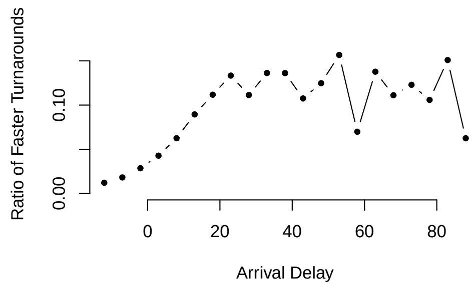

Figure 8.4: Ratio of turnarounds in less than the minimum ground time

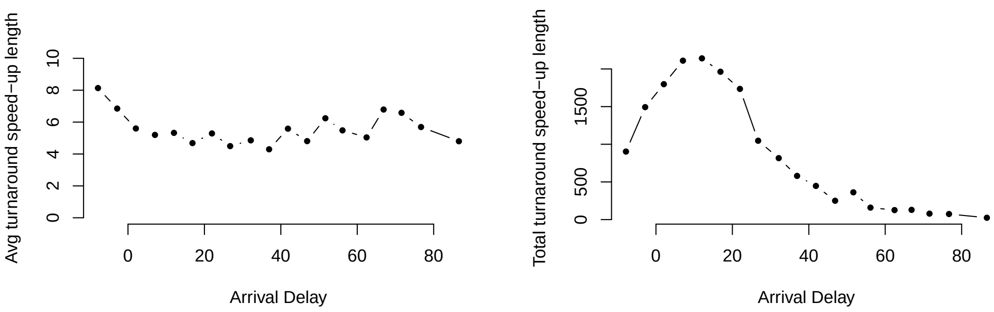

Figure 8.5: Operational relevance of turnarounds in less than the minimum ground time

the occurrence frequency of certain arrival delay values. Most substantial savings are achieved for arrival delays between 0 and 25 minutes. The dominant factor for the curve progression is that significantly large arrival delays are extremely rare and although turnarounds are performed faster in these cases, too, the operational relevance is minor in terms of a holistic assessment of rotation delays.

8.4 Summary and Implications

In the preceding analysis, the propagation model of Dück et al. (2012) has been examined in order to evaluate its prediction accuracy on propagated delays. It has turned out that 70.8% of delay propagation by aircraft rotations can be precisely estimated. Within a ±5 minute tolerance threshold, the estimation is correct for about 90.3% of all flights. The prediction provides best results for large hub airports with values up to 97.15% while at continental spokes 85.49% can be reached.

On average, rotation delays are overestimated by 0.54 minutes per flight. This value ranges from nearly 0 at Hub 2 to around 0.91 minutes at continental spoke airports. A substantial responsibility for these effects lies in the fact that actual ground times sometimes fall below scheduled target minimum ground times when delay propagation emerges. It happens in 13.33% of all turnarounds when the aircraft arrives late. On average, 5.18 minutes of propagated delay can be saved in these cases. Deeper insight in the dependency between arrival delays and turnaround speed-ups shows that operational relevance is given for arrival delays between 1 and 25 minutes. Dependencies between the average turnaround speed-up length and the arrival delay duration cannot be observed. However, due to the low incidence of larger delays, their influence on the overall delay propagation estimation is comparatively small.

Note that the analysis has not been performed for crew connections due to the lack of related data records. In consequence, our findings are not yet incorporated into the stochastic simulation during subsequent studies in the scope of this thesis. This is because considering correction terms only for aircraft connections may lead to an unbalanced assessment of crew- and aircraft-related propagation. Therefore, the evaluation of crew connections and the subsequent combination with our findings has to be reconsidered in future research.

Chapter 9

Assessing the Influence of Primary Delay Prediction Models in Robust Resource Scheduling

The potential of improving primary delay prediction for long-term robust scheduling has been examined in Chapter 7. Resulting statistical models can be used for refined primary delay prediction in robust scheduling and simulation. However, it must be assumed that the prediction accuracy improvement does not necessarily reflect in an identical improvement of the robust efficiency of generated schedules. The main reason is that not all flight connections that are likely to propagate delay can be prevented since scheduling decisions may be predetermined in previous scheduling stages.

In this regard, we examine the actual impact of primary delay prediction models on the robust efficiency of crew schedules based on existing aircraft rotations in a sequential scheduling approach. To that, crew schedules with increasing degrees of stability are generated based on varying prediction models. The robustness of resulting schedules is evaluated by simulating resulting crew schedules along with given aircraft rotations. Flexibility aspects are not considered in order to provide an uncompromised assessment without potential interference by subsequent schedule adaptions.

In the following Section 9.1, we provide details on the incorporation and application of primary delay prediction models into scheduling and simulation. In addition, suitable measures for robustness and robust efficiency are introduced. In Section 9.2, the impact on the robust efficiency that results from applying refined delay prediction models from Chapter 7 is examined. Moreover, a sensitivity analysis is performed in Section 9.3. Firstly, the impact of complete knowledge of all delays in scheduling is checked. This approach is referred to as the *oracle* setting and provides a theoretical upper bound of the improvement in robust efficiency. Besides, effects resulting from delay misestimation as well as systematic over- and underestimation are investigated. Eventually, conclusions are drawn in Section 9.4.

9.1 Application of Primary Delay Prediction Models in Scheduling and Simulation

In this section, we provide technical details on the general creation and integration of delay prediction models, based on the findings from Chapter 7, into the prototypical scheduling and simulation framework. For a given set of flights F in a schedule, a primary departure delay d_f is predicted for each flight $f \in F$. Therefore, every flight is categorized by certain attributes such as flight direction, season and weekday (best derived model *E1* from Chapter 7). For every category a regression model is fitted for the daytime variable. For illustration, the parameters of exemplary linear and cubic spline model are given in Table 9.1. $\beta_0, \ldots, \beta_5$ are the mean value estimation parameters (log scale), σ depicts the respective standard deviation. For a given daytime t, the expected primary delay $\hat{\mu}(t)$ and its standard deviation $\hat{\sigma}(t)$ is determined per category. Since we use log-scale delays, a random number $X \sim N(\mu(t); \sigma(t)^2$ can then be picked from the normal distribution. Finally, the resulting primary delay d is computed as $d = e^X$.

Concerning scenario-based stochastic scheduling approaches as presented in Yen and Birge (2006) or Dück et al. (2012), the robustness evaluation of a resource schedule takes into account a set of primary delay scenarios Ω. With a given set of flights F, a delay scenario $\omega \in \Omega_F$ represents random variables for primary delay that result in deviations from departure times of the flights $f \in F$. Now, the primary departure delay of flight f in scenario ω is $d_{f\omega}$ which can be drawn analogously from our model. For sensitivity analysis, diverse models can be applied straightforwardly by modifying $\hat{\mu}(t)$ and $\hat{\sigma}(t)$.

In the following, we describe the application and evaluation of primary delay prediction models based on Figure 9.1. Besides the obligatory *Flight Schedule*, *Primary*

Model	β_0	β_1	β_2	β_3	β_4	β_5	σ
Linear	-0.0179	-0.8980	-	-	-	-	1.024
Std. Err.	0.0151	0.0151	-	-	-	-	-
Spline	0.0805	0.2486	-0.3376	-0.0509	-0.2909	-0.2715	1.023
Std. Err.	0.0412	0.0836	0.0864	0.0866	0.1213	0.1072	-

Table 9.1: Exemplary standardized parameters of a regression model with five cubic splines for the category [S: Winter, DI: Spoke-to-Hub, W: Monday]

Delay Scenarios have to be provided as an input factor for the *Robust Scheduling Strategy.* According to the assumptions in Chapter 7, a delay scenario consists of delay values between 0 and 180 minutes for each flight of the schedule. For the generation of crew schedules, we use the crew pairing optimization approach of Dück et al. (2012). Each crew pairing is assessed concerning its delay propagation risk for each given primary delay scenario with subsequent aggregation by averaging the amount of propagated delay over all scenarios. During the pricing step of column generation, new crew pairings are created. Therefore, it is determined which flights are consecutively operated by the same crew pairing. A delay propagation and evaluation module assesses the delay propagation of these connections, depending on given primary delay scenarios. In contrast to Dück et al. (2012), we use an a minute-based rather than a punctuality-based evaluation. By doing so, a more decent way to directly address delay propagation is obtained. The resulting penalty value is then added to the costs of the related crew pairing. The usage of a propagation penalty factor parameter allows the calibration between the objectives cost minimization and robustness. With increasing propagation penalty factors, pairings with less delay propagation costs are preferred when the schedule is constructed.

In the subsequent *Schedule Simulation*, the robustness of the resulting schedule is evaluated. Delay scenarios can vary between scheduling and simulation. A schedule can therefore be exposed to varying delay scenarios in order to assess its robustness under different circumstances. The usage of one single primary delay scenario is possible when testing schedules against a set up virtual reality (introduced as *oracle*) as part of a sensitivity analysis. If systematic robustness shall be evaluated, a set of varying delay scenarios can be obtained in order to prove robustness in numerous conceivable scenarios.

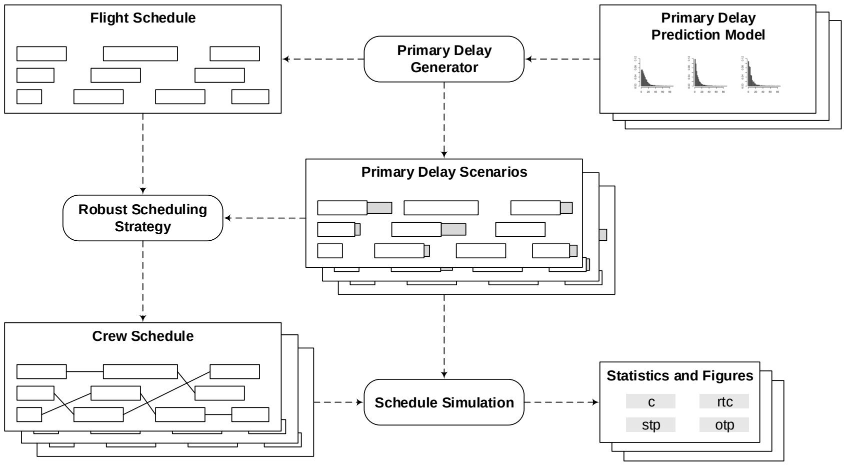

Figure 9.1: Framework for the evaluation of the influence of primary delay prediction models on nominal costs and robustness

As the key robustness measure, we use the secondary-to-primary delay ratio (stp) as a quotient between the sum of all secondary delays and the sum of all primary delays. It determines how many minutes of secondary delay are induced by one minute of primary delay. As a common robustness measure, we also provide results concerning the on-time performance (otp). In the scope of this analysis, refined delay prediction does not only impact the robustness but also nominal costs (c) may change accordingly. Therefore, we consider the quotient between the change in robustness and the change in nominal costs as a key metric for the robust efficiency. In the following, we refer to it as the robustness-to-cost (rtc) ratio.

9.2 The Impact of Refined Primary Delay Prediction

We implement two different prediction models for primary delay for the usage in the crew scheduling approach of Dück et al. (2012), namely a categorical model (CAT) and a general model (GEN), reflecting the ANCOVA analysis in Chapter 7. The GEN model assumes the same risk of primary delay occurrence for all flights, independently from any spatio-temporal aspects. Its parameters are derived from historical data without categorization – the probability of a flight delay is 47.5968%, the average delay is 10.6682 minutes with a standard deviation of 14.3418 minutes.

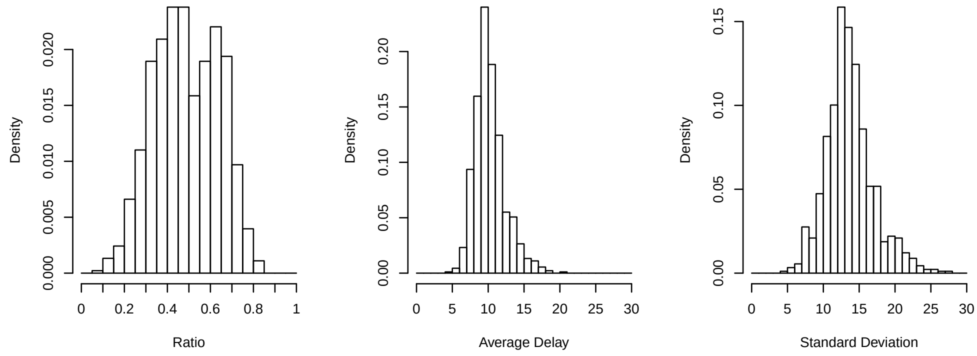

Figure 9.2: Parameter variations between categories in the CAT model

In contrast, the CAT model reflects the model E1, D(15)°S°DI°WE, see Table 7.4 of Chapter 7. Between the categories, the average delay ranges from 4.5625 to 20.94774 minutes, the delay ratio from 9.8592% to 82.069%. For a graphical illustration of parameter variations, see Figure 9.2.

We proceed to the implementation of these models into the scheduling framework, resulting in different settings for our study that are subsequently evaluated. In scheduling, either the GEN or CAT model is used for primary delay estimations. If cost-efficiency is considered as the only objective, no prediction model is applied. Once schedules are generated, they are simulated in order to assess the actual operational performance. A delay prediction model is used to set up an artificial reality that the generated schedules are tested against. For the evaluation of the benefit of refined delay prediction, we use the CAT model in simulation. In both scheduling and simulation, we use 100 delay scenarios each. Given these considerations, three basic configurations are set up for computational experiments:

- **Setting CAT-CAT** Primary delay occurrence probabilities shall be known in scheduling. Therefore, a delay generator based on the CAT model is used both in scheduling and simulation. Delay distributions can be fully adapted during scheduling. However, since the delay propagation evaluation during scheduling has to cope with 100 different delay scenarios, it cannot be declared as an oracle solution. Furthermore, delay propagation may emerge due to either a lack of degree of freedom in scheduling or a disproportionate increase of nominal costs.

Instance	1	2	3	4	5	6	7	8	9	10
Flights	395	385	396	393	420	427	419	422	338	311
Rotations	68	69	68	69	68	70	70	70	59	53
Pairings	66	63	66	64	69	70	71	70	61	58

Table 9.2: Test set for the assessment of refined delay prediction

- **Setting GEN-CAT** In contrast, only general assumptions on primary delays are known in scheduling, depicted by the usage of the GEN model in the delay generator. Resulting schedules are then simulated in a more distinct environment based on the CAT model. Unlike CAT-CAT, additional delay propagation is implied by the intentional difference in assumptions on delay occurrences between scheduling and simulation.
- **Setting CE-CAT** For comparison, we use a scheduling strategy with the only objective of cost-efficiency (CE). No delay risks are taken into account during scheduling, thus, schedules are less robust but offer least nominal costs. The evaluation is performed based on the CAT model and therefore the robustness assessment is comparable to previous settings.

For computational experiments, we use 10 one-day flight schedule instances of the A321 fleet, see Table 9.2 for details. The instances are extracted from the historical data set and are dense in terms of flights per day, allowing considerable propagation effects and high degrees of freedom for crew scheduling. For every instance and both settings CAT-CAT and GEN-CAT, we generate crew schedules with up to six different degrees of stability. This is done by altering the stability penalty factor, controlling the trade-off between nominal costs and stability[1].

In order to prevent outlier-dependent solutions, each setting is repeatedly solved with four different seeds for the random number generator of the respective delay generator in scheduling and simulation. Altogether, the following analysis is performed on 480 different crew schedules, i.e. 240 each for the CAT-CAT and GEN-CAT setting. In the CE-CAT setting, one cost-efficient solution is generated for every instance. The CE model is then simulated for every seed as a reference for evaluation.

Starting with the analysis, Table 9.3 provides results for all runs, averaged by the prediction model used in scheduling. Column c reflects the nominal cost averaged

[1]Penalty factor values are 100, 250, 500, 750, 1,000, 2,000.

Model	c	stp	otp_0
CE	970.30	1.0598	70.03
GEN	990.35	0.8947	73.08
CAT	989.39	0.8865	73.12

Table 9.3: Results aggregated by the type of prediction

over all settings, offering a cost increase of 2.07% for the GEN model and 1.97% for the GEN model, compared to the cost-efficient solutions CE. The average stp improvement is 15.58% for the GEN model and up to 16.35% for CAT. The absolute otp_0 increase is nearly identical at about 3%. An important interfering factor for our analysis becomes obvious when considering the different amount of primary delays generated by GEN and CAT for the same instances. GEN or CAT models do not solely lead to different distribution of delays but also imply differences in the total amount of primary delays the scheduling strategy has to cope with. Over all instances, the amount of primary delays generated by the CAT model are 6.97% higher than by GEN. Thus, scheduling strategies using the CAT model have to cope with significantly more primary delays. Basically, assuming higher delay risks in scheduling is supposed to lead to higher stability at increased nominal costs. However, nominal costs of CAT do not exceed their equivalents of GEN and are even slightly lower. This effect indicates that the CAT results are underestimated in comparison to GEN.

Going into more detail, we compare the changes in nominal costs and stp values for GEN and CAT models compared to CE per setting. We consider c_s as the costs of resource schedules resulting from setting $s \in S$, consisting of the stability penalty factor and seed value. Analogously, stp_s depicts the robustness measure per setting $s\ inS$. For prediction models m, we compute the relative difference of nominal costs $c^r_{s,m}$ and robustness measure $stp^r_{s,m}$ compared to their CE equivalents by:

$$c^r_{s,m} = \frac{c_{s,m} - c_{s,CE}}{c_{s,CE}}, \qquad \forall s \in S, m \in \{GEN, CAT\}, \tag{9.1}$$

$$stp^r_{s,m} = \frac{stp_{s,m} - stp_{s,CE}}{stp_{s,CE}}, \qquad \forall s \in S, m \in \{GEN, CATs\}. \tag{9.2}$$

We check the statistical significance of the results. As it follows from the preceding analysis, robustness and nominal costs must be taken into consideration simultane-

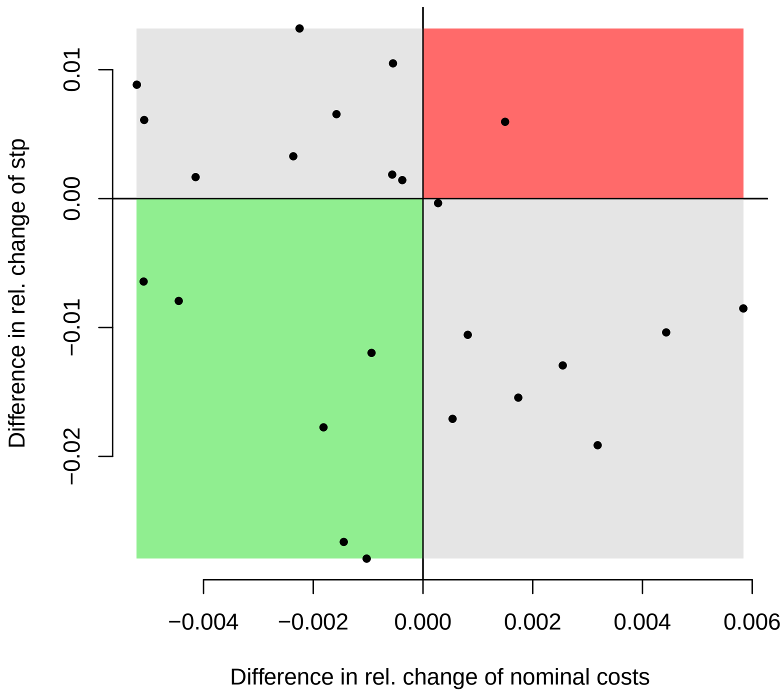

Figure 9.3: Differences in robustness and nominal costs between GEN and CAT

ously when assessing the potential benefit of CAT over GEN. As test statistics we use the $rtc_{s,m} = \frac{stp^r_{s,m}}{c^r_{s,m}}$, representing the increase of robustness per increase of nominal costs for models $m \in \{GEN, CAT\}$, where $c^r_{s,m} \neq 0$[2]. The difference between $rtc_{s,CAT}$ and $rtc_{s,GEN}$, averaged over all settings s, is 0.841%. A one-sided paired t-test[3]. provides a p-value of 0.01889 (t=2.0891, df=236), suggesting that the mean of rtc_{CAT} is significantly greater than rtc_{GEN}.

For graphical illustration, we plot relative differences of c^r and stp^r values between GEN and CAT models per setting s in Figure 9.3. The x-axis depicts the difference

[2]Note $stp^r_{s,m} < 0$ stands for less delay propagation, i.e. more robust solutions.
[3]$H_0 : \mu_{CAT} \leq \mu_{GEN}$, $H_1 : \mu_{CAT} > \mu_{GEN}$, confidence interval: (0.2232;1.9076)

in relative nominal cost change between CAT and GEN models by:

$$\Delta c^r_{CAT,GEN,s} = c^r_{s,CAT} - c^r_{s,GEN}, \qquad \forall s \in S. \tag{9.3}$$

CAT solutions are cheaper than GEN solutions if $\Delta c^r_{CAT,GEN,s}$ is negative. Analogously, the y-axis depicts the difference in stp values:

$$\Delta r^r_{CAT,GEN,s} = stp^r_{s,CAT} - stp^r_{s,GEN}, \qquad \forall s \in S. \tag{9.4}$$

CAT solutions are more robust if $\Delta r^r_{CAT,GEN,s}$ is negative. In 9 settings, the CAT model offers cheaper solutions with less robustness (upper left area). For further 8 settings, the relation is vice versa (lower right). In about 6 cases, the CAT model is dominant (lower left), mostly for settings with small propagation penalty factors. In one setting the GEN model even leads to more robust and cheaper solutions (upper right). In this setting, the highest delay propagation penalty factor is used, possibly leading to distorted results. This assumption is supported by the fact that it only affects one of the four seeds while for the remaining three seeds the results are non-dominant.

Finally, we have a closer look on the similarity between schedules. The similarity between two resource schedules is determined by the number of similar successor flights: The similarity index between schedules generated with GEN and CAT models is 79.83% on average. Schedules generated with GEN (or CAT) are similar by 76.19% (76.00%, respectively) with cost-efficient schedules (CE). Table 9.4 provides more detailed similarity values, averaged per instance. Regardless which scheduling approach is used, around three fourth of each schedule remain unchanged. Similarity is even higher between schedules generated by CAT or GEN. The observations both support and explain our previous findings – although refined delay prediction has a positive effect on robust scheduling, its influence underlies substantial restrictions regarding the degrees of freedom in scheduling.

Instance	CAT-GEN	CAT-CE	GEN-CE
1	0.7856	0.7514	0.7516
2	0.8155	0.7863	0.7734
3	0.7919	0.7463	0.7336
4	0.8291	0.7641	0.7662
5	0.8008	0.7535	0.7586
6	0.7975	0.7748	0.7658
7	0.7803	0.7569	0.7313
8	0.8027	0.7646	0.7709
9	0.7899	0.7639	0.7713
10	0.7892	0.7572	0.7775

Table 9.4: Similarity between different prediction models, averaged per instance

9.3 Sensitivity Analysis for Primary Delay Prediction

In this section, theoretical primary delay prediction models are applied for a sensitivity analysis in which the influence of varying primary delay occurrence assumptions during scheduling shall be further evaluated. Firstly, we consider the complete knowledge of all emerging primary delays during scheduling, referred to as an *oracle* prediction model configurations, in Section 9.3.1. In the subsequent misplacement analysis in Section 9.3.2, we estimate the effects of assuming delay risks for the wrong flights. Finally Section 9.3.3 deals with the under- and overestimation of primary delay duration.

9.3.1 Oracle Configurations and the Influence of the Delay Scenario Count

All delay scenarios are generated using the GEN model with parameters $\mu = 10.66817$ and $\sigma = 14.34176$. The ratio of delayed flights is set to 50% in order to obtain straightforward comparability to results in the subsequent delay misplacement analysis. Computational experiments are performed with regard to three configurations: In the main configuration OC.1.1, exactly one identical primary delay scenario is used both in scheduling and simulation. This configuration reflects a *virtual reality* that is completely known in scheduling. In related configurations OC.10.1 and OC.100.1, 10 and 100 primary delay scenarios are used in scheduling, respectively. Actual flight delay risks are then estimated based on averaging delay values of all

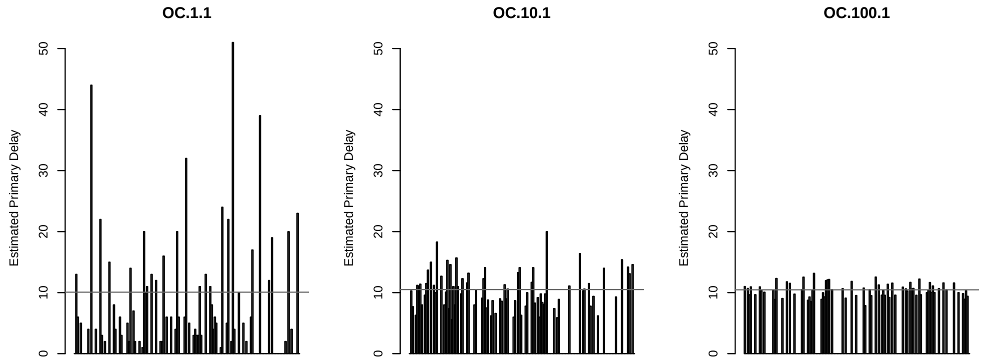

Figure 9.4: Primary delay estimations per configuration

considered scenarios. Note that in different scenarios, different flights can be expected to be delayed. As a consequence, delay risks can be distributed more evenly over all flights.

The configurations are exemplarily illustrated in Figure 9.4 for one flight schedule. The left panel shows the single-scenario configuration OC.1.1, depicting the virtual reality also used in scheduling. The the middle panel and right panel shows configurations OC.10.1 and OC.100.1, respectively. The average delay per configuration is illustrated as a gray horizontal bar in each panel. The mandatory decrease of fluctuations in the expected delay duration is apparent. As a consequence, the increase of the number of considered primary delay scenarios during scheduling leads to more evenly distributed delay risk assumptions. We evaluate the impact on the robust efficiency of the crew schedules that are generated per configuration. Similar to the previous study on refined delay prediction, experiments are repeated four times with differing seed values for random number generation and the results are averaged.

Table 9.5 shows the outcome for cost-efficient solutions (CE) as well as OC.1.1, OC.10.1 and OC.100.1. Since assumptions on expected delays are severely weakened in relation to the virtual reality scenario, is must be assumed that the distribution of buffer times is increasingly balanced. Interestingly, there is a significant increase in nominal costs between OC.1.1 and OC.10.1/OC.100.1 while both stp and otp values can be slightly improved. This effect can be attributed to the fact that additional buffers are incorporated into the schedules in OC.10.1 and OC.100.1 while they

Model	$\boldsymbol{c}$	$\boldsymbol{stp}$	$\boldsymbol{otp_0}$	$\boldsymbol{rtc}$
CE	970.30	1.0507	69.19	–
OC.1.1	985.01	0.9002	71.02	-11.0416
OC.10.1	994.52	0.8979	71.29	-8.5354
OC.100.1	994.75	0.8931	72.08	-7.8125

Table 9.5: Oracle models with varying number of delay scenarios in scheduling

are rejected as too costly in OC.1.1. In conclusion, complete knowledge on delay occurrences does not simply improve the robustness but leads to a more efficient cost-to-robustness ratio. As expected, the best rtc value of -11.416 can be achieved by the OC.1.1 model. In comparison, the rtc decreases by 22.7% for the OC.10.1 configuration and by 29.24% for the OC.100.1 configuration. It can be stated that the generalization of assumptions on delay occurrences directly affects the efficiency of a robust scheduling strategy in terms of the necessary nominal cost increase for improved robustness.

For a more detailed investigation, we compute the relative change in nominal costs $c^r_{s,m}$ and robustness $stp^r_{s,m}$ per setting s and configuration m compared to the respective cost-efficient solution:

$$c^r_{s,m} = \frac{c_{s,m} - c_{s,\mathrm{CE}}}{c_{s,\mathrm{CE}}}, \qquad \forall s \in S, m \in \{\mathrm{OC.1.1}, \mathrm{OC.10.1}, \mathrm{OC.100.1}\}, \tag{9.5}$$

$$stp^r_{s,m} = \frac{stp_{s,m} - stp_{s,\mathrm{CE}}}{stp_{s,\mathrm{CE}}}, \qquad \forall s \in S, m \in \{\mathrm{OC.10.1}, \mathrm{OC.100.1}\}. \tag{9.6}$$

Each setting s stands for a combination of a robustness penalty factor and one of four seed values as introduced in Section 9.2. Figure 9.5 provides plots of differences in c^r and stp^r values between the configurations. In the left panel, the x-axis depicts the difference in nominal cost improvement between OC.1.1 and OC.10.1 per setting s:

$$\Delta c^r_{\mathrm{OC.1.1},\mathrm{OC.10.1},s} = c^r_{s,\mathrm{OC.1.1}} - c^r_{s,\mathrm{OC.10.1}}, \qquad \forall s \in S. \tag{9.7}$$

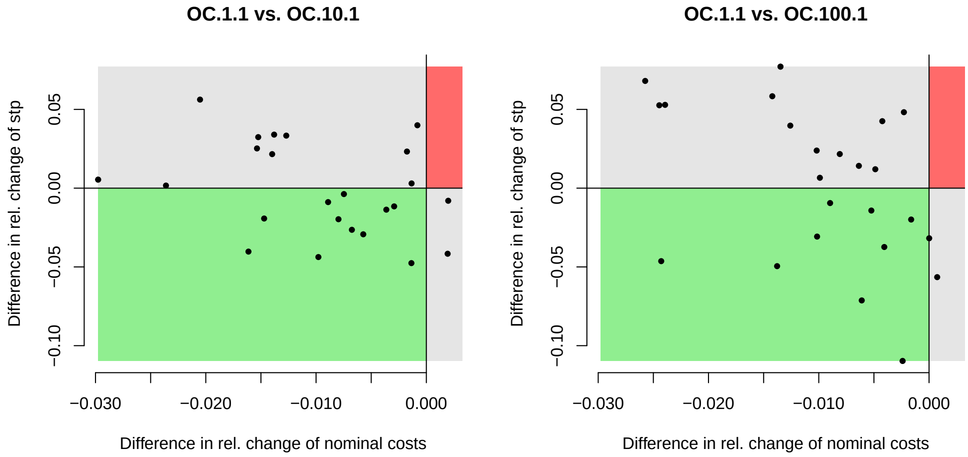

Figure 9.5: Differences in robustness and nominal costs between the oracle model OC.1.1 and its variations OC.10.1 (left) and OC.100.1 (right)

OC.1.1 solutions are cheaper than OC.10.1 solutions if $\Delta c^r_{\text{OC.1.1,OC.10.1},s}$ is negative. The y-axis depicts the respective difference concerning the robustness measure:

$$\Delta r^r_{\text{OC.1.1,OC.10.1},s} = stp^r_{s,\text{OC.1.1}} - stp^r_{s,\text{OC.10.1}}, \qquad \forall s \in S. \tag{9.8}$$

OC.1.1 solutions are more robust if $\Delta r^r_{OC.1.1,OC.10.1,s}$ is negative. In no case, OC.10.1 is dominant to OC.1.1. For eleven of 24 settings, OC.1.1 provides cheaper but less robust solutions. Remaining 13 settings, a higher level of robustness achieved, in 11 cases even at lower nominal costs. In the right panel, analogous relative differences between OC.1.1 and OC.100.1 are provided. Results show a greater spread concerning the differences in stp improvement. In contrast to the left panel, OC.1.1 obtains more robust solutions only for eleven settings. Moreover, nominal costs are smaller or at least equal in all but one settings in OC.1.1. In no setting OC.100.1 is dominant.

In conclusion, previous findings can be confirmed. More specific delay estimation leads to substantial gains in efficiency in terms of the robustness-to-cost ratio. More general and thus equally distributed delay assumptions obtain even slightly more robust solutions, however, at considerably higher nominal costs.

Model	c^r	r^r	rtc
OC.1.1	1.59	-14.05	-11.0416
REV.1.1	1.71	0.99	-1.0493
OC.10.1	2.52	-13.90	-8.5354
REV.10.1	2.33	-4.36	-1.7737
OC.100.1	2.56	-14.29	-7.8125
REV.100.1	2.52	-11.08	-5.0749

Table 9.6: Results for changes in nominal costs and robustness for REV and OC configurations

9.3.2 The Influence of Delay Misplacement

In the preceding analysis, the impact of perfect delay prediction and the influence of generalizing assumptions on delay duration have been assessed. The aim of the following experiments is to measure the impact of the correct estimation concerning *which* flights are expected to be delayed. Results from oracle solutions form the starting point. In OC.1.1, OC.10.1 and OC.100.1, for 50% of the flights F per instance a delay is assumed. We denote these flights as $F_1 \subset F$. In this analysis, we assume delays for flights $F_2, F_2 \subset F, \forall f \in F_2 : f \notin F_1$ in scheduling. Analogously to oracle configurations, related configurations are declared as REV1.1, REV.10.1 and REV.100.1. In simulation, we adhere to the usage of the respective virtual reality scenario. Also, the determination of delay duration is performed by the same GEN model with identical parameters.

Table 9.6 provides aggregated results per configuration. c^r depicts the average relative change of nominal costs compared to cost-efficient solutions. Analogously, r^r depicts the relative change of stp values. The robustness-to-cost ratio is given in the last column. Due to the fact that the same amount of delays is considered in related OC and REV settings, nominal cost changes shall be almost identical. While this holds for models OC.10.1 and OC.100.1, small differences are apparent between OC.1.1 and REV.1.1. The relative change of stp even decreases by about 1% in REV.1.1 compared to cost-efficient solutions. For nearly the half of all test runs, cost-efficient solutions outperform REV.1.1 solutions in terms of robustness (122 of 240 runs). This effect can be explained by the fact that buffer misplacement not only allocates buffer times at wrong places but also implicitly eliminates them from flight

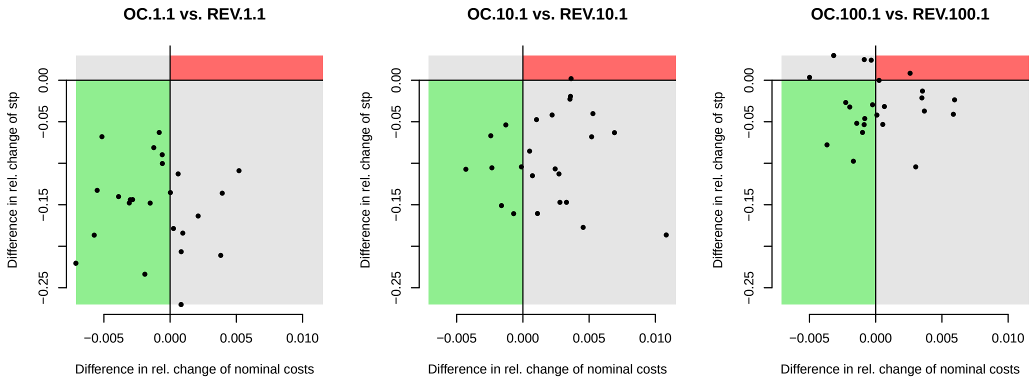

Figure 9.6: Differences between relative nominal cost change and relative stp change of related OC and REV configurations

connections where they are actually necessary. However, the rtc ratio still remains negative[4], indicating a slight improvement of the average robustness-to-cost ratio.

In models REV.10.1 and REV.100.1, r^r values converge towards approximately 14.3. Thus, the stochastic approach overcomes the misplacement of delay assumptions in the REV models at a large number of delay scenarios. However, this is at the expense of a nominal cost increase of nearly 1%.

In this regard, Figure 9.6 provides a more detailed illustration by comparing related OC and REV models for 1, 10 and 100 delay scenarios. The structure of the Figure is analogous to Figure 9.5. For the left panel, the x-axis depicts $\Delta c^r_{\text{OC.1.1,REV.1.1},s}$ per setting s. OC.1.1 solutions are cheaper than REV.1.1 solutions if $\Delta c^r_{\text{OC.1.1,REV.1.1},s}$ is negative. The y-axis depicts $\Delta r^r_{\text{OC.1.1,REV.1.1},s}$ per setting s. OC.1.1 solutions are more robust if $\Delta r^r_{\text{OC.1.1,REV.1.1},s}$ is negative. The middle and the right panel show related results for the comparison of OC.10.1/REV.10.1 and OC.100.1/REV.100.1, respectively. As expected, no model has considerable advantages regarding the nominal cost change. This is different for the difference in the relative change of stp values (Δr^r). OC.1.1 solutions are always significantly more robust than their counterparts, leading to dominant results for 14 of 24 settings.

The OC.10.1 configuration only leads to 7 of 24 dominant results. In the OC.100.1 configuration, 9 settings are dominant, however, the difference to REV.100.1 is

[4]The difference emerges because the rtc is computed on aggregated r^r and c^r values and is holds $\overline{r}/\overline{c} \neq \overline{r^r/c^r}$.

smaller. Furthermore, the REV.100.1 configuration obtains more robust results in 5 settings. In conclusion, the advantage of OC configurations disappears with increased number of considered delay scenarios and the advantage over REV models concerning the robustness increasingly disappears.

9.3.3 The Effects of Primary Delay Under- and Overestimation

In the last part of this study, we address the impact of potential under- and overestimation of primary delays during scheduling compared to the assumptions in simulation.

The following configurations are set up: In the reference configuration *OC.100.100*, μ remains set to value 10.66817. In configurations *un.50, un.40, un.30, un.20, un.10*, μ is decreased by 50%, 40%, 30%, 20% and 10%. Analogously, in configurations *ov.10, ov.20, ov.30, ov.40*, μ is increased by 10%, 20%, 30%, 40% and 50%. For reasons of comparability, delays are assumed for the same flights in every scenario while they differ in duration, according to the respective configuration.

Based on the findings from previous analyses, we use 100 delay scenarios in both scheduling and simulation simulation. Consequently, all delay scenarios used in simulation are known in scheduling in the OC.100.100 configuration. In contrast to the setup of one single virtual reality scenario in simulation, this approach enables the concentration on systematic behavior rather than on the effect of complete knowledge of delays[5]. Computational experiments are repeated four times with different seed values for every setting. In consequence, 2,440 generated and simulated crew schedules form the basis for the evaluation.

Figure 9.7 shows scatterplots with different aggregation levels for the relative change of nominal costs and stp compared to the related cost-efficient solution. Every light gray dot represents a solution for one instance at one of four seed values, one delay propagation penalty factor and one delay prediction model configuration. Dark gray dots show aggregated values per seed value, delay propagation penalty value and delay prediction model configuration, i.e. each dot stands for the average value over all instances. A crescent-shaped first pattern becomes obvious. The pattern is enforced by the last aggregation step in which the connected black dots provide

[5] In Appendix C.4, results are provided for experiments based on reference configurations OC.1.1 and OC.100.1 and derived under- and overestimation configurations. Although slight deviations can be observed in OC.1.1, the same general mechanisms as in Figure 9.7 become apparent.

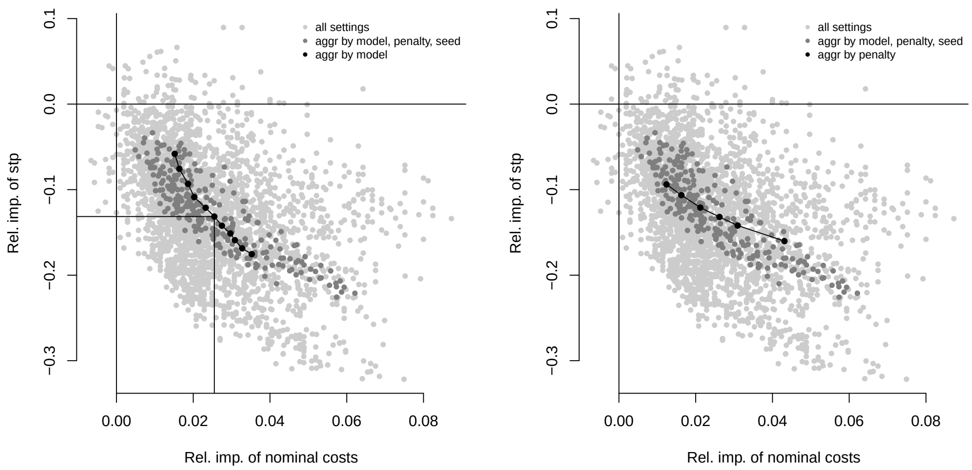

Figure 9.7: Relative change of nominal costs and stp, aggregated per prediction model configuration (left panel) and delay propagation penalty (right panel)

aggregation per delay prediction model. Auxiliary lines indicate the OC.100.100 solution. The pattern forms a Pareto front in which no configuration is dominated by another. Instead, under- and overestimation leads to a diversification of the reference configuration OC.100.100. Primary delay underestimation leads to less robust and less costly solutions while overestimation improves the robustness at the expense of increased nominal costs.

To compare the Pareto front, we aggregate changes in nominal costs and stp per delay propagation penalty factor compared to the related cost-efficient solution. Results are displayed in the right panel of Figure 9.7. Each black dot shows aggregated results for eleven different delay prediction models, 10 instances and four different seed values. Since the propgation penalty factor is the key instrument for controlling the trade-off between nominal costs and robustness, the apparent Pareto front is self-evident. Comparing the Pareto fronts of the left and the right panel, the Pareto front in the left panel is considerably steeper.

With regard to the trade-off between nominal costs and stability, Figure 9.8 provides a graphical illustration of derived rtc values aggregated per delay prediction model configuration (left panel) and propagation penalty factor (right panel). Gray

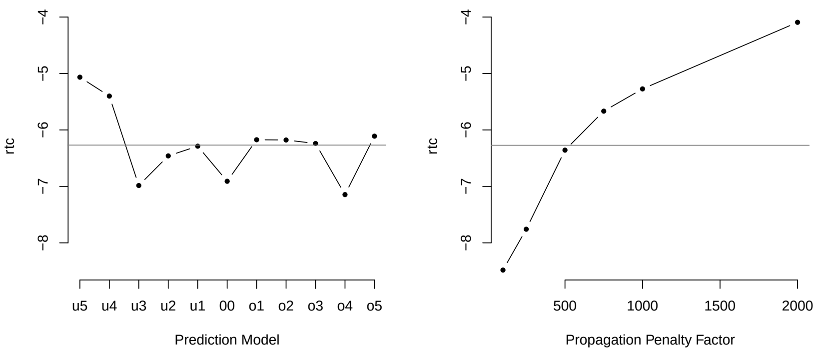

Figure 9.8: RTC ratio, averaged by prediction model (left panel) and delay propagation penalty (right panel)

horizontal lines indicate the average rtc for all solutions. The evident pattern in the right panel can be traced back to the natural effect of increasing marginal costs for robustness – while initially, the incorporation of robustness is rather cheap, further improvement of robustness becomes costly at high propagation penalty factor values.

Similar patterns cannot be observed in the left panel. Applying configurations un.50 and un.40, additional nominal costs for robustness tend to be relatively high. The rtc value of the remaining models fluctuate under the overall average. In conclusion, under- or overestimation of delay duration does not lead to systematically differing marginal costs for robustness in any direction.

9.4 Summary and Implications

In this chapter, we assessed the influence of primary delay prediction modeling on the nominal costs and robustness of crew schedules. For the robustness assessment, we concentrate on the secondary-to-primary delay ratio (stp), indicating the amount of propagated delay per primary delay. The change in robust efficiency is indicated by the robustness-to-cost ratio (rtc), indicating the gain of robustness in relation to the nominal cost increase.

Referring to refined delay prediction, the analysis in Chapter 7 provides an EEPE improvement 1.95% for the best statistical model (see Table 7.4). On average, the related CAT prediction model configuration leads to an rtc improvement of 0.841% compared to the GEN model that bases on assuming identical primary delay occurrence for all flights. As a result, the prediction accuracy improvement of statistical models is not fully reflected in an identical improvement of the robust efficiency. This is because differences in the parameters of the categorical and generalized models are comparably small. For example, alterations in mean values are mostly within a range of ± 5 minutes. In Crew Pairing Optimization, scheduling decisions must be assumed as too restricted to lead to more substantial changes. As a consequence, a straightforward necessity for future work is to obtain results in a setting with an increased degree of freedom by addressing the integrated consideration of crew and aircraft scheduling decisions.

Besides the evaluation of refined data-driven primary delay prediction models, a sensitivity analysis is performed in order to further examine the potential impact of varying primary delay prediction assumptions between scheduling and simulation. We address the complete knowledge of primary delays, the impact of delay misplacements, over- and underestimation. By far the best robustness-to-cost ratio can be achieved if primary delay occurrences are fully known, referred to as *oracle* solutions (OC.1.1 configuration). By considering additional primary delay scenarios in scheduling (OC.10.1, OC.100.1), the complete knowledge on primary delay occurrences in scheduling is weakened. As a result, delays are assumed to be more evenly distributed over all flights. While the robustness outcome remains at a comparable level in these configurations, this is at the expense of a significant rise in nominal costs. A contrary outcome can be observed when evaluating the effect of misplacing primary delays assumptions in scheduling. In the one-scenario configuration (REV.1.1), this leads to the incorporation of ineffective robustness while nominal costs are necessarily increased. About half of the resulting schedules are even outperformed by their cost-efficient counterparts concerning their stp values. Again, this behavior is superimposed by randomizing delay assumptions by considering multiple primary delay scenarios in scheduling. As a result, both stp and rtc values converge between OC and REV models. In the final study on under- and overestimation of primary delays in scheduling, Pareto-optimal solutions concerning nominal costs and robustness are revealed. Marginal costs for robustness are not increased. Notably, the reference configuration does not outperform any other configuration.

Future work has to deal with the incorporation of findings into a robust scheduling approach that provides a higher degree of freedom. Findings of Amberg et al. (2018) show that the integration of crew and vehicle scheduling in provides significantly improved solutions in the field of public bus transportation. It has to be examined in how far the results can be transferred to airline crew and aircraft scheduling. Furthermore, the question is to what extent primary delay prediction is equally relevant if flexibility-related aspects are considered in simulation so that schedule adaptions can be performed in case of delay propagation.

Chapter 10

Modeling Crew and Aircraft Recovery for the Evaluation of Flexibility

For a holistic assessment of the robust efficiency of resource schedules, it is necessary to implement recovery actions that correspond with the definition of flexibility from Section 3.3. The aim is to exploit the schedule flexibility rather than to improve the operational performance to the highest possible degree. In this regard, we provide the conception, implementation and evaluation of recovery mechanisms for crew and aircraft schedules in this chapter.

In Section 10.1, we present the general conception of swaps. Building on that, a rule-based recovery procedure is elaborated in Section 10.2. Special attention is given to the simultaneous consideration of crews and aircraft. A computational study is performed in Section 10.3, followed by conclusions and implications for future work in Section 10.4.

10.1 Operational Crew and Aircraft Swaps

In this section, we firstly discuss variations of swapping procedures for crew and aircraft focusing on constraints that maintain schedule feasibility in the long-term concerning monthly crew schedules and aircraft maintenance. Based on the results, we propose an integrated crew and recovery swapping method that addresses schedule flexibility by design. Eventually, the impact of swapping on the on-time performance and cost efficiency is analyzed by simulation experiments.

In literature, there are two main swap types that are suitable concerning our study goal. While the general process of swapping is straightforward, differences arise from the way how rerouted crews or aircraft return to originally scheduled operations. On the one hand, so called *one-point swaps* are introduced in Shebalov and Klabjan (2006) for crews and in Burke et al. (2010) for aircraft with implicit back swaps at the end of pairings or rotations. On the other hand, Ageeva (2000) defines swaps for aircraft that require explicit back swaps to the original routes. In our context, we denote them as *two-point swaps*.

According to Shebalov and Klabjan (2006), crews must only meet once at an airport within a certain time window to perform a one-point swap as illustrated in Figure 10.1. Pairings are shown in light and dark gray. After a swap, indicated by solid lines, affected crews operate different flights than originally scheduled. An implicit back-swap is performed at the end of both involved pairings (dark gray). Therefore, additional constraints have to be met for a swap. Firstly, both pairings must finish at the same day and at the same airport. Secondly, the number of remaining days in both pairings must be the same in order to prevent extensions of pairings that affect monthly crew schedules and may not be permitted by labor rules. In this regard, Lettovskỳ et al. (2000) discuss the generation of new pairings while protecting the crews' medium-term flight schedule in an optimization-based crew recovery framework. Monthly schedules must be kept consistent, also crew pairings have to end at the same crew base. Moreover, Rosenberger et al. (2000) describe an aircraft swap policy that is similar to the approach of Shebalov and Klabjan (2006), however, there are no indications concerning returning to the original schedules. Also, the definition of *single-point swaps* for aircraft of Burke et al. (2010) corresponds with the presented approach.

In contrast, the aircraft scheduling approach of Ageeva (2000) includes explicit back swaps to the original route as illustrated in Figure 10.2. Aircraft must meet at least two times within a rotation in order to perform a feasible swap. The approach is promising for both crew and aircraft in a hub-and-spoke network with short cycles. Especially, the out-and-back principle enables the swap of certain cycles after returning to the hub in relatively short intervals during the day.

Based on these approaches, we propose an integrated crew and aircraft swapping recovery technique that is capable to evaluate the respective schedule flexibility. The following aspects form the basis of our study design:

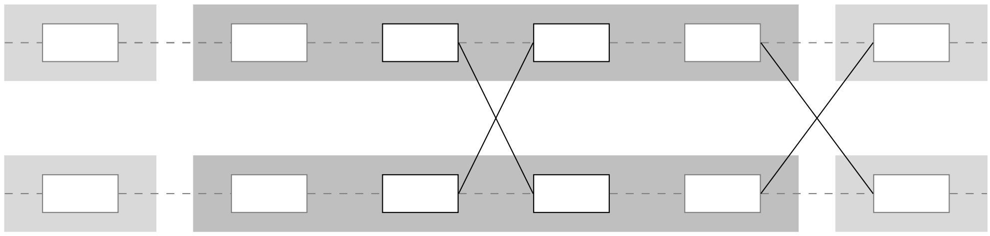

Figure 10.1: Single-point swap with implicit back swap after the crew duty or daily aircraft rotation is finished

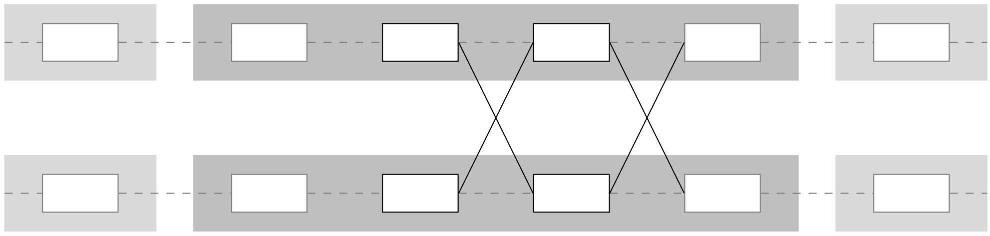

Figure 10.2: Two-point swap with explicit back swap during the crew duty or daily aircraft rotation

Maintaining Feasibility of Crew and Aircraft Schedules The concentration is on finding swaps that prevent delay propagation and lead to realizable crew sit times and aircraft turn times. The feasibility of crew pairing rules always has to be maintained. Swaps can be only performed either if the affected pairings end at the same crew base or if they can be swapped back to their original routes before. Aircraft maintenance interval compliance must be considered accordingly.

Crew Flying License Restrictions Since pairings are not anonymous but related to actual crews in operations, swaps may only be performed in case of identical flying licenses for the respective aircraft. This condition is typically met when considering sub-networks created in Fleet Assignment.

Aircraft Capacity and Configuration Aircraft may only be swapped in case of equal capacity in order to prevent passenger rerouting. Due to the already mentioned aspect of crew flying licenses, this restriction may met implicitly. However, we do not consider special aircraft seat configurations. This includes the shares between business and economy class capacities. We assume that the impact may be minor.

Relaxing Operational Swap Implementation Infeasibility Swaps are usually performed with some lead time for implementation. Passengers may be informed and redirected to other departure gates. Aircraft may swap gates and therefore taxi ways and duration can be affected. Due to the focus on crew and aircraft schedules, limitations in operational implementation are relaxed. By doing so, we must bear in mind that our approach provides an upper bound for the number of swaps that can be performed in real-world operations.

10.2 Integrated Crew and Aircraft Swapping

In this section we describe the swapping method on the basis of previously presented general assumptions and side constraints. Firstly, different connection types and potential disruption cases are illustrated with regard to simultaneous consideration of crews and aircraft. Subsequently, possible swapping routines are mapped out for every connection type in every disruption case. The section concludes with combining the individual routines to a rule-based swapping policy.

If a crew stays on the same aircraft, this is called a *Crew follows Aircraft (CFA)* connection. Hence after performing flight f, the successor flight of the crew is similar to that of the aircraft: $s^c(f) = s^a(f)$. In contrast, if the crew changes aircraft between two succeeding flights, it is referred to as a *Crew changes Aircraft (CCA)* connection, the condition $s^c(f) \neq s^a(f)$ applies. Although CCA connections actually consists of two separate connections – one for the crew (CR) and one for the aircraft (AC) – we handle them separately. This is due to the fact that CR and AC connections also appear when the crew finishes its daily duty and the rotation is continued, or vice versa. For an illustration of all four connection types, we refer to Figure 10.3. Generally spoken, a higher number of CCA connection straightforwardly leads to more interconnected networks for multiple resources. In terms of stability this means that CCA connections are supposed to have the disadvantage that a delay is likely to be propagated on two flights. This effect has been widely confirmed, e.g. by Yen and Birge (2006), Weide (2009) and Dück et al. (2012).

Different connection types imply various minimum ground times for crews. If a crew stays on the same aircraft (CFA) between flight b and its successor $s^c(b)$, the corresponding minimum sit time $mgt^{cfa}_{b,s^c(b)}$ corresponds to the minimum ground time

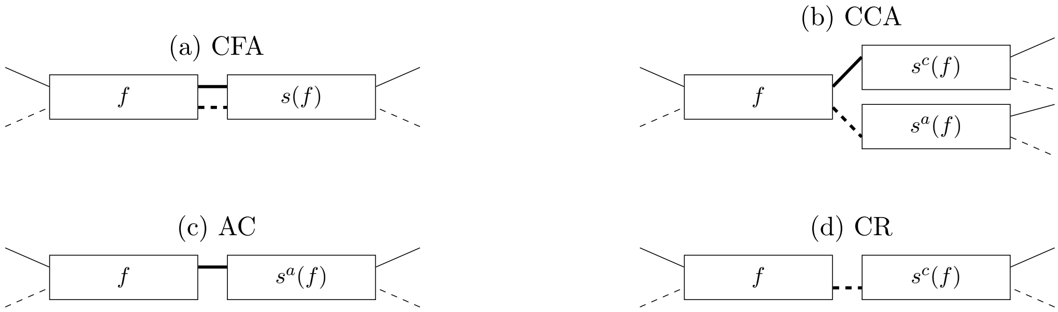

Figure 10.3: Connection types for crew and aircraft

$mgt^a_{b,s^a(b)}$ of the aircraft[1]. In case of an aircraft change (CCA), the minimum ground time of a crew is longer due to additional checks. It holds

$$mgt^a_{b,s^c(b)} < mgt^{cca}_{b,s^c(b)}. \tag{10.1}$$

In case of delayed arrivals, the actual ground time may fall below their corresponding thresholds, leading to six different cases, formally described in Figure 10.4. The x-axis depicts the ground time of the crew, the marks mgt^{cfa} and mgt^{cca} relate to minimum ground times for the crew in case of CFA or CCA connections. The y-axis depicts the ground time of the aircraft with mgt^a standing for the minimum ground time of the aircraft. A connection between f and $s^a(f)$ and $s^c(f)$ can now be classified by its actual g^a and g^c values with regard to the connection type. For generalization, we set $g^c = \infty$ for AC connections and $g^a = \infty$ for CR connections.

[1] The following notation is used in this chapter:

- f flight with arrival delay
- $s^a(f)$ successor flight of the aircraft after flight f
- $s^c(f)$ successor flight of the crew after flight f
- $s(f)$ abbreviated notation in case of $s^a(f) = s^c(f)$
- m candidate flight for a swap, $s^c(m)$ and $s^a(m)$ accordingly
- g^a actual aircraft ground time
- g^c actual crew ground time
- $mgt^a_{f,s^a(f)}$ aircraft minimum ground time between flights f and $s^a(f)$
- $mgt^{cfa}_{f,s^a(f)}$ crew minimum ground time of a CFA connection between flights f and $s^a(f)$
- $mgt^{cca}_{f,s^a(f)}$ crew minimum ground time of a CCA connection between flights f and $s^a(f)$

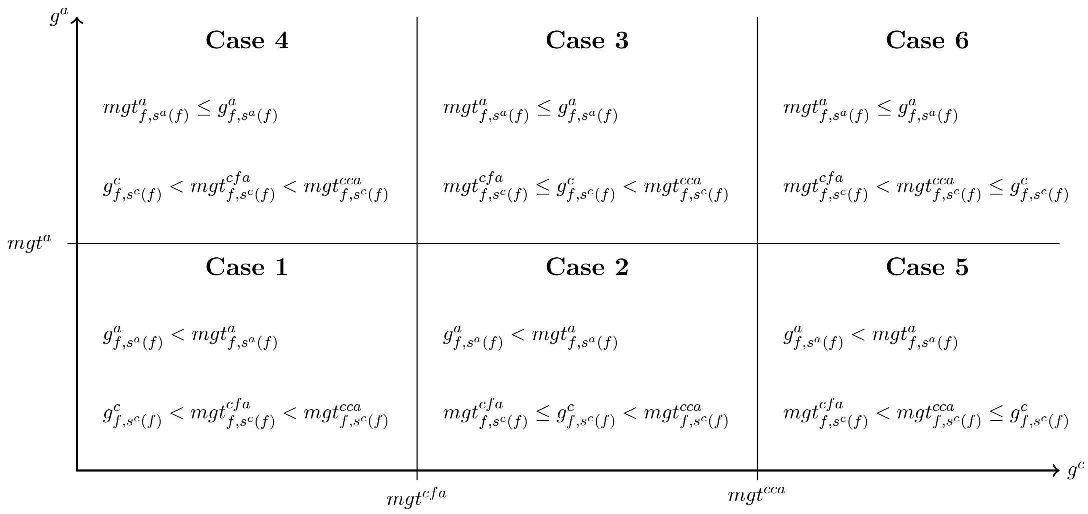

Figure 10.4: Disruption cases and related restrictions on minimal ground times

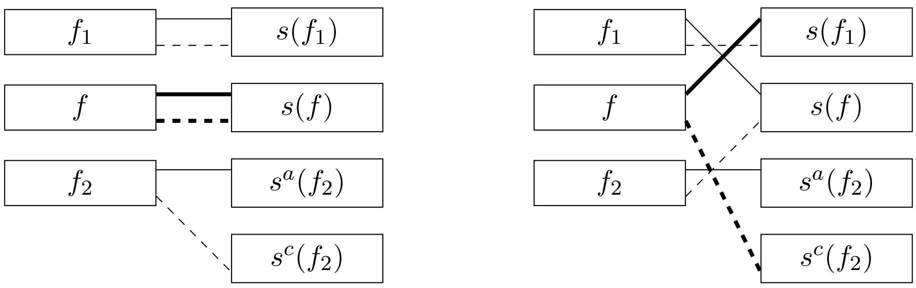

Figure 10.5: Separate swap of crew and aircraft of a CFA connection

Basically, it is always possible to swap resources separately of CFA and CCA connections, too. An exemplary swap is illustrated in Figure 10.5. The crew performs flights f and $s^c(f_2)$ in succession, the aircraft of flight f is subsequently used to carry out flight $s(f^1)$. In consequence, the original connection between flights f and $s(f)$ is avoided. Note that the crew of flight f_1 now has to perform an additional aircraft change and therefore its ground time must exceed $mgt^{cca}_{f_1,s(f_1)}$. Further separate swaps can be derived straightforwardly from the example. However, due to the increased amount of necessary changes to schedules, integrated swaps of crew and aircraft are preferred. In the following, we therefore present integrated swapping procedures for all six cases of Figure 10.4.

Figure 10.6: CFA-CFA swap (Case 1)

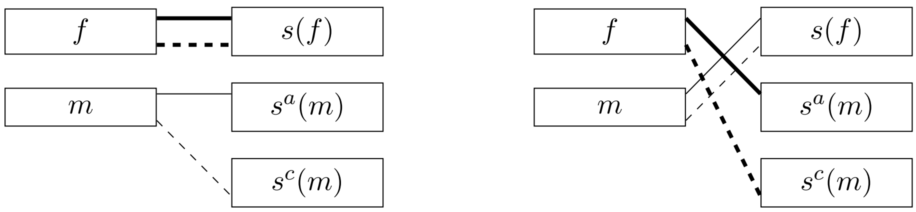

Figure 10.7: CFA-CCA swap (Case 1)

Case 1

In Case 1, both the crew and aircraft minimum ground time is violated. If the disrupted connection is of the CFA type, there are two swap possibilities besides separate swaps: CFA-CFA (Figure 10.6) and CFA-CCA (Figure 10.7). A CFA swap candidate exists when the set of Restrictions 10.2a-10.2d is met.

$$g^c_{f,s(m)} \geq mgt^{cfa}_{f,s(m)} \tag{10.2a}$$

$$g^c_{m,s(f)} \geq mgt^{cfa}_{m,s(f)} \tag{10.2b}$$

$$g^a_{f,s(m)} \geq mgt^a_{f,s(m)} \tag{10.2c}$$

$$g^a_{m,s(f)} \geq mgt^a_{m,s(f)} \tag{10.2d}$$

Note that restrictions on aircraft minimum ground times are implicitly covered by crew minimum ground times in our setting. The main advantage of this kind of swap is that schedules are mostly unaffected and no CFA connections are eliminated. Another possibility is a swap with a CCA candidate, described by Restrictions 10.3a-10.3c.

$$g^c_{f,s^c(m)} \geq mgt^{cca}_{f,s^c(m)}, \tag{10.3a}$$

$$g^c_{m,s^c(f)} \geq mgt^{cca}_{m,s^c(f)} \geq mgt^a_{m,s(f)}, and \tag{10.3b}$$

$$g^a_{m,s(f)} \geq mgt^a_{m,s(f)}. \tag{10.3c}$$

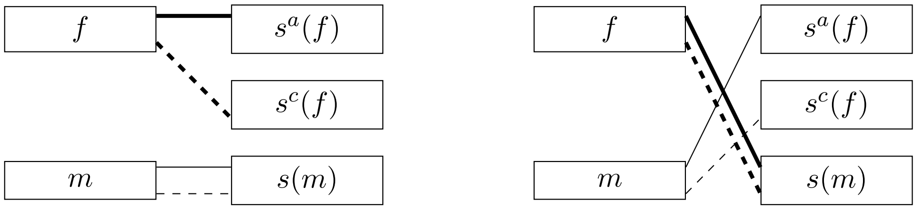

Figure 10.8: CCA-CFA swap (Case 1)

If the disrupted connection is a CCA connection, crew and aircraft may delay different flights. In this context, the CCA-CFA swap is a special case of a separate crew and aircraft swap where the crew stays on the aircraft after flight f. For an illustration, see Figure 10.8. Restrictions 10.4a-10.4c must be fulfilled.

$$g^c_{f,s(m)} \geq mgt^{cfa}_{f,s(m)} \geq mgt^a_{f,s(m)} \tag{10.4a}$$

$$g^c_{m,s^c(f)} \geq mgt^{cca}_{m,s^c(f)} \tag{10.4b}$$

$$g^a_{m,s^a(f)} \geq mgt^a_{m,s^a(f)} \tag{10.4c}$$

Another special case of separate swaps is the CCA-CCA swap (Figure 10.9) where Restrictions 10.5a-10.5d hold.

$$g^c_{f,s^c(m)} \geq mgt^{cca}_{f,s^c(m)} \tag{10.5a}$$

$$g^c_{m,s^c(f)} \geq mgt^{cca}_{m,s^c(f)} \tag{10.5b}$$

$$g^a_{f,s^a(m)} \geq mgt^a_{f,s^a(m)} \tag{10.5c}$$

$$g^a_{m,s^a(f)} \geq mgt^a_{m,s^a(f)} \tag{10.5d}$$

$$\tag{10.5e}$$

In contrast to separate swaps, the presented swap possibilities do not necessitate any additional restrictions on unaffected connections.

Case 2

The aircraft may propagate a delay as the actual ground time falls below the minimum ground time:

$$g^a_{f,s^c(f)} < mgt^a_{f,s^a(f)}. \tag{10.6}$$

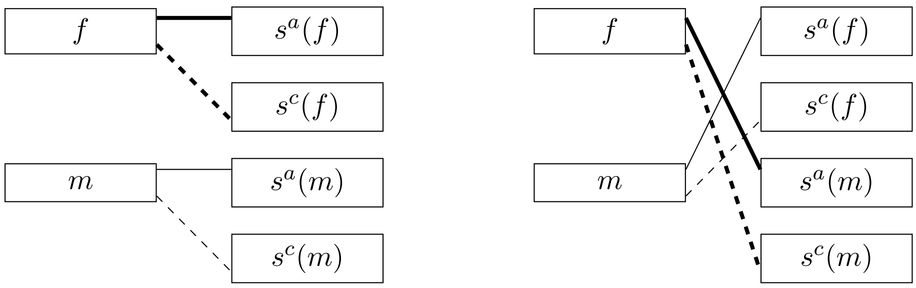

Figure 10.9: CCA-CCA swap (Case 1)

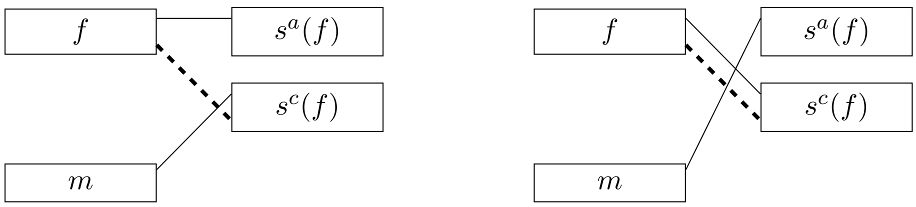

Figure 10.10: Aircraft-to-Crew swap (Case 2)

The crew is only delayed in case of a CCA connection, since

$$mgt^{cfa}_{f,s^c(f)} < g^c_{f,s^c(f)} < mgt^{cca}_{f,s^c(f)}. \tag{10.7}$$

CFA connections cannot be affected as an implication of our assumption that $mgt^a = mgt^{cfa}$ and the fact that $s^c(b) = s^a(b)$. For CCA connection, swaps from Case 1 can be applied equivalently. A special case is the *aircraft-to-crew* swap, see Figure 10.10. The crew ground time is not sufficient to perform an aircraft change. However, it may be sufficient when the crew stays on the aircraft. This can be achieved by reverting the scheduled aircraft change, if Restrictions 10.8a and 10.8b are fulfilled.

$$g^c_{f,s^c(m)} \geq mgt^{cfa}_{f,s^c(m)} \geq mgt^a_{f,s^a(m)}, \qquad s^a(m) = s^c(f) \tag{10.8a}$$

$$g^a_{m,s^a(f)} \geq mgt^a_{m,s^a(f)} \tag{10.8b}$$

By swapping the aircraft, both the delay propagation of crew and aircraft can be prevented. Besides, CCA-CCA and CCA-CFA swaps from Case 1 can be performed, too.

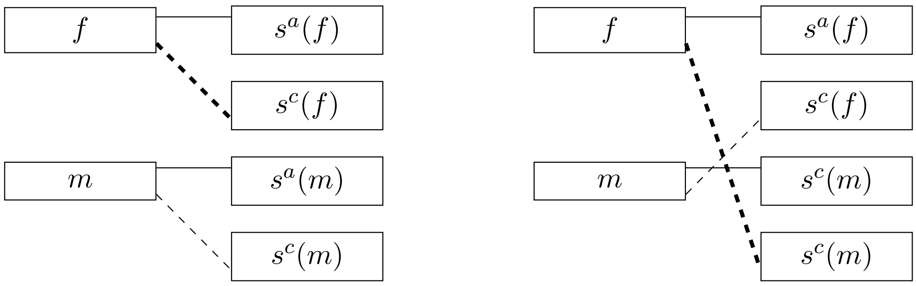

Figure 10.11: CCA-CCA swap (Case 3)

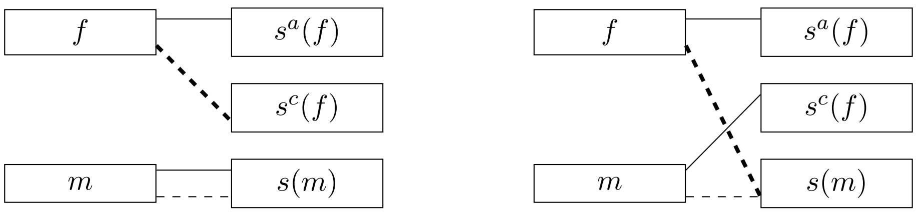

Figure 10.12: CCA-CFA swap (Case 3)

Case 3

In Case 3, only the crew is delayed in a CCA connection. Therefore, only separate crew swaps must be performed. We present two particular possibilities, where the partnering connection for the swap is a CCA or CFA connection, see Figures 10.11 and 10.12. In both cases, Restrictions 10.9a-10.9b apply.

$$g^c_{f,s^c(m)} \geq mgt^{cca}_{f,s^c(m)} \tag{10.9a}$$

$$g^c_{m,s^c(f)} \geq mgt^{cca}_{m,s^c(f)} \tag{10.9b}$$

Due to the policy to maintain CFA connections, a CCA swap candidate may be preferred. Besides, an *aircraft-to-crew* swap can be performed with the same restrictions as in Case 2. In addition, an *crew-to-aircraft* swap is possible as depicted in Figure 10.13 if Restrictions 10.10a-10.10b are fulfilled.

$$g^c_{m,s^c(f)} \geq mgt^{cca}_{m,s^c(f)} \tag{10.10a}$$

$$g^c_{f,s^c(m)} \geq mgt^{cfa}_{f,s^c(m)} \tag{10.10b}$$

Note that Restriction 10.10b is always met since $mgt^a = mgt^{cfa}$ and $s^a(f) = s^c(m)$.

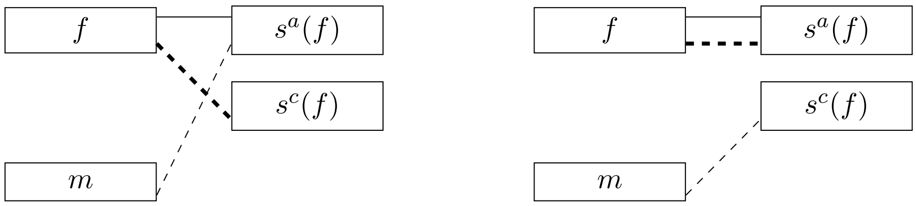

Figure 10.13: Crew-to-Aircraft swap (Case 3)

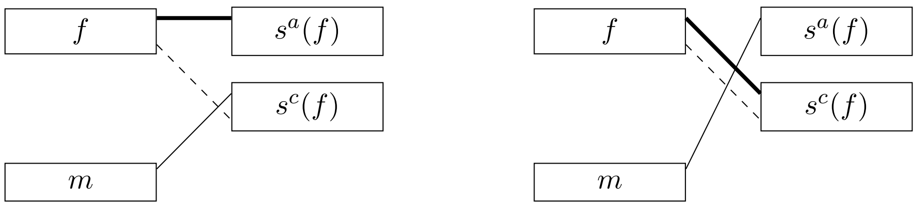

Figure 10.14: Aircraft-to-Crew swap (Case 5)

Case 4

Like in Case 3, only delay propagation by the crew must be prevented whereas the aircraft is on-time. CFA, CCA and crew-to-aircraft swap restrictions comply with the ones from Case 3. Note that in contrast to Case 3, an aircraft-to-crew swap is not useful due to $g^c_{f,s^c(f)} < mgt^c fa_{f,s^c(f)}$.

Case 5

Case 5 represents an aircraft delay while the crew is on-time. Therefore, CFA connections do not exist by definition. Basically, separate aircraft swaps have to be performed. However, a particular swap procedure is preferable, namely to swap the aircraft to the crew's next flight $s^c(f)$. For such an aircraft-to-crew swap it must hold

$$g^a_{m,s^a(f) \geq mgt^a_{m,s^a(f)}}. \tag{10.11}$$

The procedure is illustrated in Figure 10.14. The crew minimum ground time between flights f and $s^c(f) = s^a(m)$ is already sufficient due to $mgt^a_{f,s^c(f)} \leq mgt^{cca}_{f,s^c(f)} \leq g^a_{f,s^a(f)} = g^c_{f,s^c(f)}$. Since the outgoing connection of m cannot be CFA, Restriction 10.11 is the only one to consider. Separately swapping the aircraft with an aircraft of a CFA connection $(m, s(m))$, crew minimum ground time between flights m and $s(m)$ must then exceed $mgt^{cca}_{m,s(m)}$. This restriction is implicitly fulfilled when there is an aircraft change after flight m. CCA-CFA and CCA-CCA swaps from Case 1 are theoretically possible, too. Nevertheless, crews are not delayed in this case, so they do not have to be swapped in this context.

Case 6

In this case, it holds:

$$mgt^{cfa}_{b,s^c(b)} < mgt^{cca}_{b,s^c(b)} < g^{c}_{b,s^c(b)}, \text{ and} \tag{10.12a}$$

$$mgt^{a}_{b,s^a(b)} < g^{a}_{b,s^a(b)}. \tag{10.12b}$$

Both crew and aircraft ground times do not fall below their respective thresholds and therefore no recovery is needed.

The resulting Joint Rule-based Swapping Policy

The presented recovery rules for each case are eventually composed to a swapping policy in Figure 10.15. The order indicates the priority of desired swap opportunities, aiming at maintaining or creating as many CFA connections as possible. The priority guidelines follow the rule to maintain or create as much CFA connections as possible. For example, a CFA connection shall be recovered by a swap with another CFA connection, otherwise a swap with a CCA connection is attempted. In all cases, separate crew and aircraft swaps can be performed if they are feasible and priority swap options cannot be applied.

Based on these considerations, the actual recovery approach follows this general procedure: If delay propagation emerges, the respective disruption case and the connection type is determined. Based on these, candidate connections are searched for based on the priorities given in Figure 10.15. If affected pairings and rotations remain feasible, the swap is performed. Otherwise, the potential of a back swap is examined in order to meet crew rule or maintenance restrictions. If a feasible swap opportunity does not exist, the emerging delay is eventually propagated.

Suggested recovery rules are kept generic. Crew and aircraft minimum ground times depend on the affected incoming and outgoing flight. Distinct turn times at different airports can be implemented straightforwardly. Recovery procedures can be simplified if differences between mgt^{c}, mgt^{cfa} and mgt^{cca} are not given.

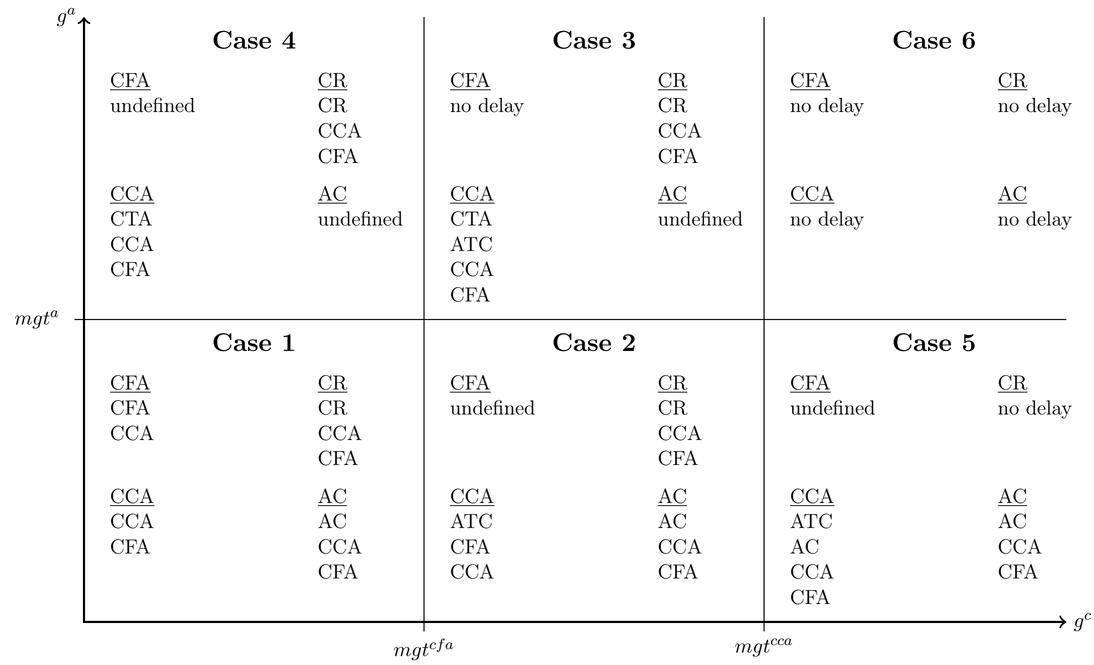

Figure 10.15: Rule-based swapping policy for integrated crew and aircraft

10.3 Computational Study

According to the presented theoretical concept, the implemented recovery procedure is used for computational experiments in order to evaluate its benefit concerning the on-time performance (otp) and secondary-to-primary delay ratio (stp). On the technical side, the frequencies of disruption cases and related successful swap types are examined.

The same set of instances as in Chapter 9 is used for this computational study. While aircraft rotations are already given from historical data, crew schedules are generated based on the following settings: For every instance, we generate nine different crew schedules with varying degrees of stability[2], based on different random generator seed values for primary delay scenario generation. As delay prediction

[2]Selected propagation penalty factors are 0 (cost-efficient solution), 100, 250, 500, 750, 1,000 and 2,000.

setting	***stp***	***otp_0***	***otp_{15}***
prop	1.0083	0.6903	0.8614
dm-15	0.6567	0.7369	0.9186
dm-0	0.4987	0.8435	0.9340

Table 10.1: Robustness measures per recovery settings prop, dm-15 and dm-0

configurations, OC.100.100 and REV.100.100 are applied, i.e. 100 delay scenarios are considered both during scheduling and simulation[3].

In summary, 490 generated schedules form the basis for the subsequent analysis. On average, schedules contain 52.21% CFA connections, followed by CCA with a share of 24.02%. There is a comparably high number of AR connections (16.48%), resulting from crew duties to be finished before the end of a rotation so that another crew takes over the aircraft for the next flight. Analogously, the crew proceeds its duty while the rotation is not continued after 1% of all flights. For further 6.28% of all flights, there is no successor flight for neither crew nor aircraft, mostly before overnight stays. Due to our focus on regular daily operations, delay propagation is not considered for these cases.

For the evaluation of generated schedules, we use recovery settings prop, dm-0 and dm-15 in simulation. No swaps are considered in prop so that every delay is propagated. In dm-0, swaps are attempted for every delay of at least one minute. In dm-15, swapping attempts are only performed if the propagated delay is larger than 15 minutes. The dm-15 setting is selected based on the common on-time performance threshold of 15 minutes. This distinction allows the examination of potential systematic differences in swapping capability of larger delays. Moreover, delays in simulation are assumed to be 20% higher than in scheduling in order to increase necessary recovery interventions.

Aggregated results concerning the stp and otp are provided in Table 10.1. In the prop setting, one minute of primary delay induces more than one minute of secondary delay. This value can significantly be reduced by more than a half in the dm-0 setting. A similar picture is shown for the on-time performance otp_0 which is increased by more than 15%. Effects are apparently smaller for otp_{15}. However, the dm-0 setting can increase its absolute value by additional 1.54% compared to

[3]See Section 9.3 for details on primary delay prediction model configurations.

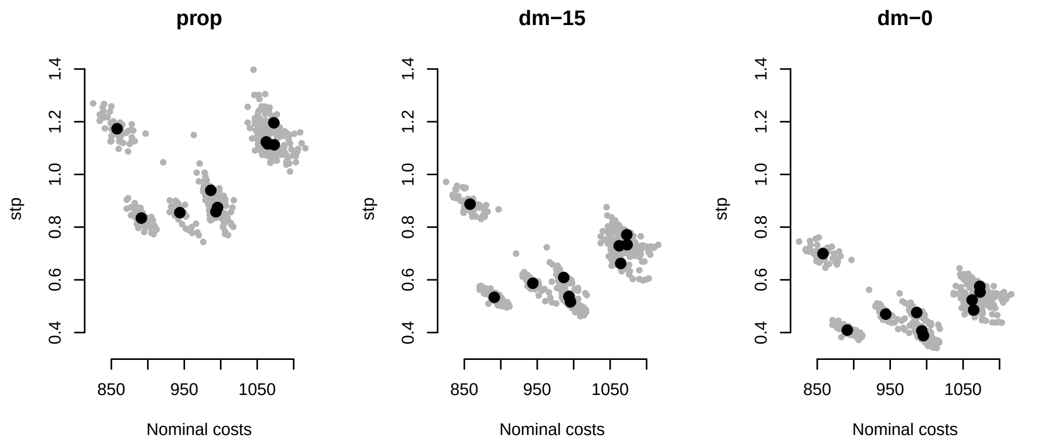

Figure 10.16: Nominal costs and secondary-to-primary delay ratio (stp) per recovery setting

dm-15. This effect can be explained by the reduction of accumulated delays smaller of maximum 15 minutes in the dm-0 setting.

Figure 10.16 illustrates the relation of nominal costs and stp outcome for every generated schedule (gray points). Black points indicate aggregations per instance. Pareto-fronts for every instance become slightly apparent, however, blurred by the impact of varying seed values and prediction model configurations. Notably, the standard deviation of stp values per instance (roughly illustrated by the cloud around each black point) decreases from 0.04412 (prop) to 0.3078 (dm-15) and finally 0.0247 (dm-0), indicating decreased impact of stability when swaps can be performed. As a consequence, Pareto fronts become less steep in dm-15 and dm-0 compared to prop.

For a more specific representation of differences between prop and dm-15/dm-0 settings, Figure 10.17 shows relative differences in nominal costs and stp values. For given generated schedules $s \in S$, the y-axis depicts

$$\Delta stp_{s,dm} = \frac{stp_{dm} - stp_{prop}}{stp_{prop}}, \qquad \forall s \in S, dm \in \{\text{dm-15}, \text{dm-0}\}. \tag{10.13}$$

Black points indicate average $\Delta stp_{s,dm}$ values per instance. The dm-0 setting provides significantly better results with an average relative improvement of 50.582% compared to 35.055% for dm-15.

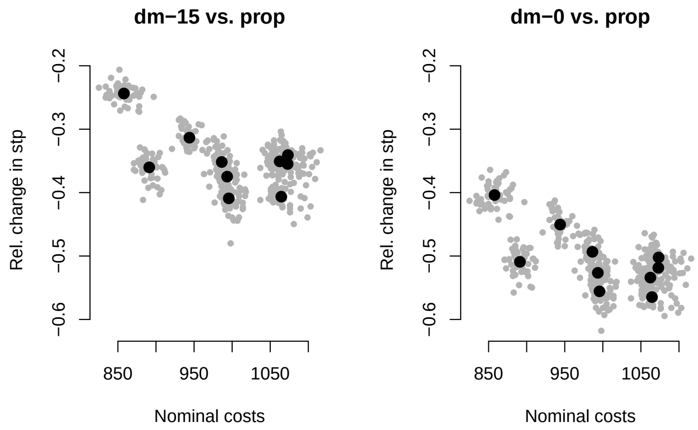

Figure 10.17: Relative stp change between dm-15/dm-0 and prop

In the following, we give insights into the success rate of swapping mechanisms. In close relation to recovery cases presented in Figure 10.15, we provide a detailed view on intended and performed swaps per disruption case in Figure 10.18. In the examination, we refer to the dm-0 setting. For comparison, results for dm-15 can be found in Appendix C.5. Right next to the case labels, the occurrence frequency of the respective case is given. So to speak, 33.07% of flights with a departure delay violate both crew and aircraft minimum ground time before the next flight(s) so that delay propagation becomes imminent. They are therefore assigned to Case 1. Case 2 (1.11%) includes CCA connections in which both crew and aircraft are likely to propagate delay. There is no sufficient time for an aircraft change of the crew. In Cases 3 and 4, propagation is imminent only for the crew (5.82% and 3.43%), respectively. Analogously, delay propagation only due to the aircraft is likely to happen for 14.75% of delayed connections (Case 5). For about 41.82% of flights departing late, no propagation on consecutive flights is expected (Case 6).

These values are further differentiated in relation to connection types (underlined). In this regard, 28.79% of all delayed connections meeting the conditions of Case 1 are of the CFA type. Consequently, remaining 4.28% of connections in Case 1 are CCA connections. Values for remaining cases can be derived accordingly.

Finally, possible swaps are listed with their particular success rate for every connection type. Note that in contrast to the occurrence frequencies of disruption cases

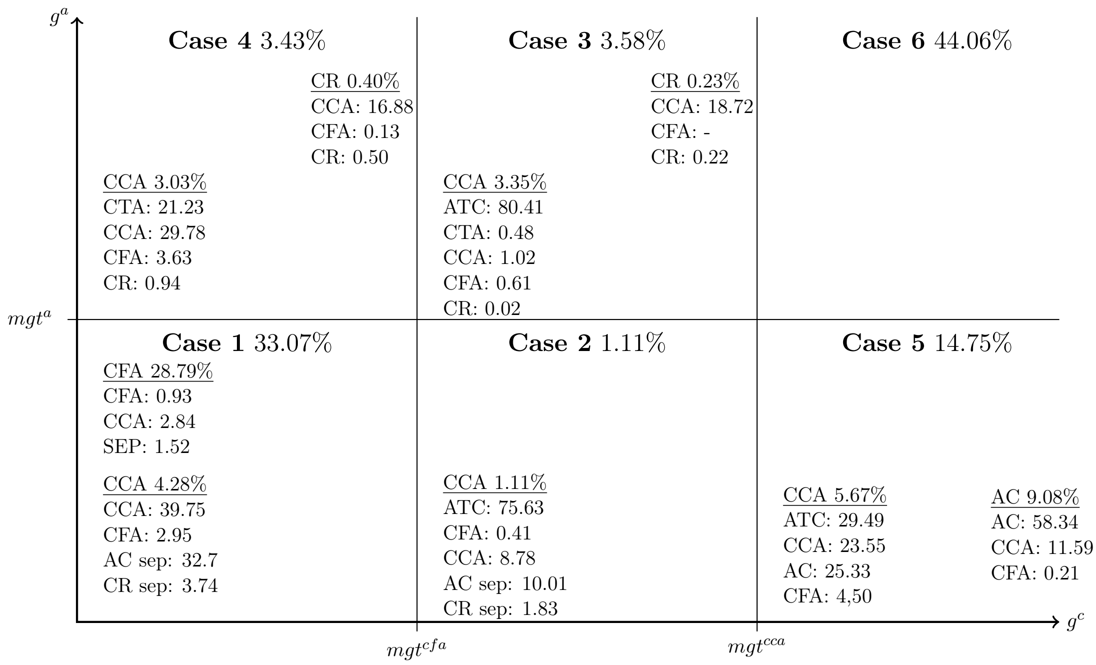

Figure 10.18: Frequency of swaps per disruption case, all figures are given in %

and connection types, the success rates are mutually independent. Although CFA is the most frequent connection type, these connections can be recovered in accumulated 4,39% of all cases. Things are different for CCA connections. They are less frequent, however, they can be recovered more often. 42,7% of these connections allow simultaneous recovery in a CCA-CCA or CCA-CFA swap. If such swaps are not possible, separate aircraft swaps are much more likely than crew swaps.

In Case 2 it becomes apparent that aircraft-to-crew swaps offer a substantial success rate of 75.63%. However, the impact is quite low with regard to the total occurrence frequency of 1.11% for Case 2. In rare cases 3 and 4, only the crew is delayed. Similar to Case 2, ATC swaps have a high success ratio for CCA connections. CTA swaps can be performed in more than one of five attempts in Case 4.

Only the aircraft can be delayed in Case 5. Again, ATC swaps are an important instrument for recovering CCA connections, however, to a lesser extent than in Cases 2 and 3. AC connections are often recovered by swapping with another AC connection, assumingly by taking the advantage of increased occurrence of AC connections when crew duties end at their crew base in a certain time window. For remaining 44.06% of all flights, departure delays do not propagate to following flights, thus, no swaps have to be performed (Case 6).

A comparison to results from the dm-15 setting shows similar results, with decreased success rates throughout all swapping attempts[4]. This behavior indicates that swaps are harder to perform when potential propagated delays are higher. A notable consequence is that ATC/CTA swaps are rarely applicable although they are highly successful in the dm-0 setting.

10.4 Summary and Implications

In this chapter, we presented a rule-based recovery framework for crew and aircraft schedules. Based on a general conception of crew and aircraft swaps in Section 10.1, mechanisms for simultaneous crew and aircraft swaps has been developed in Section 10.2. The outcome of computational experiments has been examined in Section 10.3, revealing the impact of the recovery approach on key robustness measures. If swaps are applicable every time when delay propagation emerges, the secondary-to-primary delay ratio can be reduced by about 50%. Moreover, the on-time performance can be increased by 15.32%. Obtaining the common 15 minutes threshold for the on-time performance, there is still an improvement of 7.26%. On the technical side, it has turned out that the crew-follows-aircraft (CFA) principle impedes the swapping capabilities for respective connections. In contrast, special swapping techniques (ATC/CTA) that re-establish CFA connections are highly successful for smaller delays. However, the occurrence frequency of their application is comparatively low.

The conception of swapping mechanisms can be used as groundwork for the development of scheduling techniques for improving schedule flexibility. Furthermore, the impact of changing degrees of schedule stability and flexibility on the capabilities of schedule recovery must be examined. These topics are addressed in the subsequent Part IV of this thesis.

Future work has to deal with two emerging questions concerning practical applicability of swaps. On the one hand, swaps are designed as a cost-neutral recovery technique in the scope of this study. In this regard, the workload and actual opera-

[4]For results obtained in the dm-15 setting, see Appendix C.5.

tional expense of swaps must be further examined. Derived values can be straightforwardly included into the assessment of the benefit of a swap. On the other hand, and linked to the first question, the actual benefit of swaps for relatively small delays in real-world operations has to be investigated. Organizational efforts such as re-assignment of gates and ground equipment may inevitably make swaps for small delays less appealing. As a result, the proximity to operational reality can be further improved.

Part IV

Mutual Impacts of Operational Stability and Flexibility

Chapter 11

Increasing the Flexibility of Airline Crew Schedules

In Part IV, we address the impact of robust scheduling strategies on the robust efficiency focusing on stability and flexibility aspects. While in this chapter, we introduce a necessary approach for increasing the flexibility of schedules, the core study on mutual impacts between stability and flexibility is presented in Chapter 12.

The incorporation of stability into resource schedules is performed by the reallocation of buffer times. By appropriately aligning buffer placement to well estimated delay propagation risks, the nominal cost increase for improved robustness can be held to a minimum. For the consideration of stability in our studies, we use a slightly adapted version of the approach of Dück et al. (2012) in connection with primary delay generation models from Chapter 9.

In contrast, the concept of flexibility is an important driver when the ability of schedule recovery is taken into account. Explicit swap opportunities are incorporated into schedules in order to facilitate the adaption of schedules during operations. In this regard, we propose an alteration of the approach of Dück et al. (2012) that addresses schedule flexibility in this chapter. Preliminary considerations have indicated that their iterative solution approach is not sufficiently suited to the purpose of considering flexibility in simultaneous crew and aircraft scheduling. A main reason is that more sophisticated decomposition and interaction strategies must be developed in order to synchronize the incorporation of swap opportunities that are compatible between crews and aircraft.

As a consequence, the presented solution approach and the computational study are focused on the flexibility improvement solely for crew schedules in this stage of our research. Based on this, subsequent future work has to deal with refining decomposition strategies that take into account the interconnected nature of flexibility aspects for crews and aircraft as it has become apparent in Chapter 10. In this context, the integrated set partitioning formulation and the straightforward model decomposition can be seen as preliminary work for necessary enhancements in future research.

The presented work is partially based on Ionescu and Kliewer (2011). In the modeling and optimization approach, the main goal is to increase the number of swap opportunities for crew connections that are likely to propagate delay. Objective benefits for each potential swap are applied. The main drawback of the approach is that risky connections are explicitly included in crew schedules if a swap opportunity can be provided for these connections. With reference to Scholl (2001), this approach addresses the *schedule elasticity*, describing the general ability of schedule modification. In contrast, the *schedule flexibility* targets a specific purpose which is in our case the minimization of propagated delay. We therefore propose an alternate model formulation and solution strategy. Instead of maximizing the number of swap opportunities, we minimize the amount of propagated delay. If an appropriate swap opportunity is scheduled for a certain risky connection, the related delay propagation penalty is repealed. In this way, the explicit use of risky connections is prevented.

The remainder of this chapter is organized as follows: In Section 11.1, a mathematical formulation of the crew pairing problem with stochastic recourse for the evaluation of schedule flexibility is presented. A column-generation-based solution approach is described in Section 11.2. In a computational study, the impact of in increasing degree of flexibility on both robustness and nominal costs is examined in Section 11.3. In order to provide comparable results for future integration of crew and aircraft scheduling, generated schedules are simulated along with given aircraft rotations. Eventually, a summary of findings is given in the Section 11.4.

11.1 A Set Partitioning Model Formulation with Stochastic Recourse

Equations 11.1-11.5 show the base set partitioning model for integrated Crew Pairing and Aircraft Routing with stochastic recourse of (Dück et al., 2012, p. 50).

$$\min \quad \sum_{j \in P} c_j x_j + c_Q Q(X, Y, \Omega) \tag{11.1}$$

$$\text{s.t.} \quad \sum_{j \in P} a_{fj} x_j = 1 \qquad \forall f \in F \tag{11.2}$$

$$\sum_{k \in R} b_{fk} y_k = 1 \qquad \forall f \in F \tag{11.3}$$

$$x_j \in \{0, 1\} \qquad \forall j \in P \tag{11.4}$$

$$y_k \in \{0, 1\} \qquad \forall k \in R \tag{11.5}$$

F represents the set of flights that must be covered. P is the set of all feasible crew pairings for F. The binary variable $x_j \in X \subseteq P$ determines whether pairing $j \in P$ is part of the solution or not. The vector c contains nominal costs for each pairing $j \in P$.

The objective function 11.1 consists of the crew costs and a stochastic recourse function $Q^C(X, Y, \Omega)$ that reflects additional costs for the non-robustness of schedules with regard to primary delay scenarios Ω. $X \subseteq P$ and $Y \subseteq R$ describe the subset of selected crew pairings and aircraft rotations in the solution. The trade-off between nominal cost minimization and robustness maximization can be controlled by weighing the outcome of $Q(X, Y, \Omega)$ by the penalty factor c_Q. Robustness gains more importance with higher values of c_Q. Since deployment costs for aircraft of the same type are usually assumed to be identical, they do not have to be considered in the objective function, see (Weide, 2009, p. 71).

The matrix A consists of binary entries a_{fj} with value 1 if flight f is covered by pairing j, 0 otherwise. Equation 11.2 ensures that every flight f is covered by exactly one crew pairing. Analogously, R is the set of all feasible rotations, entries b_{fk} of the matrix B determine whether rotation $k \in R$ covers flight f or not. According to Equation 11.3, every flight must be covered exactly once by an aircraft rotation. Finally, restrictions 11.4 and 11.5 describe the binary nature of variables x_j and y_k. For the sole consideration of crew pairings in our study, Restrictions 11.3 and 11.5 of the integrated formulation are neglected.

11.1.1 The Objective Function of the Decomposed Stochastic Recourse

The stochastic recourse can be decomposed into

$$Q(X, Y, \Omega) = Q^C(X, \Omega) + Q^R(Y, \Omega) \tag{11.6}$$

for separate consideration in either Crew Pairing or Aircraft Routing. No interactions between crew pairings and aircraft rotations are considered yet. For the purpose of our study, the following presentation focuses on the decomposed stochastic recourse function $Q^C(X, \Omega)$ for the Crew Pairing step and the outcome of $Q^R(Y, \Omega)$ is set to 0. The objective function of the recourse determines the amount of propagated delay for a feasible crew pairing solution X under assumption of delay scenarios Ω, see Equation 11.7.

$$Q^C(X, \Omega) = \min \sum_{\omega \in \Omega} p_\omega \sum_{f \in F} (1 - s_{f\omega})\delta^s_{f\omega} + s_{f\omega} c_{s_{f\omega}} \tag{11.7}$$

Each scenario $\omega \in \Omega$ can be associated with an occurrence probability p_ω. Propagated delay $\delta^s_{f\omega}$ per flight f and scenario ω is taken into account if no swap opportunity exists. The binary variable $s_{f\omega}$ indicates a swap opportunity for flight f in scenario ω. Note that we only consider swaps that entirely prevent delay propagation, i.e. drop $\delta^s_{f\omega}$ to zero. With the last term, individual implementation costs $c_{s_{f\omega}}$ can be applied for each swap $s_{f\omega}$.

11.1.2 Restrictions of the Decomposed Stochastic Recourse

In the following, we present the constraints of the recourse model. They address (A) the compliance with ground and flying times, (B) the correct calculation of the propagated delay $\delta^s_{f\omega}$, and (C) feasibility restrictions that must be fulfilled by valid swap opportunities.

(A) Minimum Ground Time and Flying Time Compliance

The following Constraints 11.8-11.12 reflect the propagation mechanism described by Equations 8.5-8.9 in Section 8.1. Due to the set partitioning formulation of the main problem, each flight f is associated with exactly one pairing $j, j \in X$ of a feasible solution $X \subseteq P$. Constraint 11.8 ensures compliance with minimum ground times between flight f and its predecessor $p_j(f)$ in this pairing j. The actual departure time $ATD_{f\omega}$ depends on the actual arrival time $ATA_{p_j(f)\omega}$ of $p_j(f)$, the crew minimum ground time $mgt^C_{p_j(f)f}$ and potential primary delay $\delta^p_{f\omega}$:

$$ATA_{p_j(f)\omega} + mgt^C_{p_j(f)f} + \delta^p_{f\omega} \leq ATD_{f\omega}, \quad \forall f \in F, \omega \in \Omega. \tag{11.8}$$

The difference between actual departure and arrival times must be at least as long as the planned flying time t_f:

$$ATA_{f\omega} - ATD_{f\omega} \geq t_f, \quad \forall f \in F, \omega \in \Omega. \tag{11.9}$$

Furthermore, early departures are not considered:

$$ATD_{f\omega} \geq STD_f, \tag{11.10}$$
$$ATA_{f\omega} \geq STA_f, \tag{11.11}$$
$$\forall f \in F, \omega \in \Omega.$$

(B) Calculation of Propagated Delays $\delta^s_{f\omega}$

With given $ATD_{f\omega}$, $STD_{f\omega}$ and $\delta^p_{f\omega}$, it holds for the secondary delay

$$\delta^s_{f\omega} = (ATD_{f\omega} - STD_{f\omega}) - \delta^p_{f\omega}, \quad \forall f \in F, \omega \in \Omega. \tag{11.12}$$

(C) Feasibility Restrictions for Swap Opportunities

The binary variable $s_{f\omega}$ determines if a valid swap opportunity exists for flight f under scenario ω. For setting $s_{f\omega}$ to 1, Restrictions 11.13-11.14 must be fulfilled. Firstly, a crew connection between $p_j(f)$ and f must induce a certain risk of delay propagation: only if $\delta^s_{f\omega}$ exceeds the threshold t_δ, a swap opportunity can be provided. Following, the binary variable $\rho_{p_j(f)f\omega}$ has value 1 if the connection between $p_j(f)$ and f is risky. In this context, M_δ is a sufficiently large number so that the inequality is always valid if $\rho_{p_j(f)f\omega}$ is set to 0. The condition can be formulated as follows:

$$t_\delta - (1 - \rho_{p_j(f)f\omega})M_\delta \leq \delta^s_{f\omega}, \quad \forall f \in F, \omega \in \Omega. \tag{11.13}$$

The connection to $s_{f\omega}$ is drawn as a swap opportunity can only be used to revoke the influence of $\delta^s_{f\omega}$ in the objective function if the respective connection between flights $p_j(f)$ and f is risky. It follows:

$$\rho_{p_j(f)f\omega} \geq s_{f\omega}, \quad \forall f \in F, \omega \in \Omega. \tag{11.14}$$

Consistency requirements for a feasible swap have been developed in Section 10.2. We declare variable $s^L_{f_{jk}\omega} \in \{0,1\}$ as an indicator of whether a swap is possible for flight f between pairings j and k in scenario ω. Note that determination of $s^L_{f_{jk}\omega}$ is a non-linear task due to crew pairing and duty regulations. The connection to $s_{f_{jk}\omega}$ can be drawn as follows:

$$\sum_{k \in X,\, k \neq j} s^L_{f_{jk}\omega} \geq s_{f\omega}, \quad \forall f \in F, \omega \in \Omega. \tag{11.15}$$

The formulation of Constraint 11.15 implicitly ensures that pairings j and k are part of the solution X[1].

11.2 The Column Generation Solution Approach

The proposed stochastic model can be solved by column generation[2] with an extension that explicitly handles the recourse model. The solution process is illustrated

[1] The recourse model for Aircraft Routing can be derived straightforwardly. The main difference is that the minimum ground time parameter $mgt^C_{p_j(f)f}$ must be replaced by the aircraft specific $mgt^A_{p_j(f)f}$ in Constraint 11.8. Also, the determination of $s^L_{f_{jk}\omega}$ must be adapted to aircraft rotation requirements.

[2] The fundamental column generation process is described in Appendix A.

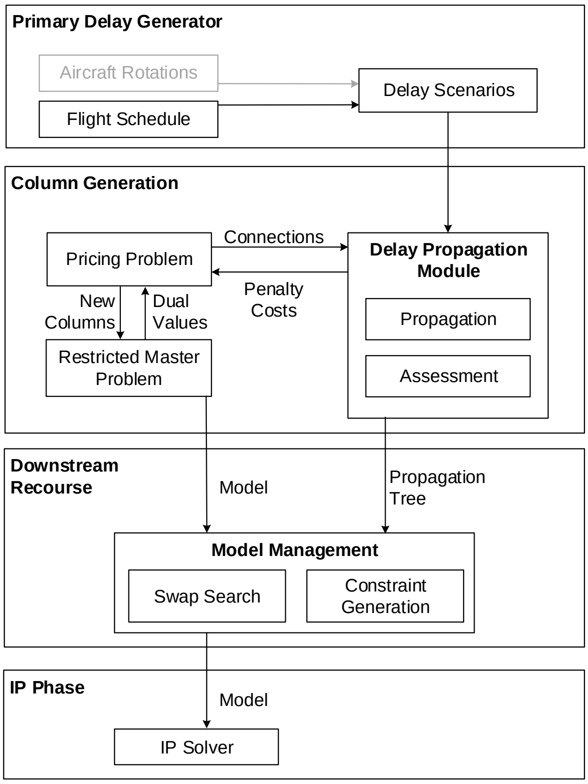

Figure 11.1: The column-generation-based solution procedure

in Figure 11.1. Initially, a set of delay scenarios Ω for the underlying flight schedule is generated by the selected *Primary Delay Generator*. As proposed in Dück et al. (2012), delays induced by aircraft connections can be taken into account in the subsequent Crew Pairing step. Ω then contains primary delays as well as delays propagated by fixed aircraft rotation connections.

The *Pricing* step is performed using a dynamic programming approach for a resource constrained shortest path problem. Pairings are constructed by gradually append flights whose dual values suggest an improvement of the current restricted master solution. During this step, delay propagation is evaluated and assessed for every connection based on Ω using the fundamental approach of Dück et al. (2012) that is encapsulated in the *Delay Propagation Module*. We use a refined evaluation strategy that focuses on the secondary-to-primary delay ratio (stp) rather than

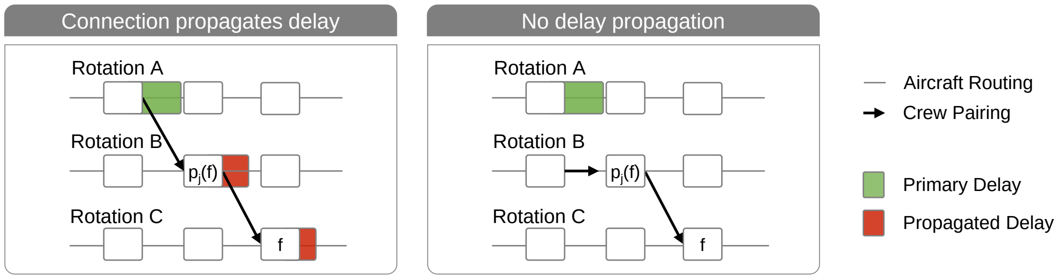

Figure 11.2: Context-dependent propagation assessment for the connection between flights $p_j(f)$ and f

the on-time performance (otp). Consequently, the original decomposed stochastic recourse $Q^P(j, Y, \Omega)$ for pairing j is reformulated to:

$$Q^P(j, \Omega) = \min \sum_{\omega \in \Omega} p_\omega \sum_{f \in F} \delta^s_{f\omega}. \tag{11.16}$$

Penalty costs $c_Q \cdot Q^P(j, \Omega)$ are added to the nominal costs of pairing j. Depending on the chosen c_Q value, generated pairings either tend to be more stable or more cost-efficient. If $c_Q = 0$, cost-efficiency is the only objective.

Due to the additive structure of delay propagation in a pairing j, the assessment of propagation risk can be performed connection-wise during the dynamic programming approach. In detail, if a flight f is appended to a currently generated pairing j, we denote the set of all preceding flights as subroute j'. The assessment of delay propagation risk depends on j'. A minimal example for this is illustrated in Figure 11.2.

The left panel illustrates a crew connection $(p'_j(f), f)$ of a subroute j' (bold black lines) with two consecutive aircraft changes between rotations A, B and C (gray lines). Primary delay (green mark) is propagated to $p'_j(f)$ of Rotation B (red mark). In the following, the connection to flight f is assumed to be risky as it propagates the delay further to Rotation C. In contrast, the assessment of the same connection between $p'_j(f)$ and f may vary due to a different precedent subroute, exemplarily shown in right panel. Since the crew initially stays on the aircraft of Rotation B, it is not affected by primary delay. In this case, the considered connection is not assumed to have a risk of delay propagation.

Naturally, many identical subroutes are considered, e.g. when duties share the same itinerary in the morning but differ for the rest of the day. For reasons of computational efficiency, we therefore store information on already evaluated subroutes. We use a follow-on based propagation forest leaning on the structure of propagation trees, seeAhmadBeygi et al. (2008). The concept of follow-ons has been initially introduced in Ryan and Foster (1981) and describes two flights $p_j(f)$ and f that are consecutively operated by one crew. Each path in an individual tree in the forest represents a pairing j. Hence, each partial tree determines a subroute j' of j. The root nodes of the trees each refer to the very first follow-on at the start of any pairing or rotation. Subsequent follow-ons may appear in different trees, depending on their predecessors.

For every follow-on, we store the amount of secondary delay that is induced by it depending on its subroute j'. Corresponding values are reflected by $Q^P(f, j', \Omega)$ as partial solutions of $Q^P(j, \Omega)$ in the dynamic programming approach for the last follow-on of j'. By storing these values, the assessment must be performed only once for each follow-on $(p_{j'}(f)$,$f)$ during pricing. The stored information is also an important input factor for the later Downstream Recourse.

New columns that relate to generated pairings are added to the *Restricted Master Problem* (RMP). It is therefore updated and solved in every column generation iteration. The RMP is formulated as an LP relaxation of the original model including all available columns of the current iteration and can be solved by a standard LP solver. The column generation process terminates if either no new columns are generated or the objective of the RMP cannot be further improved.

In the following *Downstream Recourse*, the swap opportunity search and related model modifications are encapsulated in the *Model Management*. The propagation trees built up by the Delay Propagation Module are the starting point. For tractability reasons, we sort follow-ons by their severeness of delay propagation risk, following the assumption that swaps for more risky follow-ons have the greatest impact on the stp. For every considered follow-on, candidate pairings for a swap opportunity are located within the pairing pool that has been generated in Column Generation. The swap search technique is identical to the rule-based recovery approach presented in Section 10 and applies to Restrictions 11.8-11.12 for minimum ground time and flying time. Performing a swap must result in feasible pairings and rotations, 11.15. In comparison, the computational efforts are significantly higher with an overall complexity of $\mathcal{O}(n^2)$. However, in practice the complexity can be significantly reduced because only pairings must be considered that contain flights departing at the same airport with a departure time relatively close to the respective disrupted connection.

Adhering to the notation used in the description of the Delay Propagation Module, let $(p_{j'}(f), f)$ be a follow-on in the context of subroute j' with an considerable risk of propagation. All pairings that include the follow-on and j' are part of set J. For every $j \in J$, let K_j be a set of pairings whose elements $k \in K_j$ provide a swap opportunity for the crew connection between flights $p_{j'}(f)$ and f in j. The current model is then enhanced by the following items: A newly introduced binary decision variable s_{fj} is connected with an objective value q_{fj} that refers to the costs for the delay propagation risk $\delta^s_{f\omega}$ of the follow-on $()p'_j(f); f)$:

$$q_{fj} = c_Q \cdot Q^P(f, j', \Omega). \tag{11.17}$$

Note that $Q^P(f, j', \Omega)$ can be picked from the propagation tree built in pricing. Thus, its determination does not necessitate computational effort in the Downstream Recourse. To connect variable s_{fj} with pairings j and K_j, the following constraint is inserted:

$$x_j - \sum_{k \in K_j} x_k - s_{fj} \leq 0. \tag{11.18}$$

If pairing j is selected as a part of the final IP solution, at least one pairing $k \in K$ must also be part of the solution. Otherwise penalty costs p_{fj} apply since $s_{fj} = 1$ is forced. Note that for proper functioning, the variable-constraint-pair must be inserted for all pairings $j \in J$ where $K_j \neq \varnothing$. For certain elements $j \in J$, there may be no swap opportunity, reflecting in $K_j = \varnothing$. The objective value of these pairings is increased by its corresponding rate q_{fj}, otherwise the follow-on may unintentionally be part of the final solution.

A formal description of the approach is provided in Algorithm 1. The integration into a branch-and-price framework is straightforward as every promising node must be processsed by the the Downstream Recourse. Eventually, integer solutions are obtained in the *IP Phase* by using standard IP solvers.

Algorithm 1: Downstream Recourse

input: Model M
Propagation trees Z, $z \in Z$ containing flight f, subroute j', pairing set J, recourse value $q = Q^P(f, j', \Omega)$

```
L ← Z as list, sorted by q_z in descending order
i ← 1
while no termination criteria met do
    z_act ← L[i]
    f ← f_{z_act}
    foreach j ∈ J_{z_act} do
        determine swap set K_j
        if K_j = ∅ then
            increase objective of j by q_{z_act}
        else
            insert new variable s_fj with objective q_{z_act} into model M
            insert new constraint x_j − Σ_{k∈K_j} x_k − s_fj ≤ 0 into model M
        end
    end
    i ← i + 1
    check termination criteria
end
```

output: Model M

11.3 Computational Study

In this section, we present a computational study on the impact of improved crew schedule flexibility on the nominal costs and robustness. When improving the flexibility, the straightforward assumption is that the number of scheduled swap opportunities is increased. However, in contrast to Shebalov and Klabjan (2006) and Ageeva (2000), the stochastic evaluation of swaps in the Downstream Recourse weights the benefit of each swap regarding the possibly prevented delay propagation. Thus, the absolute number of swapping opportunities may not allow straightforward conclusions to be drawn about the resulting schedule robustness.

As a consequence, we examine the benefit of incorporating swap opportunities with a supposed high risk of delay propagation in three steps. Firstly, we evaluate the secondary-to-primary delay ratio (stp) of schedules generated by the presented approach. In addition, possible trends in the robustness-to-cost ratio (rtc) are discussed. The stp is known from Section 9.1 as the main indicator for schedule robustness. The rtc, describing the relation between the robustness benefit and nominal cost increase has been defined in Section 9.1. Afterwards, the number of allocation conflicts arising from delayed flights is examined for schedules with an increase of flexibility. Eventually, technical details of the approach are provided concerning the success ratio of scheduled swaps at varying degrees of flexibility.

All experiments are carried out on a desktop PC with an Intel Core i5-2500k processor with 8GB RAM. In the column generation framework, we use CPLEX 12.2 for the restricted master problem and for the IP phase. As test instances, we use the set of flight schedules known from previous analyses in Sections 9 and 10. Since cost-neutral swaps are a basic assumption in our definition of flexibility, artificial swap costs $c_{s_{f\omega}}$ are set to zero. For the delay tolerance threshold it also holds $t_\delta = 0$, i.e. swaps are potentially incorporated if at least one minute of propagated delay can be prevented. Although such low-impact swaps do not appear to be beneficial, we leave the decision up to the optimization process and avoid unnecessary interventions.

Schedules with different degrees of flexibility are constructed by varying the penalty factor $c^Q \in \{0, 10, 50, 100, 200, 250, 500, 750, 1000, 2000, 2500, 3000, 4000\}$. Values $c^Q = 0$ indicate cost-efficient solutions without any explicit consideration of robustness. Similar to the analysis of primary delay prediction influences, every setting of c^Q is solved with four different seed values for the randomizer of the primary delay generator. With 10 test instances[3], this results in 520 parameter settings and unique scheduling solutions. For the primary delay generator, we use the same settings of the rule-based recovery approach evaluation in Chapter 10, namely an OC.100.100 prediction model configuration with parameters mean = 10.66817, sd = 14.34176, ratio = 0.5, relative correction = 0.2[4].

Computational effort is quite small and computation times do not appear to be a bottleneck. Theoretical solution times of the Downstream Recourse are $\mathcal{O}(n^2)$ with n being the number of considered pairings. However, the computational effort can be significantly reduced by straightforward concepts for the delay propagation evaluation and search for candidate swap pairings. Firstly, only flight connections must be considered that are likely to propagate delay concerning the underlying primary delay scenarios. Due to the usage of the propagation tree built in the dynamic programming approach during pricing, no computational effort is necessary. On average, 4,067.2 different follow-ons with distinct subroutes exist per instance, 1,977.08 of these follow-ons are assumed to propagate delay of at least one minute. Note that the ratio of risky connections heavily depends on primary delay generation. Secondly, pairings are only worth considering them as swap candidates when they contain a flight at nearly the same time at the same airport as the risky follow-on. For the used test set with up to 427 daily flights, solution times of 6:25 minutes on average with a standard deviation of 2:02 minutes seem negligible. We assume that for multi-day instances the complexity significantly increases. This is because many more identical subroutes and duties can be combined for the construction of pairings. However, when overnight swaps are not taken into account, the Downstream Recourse can be solved on a daily basis, assuring tractable computational effort.

For cost figures, we refer to nominal costs of generated schedules. As the main robustness measure, we adhere to the secondary-to-primary delay ratio (stp)[5]. Schedule robustness is evaluated by using the rule-based recovery approach presented in Section 10 with settings dm-15 and dm-0. In setting *dm-15*, crews and aircraft are swapped if more than 15 minutes of delay propagation can be prevented. Analogously, swaps are applied if possible for all delays of at least one minute in setting

[3]For details on the test instances, see Table 9.2.
[4]We explicitly do not use OC.1.1 configurations in order to assess the general behavior of the scheduling approach.
[5]For definition and computation details, see Section 9.1.

dm-0, corresponding to $t_\delta = 0$ in scheduling. Both approaches and their effects on robustness and cost measures have been discussed in Section 10.3. Finally, in the *prop* setting schedules are evaluated by propagating all delays while not intervening in any case. In the scope of this study, results are used as reference values. The performance of schedules with an increased degree of flexibility is evaluated in more detail in the subsequent analyses in Chapter 12. Besides the stp, we discuss the number of emerging and resolved resource allocation conflicts that come along with varying flexibility degrees of the generated schedules. All simulation results are obtained by averaging the outcome of 100 primary delay scenarios that are individually simulated.

In the following Section 11.3.1, we examine the trade-off between nominal costs and robustness. Section 11.3.2 deals with the impact on the number of emerging resource allocation conflicts. The success rate of scheduled swaps is addressed in 11.3.3.

11.3.1 Examining the Trade-off between Nominal Costs and Robustness

Starting with the trade-off between nominal costs and robustness, Figure 11.3 provides initial results concerning nominal cost changes and stp values changes from both dm-0 (left panel) and dm-15 (right panel) evaluation. Coordinates (0,0) depict the cost-efficient solution which is the reference for relative changes in nominal costs and stp. The x-axis depicts the average change of nominal costs referred to the cost-efficient solution. Since nominal costs are not affected by alternating evaluation settings, these values are identical per schedule in both panels. Analogously, the stp change is displayed on the y-axis. Note that dm-15 and dm-0 have a different impact on the stp[6]. Thus, in contrast to nominal costs the axis intercepts in both panels do not suggest similar absolute stp values.

Each light gray point depicts one particular run, i.e. one specific instance, seed and penalty factor c^Q. Similar to previous results in Section 9.3.3, a few solutions provide lower nominal costs as the actual cost-efficient solution. It mostly occurs at low values of c^Q. We assume that in these rare cases, slight perturbations in the

[6]See Figure 10.17 and the related discussion in Section 10.3.

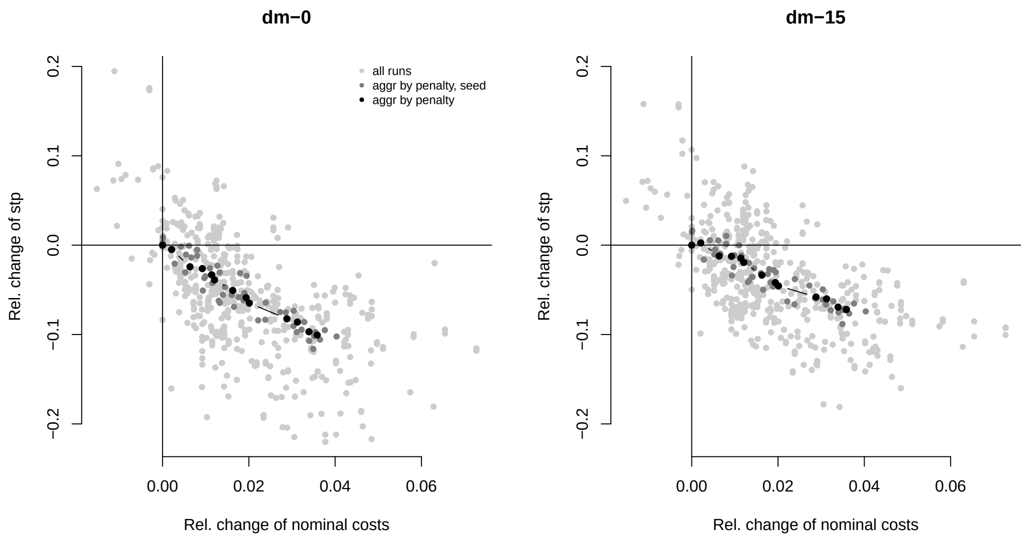

Figure 11.3: Relative change of nominal costs and stp for dm-0 (left panel) and dm-15 evaluation (right panel)

column generation process may allow the exploitation of previously unconsidered areas of the solution space.

Medium gray points indicate a first aggregation level. Each point depicts a different seed and penalty factor combination, averaging the results for 10 instances each. Thus, remaining variations are a result of randomization (seed) and systematical effects (penalty factor). Finally, the black points indicate aggregations by penalty factor for 40 instances each. Connecting lines indicate increasing values of c^Q. A systematical trend is obvious and even more considerable for dm-0. The straightforward explanation is that $t_\delta = 0$ complies with the dm-0 evaluation while in the dm-15 setting scheduled swaps cannot be used for small delays up to 15 minutes .

For a closer look at the trade-off between nominal costs and stp, we provide the robustness-to-cost ratio (rtc) per penalty value c^Q for dm-0 and dm-15 in Figure 11.4. During the evaluation of primary delay prediction, a steady decrease of the rtc has been observed for stability, see Figure 9.8. In contrast, the rtc ratio does not show a significant trend but a quite similar flat course ending up around -3 (dm-0) and -2 (dm-15). However, sharper fluctuations at lower flexibility degrees appear in contradicting direction. Since in dm-15 swaps are performed only for larger delays,

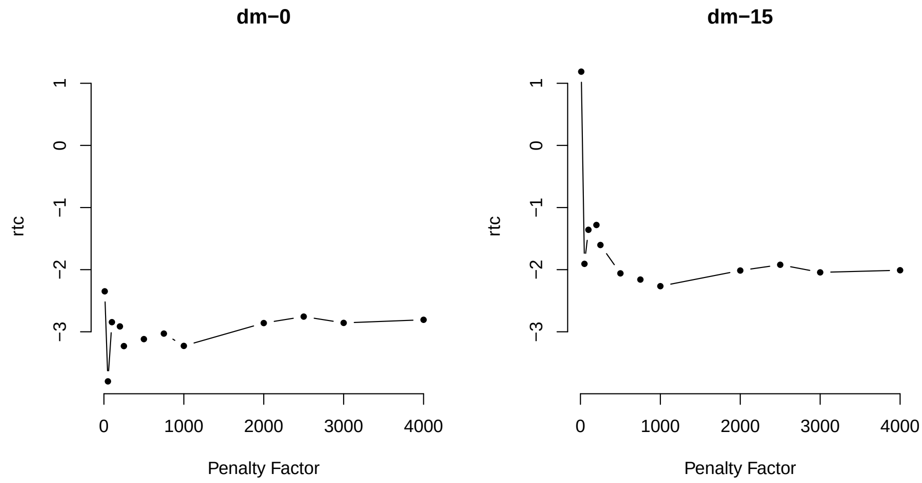

Figure 11.4: Robustness-to-cost ratio for dm-0 (left panel) and dm-15 (right panel)

we must assume that slight improvements of swapping capability are not as beneficial as for smaller delays. This observation aligns with the concave course of aggregated results in Figure 11.3 (right panel).

11.3.2 The Impact on Emerging Resource Allocation Conflicts

Going into detail, we examine the number of resource allocation conflicts and swaps that are performed to resolve them. In our definition, a resource allocation conflict results from a delayed flight arrival so that there is insufficient time for the following aircraft turnaround and/or crew connection. The resulting schedule infeasibility must be resolved by the disruption management, either by postponing the subsequent flight's departure – referred to as delay propagation – or by adapting the schedule – referred to as swapping. Note that the analysis of allocation conflicts does not consider the length of potentially propagated delay per conflict.

In the following, we examine the number of emerging conflicts in the evaluation settings prop, dm-0 and dm-15. In Figure 11.5, the relative change of conflicts per setting is illustrated. The x-axis depicts the nominal cost change, the y-axis depicts the relative change in the number of allocation conflicts. The starting point (0,0) of each curve relates to average nominal costs and the average number of conflicts

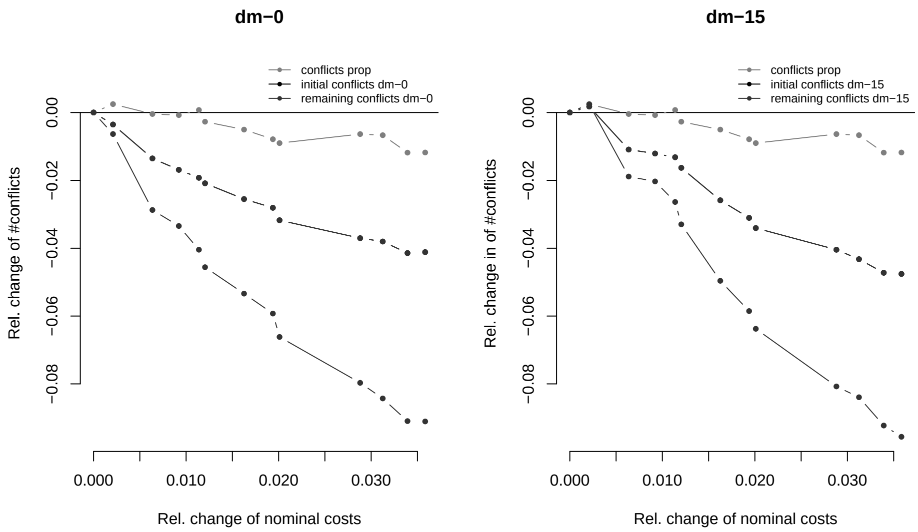

Figure 11.5: Relative change of emerging resource allocation conflicts in different evaluation settings

in the respective setting prop, dm-0 and dm-15 for cost-efficient schedules ($c^Q = 0$). Absolute values that correspond to $y = 0$ are prop: 184.99; dm-0: 120.97; dm-15: 40.07. Note that in the dm-15 setting, a resource allocation conflict denotes a potential delay propagation of more than 15 minutes. Smaller delays are propagated without a swapping attempt. The y-axis intercept at 0 serves as an auxiliary line for better identification of trends, especially for the prop setting. Each point reflects the average change of the number of conflicts per penalty factor c^Q, connecting lines indicate increasing c^Q values.

In both panels, the light gray connected points show the relative change in the number of allocation conflicts in the prop setting. Because no recovery intervention is applied, every allocation conflict inevitably leads to a delay propagation. Thus, one primary delay may cause several conflicts in case of multiple or even cascading propagation over crew pairings and aircraft rotations.

The general trend of the curve shows a slight, non-monotonous decrease. Since primary delay scenarios are identical for all experiments, a reduction of conflicts induced by secondary delays must be responsible for this trend. On average, up to 1.18% of conflicts can be prevented compared to cost-efficient solutions at the highest degree of flexibility, corresponding to the rightmost point of the light gray curve. This is a first hint on flexibility positively affecting stability. Even if no swaps are performed, delay propagation is slightly decreased with an increasing degree of flexibility[7].

Analogously, the black line indicates the relative change in the number of initial allocation conflicts in the dm-0 setting (left panel). By naming them *initial*, we refer to the fact that in contrast to prop, a conflict may be resolved by a potential swap in dm-0 and dm-15. There is a significantly larger decrease of allocation conflicts in dm-0 compared to prop – for the highest degree of flexibility, initial conflicts are reduced by 4.12%. The dm-15 curve (right panel) shows a similar behavior with a more concave progression for low c^Q values. Due to the usage of identical primary delay scenarios, the number of conflicts emerging from primary delays is identical in prop, dm-0 and dm-15 settings. In consequence, gaps between the prop curve on the one side and dm-0/dm-15 curves on the other side are a direct consequence of a decrease in conflicts emerging from secondary delays. This effect is a consequence of improved swapping capabilities, preventing further delay propagation.

In this regard, the lower dark gray line indicates the number of remaining conflicts after swapping, i.e. conflicts that are not or cannot be resolved. In these cases, delay propagation is inevitable.

In this respect it corresponds to the evaluation of prop (light gray line) where no swaps are applied. For most flexible solutions, the number of total allocation conflicts can be reduced by about 9.11% at a nominal cost increase of 3,58% compared to the average for cost-efficient solutions. For dm-15, the respective value is a comparable order (9.56%). The widening gap between initial and remaining conflicts indicates that with increasing c^Q values, the swapping capability improves and more conflicts can be resolved. For the largest c^Q values, the course of the curves is even contrary: while the number of initial conflicts slightly increases, more of these conflicts can be resolved.

For a closer look, the ratio of these resolved conflicts is depicted Table 11.1. c^Q shows the penalty factor value, Δc_n the average nominal cost change compared to cost-efficient schedules. Columns 3 and 4 show the relative change in resolved conflicts for dm-0 ($res_{dm\text{-}0}$) and dm-15 ($res_{dm\text{-}0}$). A substantial increase is obvious

[7]Further indications on the mutual impact of swapping capabilities and delay absorption capacities is analyzed in the subsequent Chapter 12.

c^Q	Δc_n	$\Delta res_{dm\text{-}0}$	$\Delta res_{dm\text{-}15}$
0	0.0000	0.0000	0.0000
10	0.0021	0.0028	-0.0007
50	0.0064	0.0152	0.0080
100	0.0092	0.0166	0.0082
200	0.0114	0.0212	0.0132
250	0.0120	0.0247	0.0167
500	0.0163	0.0279	0.0237
750	0.0194	0.0312	0.0275
1000	0.0201	0.0344	0.0297
2000	0.0288	0.0426	0.0403
2500	0.0313	0.0463	0.0407
3000	0.0339	0.0495	0.0450
4000	0.0358	0.0499	0.0480

Table 11.1: Relative change of resolved conflicts for dm-0 and dm-15, averaged per penalty factor c^Q

for both dm-0 and dm-15 with a naturally higher and more monotonous improvement for the former. This is because it is more closely adapted to the setting $t_\delta = 0$ in scheduling that allows the incorporation of swaps for delays of at least one minute.

In conclusion, the results show that the number of emerging resource allocation conflicts can be substantially reduced by means of two aspects. On the one hand, the incorporation of swapping opportunities leads to a substantial reduction of delay propagation inducing conflicts. As a beneficial side effect, the number of initial conflicts decreases, too. This can be clearly attributed to the fact that swapping allows the prevention of consecutive or even cascading propagation of one delay. On the other hand, improved flexibility goes along with a slight increase of the stability in terms of delay absorption capacity. Even if no swaps are applied, the number of resource allocation conflicts decreases slightly. The question of mutual effects between stability- and flexibility-related mechanisms forms the basis of the analysis in Chapter 12.

11.3.3 Scheduled Swaps and their Success Rate

While the previous examination has proved that the presented approach offers a controlled improvement of the schedule robustness, we now investigate the technical

c^Q	$\varnothing swaps_s$	$succ_{dm\text{-}0}$	$succ_{dm\text{-}15}$
0	–	–	–
10	52.60	0.7183	0.5620
50	57.18	0.7178	0.5640
100	57.85	0.7088	0.5478
200	58.28	0.7185	0.5565
250	58.48	0.7143	0.5578
500	60.13	0.7093	0.5638
750	60.63	0.7073	0.5660
1000	60.83	0.7155	0.5718
2000	61.78	0.7205	0.5865
2500	62.65	0.7265	0.5913
3000	62.20	0.7368	0.6045
4000	61.48	0.7395	0.6118

Table 11.2: Relative change of resolved conflicts and success rate of scheduled swaps for dm-0 and dm-15, averaged per penalty factor c^Q

efficiency of the scheduling approach. The concentration is on the key instrument for flexibility, namely the crew swaps that are explicitly incorporated into the schedules for follow-ons with assumingly high propagation risks. We consider the success rate of scheduled swaps during simulation. The success rate is defined as the share of simulation runs in which a scheduled swap can actually be performed. Related values are obtained in Table 11.2. In column $swap_s$, the average absolute number of scheduled swaps per penalty factor c^Q is provided. Interestingly, for supposedly highest degrees of flexibility the number of scheduled swaps slightly decreases, although the previously examined benefit concerning stp and the amount of resource allocation conflicts is evident. Furthermore, columns $succ_{dm\text{-}0}$ and $succ_{dm\text{-}15}$ show the success rate of scheduled swaps in the respective simulation setting. In our context, the application of scheduled swaps can be impeded by several reasons. Firstly, previously performed swaps may alter pairing and rotations in a way that subsequent swaps become infeasible. Secondly, the proposed decomposition strategy does not provide definite prediction of swap feasibility since pairing interactions cannot be entirely taken into account. Thirdly, schedules are built on the basis of 100 primary delay scenarios and the optimization approach has selected a set of swap opportunities that promises best accumulated results. Consequently, not every scheduled swap is applicable for every delay scenario.

The average success rate for scheduled swaps is 71.94% with a standard deviation of 7.44% for dm-0. With regard to the mentioned uncertainty aspects for scheduled swaps, we consider these values as appropriate. Values for dm-15 are significantly smaller. On average, only about 57.36% of all scheduled swaps are successful with a standard deviation of 10.81%. These distinct differences between dm-15 and dm-0 may support the assumption that scheduled swaps for longer delays may be less predictable and are more often exposed to the risk of becoming infeasible in simulation.

At very low c^Q values, no substantial trend can be observed, the success ratio fluctuates around 71% and 55.5%, respectively. In contrast, the success ratio constantly increases at higher c^Q levels and shows an opposed development to the number of scheduled swaps. This behavior corresponds to the propagation-risk-based assessment of swaps in the scheduling approach – instead of unconditionally increasing the total number of swapping opportunities, they are selected based on their impact concerning delay propagation prevention. Swaps that are assumed to be used in many delay scenarios are preferred, positively affecting the swap success rates with increasing flexibility degrees.

11.4 Summary and Implications

In this chapter, we have provided a stochastic scheduling approach based on Dück et al. (2012) and Ionescu and Kliewer (2011) for increasing the flexibility of crew and aircraft schedules for regular daily operations. Based on the mathematical model of the original integrated problem, a decomposition strategy has been presented. The resulting problem formulation can be solved by a column generation based approach with a Downstream Recourse that encapsulates flexibility-related issues. In contrast to Shebalov and Klabjan (2006), Burke et al. (2010) and Ageeva (2000), delay propagation risks are taken into account for evaluating the benefit of potential swaps. The solution times for the Downstream Recourse are $\mathcal{O}(n^2)$ when n is the number of generated pairings during column generation. However, they can be significantly decreased by using certain straightforward concepts for the evaluation and search of candidate swap pairings. In particular, the usage of propagation trees built in the dynamic programming approach during pricing has a substantial positive impact.

For the used test set with up to 427 daily flights, solution times of 6:25 minutes on average seem negligible.

For the assessment of the approach, Pareto-optimal solutions have been generated, describing the trade-off between nominal costs and robustness. The secondary-to-primary delay ratio can be increased by 10.06% at a nominal cost increase of 3.58% at the highest degree of flexibility in the dm-0 setting. Furthermore, the amount of emerging resource allocation conflicts can be significantly reduced by up to 9.11%. The number of swap opportunities for connections that are likely to propagate delay can be increased. The success ratio for these scheduled swaps remains quite constant but rises at high flexibility levels. Additionally, hints on a potential mutual impact of schedule flexibility on stability issues have become apparent already. For an in-depth analysis of this topic, we refer to the subsequent chapter 12.

On the technical side, future improvements of the approach are two-fold. Firstly, it has to be evaluated if the presented Downstream Recourse can be enhanced by a consideration of swapping opportunities that is directly connected with the pricing step. In this way, pairing sets can be generated that explicitly provide swap opportunities among each other. The question is, if the prediction accuracy of for delays is high enough to be worth the added amount of complexity. On the other side, stability may be more advantageous if the delay prediction accuracy is high. Secondly, the potential of the approach for simultaneous crew and aircraft scheduling has to be addressed in future work. It appears promising since crew and aircraft swaps can be better coordinated. Based on the presented integrated model of Section 11.1, sophisticated decomposition and interaction strategies must be developed to synchronize the incorporation of compatible swap opportunities.

Besides the improvement of the scheduling technique, future research has to deal with the evaluation of the approach in less predictive environments. In these cases, generalized approaches such as Shebalov and Klabjan (2006) or deterministic rule-based scheduling strategies may be competitive while offering reduced complexity in delay propagation evaluation.

Chapter 12

The Impact of Stability and Flexibility on the Robust Efficiency of Schedules

After examining the flexibility of airline crew schedules in Chapter 11, we address Research Objective R4 in the following analysis. The question is in how far flexibility and stability affect each other with regard to changes in nominal costs and robustness. Recalling the initial definitions, stability on the one hand describes a schedules' capacity to absorb delays, mostly by matching buffer time placements to delay propagation risks. Flexibility on the other hand describes the capability of swapping resources in order to avoid scheduled connections between flights that lead to delay propagation in operations. The increase of either stability or flexibility has positive impact on the schedule robustness. However, potential mutual impacts between stability and flexibility are not investigated yet.

Scientific studies in this context mostly consider these aspects separately. As one of the few examples for a simultaneous consideration, Burke et al. (2010) state that the influence of stability[1] is clearly dominant to flexibility in the multi-objective approach. However, findings do not allow to draw conclusions on a potential trade-off and its quantification.

This issue is addressed in the following study, forming important preliminary work for elaborating an integrated and simultaneous consideration of stability and flexibility in future research. The underlying four main questions for the examination are stated in Figure 12.1. The *scheduling policy* determines which strategy is used during scheduling. Schedules can be generated with an increased degree of either stability or flexibility. We use the approach of Dück et al. (2012) with a minute-based evaluation

[1]Note that Burke et al. (2010) refer to our concept of stability as *schedule reliability*.

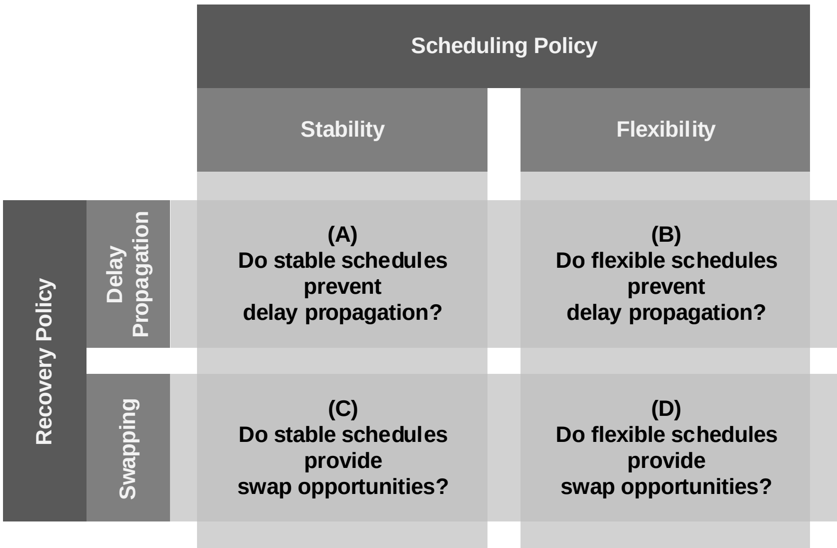

Figure 12.1: Research questions concerning mutual impacts of stability- and flexibility-related aspects

of delay propagation risks (see Section 9.1) for increasing stability of crew schedules. The approach presented in Chapter 11 is used for increasing flexibility. The *recovery policy* determines if delays are just propagated or if crew and aircraft swaps can be applied. While the former directly targets at the evaluation of stability, the latter is essential for the evaluation of flexibility.

Consequently, questions (A) and (D) address self-evident aspects of stability and flexibility. The core of the current examination is therefore formed by questions (B) and (C) which address cross-relations between stability and flexibility. Question (D) has been examined in the previous chapter 11. For answering question (B), schedules with an increased degree of flexibility are simulated while not allowing any recovery interventions. Basic findings for (B) have already been part of the discussion on the number of conflicts in schedules with increased degrees of flexibility related to Figure 11.5 and are now elaborated further. Analogously, schedules that are initially generated with an increased degree of stability are simulated by using delay propagation only (A) as well as enabling swapping (C).

The remainder of this chapter is organized as follows: In the beginning of Section 12.1, we provide details on settings and instances for the computational study. Primary concern is to provide comparability between results by the stability- and flexibility-related scheduling approaches. Afterwards, research questions (A)-(D) are tackled with regard to nominal costs and the secondary-to-primary delay ratio (stp). The impact on emerging resource allocation conflicts is addressed in Section 12.2. Finally, a summary of findings and implications for future research is provided in Section 12.3.

12.1 The Impact of Scheduling Strategies on Nominal Costs and the Secondary-to-Primary Delay Ratio

The following examination seamlessly connects to the assessment of the flx scheduling approach in Section 11.3. For carving out mutual impacts of stability and flexibility on delay absorption capacities and swapping capabilities, resulting schedules are now compared to additionally generated schedules with an increasing degree of stability. The ladder are generated by the approach of Dück et al. (2012) with a minute-based recourse model for emerging propagated delay. In the following, this approach is denoted as *stb*. In order to ensure comparability between flx and stb outcomes, computational results are generated with main settings identical to Section 11.3. Solely, selected penalty factor values c^Q differ between flx and stb in order to approximately match the range of the nominal cost outcome of schedules resulting from the flexibility approach to ensure a high degree of comparability[2]. As a result, nominal costs of schedules generated by stb increase by 3.75% at the highest level of stability, the respective value for schedules generated by flx is 3.58%.

Concerning the assessment of robustness by simulation, prop and dm-0 correspond to the known evaluation types where delays are either entirely propagated or swaps can be performed in order to prevent propagation. In order to fully exploit potential benefits of swapping capabilities, we do not use a trigger threshold for swaps such as used in dm-15.

[2]Penalty factors used in the stb approach are: 0 (cost-efficient), 10, 25, 50, 100, 200, 300, 400, 500, 600, 750, 850, 1000

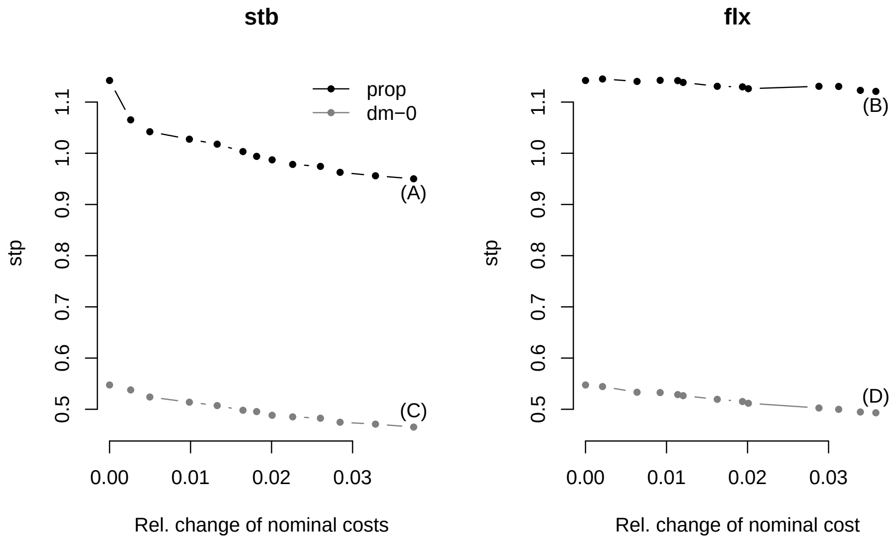

Figure 12.2: Absolute stp values for schedules generated with increasing degrees of stability (stb) and flexibility (flx)

The secondary-to-primary delay ratio (stp) is the most relevant figure for assessing the robustness of schedules in our context. It needs to be borne in mind that, besides the selected scheduling strategy, it is also be influenced by the evaluation method in simulation. In this regard, computational results in Section 10.3 have proved that about half of all propagated delays can be prevented by swaps, leading to substantial improvements of absolute stp values. Regarding the current experiments, the average stp for cost-efficient schedules ($C^Q = 0$) is 1.142 when using the prop evaluation. Applying the dm-0 evaluation leads to an average stp of 0.547, which is a reduction of about 52.1%.

Referring to the considered scheduling strategies, Figure 12.2 provides absolute stp values aggregated by penalty factors c^Q for schedules generated with stb (left panel) and flx (right panel). The x-axis depicts the relative change in nominal costs while the y-axis shows the related average stp value for all test runs with identical c^Q. The black line indicates stp values obtained by the prop evaluation, i.e. all delays are propagated. Letters (A)-(D) refer to the respective question of Figure 12.1. For schedules generated by the stb approach (A), there is an initial steep descent that

Δc_{stb}	Δstp_{stb}	Δc_{flx}	Δstp_{flx}
0.00	0.5947	0.00	0.5947
0.26	0.5275	0.21	0.6008
0.50	0.5182	0.64	0.6071
0.99	0.5137	0.92	0.6098
1.33	0.5106	1.14	0.6132
1.65	0.5053	1.20	0.6117
1.82	0.4985	1.63	0.6113
2.01	0.4989	1.94	0.6146
2.26	0.4931	2.01	0.6144
2.60	0.4919	2.88	0.6283
2.85	0.4881	3.13	0.6307
3.28	0.4850	3.39	0.6282
3.75	0.4849	3.58	0.6277

Table 12.1: Absolute differences of stp values between prop and dm-0 with related nominal cost values

flattens out in the further course. Schedules generated with the flx approach (B) do not show a similarly obvious trend at first sight. Results obtained by the dm-0 setting are mutually comparable with slightly better average stp values for schedules generated by stb (C) compared to flx (D). The initial steep descent in (A) is not apparent in (C). We assume that the lack of robustness in cost-efficient schedules can be compensated by swapping recovery actions in simulation.

Table 12.1 provides detailed figures concerning differences between prop and dm-0. The first and third row show the nominal cost increase for scheduling strategies $s \in \{stb, flx\}$. Rows two and four depict the respective absolute difference of stp values between prop and dm-0 denoted as Δstp_s. These values correspond to the gap between lines (A) and (C) for stb and (B) and (D) for flx, respectively.

For cost-efficient schedules, about 0.5947 minutes of secondary delays per one minute of primary delay can be prevented by swapping. For schedules with an increasing degree of stability, this value reduces constantly until 0.4849, i.e. the beneficial influence of swaps decreases. In the following, we have to ascertain if this effect is the expected result of improved stability or also goes along with a concurrent impact on the schedules' flexibility. In contrast, schedules with an increasing degree of flexibility are supposed to offer better swapping capabilities which reflects in a widening gap between stp values from prop and dm-0.

For a closer insight on the actual impact of scheduling strategies on the robustness, Figure 12.3 addresses the relative change of stp values for schedules generated by stb or flx, compared to respective cost-efficient solutions. The focus of the following examination is on trends in relative changes of the stp per scheduling strategy. Panels (A)-(D) correspond to the research question matrix in Figure 12.1. Accordingly, Panels (C.2) and (D.2) are intended to provide further elaboration of swapping capabilities with regard to questions (C) and (D). The y-axis intercepts at 0 represent different absolute stp values of cost-efficient schedules for different evaluation methods as reference values. Note that therefore conclusions on absolute stp values between vertical panels cannot be drawn.

Each light gray point stands for a single test run for a specific flight schedule and penalty value c^Q. Medium gray points stand for the average per penalty value c^Q and seed, i.e. one point shows the average results of 10 instances each. Finally, each black point depicts the average per penalty value, making trends apparent that base on an increase of stability or flexibility. Related underlying numerical values can be found in the Appendix C.6.

12.1.1 Stability

Panel (A) shows effects of increased stability. Delay propagation can be reduced by up to 16.78% at a maximum cost increase of 3.75% on average compared to cost-efficient schedules. Starting with an exceptional stp decrease of 6.69% for the lowest penalty factor setting, marginal costs for robustness increase steadily, affecting the robustness-to-cost ratio (rtc) in the same way (see Appendix C.6).

We proceed to results obtained by the dm-0 evaluation method that allows swapping during simulation. Results for schedules with an increasing degree of stability are presented in Panel (C). Note that no distinction can be made whether changes in stp values are attributable to stability or flexibility effects. In fact, Panel (C) illustrates changes in stp and nominal costs for the stb scheduling approach as a whole.

The relative improvement of the stp is more equally distributed compared to Panel (A) while the overall trend is comparable. Nevertheless, the initial large stp improve-

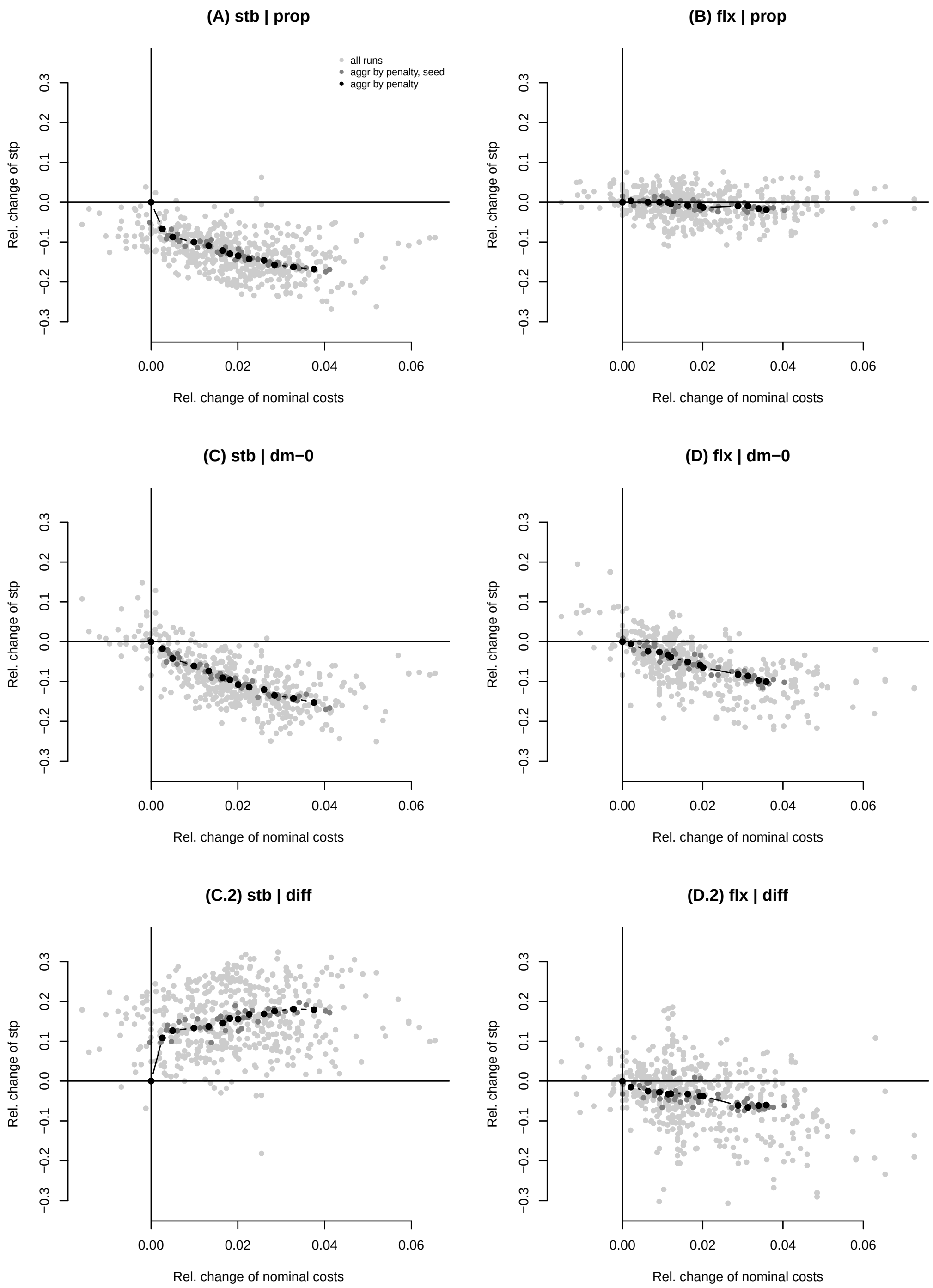

Figure 12.3: Relative change of nominal costs and stp, split up by prop (upper two panels), dm-0 evaluation (middle panels) and carved out swapping influence (lower panels) for stability (left) and flexibility (right) consideration in scheduling

ment of 6.69% in Panel (A) cannot be observed, the corresponding value in Panel (C) is 1.75%. In addition, the maximum average stp improvement of 15.28% in dm-0 for schedules with the highest degree of stability is lower than the maximum improvement of 16.78% in the prop setting (Panel (A)). The rtc ratio decreases steadily, however at a lower level compared to the prop-based evaluation (Panel (A)).

Two aspects can be presumed at this point. Firstly, swaps are a valuable instrument to compensate the lack of stability. Thus, compared to a pure prop evaluation, relative differences in stp values between cost-efficient and stable schedules diminish. Secondly, increased stability impedes the necessity and the beneficial influence of swaps on the stp decreases. Improved delay absorption capacity does not reflect in dm-0 to the same extent as in prop. In this regard, a trade-off is apparent concerning the question whether delay propagation risks can be either prevented by stability or resolved by swapping.

For a more distinctive view on the stp benefit due to swapping, Panel (C.2) illustrates the relative change of the difference between absolute stp values obtained by prop and dm-0, introduced as *diff* in Table 12.1. The stp benefit due to swapping decreases by 10.86% compared to cost-efficient schedules for the lowest c^Q value 10. Further increase of stability leads to reductions of up to 18.11%. Setting this into relation with results from Panel (A), the increased delay absorption capacity clearly is at the expense of the influence of swapping. Assumptions drawn from the discussion of Panel (C) can be confirmed.

12.1.2 Flexibility

We move on to the evaluation of schedules with an increased degree of flexibility. Analogously to Panel (A), the improvement of delay absorption for schedules generated by the flx approach is illustrated in Panel (B). A slight positive trend in terms of increased delay absorption capacity becomes apparent, however, for low c^Q values the average stp change is worse than for cost-efficient solutions. Nevertheless, 1.86% of propagated delays can be prevented only by delay absorption in the most flexible schedules (compared to 16.78% for most stable schedules, see Panel (A)). The rtc ratio fluctuates between 0 and -1 without a particular general tendency.

Panel (D) shows the overall benefit of the flx scheduling approach in the dm-0 evaluation setting. Naturally, the improved swapping capability comes in and provides a decent stp improvement compared to Panel (B). The course of the curve is nearly linear with a the maximum stp improvement is 10.06%. Results from Panels (C) and (D) allow a direct comparison of the stb and flx approach as a whole, notwithstanding whether a stp improvement results from well placed buffer times or swapping. On the basis of the same cost-efficient schedules determining reference values for relative changes that reflect in coordinates (0,0), the flx approach provides inferior solutions in both stp and rtc compared to the stb approach (Panel (C)). As indicated by the curve position and the less convex curve shape, the necessary nominal costs increase for a stp improvement is higher for flx than stb. In this regard, there is also a perceptible rise in the nominal cost change from 2.01% to 2.88% between c^Q values of 1,000 and 2,000 since on average 0.95 additional crews must be deployed per schedule.

According to the definitions of (Dück, 2010, p. 89), both stb and flx approaches are *predictable* in the dm-0 setting since propagation penalty factors correlate with the actual robustness measure outcome. Moreover, the stability-related approach is more *efficient* in both prop and dm-0 settings as it offers better solutions concerning the robustness-to-cost outcome.

Although the majority of the stp improvement in Panel (D) can be clearly attributed to improved swapping capabilities, it is also exposed to an, albeit low, influence of indirect stability deviations. In a final step, we therefore evaluate the unbiased progression of swapping capabilities analogously to the examination related to Panel (C.2). For this purpose, Panel (D.2) illustrates the relative difference of diff values. In general, the swapping impact illustrated in Panels (C.2) and (D.2) generally fluctuates to a considerably greater extent compared to the stability-related scatter plots in Panels (A) and (B). The aggregation by penalty values c^Q shows an initially rising impact on the stp. It tails off in the further course and even slightly decreases for the last two aggregation levels of $c^Q \in \{3,000, 4,000\}$. Therefore, the ongoing overall stp improvement (Panel (D)), is presumably caused by the simultaneous stability improvement that is apparent in Panel (B). Concerning this effect it can be suggested that additional swapping capabilities can hardly be incorporated. Due to the nature of the flx approach, further robustness can only be achieved by preventing connections between flights with high propagation risk – which is tantamount to an increase of stability.

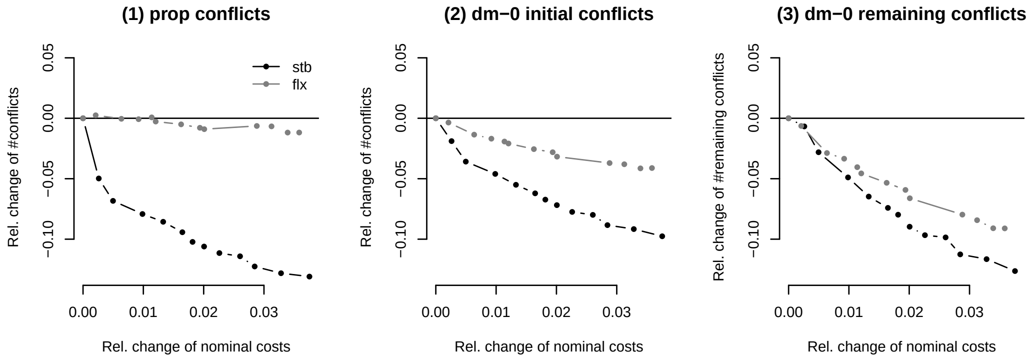

Figure 12.4: Relative change of emerging resource allocation conflicts for schedules generated with stb and flx

12.2 Examining Changes in Resource Allocation Conflict Occurrences

We compare the development of the number of emerging and resolved resource allocation conflicts for schedules with an increased degree of either stability or flexibility. On the one hand, we consider the number of conflicts in the prop setting in order to evaluate the influence of the scheduling strategies stb and flx on the delay absorption capacity. The examination is associated with research questions (A) and (B). Due to the absence of swapping, the only influential factor is the changing delay absorption capacity which simplifies the examination. On the other hand, the swapping capability is analyzed for stb and flx, reflecting research questions (C) and (D).

Figure 12.4 illustrates relevant figures concerning the relative change of the number of emerging conflicts compared to cost-efficient schedules for stb (black lines) and flx (gray lines). The x-axis shows the relative change in nominal costs while the y-axis shows the relative change in the number of conflicts, both averaged per penalty factor c^Q over all corresponding experimental runs[3].

Regarding research questions (A) and (B), Panel (1) shows the results obtained by prop setting in which per definition every conflict leads to delay propagation. The number of emerging conflicts is naturally decreasing to a large extent when

[3]To facilitate better classification of the results, the average absolute numbers of conflicts emerging in cost-efficient schedules are: prop: 184.99; dm-0 (initial): 120.97; dm-0 (remaining): 74.46.

the stability of schedules is improved (stb curve). With an initial steep decrease that flattens out for high levels of stability, the development is similar to the stp improvement shown in Panel (A) of Figure 12.3. However, since the scheduling approach directly targets at the stp rather than the number of conflicts, the curve is more angular and not always convex. The number of conflicts can be reduced by 13.1%, leading to a stp improvement of 16.78% (see Panel (A) of Figure 12.3). As a side effect of schedules with increased flexibility, the number of conflicts shows a slight but unstable decrease (flx). At the highest level of flexibility, 1.18% less conflicts occur and lead to an stp improvement of 1.86% (Panel (B) of Figure 12.3). On the whole, the observable developments are congruent with the ones obtained by the preceding stp-based analysis.

The conflict-based examination of research questions (C) and (D) is carried out for the dm-0 evaluation setting. Panel (2) illustrates the relative change of the number of initial conflicts. Every initial conflict can be resolved if a feasible swap opportunity exist and otherwise leads to delay propagation. In contrast to Panel (1), the stb curve shows a flatter and less convex course. This is mainly due to the fact that cost-efficient schedules can already be recovered to a comparably large extent. The decrease of emerging initial conflicts that goes along with an increased level of stability is necessarily smaller. Regarding flx, the recurring benefit of swaps becomes apparent – not only the number of remaining conflicts but also the number of initial conflicts is implicitly reduced, see the related discussion in Section11.3.

The relative change in the number of remaining conflicts in dm-0 is illustrated in Panel (3), including all (formerly initial) conflicts that cannot be resolved and thus entail delay propagation. For both stb and flx, the overall development is rather linear. In general, the increased delay absorption capacity of schedules generated by stb provides superior results compared to their counterparts with increased swapping capability (flx). For stb, remaining conflicts can be reduced by 12.64%, corresponding to an stp improvement of 15.28% (Panel (C) of Figure 12.3). Corresponding values for flx are 9.11% (remaining conflicts) and 10.06% (stp; according to Panel (D) of Figure 12.3).

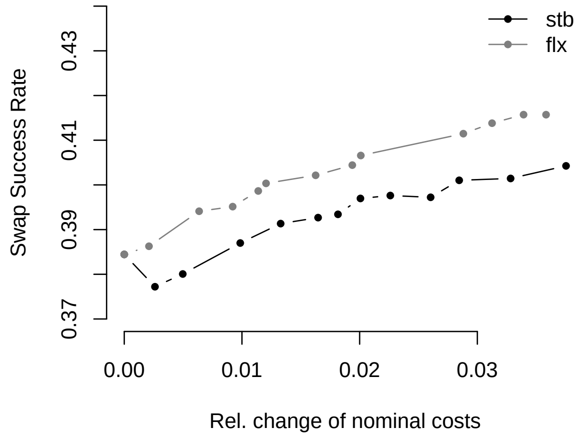

Figure 12.5: Swap success rate of schedules with an increased degree of stability and flexibility

Finally, we examine the swap success rate for both robust scheduling strategies. It is defined as based on the number of resolved conflicts γ_r and the number of initial conflicts γ_i as follows:

$$(\gamma_i - \gamma_r)/(\gamma_i) \tag{12.1}$$

Its development is illustrated in Figure 12.5 for stb (black line) and flx (gray line). Figures are averaged per penalty factor value c^Q. As reference value, 38.44% of conflicts can be resolved in cost-efficient schedules. For most flexible schedules, the average swap success rate increases up to 41.57%, the respective value for the most stable schedules is 40.26%. This corresponds to a relative improvement of 8.13% and 5.15%, respectively. Flexible schedules offer superior swap success rates, nevertheless, the increasing course of both curves is comparable.

In conclusion, the impact of stability and flexibility on swapping capabilities can be summarized as follows: Schedules with an increased degree of stability induce less interventions during simulation, leading to fewer initial conflicts. However, the swap success rate for these conflicts increases at high degrees of stability. To compare, significantly more swaps are performed in dm-0 at an even higher success rate. Moreover, flx generated schedules indirectly reduce the number of initial conflicts that may result from emerging delay propagation. While the number of initial conflicts is only slightly improved compared to cost-efficient schedules in the prop setting, it is significantly reduced in dm-0 as a result of swaps that prevent subsequent conflicts.

12.3 Summary and Implications

In the preceding analysis, the impact of increased stability and flexibility of schedules on their delay absorption capacity and swapping capability has been examined. Summarized findings are presented in the following by addressing the research questions (A)-(D).

(A) *Do stable schedules prevent delay propagation?*
Schedules with an improved degree of stability offer the best results concerning delay propagation. The secondary-to-primary delay ratio is improved by 16.78% at a nominal cost increase of 3.75%. In that respect, stable schedules offer the best robustness-to-cost ratio at low degrees of stability. Marginal costs for further improvement rise significantly. Moreover, emerging delay propagation conflicts can be reduced by up to 13.1%.

(B) *Do flexible schedules prevent delay propagation?*
In contrast to (A), the prevention of delay propagation is a minor side effect in flexible schedules. For low degrees of flexibility, results are undecided while at higher degrees delay propagation can be reduced by up to 1.86%. We assume that this behavior is a result of the scheduling approach that necessarily penalizes flight connections with delay propagation risks if no swap opportunity can be incorporated. Moreover, schedules with increased nominal costs inherently contain additional slack times compared to their cost-efficient counterparts. Concerning the robustness-to-cost ratio, corresponding rtc values only indicate a slightly positive robustness-to-cost ratio. The results are almost identical when it comes to emerging delay propagation conflicts.

(C) *Do stable schedules provide swap opportunities?*
This question deals with the performance of schedules when swaps are allowed in simulation. Compared to (A), the difference between cost-efficient and slightly stable schedules is significantly reduced since low delay absorption capacity levels of cost-efficient schedules can be compensated to a substantial extent by recovery interventions in simulation. The actual swapping capability is hard to capture and interferes with the impact of stability. Nevertheless, it becomes apparent that less swaps are necessary as a result of an improved delay absorption capacity. In addition, an increasing degree of stability leads to higher success rates for intended swaps.

(D) *Do flexible schedules provide swap opportunities?*
By design, swapping capabilities are increased by the flx approach. The highest stp improvement of 10.06% can be achieved at a nominal cost increase of 3.58%. The swapping success rate is constantly higher than in stability-based schedules. In addition, the number of emerging delay propagation conflicts can be implicitly reduced by preventing follow-up conflicts that may result from successive propagation over crew pairings.

The presented findings form important groundwork for the simultaneous consideration of stability and flexibility during scheduling in future research. Moreover, it has to be determined if a (partial) integrated consideration of crew and aircraft schedules leads to comparable effects. In this regard, the main question is if increased degrees of freedom impact both stability- and flexibility-related aspects to the same extent. Analogously, the potentially changing impact of stability and flexibility in increasingly unforeseeable environments can be investigated in subsequent studies.

Chapter 13

Summary & Outlook

In this thesis, we addressed influential factors and effect mechanisms of schedules in the field of robust efficiency of airline resource scheduling. While for the former we concentrated on the influence of exogenous delay prediction modeling on the robust efficiency, the latter dealt with the examination of mutual impacts between the stability and flexibility of crew schedules. For this purpose, we combined methods of data analysis (Part II), stochastic simulation of airline operations (Part III) and mathematical optimization for the crew scheduling (Part IV).

In the following, we provide a summary of the thesis in Section 13.1 with respect to the research objectives that were specified in Chapter 5. Afterwards, practical implications of our findings are presented in Section 13.2. Eventually, directions for future research are proposed in Section 13.3.

13.1 Summary

Chapter 1 provided an introduction into the field of robust airline resource scheduling and outlined the motivation for directions chosen for our research. The airline planning process was described in Chapter 2. Based on an overview of strategic and tactical planning, a particular focus was on aircraft and crew scheduling. Moreover, the impact of chronological decision making in the sequential scheduling approach was outlined, followed by a discussion on the potential of (partial) integration of scheduling stages.

Chapter 3 demonstrated crew- and turnaround-related process in flight operations. Inevitably emerging disruptions and delays were discussed and classified. Based on this, we derived implications for reactive and predictive operational schedule recovery as well as for the predictive adaption of delays during the scheduling process. Concerning the latter, the concept of robust efficiency was introduced, aiming at the adaption of operational requirements already during the scheduling phase.

Recent studies from related scientific literature were examined in Chapter 4. At first, we addressed optimization techniques that influence the stability or flexibility of schedules. It became apparent that most publications investigate stability-related issues while there are only a few studies dealing with schedule flexibility. Moreover, mutual influences and effect mechanisms between stability and flexibility were not sufficiently investigated although their understanding is important for a holistic view on the robust efficiency of schedules. Furthermore, the state-of-the-art in primary delay prediction for stochastic robust scheduling was presented. The necessity of examining the potential of interpretable decision rules and applying derived delay prediction models in scheduling and simulation became apparent. In consequence, four particular imminent research objectives were stated in Chapter 5:

R1 *Examining the potential of data-driven primary delay prediction*
The stochastic nature of exogenous delays is taken into account by an increasing number of recent robust scheduling approaches. Understanding delay occurrence mechanisms is inevitable for obtaining a better trade-off between cost-efficiency and robustness. This topic was therefore addressed in Part II. At first, the historical data set is introduced in Chapter 6. In the hub-and-spoke network configuration, a significant amount of domestic point-to-point connections could be observed. Moreover, considerable time-dependent patterns by the time of day and weekday as well as strong seasonal components exist.

Findings were taken into account in the subsequent data-driven examination of delay occurrence mechanisms and an ANCOVA-based statistical modeling and assessment step of derived decisions rules in Chapter 7. Taking into account the requirements of robust scheduling, the focus was on interpretable decision rules for long-term delay prediction based on spatio-temporal flight attributes rather than direct cause-and-effect relations. On the methological side, it became apparent that standard errors of statistical estimators were too small for conclusive inference on the large data set. Results must therefore be pared down to the prediction accuracy which is a general challenge in data-driven approaches on large data sets. As a key result, it turned out by cross-validation that the estimated extra-sample prediction error (EEPE) can be improved by 1.95% for the best prediction model compared to the mean value. This is equivalent, in absolute terms, to a prediction accuracy improvement of nearly one minute per flight. Interestingly, the best prediction model is based on a distinction per season rather than more detailed splits by months or weeks. Results obtained by the application of non-parametric random forests are slightly better but still in the same order of our modeling approach.

In general, it must be stated that primary delays are inherently hard to predict in the long-term context of robust scheduling due to two main reasons: Delay recording mechanisms underlie constraints that in our case leads to an underestimation of primary delays. In particular, primary delay occurrences are not tracked if secondary delays emerge at the same time, e.g. due to a late aircraft arrival. In fact this is not a simple matter since delay occurrences are not always independent and may be subject to a more complex What-If analysis. Moreover, it must be assumed that predictable patterns in delay occurrences are already taken into account in continuously optimized resource schedules of an airline. Both aspects together result in a decrease of the signal-to-noise ratio in the recorded data.

R2 *Assessing the influence of refined delay prediction on the robust efficiency of schedules*

This research objective is a direct consequence of R1 and aims at the examination of the actual influence of delay prediction models on the robust efficiency of resource schedules. R2 was addressed in Chapter 9 (Part III). The groundwork for the analysis is formed by an event-driven stochastic simulation of airline operations as described in Chapter 8. Furthermore, the potential refinement of the delay propagation model of Dück et al. (2012) was discussed in 8. It turned out that propagated delays are overestimated in theory, neglecting potential speed-ups in turnaround processes when the aircraft arrives late. Since occurrence frequencies decrease for larger delays, operational relevance is given especially for delays up to 25 minutes.

The actual study on the influence of primary delay prediction in Chapter 9 was performed in two steps. We examined the benefit of refined delay prediction based on delay prediction models obtained in Chapter 7. In this regard, a categorical prediction model with individual parameters per season, flight direction, weekday and time of day was compared to a generalized model that assumes the same delay risk for all flights. Delay prediction does not only affect the robustness of a schedule but also its nominal costs. Therefore, we introduced the robustness-to-cost ratio (rtc) as a key measure, indicating the gain in robustness per nominal cost increase. The results are obtained by applying different prediction models in crew scheduling. A key finding is that in this setting, the theoretical delay prediction accuracy improvement cannot be fully transferred to an improvement of the trade-off between nominal costs and robustness of a schedule. The categorical model shows an rtc improvement of 0.841% compared to the generalized model. It must be assumed that scheduling decisions are too restricted in sequential scheduling for reflecting the theoretical delay prediction accuracy improvement. Although the categorical model provides more exact predictions, parameters are often in the same order as in the generalized model. In particular, deviations in mean values are mostly in the interval of ± 5 minutes and therefore often too small to have a severe impact on scheduling decisions.

In order to provide more specific findings on the impact of delay prediction models, a sensitivity analysis was performed in a second step. More precisely, we investigated the impact of either complete knowledge or misplacement of delay occurrences, the impact of the number of considered delay scenarios in scheduling and simulation as well as under- and overestimation of delays. As expected, complete knowledge of delay occurrences – referred to as oracle solutions – leads by far to the best rtc improvement. If, in contrast, the wrong flights are assumed to be delayed, nominal costs are increased but severely ineffective robustness is incorporated into schedules. These extreme assumptions can be weakened by generalizing both delay placement and duration when additional delay scenarios are considered in scheduling. Interestingly, the robustness outcome of oracle solutions remains almost identical, however, at a substantial increase of nominal costs. In addition, under- and overestimation of the delay duration leads to Pareto-optimal outcomes that are not dominated by the reference model. Marginal costs for additional robustness do not increase when higher delays are assumed.

R3 *Developing recovery instruments that are suitable to evaluate the flexibility of crew and aircraft schedules*

The evaluation of schedule stability requires delay propagation as the only instrument for schedule recovery in simulation. In contrast, a holistic evaluation of the robust efficiency in simulation must also take into account recovery techniques that in particular make use of schedule flexibility properties. For this purpose, we provided the conception of a rule-based recovery approach for crews and aircraft in Chapter 10. The approach aims at the evaluation of schedule flexibility rather than providing the highest possible degree of operational performance. Nevertheless, secondary delays can be reduced by about 50% in contrast to sole delay propagation. Furthermore, it became apparent that synchronicity between crew and aircraft schedules impede the swapping capability. While crew-follows-aircraft connections are proven to have a positive impact on the schedule stability by preventing cascading delay propagation, they offer less abilities to be recovered in simulation.

R4 *Investigating mutual impacts between stability and flexibility as groundwork for a holistic view on the robust efficiency of schedules*
The concept of robust efficiency aims at a preferably close interconnection between scheduling and operations. This can be achieved by improving the stability and flexibility during the construction of schedules. Concerning this matter, R4 aims at investigating potential mutual impacts between these schedule properties. In Chapter 11, we developed an optimization approach that increases the swapping capabilities of schedules in close connection to the swapping method presented in Chapter 10. We proposed an integrated mathematical formulation for crew and aircraft schedules with a stochastic recourse function for flexibility assessment. Based on a decomposition and column generation solution strategy, computational results are obtained for the crew scheduling step with respect to given aircraft rotations.

For the investigation of mutual impacts in Chapter 12, we compared schedules generated by the proposed approach with schedules generated by the stability-related approach of Dück et al. (2012). R4 was discussed by addressing four specific research questions:

(A) *Do stable schedules prevent delay propagation?*
Schedules with an improved degree of stability offer best results in terms of preventing delay propagation. Slight improvements of the stability can be achieved for a comparably low increase of nominal costs while marginal costs for further improvement rise significantly.

(B) *Do flexible schedules prevent delay propagation?*
At higher degrees of flexibility, the side effect of a slightly improved delay absorption capacity can be observed. This can be attributed to two factors: On

the one hand, schedules with increased nominal cost levels inherently contain more slack times. On the other hand, the scheduling approach implicitly penalizes flight connections with high delay propagation risks when no feasible swap opportunity can be incorporated.

(C) *Do stable schedules provide swap opportunities?*
In contrast to (A), it can be observed that cost-efficient schedules can already be recovered to a certain extent, compensating their low level or delay absorption capacity. In consequence, the difference to slightly more stable schedules is significantly reduced. Self-evidently, increased stability makes swaps less necessary. However, for remaining intended swaps slightly better success rates can be achieved.

(D) *Do flexible schedules provide swap opportunities?*
As expected, flexible schedules provide the best results in terms of improved swapping capabilities. With increasing degrees of flexibility, intended swaps become more likely to be successful. As an indirect result of preventing delay propagation chains, the number of emerging delay propagation conflicts that require intervention is significantly reduced. In terms of actual robustness outcome, stability is generally dominant to flexibility in the setting of our study, aiming at regular daily operations.

13.2 Practical Implications

In times of a continuously growing demand for air transportation in a highly competitive market, cost-efficient and reliable operations are of high relevance. In this field of tension, we tackled several aspects dealing with the concept of robust efficiency in airline resource scheduling. The following implications for the practice can be derived from our research:

At first, we refer to the proposed data-driven approach for the retrieval of primary delay occurrence patterns. We have provided interpretable decision rules whose transfer to different real-world settings is possible. In addition, the modeling approach itself is adaptable to different data sets. Thus, it can be used to assess the prediction accuracy of particular decision rules from practitioners. Derived statistical prediction models can subsequently be taken into consideration during scheduling.

In general, it turned out that the accuracy of prediction models highly depends on the way of delay recording: The sole tracking of flight departure delays inevitably leads to delay underestimations. This is especially the case when schedules are already adapted to potential delay risks and the signal-to-noise ratio in the data is low. Under these circumstances, the worthiness of refined prediction over general assumptions for primary delay occurrences must be evaluated. In this regard, the results of the sensitivity analysis for delay misestimation offer substantial groundwork for What-if analyses of decision makers.

At second, we presented a swapping method and a related scheduling approach for increasing the schedule flexibility. Our results show that an improved schedule flexibility has a significant positive impact on the robustness, going along with a certain increase of the nominal costs. For an implementation in real-world productive systems, the approach necessarily has to be adapted to airline-specific requirements. They are not considered yet in order to examine the full theoretical potential of schedule flexibility. It must be stated that besides methodical and technical aspects, this is also an organizational task since workflows and procedures must also be adapted.

The presented approach was used for the examination of mutual impacts between stability and flexibility. In particular, the following decision rules can be derived for the practice: While increased schedule stability reduces the number of necessary interventions during operations, an increase of schedule flexibility may offer simple and manageable actions for crew and aircraft recovery in regular daily operations. This is not at the expense of the schedules' stability. However, a superior robustness outcome can be achieved by improving the stability. It therefore cannot be replaced by the solely making use of schedule flexibility. The obtained fundamental effect mechanisms can be used as guidance for decision makers when dealing with the incorporation of robustness into aircraft and crew schedules.

13.3 Directions for Future Research

The addressed research objectives offer several aspects for future research. In the following, we present the most relevant ones concerning delay prediction modeling and application, rule-based recovery and robust crew and aircraft scheduling.

When dealing with the potential of primary delay prediction for robust resource scheduling, the impairment of delay recording mechanisms became apparent. In order to **examine and quantify the underestimation of exogenous delays**, partial processes and critical paths of ground operations can be analyzed[1]. Such findings may offer further improvement of the delay prediction accuracy. In addition, it must be examined to what extent the **findings may be generalized for other airlines**. It can be assumed that for less optimized schedules the results show a higher level of prediction accuracy improvement. In addition, point-to-point networks may underlie different delay occurrence mechanisms due to the lack of workload concentration and control procedures at important hubs.

Regarding the application of prediction models in scheduling, **the impact of refined primary delay prediction at a higher degree of freedom for scheduling decisions** has to be examined. In addition, findings from the evaluation of delay propagation mechanisms for aircraft can be transferred to crew itineraries on available data. Altogether, **refined delay propagation modeling** can be applied in scheduling and simulation.

While the integrated rule-based recovery for crews and aircraft were elaborated with a focus on its technical potentials from the schedulers' view, its practical applicability must be further examined. In particular, it is essential to **reconsider the assumption of cost-neutrality of swaps** by quantifying indirect expenses and organizational efforts for their implementation. Furthermore, the applicability of swaps for relatively small delays must be evaluated. The inevitable impact on ground operations and processes, e.g. due to gate changes, must be taken into account. Since complete reproduction of operational reality to the last detail is not expedient, abstracted decision rules can be derived.

Referring to the scheduling approach for increasing flexibility, future research has to address the **impact of flexibility in less predictive environments**. It must be examined if less complex indicator-based approaches are competitive since the adaption of particular propagation risks is negligible in these cases. Moreover, the stochastic recourse can be adapted for a (partially) **integrated crew and aircraft scheduling with regard to their flexibility**. This appears promising since scheduling of integrated swaps can be better coordinated. For that matter, more sophisticated decomposition and interaction strategies must be developed to synchronize the incorporation of compatible swap opportunities for crews and aircraft. It has

[1]Relevant data sets are often available for particular (hub) airports, see Schlegel (2010).

to be evaluated if an adaption of the iterative optimization approach of Weide et al. (2010) and Dück et al. (2012) is promising compared to a Bender's Decomposition, see Mercier et al. (2005). Eventually, the benefit of **simultaneous consideration of stability and flexibility** can be assessed, enhancing the findings of Burke et al. (2010) on the general impact of stability and flexibility on the schedule robustness. In this regard, findings from this thesis look promising since severe trade-offs between stability and flexibility are not apparent.

Bibliography

Abdelghany, A., G. Ekollu, R. Narasimhan, and K. Abdelghany (2004). A proactive crew recovery decision support tool for commercial airlines during irregular operations. *Annals of Operations Research 127*(1), 309–331.

Adler, N. and B. Golany (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western europe. *European Journal of Operational Research 132*(2), 260–273.

Ageeva, Y. (2000). Approaches to incorporating robustness into airline scheduling. Master's thesis.

AhmadBeygi, S., A. Cohn, Y. Guan, and P. Belobaba (2008). Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management 14*(5), 221–236.

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

Amberg, B. (2017). *Robuste Effizienz des Ressourceneinsatzes im öffentlichen Personennahverkehr*. Ph. D. thesis, Freie Universität Berlin.

Amberg, B., B. Amberg, and N. Kliewer (2018). Robust efficiency in urban public transportation: Minimizing delay propagation in cost-efficient bus and driver schedules. *Transportation Science*. Published online in Articles in Advance 09 Mar 2018, `https:/doi.org/10.1287/trsc.2017.0757`.

Anbil, R., J. J. Forrest, and W. R. Pulleyblank (1998). Column generation and the airline crew pairing problem. *Documenta Mathematica 3*, 677–686.

Anbil, R., E. Gelman, B. Patty, and R. Tanga (1991). Recent advances in crew-pairing optimization at american airlines. *Interfaces 21*(1), 62–74.

Arıkan, M., V. Deshpande, and M. Sohoni (2013). Building reliable air-travel infrastructure using empirical data and stochastic models of airline networks. *Operations Research 61*(1), 45–64.

Ball, M., C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, and B. Zou (2010). *Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States*. Nextor.

Ball, M., C. Barnhart, G. Nemhauser, and A. Odoni (2007). Air transportation: Irregular operations and control. *Handbooks in Operations Research and Management Science 14*, 1–73.

Barnhart, C., N. L. Boland, L. W. Clarke, E. L. Johnson, G. L. Nemhauser, and R. G. Shenoi (1998). Flight string models for aircraft fleeting and routing. *Transportation Science 32*(3), 208–220.

Barnhart, C., A. M. Cohn, E. L. Johnson, D. Klabjan, G. L. Nemhauser, and P. H. Vance (2003). Airline crew scheduling. In R. Hall (Ed.), *Handbook of Transportation Science*, pp. 517–560. New York: Springer.

Barnhart, C., A. Farahat, and M. Lohatepanont (2001). Airline fleet assignment: An alternative model and solution approach based on composite variables. *submitted to Operations Research*.

Barnhart, C., A. Farahat, and M. Lohatepanont (2009). Airline fleet assignment with enhanced revenue modeling. *Operations research 57*(1), 231–244.

Barnhart, C., E. L. Johnson, G. L. Nemhauser, M. W. Savelsbergh, and P. H. Vance (1998). Branch-and-price: Column generation for solving huge integer programs. *Operations research 46*(3), 316–329.

Barnhart, C., T. S. Kniker, and M. Lohatepanont (2002). Itinerary-based airline fleet assignment. *Transportation Science 36*(2), 199–217.

Barnhart, C. and B. Smith (2012). *Quantitative Problem Solving Methods in the Airline Industry*. Springer.

Barnhart, C. and K. Talluri (1997). Airline operations research. *Design and Operations of Civil and Environmental Engineering Systems*, 435–469.

Bazargan, M. (2010). *Airline operation and scheduling.* Farnham: Ashgate Publishing Limited.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik 4*(1), 238–252.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association 33*(203), 526–536.

Borndörfer, R., I. Dovica, I. Nowak, and T. Schickinger (2010). Robust tail assignment. In *Proceedings of the fiftieth annual symposium of AGIFORS.*

Borndörfer, R., U. Schelten, T. Schlechte, and S. Weider (2006). A column generation approach to airline crew scheduling. In *Operations Research Proceedings 2005*, pp. 343–348. Springer.

Bratu, S. and C. Barnhart (2006). Flight operations recovery: New approaches considering passenger recovery. *Journal of Scheduling 9*(3), 279–298.

Breiman, L. (2001a). Random forests. *Machine learning 45*(1), 5–32.

Breiman, L. (2001b). Statistical modeling: The two cultures. *Statistical Science 16*(3), 199–231.

Burghouwt, G., J. Hakfoort, and J. R. van Eck (2003). The spatial configuration of airline networks in europe. *Journal of Air Transport Management 9*(5), 309–323.

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

Clarke, L., E. Johnson, G. Nemhauser, and Z. Zhu (1997). The aircraft rotation problem. *Annals of Operations Research 69*, 33–46.

Clarke, L. W., C. A. Hane, E. L. Johnson, and G. L. Nemhauser (1996). Maintenance and crew considerations in fleet assignment. *Transportation Science 30*(3), 249–260.

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

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician 35*(1), 54.

CODA (2011). Planning for delay: influence of flight scheduling on airline punctuality. *EUROCONTROL Trends in Air Traffic 7*.

CODA (2017). Coda digest – all-causes delay and cancellations to air transport in europe – 2016. https://www.eurocontrol.int/sites/default/files/content/documents/official-documents/facts-and-figures/coda-reports/coda-digest-q4-2016.pdf. Last access: 30th August, 2017.

Cohn, A. M. and C. Barnhart (2003). Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research 51*(3), 387–396.

Conrady, R., F. Fichert, and R. Sterzenbach (2013). *Luftverkehr: Betriebswirtschaftliches Lehr- und Handbuch*. Walter de Gruyter.

Cook, A. J. and G. Tanner (2011). European airline delay cost reference values.

Cordeau, J.-F., G. Stojković, F. Soumis, and J. Desrosiers (2001). Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation science 35*(4), 375–388.

Cox, D. R. (2006). *Principles of statistical inference*. Cambridge University Press.

Cox, D. R. and N. Wermuth (1996). *Multivariate dependencies: Models, analysis and interpretation*, Volume 67. CRC Press.

Dantzig, G. B. and P. Wolfe (1960). Decomposition principle for linear programs. *Operations research 8*(1), 101–111.

Desaulniers, G., J. Desrosiers, Y. Dumas, S. Marc, B. Rioux, M. M. Solomon, and F. Soumis (1997). Crew pairing at air france. *European Journal of Operational Research 97*(2), 245–259.

Desaulniers, G., J. Desrosiers, Y. Dumas, M. M. Solomon, and F. Soumis (1997). Daily aircraft routing and scheduling. *Management Science 43*(6), 841–855.

Deshpande, V. and M. Arikan (2012). The impact of airline flight schedules on flight delays. *Manufacturing & Service Operations Management 14*(3), 423–440.

Domschke, W., A. Drexl, R. Klein, and A. Scholl (2015). *Einführung in Operations Research.* Springer-Verlag.

Dück, V. (2010). *Increasing stability of aircraft and crew schedules.* Ph. D. thesis, Universität Paderborn, DS & OR Lab.

Dück, V., L. Ionescu, N. Kliewer, and L. Suhl (2012). Increasing stability of crew and aircraft schedules. *Transportation research part C: emerging technologies 20*(1), 47–61.

Dück, V., F. Wesselmann, and L. Suhl (2011). Implementing a branch and price and cut method for the airline crew pairing optimization problem. *Public Transport 3*(1), 43–64.

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

Dunbar, M., G. Froyland, and C.-L. Wu (2014). An integrated scenario-based approach for robust aircraft routing, crew pairing and re-timing. *Computers & Operations Research 45*, 68–86.

Dunlay, W. and R. Horonjeff (1976). Stochastic properties of enroute air traffic-an empirical investigation. *Journal of Aircraft 13*(5), 376–381.

Duquesne, C.-M. (2013). *Intégration du déploiement de flotte et du service aux passagers dans la gestion de la planification pour compagnie aérienne.* Ph. D. thesis, Université de Grenoble.

Efron, B. (2010). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, Volume 1. Cambridge University Press.

Eggenberg, N., M. Salani, and M. Bierlaire (2010a). Constraint-specific recovery network for solving airline recovery problems. *Computers & Operations Research 37*(6), 1014–1026.

Eggenberg, N., M. Salani, and M. Bierlaire (2010b). Robust and recoverable maintenance routing schedules. *Tristan VII*.

Ehrgott, M. and D. M. Ryan (2002). Constructing robust crew schedules with bicriteria optimization. *Journal of Multi-Criteria Decision Analysis 11*(3), 139–150.

Erdmann, A., A. Nolte, A. Noltemeier, and R. Schrader (2001). Modeling and solving an airline schedule generation problem. *Annals of Operations Research 107*(1-4), 117–142.

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

EUROCONTROL (2013). Seven-year forecast: Flight movements and service units 2013-2019, september 2013.

EUROCONTROL (2016a). EUROCONTROL Annual Report 2015. https://www.eurocontrol.int/publications/eurocontrol-annual-report-2015. last access: 14th April, 2017.

EUROCONTROL (2016b). Market Segments in European Air Traffic 2015. https://www.eurocontrol.int/sites/default/files/content/documents/official-documents/facts-and-figures/statfor/doc581-market-segments-2015-v1.0.pdf. last access: January 8^{th}, 2018.

Ferguson, J., A. Q. Kara, K. Hoffman, and L. Sherry (2013). Estimating domestic us airline cost of delay based on european model. *Transportation Research Part C: Emerging Technologies 33*, 311–323.

Fink, A., N. Kliewer, D. Mattfeld, L. Mönch, F. Rothlauf, G. Schryen, L. Suhl, and S. Voß (2014). Model-based decision support in manufacturing and service networks. *Business & Information Systems Engineering 6*(1), 17–24.

Fricke, H. and M. Schultz (2009). Delay impacts onto turnaround performance. In *ATM Seminar*.

Froyland, G., S. J. Maher, and C.-L. Wu (2014). The recoverable robust tail assignment problem. *Transportation Science 48*(3), 351–372.

Gao, C., E. Johnson, and B. Smith (2009). Integrated airline fleet and crew robust planning. *Transportation Science 43*(1), 2–16.

Gopalakrishnan, B. and E. L. Johnson (2005). Airline crew scheduling: state-of-the-art. *Annals of Operations Research 140*(1), 305–337.

Gopalan, R. and K. T. Talluri (1998). The aircraft maintenance routing problem. *Operations Research 46*(2), 260–271.

Grönkvist, M. (2005). *The tail assignment problem.* Citeseer.

Grönkvist, M. (2006). Accelerating column generation for aircraft scheduling using constraint propagation. *Computers & Operations Research 33*(10), 2918–2934.

Gwiggner, C. (2007). *Analyse ds intercitudes dans les flux du trafix aérien.* Ph. D. thesis, Ecole Polytechnique.

Hand, D. J., H. Mannila, and P. Smyth (2001). *Principles of data mining.* MIT press.

Hane, C. A., C. Barnhart, E. L. Johnson, R. E. Marsten, G. L. Nemhauser, and G. Sigismondi (1995). The fleet assignment problem: Solving a large-scale integer program. *Mathematical Programming 70*(1-3), 211–232.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference and prediction* (2 ed.). Springer.

Hirata, T., A. Shimizu, D. MIMURA, and T. YAI (2013). Analysis of airborne delay characteristics of flights controlled by ground-holding: Case of the flights to tokyo international airport. *Journal of the Eastern Asia Society for Transportation Studies 10*, 2210–2222.

Hsiao, C.-Y. and M. Hansen (2006). Econometric analysis of us airline flight delays with time-of-day effects. *Transportation Research Record: Journal of the Transportation Research Board 1951*(1), 104–112.

Hyndman, R. J. and G. Athanasopoulos (2014). *Forecasting: principles and practice.* OTexts.

IATA (2008). *Airport Handling Manual.* International Air Transport Association.

Ionescu, L., C. Gwiggner, and N. Kliewer (2016). Data analysis of delays in airline networks. *Business & Information Systems Engineering 58*(2), 119–133.

Ionescu, L. and N. Kliewer (2011). Increasing flexibility of airline crew schedules. *Procedia-Social and Behavioral Sciences 20*, 1019–1028.

Jiménez, J. L. and O. Betancor (2012). When trains go faster than planes: the strategic reaction of airlines in spain. *Transport Policy 23*, 34–41.

Kasirzadeh, A., M. Saddoune, and F. Soumis (2015, February). Airline crew scheduling: models, algorithms, and data sets. *EURO Journal on Transportation and Logistics*, –.

Klabjan, D. (2005). Large-scale models in the airline industry. In *Column generation*, pp. 163–195. Springer.

Klabjan, D., E. L. Johnson, G. L. Nemhauser, E. Gelman, and S. Ramaswamy (2001). Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optimization and Applications 20*(1), 73–91.

Klabjan, D., E. L. Johnson, G. L. Nemhauser, E. Gelman, and S. Ramaswamy (2002). Airline crew scheduling with time windows and plane-count constraints. *Transportation science 36*(3), 337–348.

Kohl, N. and S. E. Karisch (2004). Airline crew rostering: Problem types, modeling, and optimization. *Annals of Operations Research 127*(1-4), 223–257.

Kohl, N., A. Larsen, J. Larsen, A. Ross, and S. Tiourine (2007). Airline disruption management – perspectives, experiences and outlook. *Journal of Air Transport Management 13*(3), 149–162.

Lan, S., J.-P. Clarke, and C. Barnhart (2006). Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. *Transportation Science 40*(1), 15–28.

Lapp, M. and A. Cohn (2012). Modifying lines-of-flight in the planning process for improved maintenance robustness. *Computers & Operations Research 39*(9), 2051–2062.

Lettovský, L., E. L. Johnson, and G. L. Nemhauser (2000). Airline crew recovery. *Transportation Science 34*(4), 337–348.

Liberatore, M. J. and W. Luo (2010). The analytics movement: Implications for operations research. *Interfaces 40*(4), 313–324.

Liebchen, C., M. Lübbecke, R. Möhring, and S. Stiller (2009). The concept of recoverable robustness, linear programming recovery, and railway applications. In *Robust and online large-scale optimization*, pp. 1–27. Springer.

Lindsey, J. K. (2004). *Statistical analysis of stochastic processes in time*, Volume 14. Cambridge University Press.

Lohatepanont, M. and C. Barnhart (2004). Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment. *Transportation Science 38*(1), 19–32.

Lučic, P. and D. Teodorovic (1999). Simulated annealing for the multi-objective aircrew rostering problem. *Transportation Research Part A: Policy and Practice 33*(1), 19–45.

Maher, S. (2013). *The application of recoverable robustness to airline planning problems.* Ph. D. thesis, The University of New South Wales.

Maher, S. J., G. Desaulniers, and F. Soumis (2014). Recoverable robust single day aircraft maintenance routing problem. *Computers & Operations Research 51*, 130–145.

Maindonald, J. and W. J. Braun (2010). Data analysis and graphics using r: An example-based approach.

Marla, L., B. Vaaben, and C. Barnhart (2016). Integrated disruption management and flight planning to trade off delays and fuel burn. *Transportation Science 51*(1), 88–111.

Meffert, H., J. Nießing, and B. Ballensiefen (2005). Der markteintritt der low-cost-airlines in deutschland: Bedrohungspotenzial und handlungsempfehlungen für die deutsche bahn ag. *ZEV rail Glasers Annalen 129*(3), 68–74.

Mercier, A., J.-F. Cordeau, and F. Soumis (2005). A computational study of benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers & Operations Research 32*(6), 1451–1476.

Mercier, A. and F. Soumis (2007). An integrated aircraft routing, crew scheduling and flight retiming model. *Computers & Operations Research 34*(8), 2251–2265.

Midkiff, A. H., R. J. Hansman, and T. G. Reynolds (2015). Airline flight operations. In P. Belobaba, A. Odoni, and C. Barnhart (Eds.), *The global airline industry*, Chapter 8, pp. 213–252. John Wiley & Sons.

Mortenson, M. J., N. F. Doherty, and S. Robinson (2015). Operational research from taylorism to terabytes: A research agenda for the analytics age. *European Journal of Operational Research 241*(3), 583–595.

Papadakos, N. (2009). Integrated airline scheduling. *Computers & Operations Research 36*(1), 176–195.

Petersen, J. D., G. Sölveling, J.-P. Clarke, E. L. Johnson, and S. Shebalov (2012). An optimization approach to airline integrated recovery. *Transportation Science 46*(4), 482–500.

Rasmussen, M., R. Lusby, D. Ryan, and J. Larsen (2011). A subsequence generation approach for the airline crew pairing problem. Technical report, Department of Management Engineering, Technical University of Denmark, Denmark.

Rexing, B., C. Barnhart, T. Kniker, A. Jarrah, and N. Krishnamurthy (2000). Airline fleet assignment with time windows. *Transportation Science 34*(1), 1–20.

Rosenberger, J. M., E. L. Johnson, and G. L. Nemhauser (2003). Rerouting aircraft for airline recovery. *Transportation Science 37*(4), 408–421.

Rosenberger, J. M., E. L. Johnson, and G. L. Nemhauser (2004). A robust fleet-assignment model with hub isolation and short cycles. *Transportation Science 38*(3), 357–368.

Rosenberger, J. M., A. J. Schaefer, D. Goldsman, E. L. Johnson, A. J. Kleywegt, and G. L. Nemhauser (2000). Air transportation simulation: Simair: a stochastic model of airline operations. In *Proceedings of the 32nd conference on Winter simulation*, pp. 1118–1122. Society for Computer Simulation International.

Rosenberger, J. M., A. J. Schaefer, D. Goldsman, E. L. Johnson, A. J. Kleywegt, and G. L. Nemhauser (2002). A stochastic model of airline operations. *Transportation science 36*(4), 357–377.

Rushmeier, R. A. and S. A. Kontogiorgis (1997). Advances in the optimization of airline fleet assignment. *Transportation science 31*(2), 159–169.

Ruther, S., N. Boland, and F. Engineer (2013). Integrated aircraft routing, crew pairing, and tail assignment: branch-and-price with many pricing problems.

Ryan, D. M. (1992). The solution of massive generalized set partitioning problems in aircrew rostering. *Journal of the operational research society*, 459–467.

Ryan, D. M. and B. A. Foster (1981). An integer programming approach to scheduling. *Computer scheduling of public transport urban passenger vehicle and crew scheduling*, 269–280.

Sandhu, R. and D. Klabjan (2007). Integrated airline fleeting and crew-pairing decisions. *Operations Research 55*(3), 439–456.

Sarmadi, S. (2004). Minimizing airline passenger delay through integrated flight scheduling and aircraft routing. Master's thesis, Massachusetts Institute of Technology.

Schaefer, A. J., E. L. Johnson, A. J. Kleywegt, and G. L. Nemhauser (2005). Airline crew scheduling under uncertainty. *Transportation Science 39*(3), 340–348.

Schlegel, A. (2010). *Bodenabfertigungsprozesse im Luftverkehr: Eine statistische Analyse am Beispiel der Deutschen Lufthansa AG am Flughafen Frankfurt/Main.* Springer-Verlag.

Scholl, A. (2001). *Robuste Planung und Optimierung.* Springer DE.

Shebalov, S. and D. Klabjan (2006). Robust airline crew pairing: Move-up crews. *Transportation Science 40*(3), 300–312.

Sherali, H. D., E. K. Bish, and X. Zhu (2006). Airline fleet assignment concepts, models, and algorithms. *European Journal of Operational Research 172*(1), 1–30.

Sherali, H. D. and X. Zhu (2008). Two-stage fleet assignment model considering stochastic passenger demands. *Operations Research 56*(2), 383–399.

Smith, B. C. and E. L. Johnson (2006). Robust airline fleet assignment: Imposing station purity using station decomposition. *Transportation Science 40*(4), 497–516.

Sodi, S. (2011). Snow disruption. AGIFORS Airline Operations, London.

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

Tam, B., M. Ehrgott, D. Ryan, and G. Zakeri (2011). A comparison of stochastic programming and bi-objective optimisation approaches to robust airline crew scheduling. *OR Spectrum 33*(1), 49–75.

Tam, M. B. B. (2011). *Optimisation approaches for robust airline crew scheduling.* Ph. D. thesis, University of Auckland.

Teodorović, D. and E. Krcmar-Nozić (1989). Multicriteria model to determine flight frequencies on an airline network under competitive conditions. *Transportation Science 23*(1), 14–25.

Tu, Y., M. O. Ball, and W. S. Jank (2008). Estimating flight departure delay distributions – a statistical approach with long-term trend and short-term pattern. *Journal of the American Statistical Association 103*(481), 112–125.

Vance, P. H., A. Atamturk, C. Barnhart, E. Gelman, E. L. Johnson, A. Krishna, D. Mahidhara, G. L. Nemhauser, and R. Rebello (1997). A heuristic branch-and-price approach for the airline crew pairing problem. *preprint*.

Vance, P. H., C. Barnhart, E. L. Johnson, and G. L. Nemhauser (1997). Airline crew scheduling: A new formulation and decomposition algorithm. *Operations Research 45*(2), 188–200.

Weide, O. (2009). *Robust and integrated airline scheduling.* Ph. D. thesis, University of Auckland.

Weide, O., D. Ryan, and M. Ehrgott (2010). An iterative approach to robust and integrated aircraft routing and crew scheduling. *Computers & Operations Research 37*(5), 833–844.

Weisberg, S. (2005). *Applied linear regression.* John Wiley & Sons.

Wesonga, R., F. Nabugoomu, and P. Jehopio (2012). Parameterized framework for the analysis of probabilities of aircraft delay at an airport. *Journal of Air Transport Management 23*, 1–4.

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

Xu, N., L. Sherry, and K. B. Laskey (2008). Multifactor model for predicting delays at us airports. *Transportation Research Record: Journal of the Transportation Research Board 2052*(1), 62–71.

Yen, J. W. and J. R. Birge (2006). A stochastic programming approach to the airline crew scheduling problem. *Transportation Science 40*(1), 3–14.

Yu, G., M. Argüello, G. Song, S. M. McCowan, and A. White (2003). A new era for crew recovery at continental airlines. *Interfaces 33*(1), 5–22.

Appendix A

Column Generation Solution Approach for Set Partitioning Problems

In the following, we describe the Dantzig-Wolfe decomposition and the fundamental column generation solution approach for the Crew Pairing Problem. The Dantzig-Wolfe decomposition splits an original problem into a pricing problem and a restricted master problem (RMP), see Dantzig and Wolfe (1960). The RMP can be formulated as a LP-relaxation of a set-partitioning formulation, see Model A.1a-A.1c.

$$\min \sum_{j \in J} c_j x_j \tag{A.1a}$$

$$s.t. \sum_{j \in J} a_{ij} x_j = b_i \qquad \forall i \in I \tag{A.1b}$$

$$0 \leq x_j \leq 1 \qquad \forall j \in J \subseteq N \tag{A.1c}$$

Columns $j \in J$ of the model correspond to the currently considered set of feasible crew pairings. It holds $J \subseteq N$ where N is the set of all theoretically possible crew pairings. c_j depict the pairing costs that are connected to the continuous decision variables x_j in the objective function. Each row $i \in I$ refers to one flight that has to be covered exactly once when the right hand side values b_i are set to 1. $a_{ij} \in A$ has the value 1 if the related flight is covered by pairing j, otherwise 0.

In every iteration k, the RMP is solved to optimality based on the current column set $J \subseteq N$. Dual values μ_k^i for every row i can be derived from A.1b. Subsequently, μ_k^i values can be used in pricing to generated additional columns that provide an improvement of the solution. A column j can improve the solution if its reduced costs are negative. The reduced costs rc_j are defined as

$$rc_j = c_j - \sum_{i \in M} a_{ij}\mu_i. \tag{A.2}$$

The pricing problem then states

$$J_{new} = \min_{j \in N} rc_j \tag{A.3}$$

where J_{new} is the set of newly generated columns. In the scope of the Crew Pairing Problem, it is usually formulated as a resource constrained shortest path problem that can be solved by dynamic programming.

The overall column generation process is illustrated in Algorithm 2. In column generation, linear programs (LP) are solved to optimality when no additional columns with reduced costs can be found in pricing. The application of column generation for integer programs (IP) like the Crew Pairing Problem necessitates an integration into a branch-and-bound procedure, referred to as branch-and-price. At every node of the branch-and-bound tree, a relaxed formulation of the original problem is solved by column generation. In a sequential approach, only the root node is solved to optimality regarding the relaxed LP. Afterwards, an IP solution is searched for based on the set of generated columns. This can be done by a standard IP solver. For details on the fundamental branch-and-price technique, we refer to Barnhart et al. (1998).

Algorithm 2: The Column Generation Solution Approach

```
Generate an initial column set J_0
k ← 0
while |J_new| > 0 do
    Solve the restricted master problem RMP with the column set J_k μ_k ←
      dual values of the current optimal RMP solution
    solve the pricing problem based on μ_k and obtain new columns J_new with
      negative reduced costs
    J_{k+1} = J_k ∪ J_new
    k = k + 1
end
```

Appendix B

Standard IATA Delay Codes

The following list of standardized IATA Delay Codes bases on IATA (2008), Chapter AHM 011.

Others

00-05	AIRLINE INTERNAL CODES
06 (OA)	NO GATE/STAND AVAILABILITY DUE TO OWN AIRLINE ACTIVITY
09 (SG)	SCHEDULED GROUND TIME LESS THAN DECLARED MINIMUM GROUND TIME

Passenger and Baggage

11 (PD)	LATE CHECK-IN, acceptance after deadline
12 (PL)	LATE CHECK-IN, congestions in check-in area
13 (PE)	CHECK-IN ERROR, passenger and baggage
14 (PO)	OVERSALES, booking errors
15 (PH)	BOARDING, discrepancies and paging, missing checked-in passenger
16 (PS)	COMMERCIAL PUBLICITY/PASSENGER CONVENIENCE, VIP, press, ground meals and missing personal items
17 (PC)	CATERING ORDER, late or incorrect order given to supplier
18 (PB)	BAGGAGE PROCESSING, sorting etc.

Cargo and Mail

21 (CD)	DOCUMENTATION, errors etc.
22 (CP)	LATE POSITIONING
23 (CC)	LATE ACCEPTANCE
24 (CI)	INADEQUATE PACKING
25 (CO)	OVERSALES, booking errors
26 (CU)	LATE PREPARATION IN WAREHOUSE
27 (CE)	DOCUMENTATION, PACKING etc (Mail Only)
28 (CL)	LATE POSITIONING (Mail Only)
29 (CA)	LATE ACCEPTANCE (Mail Only)

Aircraft and Ramp Handling

31 (GD)	AIRCRAFT DOCUMENTATION LATE/INACCURATE, weight and balance, general declaration, pax manifest, etc.
32 (GL)	LOADING/UNLOADING, bulky, special load, cabin load, lack of loading staff
33 (GE)	LOADING EQUIPMENT, lack of or breakdown, e.g. container pallet loader, lack of staff
34 (GS)	SERVICING EQUIPMENT, lack of or breakdown, lack of staff, e.g. steps
35 (GC)	AIRCRAFT CLEANING
36 (GF)	FUELLING/DEFUELLING, fuel supplier
37 (GB)	CATERING, late delivery or loading
38 (GU)	ULD, lack of or serviceability
39 (GT)	TECHNICAL EQUIPMENT, lack of or breakdown, lack of staff, e.g. pushback

Technical and Aircraft Equipment

41 (TD) AIRCRAFT DEFECTS.
42 (TM) SCHEDULED MAINTENANCE, late release.
43 (TN) NON-SCHEDULED MAINTENANCE, special checks and/or additional works beyond normal maintenance schedule.
44 (TS) SPARES AND MAINTENANCE EQUIPMENT, lack of or breakdown.
45 (TA) AOG SPARES, to be carried to another station.
46 (TC) AIRCRAFT CHANGE, for technical reasons.
47 (TL) STAND-BY AIRCRAFT, lack of planned stand-by aircraft for technical reasons.
48 (TV) SCHEDULED CABIN CONFIGURATION/VERSION ADJUSTMENTS.

Damage to Aircraft & EDP/Automated Equipment Failure

51 (DF) DAMAGE DURING FLIGHT OPERATIONS, bird or lightning strike, turbulence, heavy or overweight landing, collision during taxiing
52 (DG) DAMAGE DURING GROUND OPERATIONS, collisions (other than during taxiing), loading/off-loading damage, contamination, towing, extreme weather conditions
55 (ED) DEPARTURE CONTROL
56 (EC) CARGO PREPARATION/DOCUMENTATION
57 (EF) FLIGHT PLANS

Flight Operations and Crewing

61 (FP)	FLIGHT PLAN, late completion or change of, flight documentation
62 (FF)	OPERATIONAL REQUIREMENTS, fuel, load alteration
63 (FT)	LATE CREW BOARDING OR DEPARTURE PROCEDURES, other than connection and standby (flight deck or entire crew)
64 (FS)	FLIGHT DECK CREW SHORTAGE, sickness, awaiting standby, flight time limitations, crew meals, valid visa, health documents, etc.
65 (FR)	FLIGHT DECK CREW SPECIAL REQUEST, not within operational requirements
66 (FL)	LATE CABIN CREW BOARDING OR DEPARTURE PROCEDURES, other than connection and standby
67 (FC)	CABIN CREW SHORTAGE, sickness, awaiting standby, flight time limitations, crew meals, valid visa, health documents, etc.
68 (FA)	CABIN CREW ERROR OR SPECIAL REQUEST, not within operational requirements
69 (FB)	CAPTAIN REQUEST FOR SECURITY CHECK, extraordinary

Weather

71 (WO)	DEPARTURE STATION
72 (WT)	DESTINATION STATION
73 (WR)	EN ROUTE OR ALTERNATE
75 (WI)	DE-ICING OF AIRCRAFT, removal of ice and/or snow, frost prevention excluding unserviceability of equipment
76 (WS)	REMOVAL OF SNOW, ICE, WATER AND SAND FROM AIRPORT
77 (WG)	GROUND HANDLING IMPAIRED BY ADVERSE WEATHER CONDITIONS

Air Traffic Flow Management Restrictions

81 (AT)	ATFM due to ATC EN-ROUTE DEMAND/CAPACITY, standard demand/capacity problems
82 (AX)	ATFM due to ATC STAFF/EQUIPMENT EN-ROUTE, reduced capacity caused by industrial action or staff shortage, equipment failure, military exercise or extraordinary demand due to capacity reduction in neighbouring area
83 (AE)	ATFM due to RESTRICTION AT DESTINATION AIRPORT, airport and/or runway closed due to obstruction, industrial action, staff shortage, political unrest, noise abatement, night curfew, special flights
84 (AW)	ATFM due to WEATHER AT DESTINATION

Airport and Governmental Authorities

85 (AS)	MANDATORY SECURITY
86 (AG)	IMMIGRATION, CUSTOMS, HEALTH
87 (AF)	AIRPORT FACILITIES, parking stands, ramp congestion, lighting, buildings, gate limitations, etc.
88 (AD)	RESTRICTIONS AT AIRPORT OF DESTINATION, airport and/or runway closed due to obstruction, industrial action, staff shortage, political unrest, noise abatement, night curfew, special flights
89 (AM)	RESTRICTIONS AT AIRPORT OF DEPARTURE WITH OR WITHOUT ATFM RESTRICTIONS, including Air Traffic Services, start-up and push-back, airport and/or runway closed due to obstruction or weather, industrial action, staff shortage, political unrest, noise abatement, night curfew, special flights

Reactionary

91 (RL) LOAD CONNECTION, awaiting load from another flight

92 (RT) THROUGH CHECK-IN ERROR, passenger and baggage

93 (RA) AIRCRAFT ROTATION, late arrival of aircraft from another flight or previous sector

94 (RS) CABIN CREW ROTATION, awaiting cabin crew from another flight

95 (RC) CREW ROTATION, awaiting crew from another flight (flight deck or entire crew)

96 (RO) OPERATIONS CONTROL, re-routing, diversion, consolidation, aircraft change for reasons other than technical

Miscellaneous

97 (MI) INDUSTRIAL ACTION WITH OWN AIRLINE

98 (MO) INDUSTRIAL ACTION OUTSIDE OWN AIRLINE, excluding ATS

99 (MX) OTHER REASON, not matching any code above

Appendix C

Supplementary Computational Results

C.1 Flight Movements during the Course of the Day

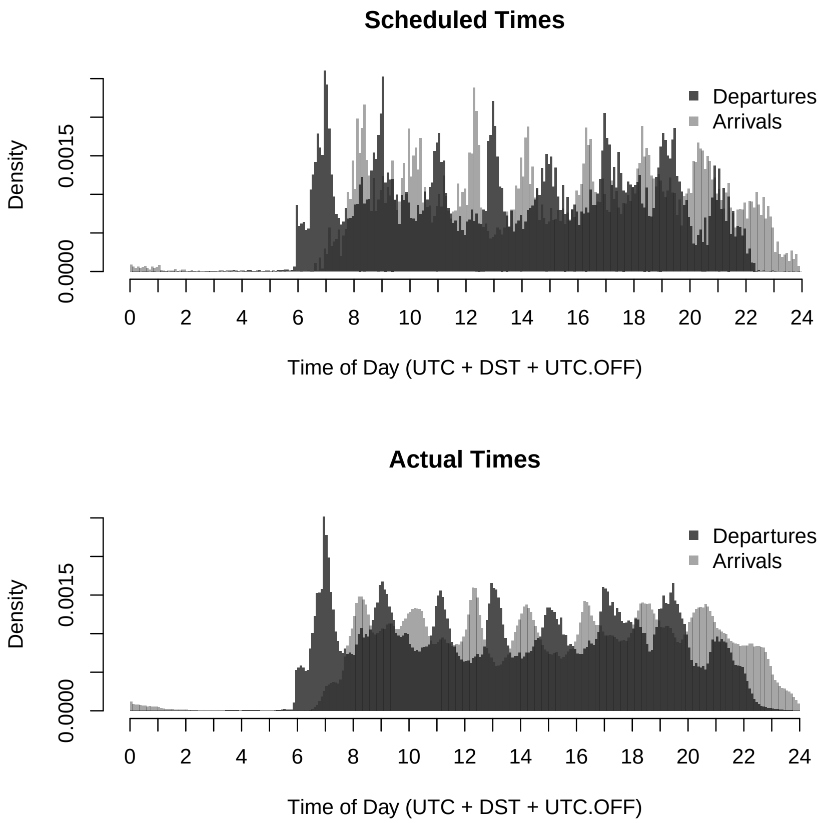

Figure C.1: Departures and arrivals per local time of day

C.2 Time Course Analysis

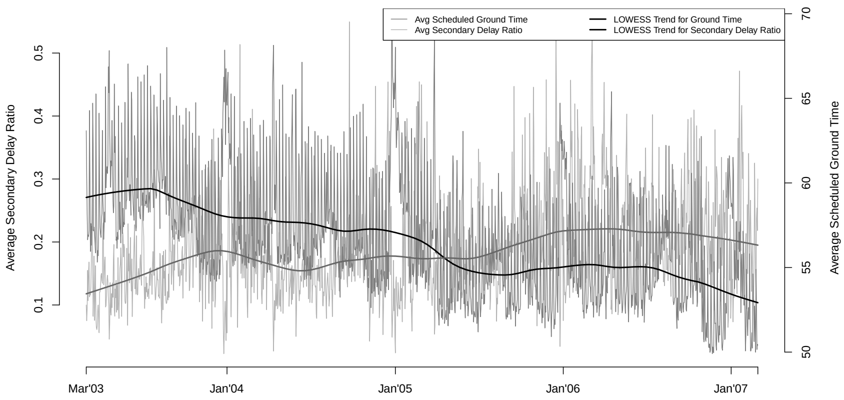

Figure C.2: Development of Secondary Delay Ratio and Scheduled Ground Times

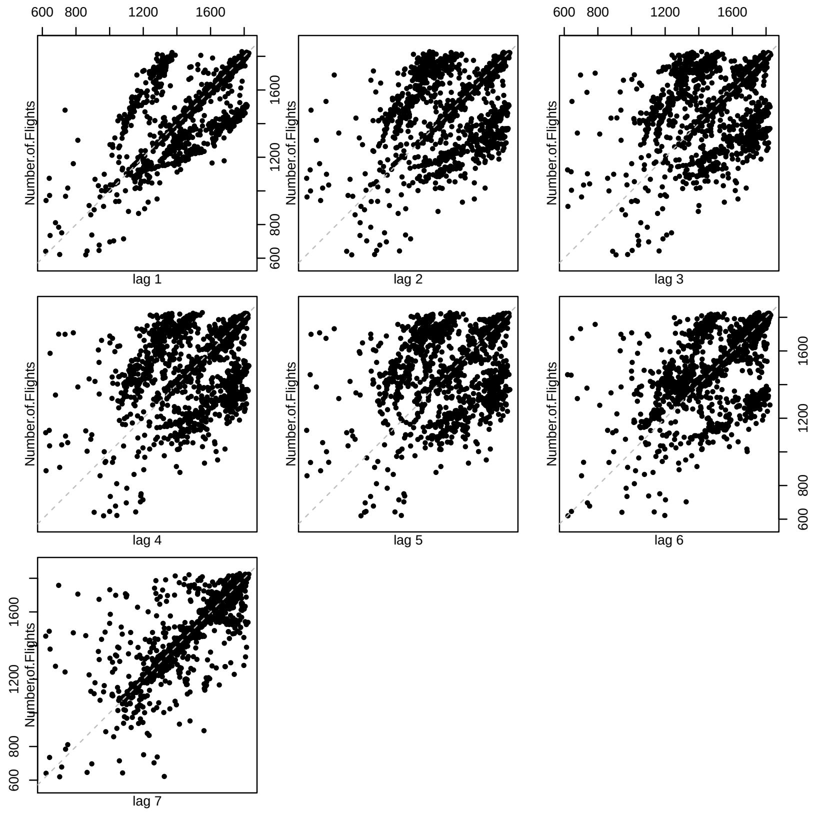

Figure C.3: Lag Plots for the number of daily flights

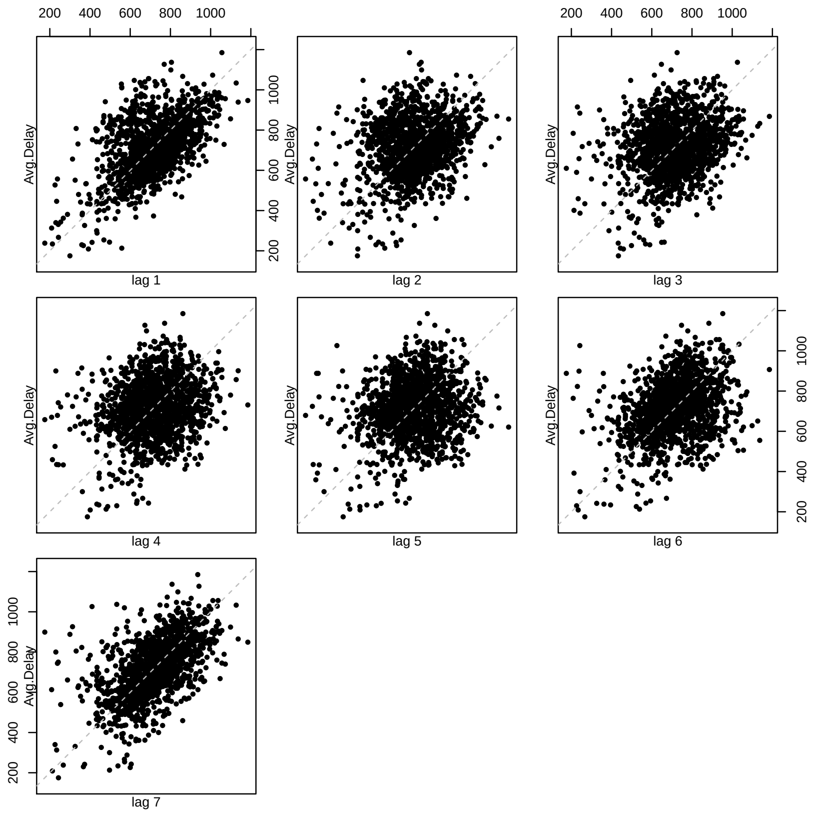

Figure C.4: Lag Plots for the average primary delay per day

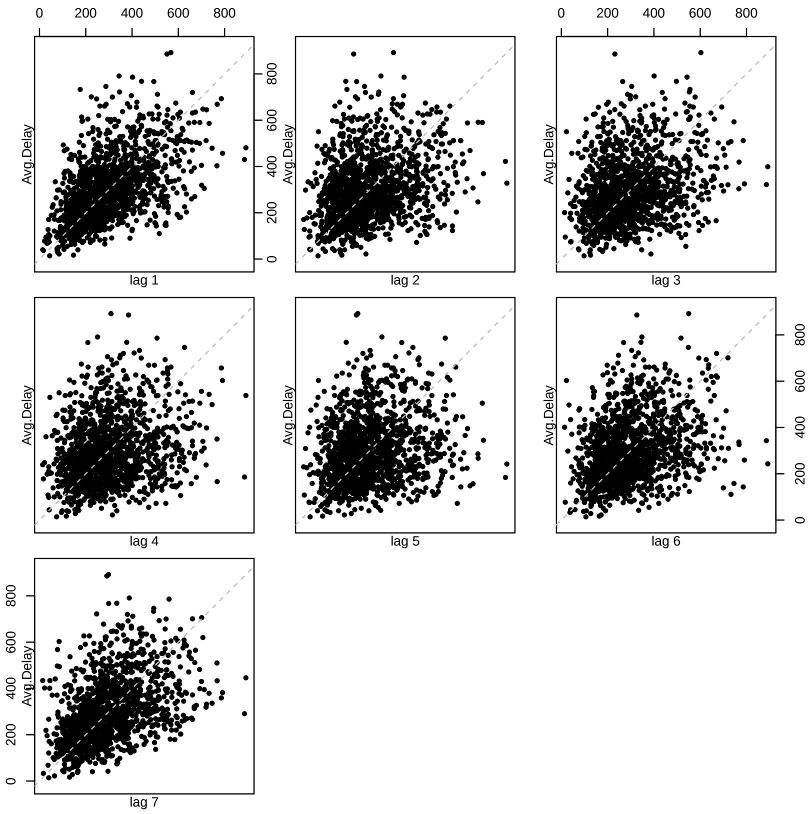

Figure C.5: Lag Plots for the average secondary delay per day

C.3 Estimators for the Determination of Minimum Ground Times

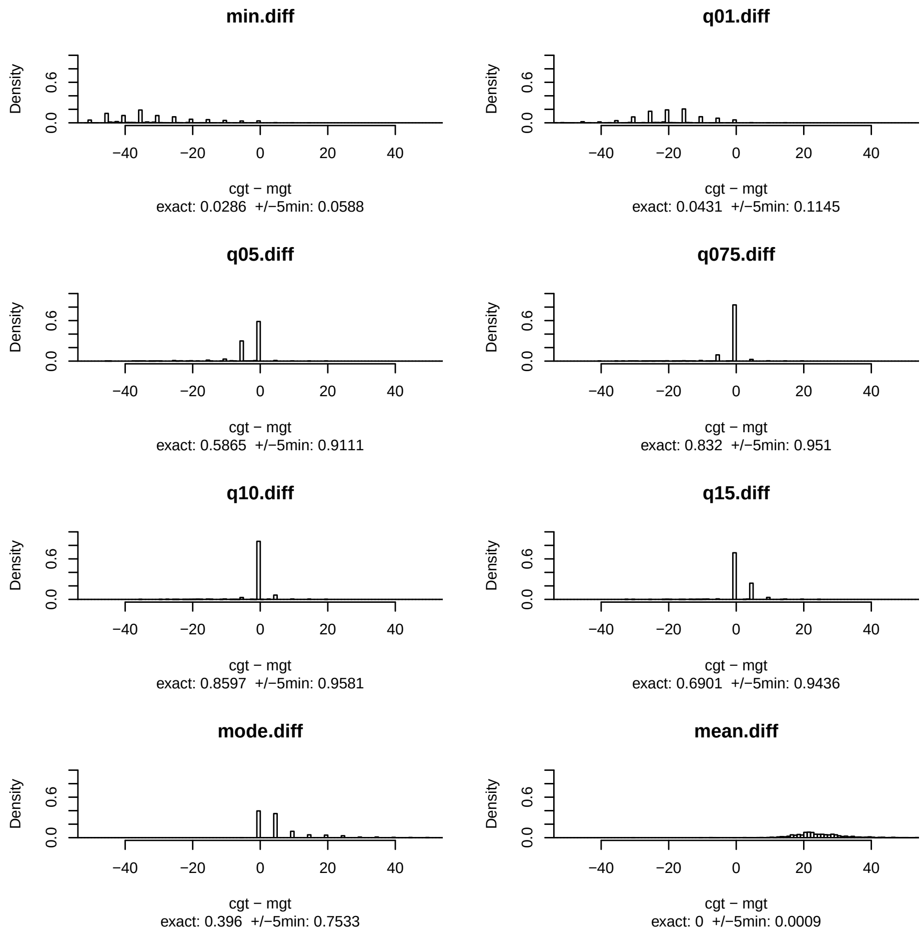

Figure C.6: Histograms for considered mgt estimators

C.4 Additional Results on Under- and Overestimation of Primary Delays in Scheduling

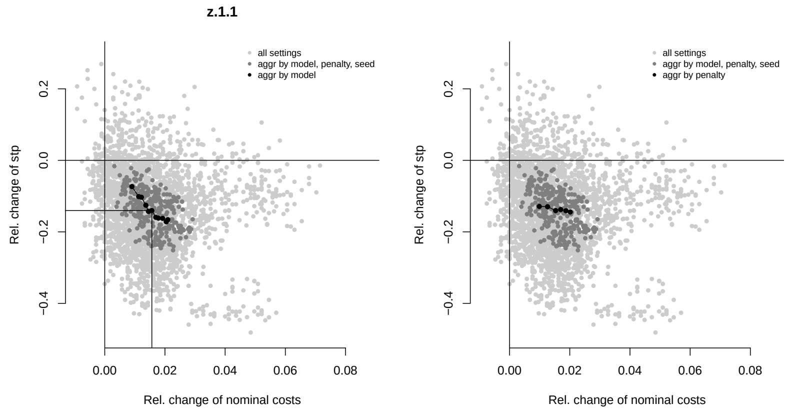

Figure C.7: Relative change of nominal costs and stp when considering one delay scenario in scheduling and simulation

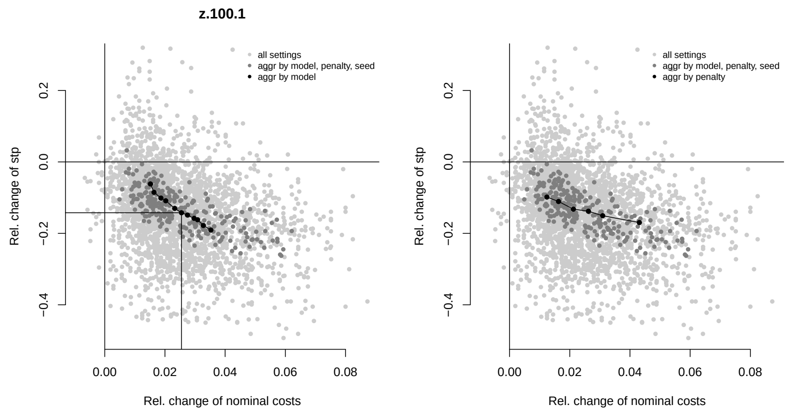

Figure C.8: Relative change of nominal costs and stp when considering 100 delay scenarios in scheduling and one delay scenario in simulation

C.5 Results for Rule-based Recovery in the dm-15 Setting

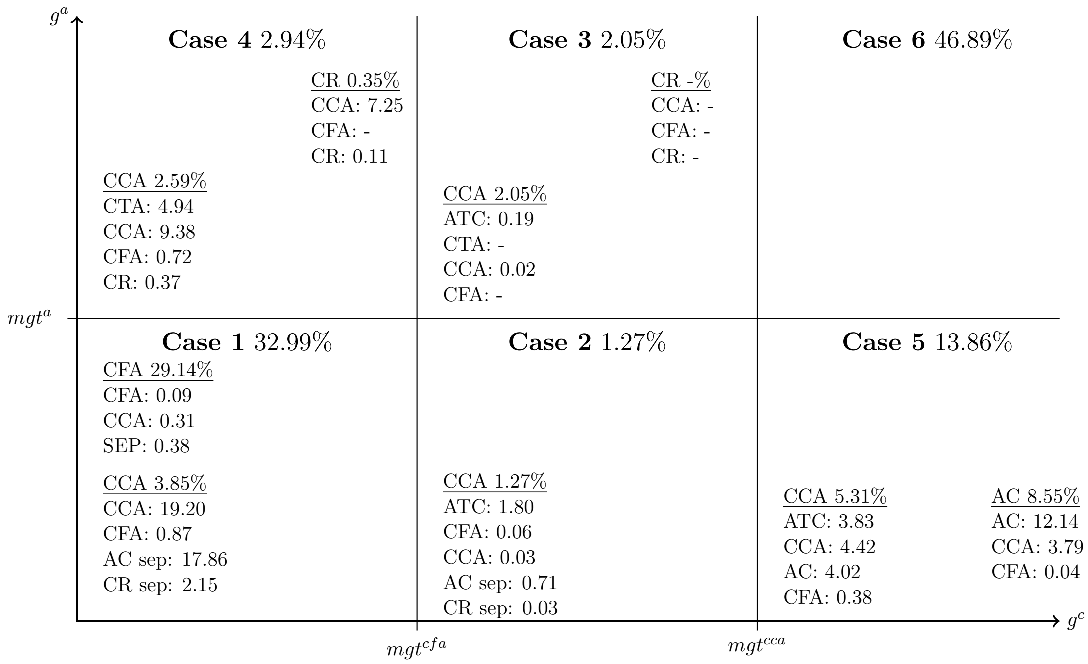

Figure C.9: Frequency of swaps per disruption case, all figures are given in %

C.6 Changes in Nominal Costs and stp Values for Stability- and Flexibility-related Approaches in prop and dm-0 Evaluation

c^Q	Δc_n	Δstp	rtc
0	0.00	0.00	–
10	0.26	-6.69	-25.64
25	0.50	-8.74	-17.58
50	0.99	-10.02	-10.17
100	1.33	-10.88	-8.19
200	1.65	-12.13	-7.36
300	1.82	-12.95	-7.13
400	2.01	-13.43	-6.69
500	2.26	-14.27	-6.31
600	2.60	-14.62	-5.62
750	2.85	-15.71	-5.52
850	3.28	-16.26	-4.95
1,000	3.75	-16.78	-4.47

(A) stb - prop

c^Q	Δc_n	Δstp	rtc
0	0.00	0.00	–
10	0.21	0.36	1.72
50	0.64	0.01	0.01
100	0.92	0.03	0.03
200	1.14	-0.05	-0.04
250	1.20	-0.36	-0.30
500	1.63	-0.82	-0.50
750	1.94	-0.96	-0.49
1,000	2.01	-1.31	-0.65
2,000	2.88	-0.95	-0.33
2,500	3.13	-0.94	-0.30
3,000	3.39	-1.62	-0.48
4,000	3.58	-1.86	-0.52

(B) flx - prop

c^Q	Δc_n	Δstp	rtc
0	0.00	0.00	–
10	0.26	-1.75	-6.70
25	0.50	-4.21	-8.48
50	0.99	-6.09	-6.18
100	1.33	-7.37	-5.55
200	1.65	-9.07	-5.51
300	1.82	-9.54	-5.25
400	2.01	-10.78	-5.37
500	2.26	-11.42	-5.05
600	2.60	-12.03	-4.62
750	2.85	-13.49	-4.74
850	3.28	-14.20	-4.32
1,000	3.75	-15.28	-4.07

(C) stb - dm-0

c^Q	Δc_n	Δstp	rtc
0	0.00	0.00	–
10	0.21	-0.49	-2.35
50	0.64	-2.42	-3.80
100	0.92	-2.62	-2.85
200	1.14	-3.32	-2.91
250	1.20	-3.89	-3.23
500	1.63	-5.07	-3.12
750	1.94	-5.87	-3.03
1,000	2.01	-6.49	-3.23
2,000	2.88	-8.24	-2.86
2,500	3.13	-8.61	-2.76
3,000	3.39	-9.69	-2.86
4,000	3.58	-10.06	-2.81

(D) flx - dm-0

Table C.1: Detailed results for changes in nominal costs and stp values and related robustness-to-cost ratios per setting

List of Publications

- Ionescu, L., Kliewer, N., Schramme, T. (2011). *A comparison of recovery strategies for crew and aircraft schedules.* Operations Research Proceedings 2010, 269-274. Springer, Berlin/Heidelberg.
- Ionescu, L., Kliewer, N. (2011). *Increasing flexibility of airline crew schedules.* Procedia-Social and Behavioral Sciences 20, 1019-1028.
- Dück, V., Ionescu, L., Kliewer, N. and Suhl, L. (2012). *Increasing stability of crew and aircraft schedules.* Transportation Research Part C: Emerging Technologies 20(1), 47-61.
- Ionescu, L., Gwiggner, C., Kliewer, N. (2014). *Empirical and mechanistic models for flight delay risk distributions.* Operations Research Proceedings 2012, 577-582. Springer, Cham.
- Ionescu L., Kliewer N. (2015). *Stability and Flexibility of Airline Resource Schedules.* Proceedings of the 7th Multidisciplinary International Conference on Scheduling: Theory and Applications.
- Ionescu, L., Kliewer, N. (2015). *Delay prediction models for robust airline resource scheduling.* 12th Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP).
- Ionescu, L., Gwiggner, C., Kliewer, N. (2016). *Data analysis of delays in airline networks.* Business & Information Systems Engineering 58(2), 119-133.
- Amberg, B., Ionescu, L., Kliewer, N. (2017). *Robust Efficiency in Public Bus Transport and Airline Resource Scheduling.* Operations Research Proceedings 2015, 259-264. Springer, Cham.
- Ionescu, L., Kliewer, N. (2018). *Examining Delay Propagation Mechanisms for Aircraft Rotations.* Proceedings of Multikonferenz Wirtschaftsinformatik (MKWI), Vol. I, 10-19, Leuphana Universität Lüneburg.

Working Papers

- Ionescu, L., Kliewer, N. (2018). *The Influence of Primary Delay Prediction Models in Robust Airline Resource Scheduling.* Working Paper, Information Systems Department, Freie Universität Berlin, Germany.

- Ionescu, L., Kliewer, N. (2018). *Recovery Strategies for the Evaluation of Schedule Flexibility.* Working Paper, Information Systems Department, Freie Universität Berlin, Germany.

- Ionescu, L., Kliewer, N. (2018). *The Impact of Stability and Flexibility on the Robust Efficiency of Airline Resource Schedules.* Working Paper, Information Systems Department, Freie Universität Berlin, Germany.

Status: March 21$^{\text{th}}$, 2018. Some results of this thesis are part of the publications listed above.