UNIVERSITÄT HAMBURG

UNIVERSITÀ DEGLI STUDI DELL'AQUILA

UNIVERSITAT AUTÒNOMA DE BARCELONA

Erasmus Mundus Consortium
MathMods

Joint Degree of Master of Science in
Mathematical Modelling in Engineering: Theory, Numerics, Applications

In the framework of the
Consortium Agreement and Award of a Joint/Multiple Degree 2013-2019

Master's thesis

The Air Traffic Flow Management Problem under Capacity Uncertainty: A Robust Optimization Approach

Supervisor

Prof. Stefano Smriglio

Candidate

Miriam Sroková
Matricola: 228475

2014/2015

Laurea Magistrale in Ingegneria Matematica
Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica
Università degli Studi dell'Aquila

To my loving parents Eva and Ladislav, and brother Tomáš.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has less than 150 figures.

Miriam Sroková
2015

Acknowledgements

I would like to express my sincere gratitude to my research supervisor Prof. Stefano Smriglio who, as my teacher to start with, introduced me to the fields of Scheduling and Optimization and later on to the research environment. I am very thankful for having a chance to work with him in such a supporting atmosphere. I thank him for his patient guidance, valuable advices and support throughout the development of this thesis.

My special thanks belong to Prof. Bruno Rubino and Prof. Corrado Lattanzio with Dott. Larivera for all their time spent on coordination of MathMods programme and making this all possible. I wish to thank all the professors which I have encountered during my studies and which have been a great motivation to me, especially Prof. Claudio Arbib, Prof. Zbyněk Kubáček and Prof. Vladimír Toma.

I owe a lot to all my friends who shared all those great moments of this journey with me. I am particularly grateful for the assistance and favorable suggestions given by my friend Masoud Ghaderi Zefreh. A special thanks goes to my "L'Aquila family", namely Mariia Dvoriashyna, Diego Pérez and Osvaldo Kim for uncountable BANG! sessions, dinners, coffee breaks, trips and laughs which kept me joyful. I wish to thank my MathMods roommates Ilden Gemil and Milica Tomaševič for all the studying sessions and wonderful conversations, your friendship has been essential and I have learnt a lot from you. I thank the fellow students of my generation: Yuliia, Lena, Alex, Ali, Nauman, Masha, Anirudh, C.K., Sudip, Emanuel, Ankan, Sandeep and the rest of MathMods crew for creating such a stimulating atmosphere and support in both academic and non-academic ways. Finally, I thank my best friend Veronika Horváthová, which has always supported me. Big thanks for being there for me.

The studies in MathMods would have been impossible for me without the great support of my family. I would like to thank my parents for their support and encouragement throughout my study. Thank you for always providing me with the best of opportunities and standing by me in tough times. I thank my brother Tomáš for his unconditional support and love. Without him, this would never be possible. Thank you. Ďakujem.

Abstract

Air Traffic Flow Management (ATFM) represents a set of activities performed by central authorities in order to reduce air traffic delays and costs. The purpose of ground holding policies is to delay departure times of flights if congestion is expected to appear on their route or landing airport. In this thesis, we study the ATFM problem with uncertainty in landing capacities of the airports. We investigate the application of robust and second-stage robust optimization in the nominal formulation of the problem and we formulate the consequent ATFM problem under capacity uncertainty. We present numerical results of the computational experiments on self-produced data sets. Finally, we provide an analysis of possible profitability of this approach. The key observations of our computational results are: a) all the problems were solved in fast computational times; b) the robust solutions preserve integrality properties; c) while applying a second-stage model, the costs reduced significantly.

Key words: robust optimization; two-stage robust optimization; air traffic flow management; ground holding problem

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Flight delays and cancelations occur on a daily basis. Each one of us has most probably experienced a late arrival of flight, or missed a flight connection and therefore is aware of personal consequences. However, while having an impact on social welfare, flight delays increase substantially costs of the air traffic industry and therefore have a considerable impact on the economy. In the United States, flight delays reached an all-time peak in 2007 and induced nationwide costs of over $32,9 billion during that year [35].

Due to the constant growth of the air traffic and limitations of airport capacities, optimal scheduling of flights on overcrowded airports has become an important concern worldwide. Congestion is persistent and arises on an almost daily basis as a consequence of even minor weather disturbances that cause reductions in nominal capacities [21]. Airlines are scheduling more flights than can be handled by busy airports and imbalances between demand and capacity result into the propagation of delays in a network of airports [39]. From the 2014 annual CODA (Central Office for Delay Analysis) report, 34,3% of all the flights in the EUROCONTROL Statistical Reference Area were delayed on arrival, with an average delay time 27,2 minutes per flight [16]. Overall, around 1,5% of the flights were cancelled in European airspace in 2014. Reactionary delay, caused by delay which could not be absorbed on subsequent flight legs, remains the largest single delay group (44.5%) in 2014, followed by delays due to turn round issues (37%) [17]. The situation is similar in the United States. During last decade, the percentage of delayed flights range between 18% and 25%, which highlights that approximately one out of four flights does not arrive on time (Fig. 1.1 Left [39]). Additionally, approximately 2% of flights are cancelled every year (Fig. 1.1 Right).

One of the solutions to the problem of congested networks of airports is to expand the capacity of airports in most affllicted regions. Nevertheless, expansions are costly, complex and controversial. For example, the costs of building Heathrow's North-west

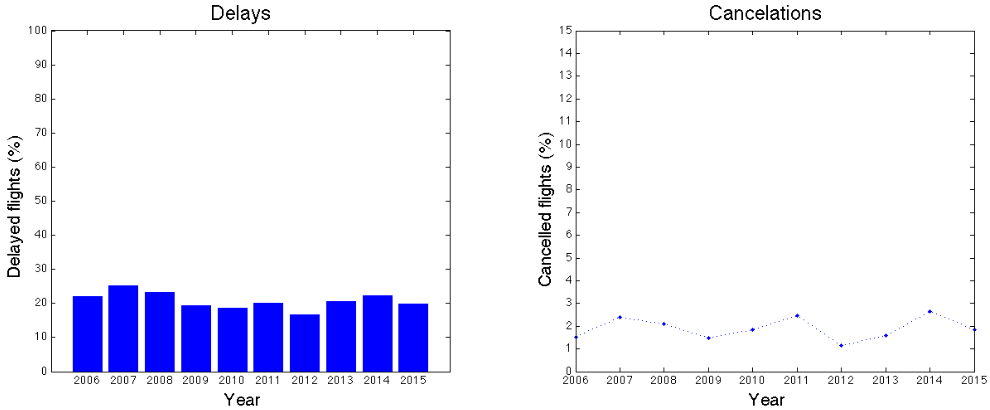

Figure 1.1: On-time performance in the U.S. in 2006-2015.
Left: Percentage of delayed flights; **Right:** Percentage of cancelled flights (Data Source: Bureau of Transportation Statistics [39]).

runaway and new sixth terminal are estimated to be around £18,6 billion; £4 billion higher than Heathrow's own estimate [24]. Another solution is shifting towards the usage of "wide-body" aircrafts. Those are aircrafts with higher capacity of passengers. This technique is slowly, but surely becoming a trend in Chinese and Japanese airline industries. However, the price of this practice is decreasing comfort as well as safety factors. Therefore, there is an observed need for new solution methodologies, which would lower the costs of congestions on the airports.

The majority of the total air traffic delays is caused due to unfavorable weather conditions. In the United States, the official source of the National Airspace System (NAS) air traffic operations and delay data is the Operations Network (OPSNET). The finalized data is freely accessible to general public 20 days after the end of each month and can be downloaded from [3]. As indicated in Fig.1.2 (Left), approximately 60-75% of total delays during the last 15 years are due to weather conditions. Moreover, the average time of delayed flights because of bad weather conditions is higher than the delay caused by other reasons. Therefore, while looking at percentage of delayed minutes, a significant dominance of those caused by weather (around 80%) can be observed (Fig.1.2 Right). This highlights the importance of incorporating weather induced capacity uncertainty in current Air Traffic Flow Management (ATFM) methodologies. The central focus of this thesis is to address exactly this problem: the problem of the ATFM, specifically the issue of managing delays due to dynamic weather conditions and minimizing overall costs. We believe that this can be achieved by proper mathematical framework.

Different measures, at different planning phases, are taken by central authorities (Eurocontrol in Europe) in order to prevent congestion [40]. Our aim in this work is to

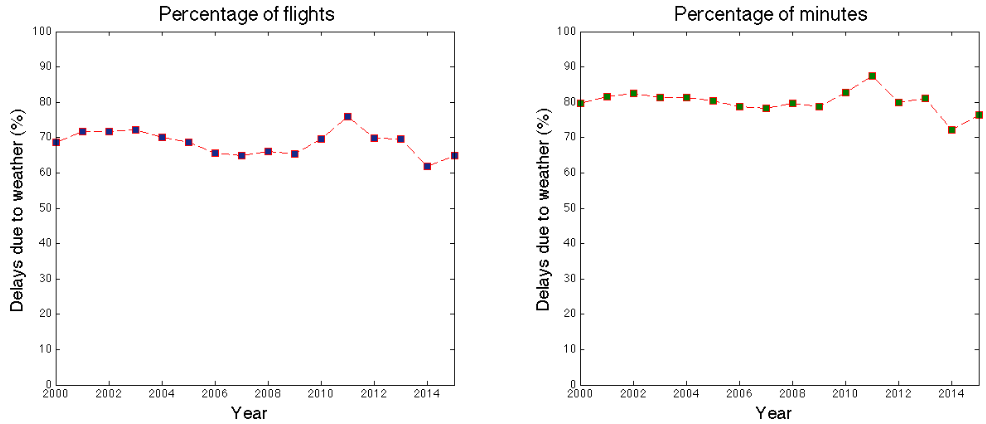

Figure 1.2: Delays caused by weather in the U.S. during the last 15 years. **Left:** Percentage of flights delayed due to weather; **Right:** Percentage of all minutes delayed due to weather (Data Source: Bureau of Transportation Statistics [39]).

focus on the planning phase performed on the day of operations, until the departure of the flight; called tactical planning. During this phase, departure times are assigned to all the flights, while taking into account local regulations and possible capacity reductions of airports caused by bad weather conditions. This is performed by ground holding policies. Their purpose is to delay the departure of a flight (hold it on the ground in the departure airport), whenever the congestion is expected to appear on its route or landing airport. These policies are motivated by the fundamental fact that airborne delays are much costlier than ground delays, because the former include fuel, maintenance, depreciation, and safety costs [8].

The ground holding problem was firstly formulated in mathematical terms by Odoni in 1987 [38]. Since then, several models and algorithms have been proposed to handle different versions of this problem (see [8], [22], [40] or [21]). We provide an overview of various contributions in Section 3.1. However, dealing with uncertain capacities in the network of airports is quite new interest of research community. A first attempt to model the network ATFM problem in a stochastic setting was done by Bertsimas and Gupta in 2011 [41]. The authors proposed a model of weather-induced uncertain capacity as well as tractable solution methodologies for the robust and adaptive ATFM problem. Our work is motivated by their approach of modelling weather-front uncertainty, but we focus more on the applicability of two-stage robust model and analysis of its profitability.

Optimization affected by parameter uncertainty has long been a focus of the mathematical programming community [19]. There are primarily two approaches in the literature to address decision-making under uncertainty, namely, i) Stochastic Programming; and ii) Robust Optimization [41]. In contrast to stochastic programming, whose frameworks are generally intractable on large-scale settings, the main advantage of robust optimization is the tractability of its solutions. The key idea is to capture the probabilistic properties of the problem by constructing relevant uncertainty sets and finding an optimal solution, which stays feasible for any realization of the uncertainty set. In order to fulfill the last property, proposed algorithms are usually dealing with the "worst-case" objective and therefore produce overly conservative solutions (see [5], [20], or [6]). While staying protected under any weather realization of the uncertainty set, the cost of objective may increase significantly and results into unused, but available runway capacities.

Consequently, there is an alternative paradigm for multi-period decision-making called adaptive optimization wherein decisions are adapted to capture the progressive information revealed over time [41]. The approach of implementing the robust frameworks on a rolling horizon basis captures the dynamic behavior of weather and yet follows the worst-case nature. Therefore, there is a guaranteed feasibility of produced schedule, while the costs of objective decrease with high probability. We are focusing on investigation advisability of this approach and possible benefits which it can bring into the air traffic industry.

1.1 Contributions and thesis outline

The contribution of this thesis is the development of two-stage robust optimization model for the nominal formulation of the ATFM problem. The model, which we propose is addressing capacity uncertainty of airports, which is caused by unfavorable weather conditions. As already mentioned, the calamitous weather is liable for majority (around 80%) of the total delayed minutes for the last 15 years. Flight delays and cancelations bring additional costs to the air traffic industry. Our overall aim therefore is to analyze possible profitability of the second-stage robust ATFM model. Our approach minimizes costs of delays and cancelations for all the flights, by allowing changes in schedule on a rolling horizon basis, while staying feasible for bad weather conditions.

A brief summary of all the chapters is as follows:

- **Chapter 2: Mathematical Theory and Background**

  In Chapter 2 we present a general overview of the mathematical theories used in our robust ATFM model. The main focus is on the theory of robust optimization and a historical overview of this concept is included. Furthermore, we introduce key definitions and several approaches to model robust problems. Additionally, the discussion on the influence of various types of uncertainty sets is provided. Also, the notions of right hand side uncertainty and multistage robust optimization are described.

- **Chapter 3: The Air Traffic Flow Management Problem**

  In this chapter, we introduce the mathematical framework of the ATFM problem. We begin with historical overview of the problem and continue by introducing the nominal formulation. Later, we describe the set packing formulation coming from the work of Prof. Smriglio and Prof. Rossi. The key idea of this formulation is that the set packing problem can be solved by solving the associated stable set problem on the intersection graph of a $\{0,1\}$ matrix A. It gives an advantage of solving a MIP problem by LP and under some assumptions assures the integrality of decision variables. Later, we incorporate weather-front induced capacity uncertainty model into both formulations. We show that the single-stage robust ATFM problem can be in both cases transformed into the deterministic one, by taking into account the worst-case values of the right hand side capacity vector. In addition, we propose a two-stage robust formulation of the ATFM problem in the end of this chapter.

- **Chapter 4: Implementation in Practice**

  In this chapter, numerical results of the computational experiments are presented. Firstly, the setup of data and scenario configuration are explained. We utilize twelve data sets of daily flight schedules to present proof-of-concept of the usefulness of the mathematical optimization methodologies. The results were obtained using the open source optimization software GLPK. In order to optimize the computational time and effort, the data was preprocessed accordingly in MATLAB. Later, we report empirical results from the proposed models. The main observations are: a) all the problems were solved in fast computational times; b) the robust solutions preserve integrality properties; c) while applying a second-stage model, the costs reduced significantly.

- **Chapter 5: Future work and conclusions**

  In this chapter, we conclude and present an overview on contributions of this thesis. In addition to this, we provide the discussion on directions for future work, which are as follows:

  - Implementation of the set packing robust model

- Extensions of the formulations
- Computational experiments with real data sets.

Chapter 2

Mathematical Theory and Background

In this chapter, we introduce mathematical theory, which is crucial for development of the air traffic flow management model under capacity uncertainty. We mainly focus on the theory of robust optimization, which we believe is the proper framework to use in order to model this problem. We provide a discussion of key definitions and the influence of various types of uncertainty sets in the robust formulation of the problem. Later, we focus on robust optimization with right hand side uncertainty and robust optimization with uncertainty on the left side of constraints. Additionally, we describe the notion of multistage robust optimization and adaptability, since this framework is going to be implemented in the model proposed in Chapter 3.

2.1 Optimization under uncertainty

Modelling uncertainty has been studied for a long time and it is considered nowadays as a crucial problem in the field of optimization [19]. It may not be obvious from the first sight, but we can say that almost all the optimization problems are influenced by uncertainty up to some degree. According to the size of the influence and its source we can categorize these uncertainties into two groups: (a) microscopic uncertainties, and (b) macroscopic uncertainties. Under microscopic uncertainties we understand uncertainties, which do not have a big impact on the solution of the problem, such as measurement errors and numerical errors. On the other hand, in the group of macroscopic uncertainties are those, which can highly influence behavior of the optimal solution, for instance forecast errors or disturbances in the environment.

In the current literature, there are two major approaches of modelling optimization problems under uncertainty: (a) stochastic optimization, and (b) robust optimization [41]. In stochastic optimization, the data uncertainty is modelled assuming its stochastic nature. More specifically, this usually requires the knowledge of exact distribution of the data and enumeration of scenarios that capture this distribution. A big disadvantage and challenge of this approach is the fact that the size of the problem increases significantly as a function of the amount of scenarios. This results in a lot of cases in computationally expensive problems. On the other hand, in robust optimization, the uncertainty is modelled in a deterministic way. The main idea is to optimize the problem against the worst-case scenario by using a min-max objective. Consequently, according to Laguna [32], there are several advantages in choosing robust optimization instead of stochastic programming:

- Normally, robust optimization finds a solution very near the optimal solutions for all scenarios.
- There is no need of knowing the exact probabilistic distribution for each scenario.
- The worst-case scenario will be feasible for the solution and also let this scenario have a fairly good optimal value.

In our work, we consider the robust optimization as a crucial approach in the development of the ATFM model with uncertain capacities. Hence, we introduce the notion of this framework in the following section.

2.2 Robust optimization

In this section, we firstly focus on introducing the concept of robust optimization. Later, we provide a historical overview of the development of this notion and we introduce several approaches to model robust problems. Various researchers from the field of robust optimization consider the work of Soyster from 1973 [42] as a first introduction into this subject. He has showed, that even small changes of uncertain parameters in the optimization problem can give highly infeasible solutions and therefore from a practical point of view the nominal optimal solution can become completely meaningless. A solution, which was optimal when only considering one scenario could be, and often is, infeasible for all other scenarios that can occur. Therefore, Soyster proposed a solution, which considers all the possible realizations. However, this formulation is considered to be too conservative and leads in most of the cases to a huge optimality loss.

Major development hasn't been seen until the middle of 90s, when various researchers started to emphasize on this problem. A significant step forward for developing a theory for robust optimization was taken independently by Ben-Tal and Nemirovski and El-Ghaoui et al.[13]. In their works, they are dealing with the issue of over-conservatism, more specifically they propose nonlinear, but convex models with ellipsoidal uncertainties. Although their frameworks give less conservative solutions, the cost of this advantage is hidden in computational difficulty (see in [5]). A lot of work has been done during the last decade by various researchers focusing mainly on properties of the solutions and tractability of different formulations.

2.2.1 General definitions

In robust optimization instead of looking for an optimal solution of a nominal deterministic problem, we search a solution that will have an acceptable performance under most realizations of the uncertain inputs of the problem [6]. Usually, no assumptions on the distribution of uncertain parameters are made, but in cases where such a distribution is known, it can be applied beneficially. In general it is considered to be a conservative (worst-case oriented), but practically useful methodology in problems where it is necessary to satisfy constraints "no matter what".

Before stating the definition of robust optimization, we define the notion of uncertainty set and describe its potential structures.

Uncertainty set

Uncertainty set is a set of all the possible realizations of a given event. The structure of an uncertainty set highly influences the existence and tractability of the solution of robust optimization problem. In the current work in the field of robust optimization the researchers are mostly dealing with following types of uncertainty sets [36]:

1. Finite uncertainty: $\Omega = \{\omega_1, \dots, \omega_N\}$
2. Interval-based uncertainty: $\Omega = [\underline{\omega}_1, \overline{\omega}_1] \times \dots \times [\underline{\omega}_M, \overline{\omega}_M]$
3. Polytopic uncertainty: $\Omega = conv\,\{\omega_1, \dots, \omega_N\} = \left\{\sum_{i=1}^{N} \lambda_i \omega_i;\ s.t. \sum_{i=1}^{N} \lambda_i = 1;\ \lambda \,\epsilon\, \boldsymbol{R}_+^N\right\}$
4. Norm-based uncertainty: $\Omega = \left\{\omega \in \boldsymbol{R}^M : \parallel \omega - \hat{\omega} \parallel \leq \alpha\right\}$
5. Ellipsoidal uncertainty: $\Omega = \left\{\omega \in \boldsymbol{R}^M : \sqrt{\sum_{i=1}^{M} \frac{\omega_i^2}{\sigma_i^2}} \leq \Omega\right\}$
6. Constraint-wise uncertainty: $\Omega = \Omega_1 \times \dots \times \mathit{\Omega}_m$, where $\mathit{\Omega}_i$ affects only i-th constraint.

Furthermore, it is important to point out that one uncertainty set can in the same time belong to several types of those mentioned above. Next, we will provide some important mathematical definitions coming from the work of Ben-Tal, El Ghaoui and Nemirovski.

Definition 1: ([6]) An uncertain Linear Optimization problem is a collection

$$\left\{\min_x \left\{c^T x + d \,:\, Ax \leq b\right\}\right\}_{(c,\, d,\, A,\, b)\, \in \mathcal{U}}$$

of LO problems (instances) $\min_x \left\{c^T x + d \,:\, Ax \leq b\right\}$ of common structure (i.e., with common numbers m of constraints and n of variables) with the data varying in a given uncertainty set $\mathcal{U} \subset \boldsymbol{R}^{(m+1)\times(n+1)}$.

The authors assume that the uncertainty set is parameterized by a perturbation vector ζ varying in a given perturbation set $\mathcal{Z}$. It is important to remark that in this formulation all decision variables should be assigned numerical values before the actual data is known. That means, that the decision maker has to make a decision while the knowledge of the actual data is given only by uncertainty set $\mathcal{U}$. In addition, all the constraints in the formulation are "hard" constraints, meaning that non of them can be violated.

Definition 2: ([6]) A vector $x \in \boldsymbol{R}^n$ is a **robust feasible solution** to an uncertain Linear Optimization problem, if it satisfies all realizations of the constraints from the uncertainty set, that is, $Ax \leq b \;\forall\, (c,\, d,\, A,\, b) \in \mathcal{U}$.

Definition 3: ([6]) Given a candidate solution x, the **robust value** $\hat{c}(x)$ of the objective in an uncertain Linear Optimization problem at x is the largest value of the "true" objective $c^T x + d$ over all realizations of the data from the uncertainty set:

$$\hat{c}(x) = \sup_{(c,\, d,\, A,\, b)\, \in \mathcal{U}} \left[c^T x + d\right].$$

In other words, in the group of all the feasible solutions of the problem we want to find the best robust value of the objective function. Next, we provide a definition of robust counterpart of the uncertain linear optimization problem.

Definition 4: ([6]) The **Robust Counterpart** of the uncertain Linear Optimization problem is the optimization problem

$$\min_x \left\{\hat{c}(x) = \sup_{(c,\, d,\, A,\, b)\, \in \mathcal{U}} \left[c^T x + d\right] \,:\, Ax \leq b \;\; \forall (c,\, d,\, A,\, b) \in \mathcal{U}\right\}$$

of minimizing the robust value of the objective over all robust feasible solutions to the uncertain problem. An optimal solution to the Robust Counterpart is called a robust optimal solution to an uncertain Linear Optimization problem, and the optimal value of the Robust Counterpart is called the robust optimal value of the uncertain linear optimization problem.

Definition 5: An optimal (feasible) solution of an uncertain optimization problem is robust if it stays optimal (feasible) under any realization of the uncertainty.

As we have already mentioned, robsut optimization has been studied by several researchers and therefore this problem has been approached from different perspectives with applications in different fields. Next we summarize those, which had most significant impact for the development of this field and those, which we consider important for our ATFM model.

2.2.2 The robust formulation of Soyster

Soyster in his work from 1973 [42] proposes a linear optimization model to construct a solution that is feasible for all the data that belong to a convex set [13]. He formulates the problem as follows:

$$\begin{array}{ll} \text{maximize } \boldsymbol{c'x} & \\ \text{subject to } \sum_{j=1}^{n} \boldsymbol{A}_j x_j \leq b & \quad \forall \boldsymbol{A}_j \in K,\ j = 1, \dots, n \\ \qquad \boldsymbol{x} \geq \boldsymbol{0}. & \end{array}$$

Suppose that the activity vectors $\boldsymbol{a}_j$ are only estimates of the true activities; all that is known with certainty is that the true $j-$th activity vector lies in a hypersphere with center at $\boldsymbol{a}_j$ and radius whose magnitude is ρ_j [42]. Convex uncertainty sets K_j are defined as $K_j = \{\boldsymbol{a} \in \boldsymbol{R}^m\ s.t.\ \parallel \boldsymbol{a} - \boldsymbol{a}_j \parallel \leq \rho_j\}$. This formulation is also known as column-wise uncertainty formulation. In addition, Soyster defines $\bar{a}_{ij} = \sup_{\boldsymbol{A}_j \in K_j}(A_{ij})$ and shows that the problem is equivalent to the following

$$\begin{array}{l} \text{maximize } \boldsymbol{c'x} \\ \text{subject to } \sum_{j=1}^{n} \overline{\boldsymbol{A}}_j x_j \leq b \\ \qquad \boldsymbol{x} \geq \boldsymbol{0}. \end{array}$$

That means, that in order to secure that all the constraints will be fulfilled under any realization, he takes the maximal possible value for each entry of the matrix and solves the problem under these “worst-case” constraints. Despite of the feasibility guarantee for all the uncertainty realizations of the model proposed by Soyster, it produces too conservative solutions. More precisely, Ben- Tal and Nemnirovski point out in their work that due to protection of robustness of the solution there is a huge optimality loss (read more in [5]).

2.2.3 The robust formulation of Ben-Tal and Nemirovski

In the end of 90s Ben-Tal and Nemirovski have dedicated a lot of work on the development of robust optimization frameworks. In the methodology from [5], they introduced two ways of treating the uncertainty: (a) unknown-but-bounded uncertainty and (b) random symmetric uncertainty. Additionally, they proposed the following formulation of robust optimization problem:

$$\begin{array}{lll}
\text{maximize} & \boldsymbol{c}'\boldsymbol{x} & \\
\text{subject to} & \sum_j a_{ij}x_j + \sum_{j\epsilon J_i} \hat{a}_{ij}y_{ij} + & \\
& +\Omega_i\sqrt{\sum_{j\epsilon J_i} \hat{a}_{ij}^2 z_{ij}^2} \leq b_i + \delta \max\left[1, |b_i|\right] & \forall i \\
& \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} & \\
& \boldsymbol{E}\boldsymbol{x} = \boldsymbol{e} & \\
& -y_{ij} \leq x_j - z_{ij} \leq y_{ij} & \forall i,\ j \in J_i \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} & \\
& \boldsymbol{y} \geq \boldsymbol{0}. &
\end{array}$$

In this model, $\varepsilon > 0$ represents magnitude affecting uncertain coefficients, the true values of $\hat{a}_{ij},\ j \in J_i$ belong to interval $[a_{ij} - \varepsilon|a_{ij}|,\ a_{ij} + \varepsilon|a_{ij}|]$ and $\boldsymbol{x}$ must satisfy the $i-$th constraint with an error of at most $\delta \max\left[1, |b_i|\right]$. Moreover, they provide a bound of the probability that the $i-$th constraint is violated by $\exp\left\{-\Omega_i^2/2\right\}$ [5].

2.2.4 The robust formulation of Bertsimas

Bertsimas has done a lot of work in various fields of robust optimization and its applications (see for example [13], [20], or [23]). In [19] he provides together with Brown and Caramanis a nice overview of the previous and current research in this field.

To introduce the notion, they consider an optimization problem with objective function $f_0(\boldsymbol{x})$ and m constraint functions $f_1, \ldots, f_m$ with uncertain parameters $\{\boldsymbol{\omega}_i\}$, $i = 1, \ldots, m$. Namely, $\boldsymbol{x} \in \boldsymbol{R}^N$ is a vector of decision variables, $f_0, f_1, \ldots, f_m : \boldsymbol{R}^N \to \boldsymbol{R}$ are functions and $\boldsymbol{\omega}_i \in \boldsymbol{R}^k$ are uncertainty parameters taking values in uncertainty sets $\boldsymbol{\Omega}_i \in \boldsymbol{R}^k$. As we know, the purpose of all constrained optimization problems is to find an optimal solution to a given objective function fulfilling all of the constraints. In the robust optimization, we are looking for a minimum cost solution $\boldsymbol{x}^*$ which is feasible for all realizations of the uncertainties $\boldsymbol{\omega}_i$ within $\boldsymbol{\Omega}_i$. This leads to mathematical formulation given by

$$\begin{aligned} &\text{minimize } f_0(\boldsymbol{x}) \\ &\text{subject to } f_i(\boldsymbol{x}, \boldsymbol{\omega}_i) \leq 0 \qquad \forall \boldsymbol{\omega}_i \in \Omega_i,\ i = 1, ..., m. \end{aligned} \tag{2.1}$$

The authors consider the uncertainty sets to be always closed sets and they point out that intuitively, if some of the $\boldsymbol{\Omega}_i$ are continuous sets, the problem (2.1) would end up having an infinite number of constraints. Further, they point out some straightforward observations, which hold for this statement of the problem:

- The fact, that the objective function is unaffected by parameter uncertainty is without loss of generality; we may always introduce an auxiliary variable, call it t, and minimize t subject to the additional constraint $\max_{\omega_0 \in \Omega} f_0(\boldsymbol{x}, \boldsymbol{u}_0) \leq t$ [19].

- The assumption of constraint-wise uncertainty set is as well without loss of generality.

- Moreover, deterministically given constraints (those, which do not include uncertainties) are included in this model by assuming the corresponding Ω_i to be singletons.

- This problem statement also contains the instances when the decision or disturbance vectors are contained in more general vector spaces than $\boldsymbol{R}^n$ or $\boldsymbol{R}^k$ (e.g., $\boldsymbol{S}^n$ in the case of semidefinite optimization) with the definitions modified accordingly [20].

We have already mentioned before some general definitions of robust linear optimization, its feasible solution and robust counterpart. Now, we will provide a little intuition on them. We begin by considering a regular constrained linear optimization problem:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{c^T x} \\ \text{subject to} \quad & \boldsymbol{Ax} \leq \boldsymbol{b} \end{aligned}$$

In ordinary problems, parameters $(\boldsymbol{c}, \boldsymbol{A}, \boldsymbol{b})$ are known deterministically, that means, that in general linear optimization problem we would know their exact values. Moreover, we know, that if the objective function is linear and feasible region is convex, we will always be able to find an optimal solution to this problem and it will lie on the boundary of feasible region. Intuitively, if we would change parameters $(\boldsymbol{c}, \boldsymbol{A}, \boldsymbol{b})$ only a little bit, we would get another, different feasible set and the previous solution would probably become infeasible. By its nature then, the solution is not designed to be robust in perturbations in the feasible set [15].

Let us investigate now what would happen, if $\boldsymbol{c}$, $\boldsymbol{A}$ and $\mathbf{b}$ would be uncertain. For doing so, we introduce uncertainty set Ω and a new parameter $\boldsymbol{\omega} \in \Omega$, which will capture the uncertainty. Now, we are able to look at the parameters $\mathbf{c}, \mathbf{A}, \boldsymbol{b}$ as functions of $\boldsymbol{\omega}$: $\boldsymbol{c}(\boldsymbol{\omega}), \boldsymbol{A}(\boldsymbol{\omega}), \boldsymbol{b}(\boldsymbol{\omega})$. We assume that Ω is a bounded set of all possible outcomes of uncertainty and it is known a priori. Without loss of generality, we conclude that parameters $\boldsymbol{c}$ and $\boldsymbol{b}$ are known deterministically, since we can always transform the problem by for instance adding an additional variable for both objective function and vector $\boldsymbol{b}$. We are looking for a solution that will stay feasible for any realization $\boldsymbol{\omega} \in \boldsymbol{\Omega}$. Under these assumptions our problem will become:

$$\begin{array}{lll} \text{minimize} & \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x} & \\ \text{subject to} & \boldsymbol{A}(\boldsymbol{\omega}) \leq \boldsymbol{b} & \forall \boldsymbol{\omega} \in \boldsymbol{\Omega}. \end{array}$$

Following lemma was stated by Constantine Caramanis in his PhD work "Adaptable Optimization: Theory and Algorithms". It provides an obvious, but important conclusion.

Lemma: ([15]) Let the rows of the matrix $\boldsymbol{A}(\boldsymbol{\omega})$ be denoted by $\boldsymbol{a}_i(\boldsymbol{\omega})$. If $\boldsymbol{A}$ has m rows, make m copies of Ω, so that $\Omega^{(i)} = \Omega$. Then the optimization problem is equivalent to the following formulation:

$$\begin{array}{lcl} \min: & \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x} & \\ s.t.: & \boldsymbol{a}_1(\boldsymbol{\omega}^{(1)})^T\boldsymbol{x} \leq b_1 & \forall \boldsymbol{\omega}^{(1)} \in \Omega^{(1)} \\ & \vdots & \\ & \boldsymbol{a}_m(\boldsymbol{\omega}^{(m)})^T\boldsymbol{x} \leq b_m & \forall \boldsymbol{\omega}^{(m)} \in \Omega^{(m)}. \end{array}$$

Clearly, $\boldsymbol{a_i}(\boldsymbol{\omega_i})^{\boldsymbol{T}}\boldsymbol{x} \leq b_i, \forall \boldsymbol{\omega_i} \in \Omega_i$ if and only if $\max_{\{\omega_i \in \Omega_i\}} \boldsymbol{a_i}(\boldsymbol{\omega_i})^{\boldsymbol{T}}\boldsymbol{x} \leq b_i, \forall i$. As a consequence we obtain a subproblem of robust optimization problem, or in literature also called the inner problem of robust optimization:

$$\left[\begin{array}{c} \max \boldsymbol{a_i}(\boldsymbol{\omega})^{\boldsymbol{T}}\boldsymbol{x} \\ s.t. \, \boldsymbol{\omega} \in \Omega \end{array}\right] \leq b_i.$$

Let us remark, that the complexity and solvability of any robust optimization problem is determined by the nature of its inner problem [15]. Hence, it is highly convenient to replace this subproblem by its dual counterpart. The structure of the dual problem is determined by both the design of the constraints as well as uncertainty set. Therefore the structure of uncertainty set and the structure of the robust optimization problem are highly correlated. If we want the robust optimization problem to be solvable, it is neccessary to build an inner problem in a solvable form. This condition however requires a simple geometrical structure of the uncertainty set Ω. On the contrary, building a too simple uncertainty set may end up in loosing either robustness or optimality of the solution. To follow we discuss deeper some of the structural properties of uncertainty sets and robust optimization problems related to those sets.

Ellipsoidal Uncertainty

Ben- Tal and Nemirovski provide in [9] several arguments why it is reasonable to choose an "ellipsoidal" uncertainty set. They define an ellipsoid in $\boldsymbol{R}^k$ as a set of the form

$$U = \left\{ \prod(u) \mid || Qu || \leq 1 \right\},$$

where $u \to \prod(u)$ is an affine embedding of certain $\boldsymbol{R}^L$ into $\boldsymbol{R}^k$ and Q is an $M \times L$ matrix.

In addition, they claim that $\Omega \in \boldsymbol{R}^{m \times n}$ is an ellipsoidal uncertainty set if it fulfills following conditions:

- Ω is given as an intersection of finitely many ellipsoids
- Ω is bounded
- ("Slater condition") there is at least one matrix $A \in \Omega$ which belongs to the "relative interior" of every ellipsoid $\Omega(\prod_l, Q_l)$, $l = 1, \ldots, k$:

  $$\forall l \leq k \quad \exists \omega_l : \quad A = \prod_l (\omega^l) \ \& \ || Q_l \omega_l || < 1.$$

  [9]

The subproblem of robust optimization problem with ellipsoidal uncertainty yields a maximization over quadratically defined set. As a consequence, the resulting dual problem is not linear [15].

Theorem 1: ([9]) The robust counterpart of an uncertain LP problem with general ellipsoidal uncertainty can be converted to a conic quadratic program.

Curious reader can find the proof of this theorem in the appendix of [9]. Conic quadratic problems can be solved by polynomial time interior point methods at basically the same computational complexity as LP problems of similar size [9].

Polyhedral uncertainty

Polyhedral uncertainty can be viewed as a special case of ellipsoidal uncertainty [9]. The use of polyhedral uncertainty sets was used with great success in Bertsimas and Sim, as in the case where the uncertainty affects the constraints in an affine manner, the dual to the subproblem is again a linear program [15]. To explain this, the authors consider the uncertainty sets directly in the parameters of the uncertain matrix $\boldsymbol{A}$, i.e.:

$$\Omega_i = \{\boldsymbol{a}_i \,:\, \boldsymbol{D}_i \boldsymbol{a}_i \leq \boldsymbol{d}_i\}\,.$$

Afterwards they rewrite the problem as:

$$\begin{aligned} &\text{minimize } \boldsymbol{c^T x} \\ &\text{subject to } \max_{\{\boldsymbol{a}_i \in \Omega_i\}} \boldsymbol{a}_i^T\, \boldsymbol{x} \leq b_i \qquad i = 1, ..., m. \end{aligned}$$

From the theory of duality we know that we can rewrite the dual problem to the inner problem as:

$$\begin{aligned} \text{minimize } & \boldsymbol{p_i^T d_i} \\ \text{subject to } & \boldsymbol{p_i^T D_i = x} \\ & \boldsymbol{p_i \geq 0.} \end{aligned}$$

Therefore we can formulate the robust optimization problem with polyhedral uncertainty set as:

$$\begin{aligned} \text{minimize } & \boldsymbol{c^T x} \\ \text{subject to } & \boldsymbol{p_i^T d_i} \leq b_i && i = 1, ..., m \\ & \boldsymbol{p_i^T D_i = x} && i = 1, ..., m \\ & \boldsymbol{p_i \geq 0} && i = 1, ..., m. \end{aligned}$$

Thus the size of such problems grows polynomially in the size of the nominal problem and the dimensions of the uncertainty set [19].

Cardinality constrained uncertainty

Bertsimas and Sim [13] in their work from 2002 presented "The New Robust Approach". This was a crucial approach in terms of applications of robust optimization into discrete optimization problems. The main trick is that they consider such a family of polyhedral uncertainty sets, which controls the amount of uncertain parameters, i.e. parameters that are allowed to change their nominal values.

We describe their formulation as follows. Let $\boldsymbol{a'}_i\, \boldsymbol{x} \leq b_i$ be the $i-$th constraint of the nominal robust problem, where $\boldsymbol{a}_i$ represents the $i-$th row of the matrix $\boldsymbol{A}$. For every

i, $i = 1, ..., m$ they introduce J_i, the set of coefficients a_{ij}, $j \in J_j$ that are subject to parameter uncertainty and take values according to a symmetric distribution with mean equal to the nominal value $\boldsymbol{a}_{ij}$ in the interval $[a_{ij} - \hat{a}_{ij},\ a_{ij} + \hat{a}_{ij}]$ [13]. Let $\mathit{\Gamma}_i \in [0,\ |J_i|]$ be a parameter adjusting the robustness. In reality, it doesn't happen often that all of the parameters a_{ij}, $j \in J_i$ are going to change. Our goal is to be protected against all cases that up to $\lfloor \Gamma_i \rfloor$ of these coefficients are allowed to change, and one coefficient a_{it} changes by $(\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it}$ [13]. In fact the authors consider following nonlinear formulation of the problem:

$$
\begin{array}{rcl}
& \text{maximize } \boldsymbol{c^T x} & \\
& \text{subject to } \sum_j a_{ij}x_j + & \\
+ \max_{\{S_i \cup \{t_i\} |\, S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor,\, t_i \in J_i \S i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} y_j + (\Gamma_i - \lfloor \Gamma_i \rfloor) \hat{a}_{it_i} y_t \right\} & \leq b_i & \forall i \\
& -y_j \leq x_j \leq y_j & \forall j \\
& \boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u} & \\
& \boldsymbol{y} \geq \boldsymbol{0}. &
\end{array}
\tag{2.2}
$$

Obviously, if $\mathit{\Gamma}_i = 0$ the problem reduces to the nominal form, and if $\mathit{\Gamma}_i$ is an integer, the $i-$th constraint will be in the form $\sum_j a_{ij}x_j + \max_{\{S_i \cup \{t_i\} |\, S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} y_j \right\} \leq b_i$. In addition, if $\mathit{\Gamma}_i = |J_i|$ the formulation will be equivalent to that provided by Soyster. Relaxing and taking the dual of the inner maximization problem, one can show that the above is equivalent to the following linear formulation, and therefore is tractable (and, moreover, is a linear optimization problem) [19].

Theorem: ([13]) Model (2.2) has an equivalent linear formulation as follows:

$$
\begin{array}{rl}
\text{maximize } \boldsymbol{c^T x} & \\
\text{subject to } \sum_j a_{ij}x_j + z_i\Gamma_i + \sum_{j \in J_i} p_{ij} \leq b_i & \forall i \\
z_i + p_{ij} \geq \hat{a}_{ij} y_j & \forall i,\ j \in J_i \\
-y_j \leq x_j \leq y_j & \forall j \\
l_j \leq x_j \leq u_j & \forall j \\
p_{ij} \geq 0 & \forall i,\ j \in J_i \\
y_j \geq 0 & \forall j \\
z_i \geq 0 & \forall i.
\end{array}
\tag{2.3}
$$

2.3 Robust discrete optimization

There has been some work done regarding the issue of extending robust otpimization formulations into discrete optimization problems. Kouvelis and Yu [30] proposed a framework, which minimizes the worst case scenario. Unfortunately, under their approach, the robust counterpart of many polynomially solvable discrete optimization problems becomes NP-hard [11]. Bertsimas and Sim present in their paper [11] a framework, which extends (2.3) into a discrete setting and they propose and proove its MIP formulation:

$$
\begin{array}{rll}
\text{maximize } & \boldsymbol{c^T x} + z_0\Gamma_0 + \sum_{j\in J_0} p_{0j} & \\
\text{subject to } & \sum_j a_{ij}x_j + z_i\Gamma_i + \sum_{j\in J_i} p_{ij} \le b_i & \forall i \\
& z_i + p_{0j} \ge d_j y_j & \forall j \in J_0 \\
& z_i + p_{ij} \ge \hat{a}_{ij} y_j & \forall i \ne 0,\ j \in J_i \\
& p_{ij} \ge 0 & \forall i,\ j \in J_i \\
& y_j \ge 0 & \forall j \\
& z_i \ge 0 & \forall i \\
& -y_j \le x_j \le y_j & \forall j \\
& l_j \le x_j \le u_j & \forall j \\
& x_i \in \boldsymbol{Z} & \forall i = 1, ..., k.
\end{array}
\tag{2.4}
$$

2.4 Multistage robust optimization

Multistage robust optimization (in literature also adjustable or adaptable robust optimization) is a special approach to model adjustability of the solution under uncertain data. The main idea is to divide the decision process in multiple stages, under assumption that in some of the real world problems the decision maker is provided by updated information on uncertainty while the time passes. More specifically, in the beginning, he gets the first information about the structure of uncertainty set and later he gets a new, updated information and this can happen multiple times. General assumption is that in most of the cases the later information is the better one, due to a better awareness of the situation.

In the previous text we were studying the structure of problems where the decision is being made only once - single stage (or static) problems. This type of problems lead to solutions, which are protected over all of the realizations of uncertainty, but the trade-off for this is a too high conservatism of an optimal value. If we would take into account that the decision maker will be provided by another, updated versions of the uncertainty, we could get a better solution (perhaps at some computational cost). This gives the possibility to select in the beginning k contingency plans ($k \in \boldsymbol{Z}^+$) for the second stage solution, in which all the cases of the realization of uncertainty would be covered. Then, in the second stage, we can neglect some of those plans or possibly choose the one, which protects the updated uncertainty set.

In literature ([15], [41]), the authors distinguish two types of adaptability: (a) complete adaptability and (b) finite adaptability. Completely adjustable problems are such problems, for which the realization of the uncertainty set is known completely and therefore in the second stage the decision maker is provided by the true information. The completely adaptable formulation is known to be NP-hard to solve in general ([15]). This, however, is in reality almost never the case. In the later stages of the decision we are usually able to predict better the future outcome of uncertainty, but we can almost never say with complete knowledge what is going to happen. There has been some work done on the topic of finite adaptability and it is known that under some assumptions, this kind of problems is polynomially solvable.

In the finite k-adaptability problem the decision maker chooses in the second stage k solutions $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$ which cover all the realizations of uncertainty. Then, after knowing the realization, he is sure that one of $\boldsymbol{y}_i$, $i = 1, ..., k$ is feasible and he chooses the best one. Bertsimas and Caramanis formulate this second stage robust optimization problem followingly:

$$
\begin{array}{lll}
\min: & \boldsymbol{c}^T\boldsymbol{x} + \max\left\{\boldsymbol{d}^T\boldsymbol{y_1}, ..., \boldsymbol{d}^T\boldsymbol{y_k}\right\} & \\
\text{s.t.:} & \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{y_1} \geq \boldsymbol{b} & \forall\omega \in \Omega \\
& \quad\quad\quad\quad or & \\
& \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{y_2} \geq \boldsymbol{b} & \forall\omega \in \Omega \\
& \quad\quad\quad\quad or & \\
& \quad\quad\quad\quad \vdots & \\
& \quad\quad\quad\quad or & \\
& \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{y_k} \geq \boldsymbol{b} & \forall\omega \in \Omega.
\end{array}
\tag{2.5}
$$

Furthermore, they prove that in case of polyhedral uncertainty set the k-adaptability problem becomes a k-partition problem, where the decision maker is looking for an optimal partition of the uncertainty set Ω into k subsets such that $\Omega = \Omega_1 \cup ... \cup \Omega_k$. The authors provide and prove that the previous formulation is equal to (see section 3.2.1 in [15]):

$$
\min_{\Omega = \Omega_1 \cup ... \cup \Omega_k}
\begin{bmatrix}
\min: & \boldsymbol{c}^T\boldsymbol{x} + \max\left\{\boldsymbol{d}^T\boldsymbol{y_1}, ..., \boldsymbol{d}^T\boldsymbol{y_k}\right\} & \\
\text{s.t.:} & \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{y_1} \geq \boldsymbol{b}, & \forall(\boldsymbol{A}, \boldsymbol{B}) \in \Omega_1 \\
& \vdots & \\
& \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{y_k} \geq \boldsymbol{b}, & \forall(\boldsymbol{A}, \boldsymbol{B}) \in \Omega_k.
\end{bmatrix}
$$

Additionally, the authors propose a bilinear optimization formulation to compute an optimal 2-adaptabillity value ($k = 2$) and the optimal two contingency plans Ω_1, Ω_2. However, we will focus more on different approach of modeling adaptability of robust problems - affine policies. This is mainly due to future application of affine policies (in literature also called Linear Decision Rules) to our ATFM model. The other reason is that in most of the practical cases the adjustable robust counterpart is computationally intractable. This difficulty can be addressed by restricting the adjustable variables to be affine functions of the uncertain data [7].

2.4.1 Affine Policies

Ben- Tal et al. distinguish in their work [7] two types of variables: (a) "here and how" and (b) "wait and see". "Here and how" variables are those for which any knowledge of the realization of the uncertainty does not give extra value in the decision process. In the majority of real-world optimization problems only some of the variables belong to the first class. In contrast, "wait and see" variables are those, which can be made after some part of uncertainty is released and therefore they can adjust themselves to the new situation. Variables belonging to the group (a) are also called non-adjustable variables and those belonging to (b) adjustable variables.

We divide the vector of decision variables $\boldsymbol{x}$ into two sub-vectors $\boldsymbol{x} = (\boldsymbol{u}^T, \boldsymbol{v}^T)^T$, where $\boldsymbol{u}$ represents non-adjustable and $\boldsymbol{v}$ adjustable variables. Then, according to formulation of Ben-Tal [7], adjustable robust counterpart becomes:

$$\min_{(s,u,v)} \left\{ s : c^T \begin{pmatrix} u \\ v \end{pmatrix} \leq s,\ Au + Bv \leq b \right\}_{[A,B,b,c] \in \Omega}.$$

The authors follow by assuming the uncertainty set to be affinely parametrized by a "vector of perturbations" ξ, which belongs to a nonempty convex compact perturbation set, so that

$$\Omega = \left\{ [A,\, B,\, b] = \left[A^0,\, B^0,\, b^0\right] + \sum_{t=1}^{T} \xi_t \left[A^t,\, B^t,\, b^t\right] \right\}.$$

This assumption allows them to provide a restriction on the adjustable variables to be affine functions of ξ (for more explanation see in [7]):

$$\boldsymbol{v} = \boldsymbol{v}(\xi) = \boldsymbol{v}^0 + \sum_{t=1}^{T} \xi_t \boldsymbol{v}_t.$$

Hence, the adaptive problem with affine recourse enforced a priori becomes:

$$\begin{array}{lll} \min_{\boldsymbol{u},\boldsymbol{v^0},\boldsymbol{v_t}} & \left\{\boldsymbol{c^T u} + \max_{\xi_t\in\Omega} \boldsymbol{d^T(v^0} + \sum_{t=1}^{T} \boldsymbol{\xi_t v_t)}\right\} & \\ \text{subject to} & \boldsymbol{Au} + \boldsymbol{B}(v^0 + \sum_{t=1}^{T} \xi_t \boldsymbol{v}_t) \le \boldsymbol{b} & \forall A,\ B,\ ,b,\ \xi \in \mathit{\Omega}. \end{array}$$

Furthermore, they provide a broad discussion on computational tractability of various cases of perturbation sets (see [7]).

Bertsimas and Maes show that a two-stage adaptive optimization problem with affine recourse can be converted to a single deterministic linear program [10] . For our future work, it is important to remark, that linear decision rules are optimal only in very rare occasions. In fact the main motivation for the use of linear decision rules is its tractability. In reality, it is not always possible to compute the optimal solution by any other approach, and therefore having a tool computing fairly good approximation is crucial.

Now, we would like to focus on a class of problems for which the uncertainty occurs only on the right hand side of constraints. This class is called right hand side uncertainty problems.

2.5 Right hand side uncertainty

Many real world optimization problems contain uncertainties in the right hand side of the constraints (for example uncertainties in capacities, demand). Firstly, let us discuss the single stage (static) robust optimization problem with right hand side (RHS) uncertainty, that is:

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c^T x} \\ \text{subject to} & \boldsymbol{A_i x} \le \boldsymbol{b_i(\omega)} \quad \forall \omega \in \Omega,\ i = 1, ..., m. \end{array}$$

If Ω is a bounded set, then intuitively for every row i we can find $\hat{b}_i = \min\, b_i(\omega)$ and clearly, the previous formulation is equal to:

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c^T x} \\ \text{subject to} & \boldsymbol{A_i x} \le \hat{b}_i \quad i = 1, ..., m. \end{array}$$

A lot of work on 2-stage robust linear programming with right hand side uncertainty has been done by Minoux and is presented in [37]. Here he discusses complexity results and applications of this framework into real-world problems. He defines a class of linear problems with “Finite Pollynomially Bounded Uncertainty Set” (FPBU) and proves the following proposition:

Proposition: ([37]) All problem instances in $R - LP - RHSU - FPBU$ are polynomially solvable.

Here R stays for robust, intuitively LP means linear programming and $RHSU$ stands for right hand side uncertainty. Proof of this proposition can be found in [37]. From the perspective of application, it is important to point out that FPBU includes the class of problems for which the cardinality of their uncertainty sets is bounded by some K and therefore this set of problems is polynomially solvable under $R - LP - RHS$ as well.

Chapter 3

The Air Traffic Flow Management Problem

During last decades, the air traffic industry has been growing continuously and hence the air system's infrastructure has been under a constant pressure. Congestion phenomena are persistent and arise almost on a daily basis as a consequence of bad weather conditions which cause sudden capacity reductions [18]. Most of the delays are created by imbalancies between demand and capacity resulting from airlines scheduling more flights than available capacity at busy airports and by the propagation of delays in a network of airports [29]. In Europe, around 34% of all the flights were delayed on arrival and additional 1,5% was cancelled in 2014 [16]. The situation is similar in the United States, where flight delays reached an all-time peak in 2007 and induced nationwide costs of over $32,9 billion during that year [35]. Therefore, there is an evident need for solution methodologies, which would decrease the costs of congestions on the airports.

Suppose that a flight f is flying from a place A to B and it is scheduled to land in B at time t. If the decision maker at the time of its departure knows, that the landing capacities will be violated in airport B at t, meaning that there will be no landing spot for f in B at t, it is cheaper to hold f in A and wait until a free landing spot in B at some $t + k$ appears. This procedure is called Ground Holding and it has been studied by several researchers over past decades. The ground holding policies are nowadays widely used in order to minimize the costs of congestions and airborne delays, since it is much safer and cheaper to hold an aircraft on the ground than in the airspace. The main difficulty of this problem is to assign ground delays optimally.

3.1 History and overview of the problem

The air traffic flow management problem has been studied for almost three decades. A pioneer work on this topic was presented by Odoni in 1987 (see [38]). His main idea was to develop a real time flights scheduling model in order to minimize congestion costs in the air traffic industry. The concept of this work was essential for the models, which were developed afterwards. Through the continuous development of air transportation industry, different versions of ATFM problem have been studied and different types of models have been proposed.

Firstly, Terrab and Odoni [43] were dealing with the Single-Airport Ground-Holding Problem (SAGHP) as the simplest methodology, proposing an optimal schedule for the airport while taking into account limitations of possible landing and departing operations for every unit of time. Strategies coming from models of Odoni and Bianco (see [14] and [33]) were applied at some Italian airports. Notable research on SAGHP has been conducted by the Institute of Flight Guidance for Airports in Germany (see [44]) and Bianco et al. in [31]. Hoffman and Ball proposed deterministic formulations, which have been efficiently implemented in the USA [28]. In most of these SAGHP approaches, arrival and departures are treated as independent variables [4]. That means, that the number of flights assigned to take off does not depend on the number of flights, which are assigned to arrive in the same time period, which does not describe reality and can result in very conservative solutions. Another disadvantage of these models is that the capacities are assumed to be deterministically known in advance. In reality, this is not always true due to the number of possible unpredictable events, which may result in capacity reductions.

Subsequently, the Multiple-Airport Ground-Holding Problem (MAGHP) in the air traffic industry started to be discussed. In 1993, Vranas, Bertsimas and Odoni formulated and studied problem of assigning ground-holding delays optimally in general network of airports [8]. They presented several integer programming formulations of the problem, which minimize the total delay cost of all flights in the network. Also, they consider the departure and arrival capacities of airports as deterministic functions of time and space. Further, the sectors of the airspace are assumed to have unlimited capacities, meaning that no routings of the flights are considered. Moreover, they take into account successive flights performed by the same aircraft and they are allowing canellations of flights, in a way that there are no upper bounds on the delay, nor on ground holding.

As a crucial point in the development of ATFM is considered the work of Bertsimas and Stock from 1994, in which they presented the model with the possibility of rerouting flights [12]. They divided the airspace into sectors and took into account the capacities of all sectors individually. If the available capacity of a sector of airspace is violated, a flight which was assigned to fly through this sector may be rerouted, thus a different flight path may be assigned to it in order to reach its destination (see in[12]). The main difference in this model arises from the interpretation of decision variables w_{ft}^{j} which are equal to 1 if the flight f arrives to sector j by the time t. The authors managed to solve large scale, realistic size problems with competitive results.

However, due to different nature of European airspace comparing to the U.S., reformulation of this problem should be considered while applying in Europe. In contrast with the U.S., congestions may occur in the airspace as well and therefore sector capacities have to be taken into account. The en route sector capacity constraints, in turn, generate complex interactions among traffic flows [34]. In 2007, Lulli and Odoni proposed the model, which illustrates the complex nature of European ATFM, taking into account combinations of ground and airborne holding (see [34]).

Bertsimas, Lulli and Odoni present a new integer programming approach for the ATFM in [18]. They propose a model that covers all the phases of each flight (take off, rerouting, landing), taking into account successive flights and the possibility of cancellations. The core of their idea is to propose an objective function minimizing the sum of ground and air delay costs in such a way, that the ground holding is prioritized over the air holding. Moreover, they solve problems of size comparable to real situations in the American airspace in short computational times (see Computational Experience in [18]).

In our work, we formulate a simple version of the ground holding problem and afterwards we propose its extensions. Our purpose is to discuss the possibility of applying the robust optimization frameworks to the ATFM problem formulations and to analyze the profitability of the new robust model.

3.2 The nominal formulation

In this section, we introduce the formulation of the air traffic flow scheduling problem with deterministic landing capacities of the airports. We assume, that the schedule is made only once, in the beginning, and no further changes are allowed. Also, without the loss of generality, we take into account only the case of congestions on arrival airports.

The structure of our model comes from the work of prof. Smriglio and prof. Rossi. Firstly, they propose the nominal formulation of the ATFM problem and after that, they investigate a set packing formulation of the ground holding problem and design a branch-and-cut algorithm to solve the problem in high congestion scenarios, i.e., when lack of capacity induces flights cancellation (see in [40]).

Let us consider a set of airports $\mathcal{K} = \{1, \dots, K\}$ and a set of flights $\mathcal{F} = \{1, \dots, F\}$. We suppose, that each flight will be performed according to a planned timetable. Furthermore, we consider an ordered set of time periods $\mathcal{T} = \{1, \dots, T\}$. Each time period represents 15 minutes long time interval. For the nominal formulation of the model, we consider the following input data:

$h_i \in \mathcal{K}$ = the departure airport of flight $i \in \mathcal{F}$
$k_i \in \mathcal{K}$ = the arrival airport of flight $i \in \mathcal{F}$
$d_i \in \mathcal{T}$ = the scheduled departure time of flight $i \in \mathcal{F}$
$r_i \in \mathcal{T}$ = the scheduled arrival time of flight $i \in \mathcal{F}$
$R_k(t)$ = the capacity of airport $k \in \mathcal{K}$ at time $t \in \mathcal{T}$
$A_i \in \boldsymbol{Z}_+$= the maximum allowed ground delay for flight $i \in \mathcal{F}$
c_i = cancelation cost of flight $i \in \mathcal{F}$
w_i= delay cost per unit time period for flight $i \in \mathcal{F}$
δ_i = maximum delay allowed for i to land such that $j = succ(i)$ is not delayed.

Next, we define a set of time periods in which flight i can be assigned to land as

$$\mathcal{T}_i = \{t \in \mathcal{T} \colon r_i \leq t \leq r_i + A_i\},$$

and a set of flights assigned to land at airport k as

$$\mathcal{F}(k) = \{i \in \mathcal{F}, \text{ s.t. } k = k_i\}.$$

Under the maximum ground delay we understand the maximum number of time periods which flight $i \in \mathcal{F}$ can be hold on the ground. Without the loss of generality, we assume that the congestion may occure only on the arrival airports. The flight times $r_i - d_i$, $i \in \mathcal{F}$ are fixed quantities and all the delays occur only on the ground [40].

The capacity $R_k(t)$ of airport $k \in \mathcal{K}$ at time $t \in \mathcal{T}$ represents the maximum number of aircrafts which can be assigned to land at time $t \in \mathcal{T}$. We assume, that $R_k(t)$ is a fixed number for every $k \in \mathcal{K}$, $t \in \mathcal{T}$. Finally, we suppose that cancellations of flights are allowed and in practice cancellation costs always dominate the delay costs, i.e., for every $i \in \mathcal{F}$, $c_i \gg A_i w_i$ [40].

Pair of successive flights

When a flight $j \in \mathcal{F}$ is performed after flight $i \in \mathcal{F}$ and by the same aircraft, flights $i, j \in \mathcal{F}$ are said to be connected and we denote $j = succ(i)$ [40]. Set $\mathcal{F}' \subset \mathcal{F}$ is a set of all flights that have a successor. For a pair of connected flights $i, j \in \mathcal{F}$, such that $i \in \mathcal{F}', j = succ(i)$, and a pair of time periods $t' \in \mathcal{T}_i, t'' \in \mathcal{T}_j$ we have: $t''-t' = r_j - r_i - \delta_i - 1$. We denote landing times t' and t'' as coupled landing times for a pair of flights (i, j).

Decision variables

The decision variables specify if the flight $i \in \mathcal{F}$ is assigned to arrive at time slot $t \in \mathcal{T}_i$. If this is the case, then the decision variable is set to 1: $y_{it} = 1$. Otherwise, we set $y_{it} = 0$.

Objective function

The objective is to minimize delay and cancellation costs:

$$\min \sum_{i \in \mathcal{F}} \left(\left(1 - \sum_{t \in \mathcal{T}_i} y_{it} \right) (c_i + w_i r_i) + w_i \left(\sum_{t \in \mathcal{T}_i} t y_{it} - r_i \right) \right).$$

Constraints

1. Capacity constraints for the arrivals at the airport k at time t:

$$\sum_{i \in \mathcal{F}(k)} y_{it} \leq R_k(t) \qquad \forall k \in \mathcal{K}, \forall t \in \mathcal{T}.$$

2. Each flight i is assigned to land at most at one time slot t:

$$\sum_{t \in \mathcal{T}_i} y_{it} \leq 1 \qquad \forall i \in \mathcal{F}.$$

3. Connectivity between flights:

$$\sum_{t \geq t'} y_{it} + \sum_{t \leq t''} y_{jt} \leq 1,$$

$$\forall i \in \mathcal{F}', j = succ(i) \text{ and } t', t'' \text{coupled for} (i, j).$$

4. Finally, we present the integrality constraints:

$$y_{it} \in \{0, 1\} \qquad \forall i \in \mathcal{F}, \ \forall t \in \mathcal{T}.$$

3.2.1 Size of the nominal formulation

The size of the nominal formulation given above depends on the number of flights, airports and time periods. Let us denote by $\mathcal{F}$ a set of all the flights considered in our problem and by $\mathcal{F}'$ a subset of all the flights, for which successive flight exists. Furthermore, $\mathcal{K}$ stands for the set of all the airports and $\mathcal{T}$ the set of all the time periods considered. Indeed, it is possible to bound the number of variables of the nominal problem by

$$|\mathcal{K}| \times \max_{k \in \mathcal{K}} |\mathcal{F}(k)| \times \max_{f \in \mathcal{F}} |\mathcal{T}_i|.$$

Additionally, the number of constraints can be upper bounded by

$$|\mathcal{K}| \times |\mathcal{T}| + |\mathcal{F}| \times \max_{f \in \mathcal{F}} |\mathcal{T}_i| + |\mathcal{F}'| \times \max_{f \in \mathcal{F}} |\mathcal{T}_i|.$$

Let us show how this would look like in reality. We consider 30 000 flights, which represents a number of flights flying across European airspace on a typical summer day [25]. According to [1], there are currently 153 large airports in Europe. Those are the airports, which operate international, as well as domestic flights daily. From the interpretation of our problem, large airports are those we are interested in, since they are main victims of congestions. For example, for the busiest airport in Europe, London-Heathrow Airport, there were in average 1290 scheduled flights per day in 2014. Since we have divided time into a set of time slots, where each slot represents 15 minutes interval, to one day correspond 96 time slots. By allowing each flight to be delayed at most by 2 hours, we set $\max_{f \in \mathcal{F}} |\mathcal{T}_i| = 8$. Therefore, for a regular day in European airspace an upper bound for variables of our nominal problem would be:

$$153 \times 1\,290 \times 8 = 1\,578\,960.$$

Moreover, assuming that one third of all the flights have successors, the number of constraints for a regular summer day in European airspace would be upper bounded by:

$$153 \times 96 + 30\,000 \times 8 + 10\,000 \times 8 = 14\,688 + 240\,000 + 80\,000 = 334\,688.$$

3.3 The set packing formulation

In this section, we will derive the set packing formulation of the nominal problem, which was presented in the work of Prof. Smriglio and Prof. Rossi [40]. Let us firstly remind the objective of the set packing problem. Given a finite set $\mathcal{V}$ and a list of subsets of

$\mathcal{V}$ called $\mathcal{V}'$, the set packing problem investigates whether some k elements of $\mathcal{V}'$ are pairwise disjoint, that is, no two of them share an element.

Our aim is now to reformulate the ground holding problem described above as a general set packing problem, namely $\min\left\{c^T x : Ax \leq \mathbf{1},\ x \in \{0,1\}^n\right\}$, where A is a $m \times n$ binary matrix. In order to do so, for each airport and for each time slot $(k,t) \in \mathcal{K}\times\mathcal{T}$ we create a set of arrival slots $S_{kt} = \left\{S^1_{k,t}, \ldots, S^{R_k(t)}_{k,t}\right\}$, that is a set of $R_k(t)$ single capacity arrival slots. For example, if an airport k_i is able to accommodate 8 flights at time t_i, $S_{k_i t_i} = \left\{S^1_{k_i,t_i}, \ldots, S^8_{k_i,t_i}\right\}$. Under this assumption, we are able to reformulate the nominal problem followingly.

Objective function

Let φ be the set of all capacity slots $S_{k,t}$ of all the airports $k \in \mathcal{K}$ for each of the times $t \in \mathcal{T}$, namely let $\varphi = \bigcup_{k\in\mathcal{K}, t\in\mathcal{T}} S_{k,t}$. Then the objective becomes:

$$\min \sum_{i\in\mathcal{F}} \left(\left(1 - \sum_{t\in\mathcal{T}_i, s\in S_{k_i t}} x_{is} \right) (c_i + w_i r_i) + w_i \left(\sum_{t\in\mathcal{T}_i, s\in S_{k_i t}} t x_{is} - r_i \right) \right).$$

Constraints

1. For each of the capacity slots must hold that at most one flight can land on it:

$$\sum_{i\in\mathcal{F}} x_{is} \leq 1 \qquad \forall s \in \varphi.$$

2. Each flight i can land at most once:

$$\sum_{t\in\mathcal{T}_i, s\in S_{k_i t}} x_{is} \leq 1 \qquad \forall i \in \mathcal{F}.$$

3. Connectivity between flights:

$$\sum_{t\geq t', s\in S_{k_i t}} x_{is} + \sum_{t\leq t'', q\in S_{k_j t}} x_{jq} \leq 1,$$

$$\forall i \in \mathcal{F}', j = succ(i) \text{ and } t', t'' \text{coupled for} (i,j).$$

4. Integrality constraints:

$$x_{is} \in \{0,1\} \qquad \forall i \in \mathcal{F},\ \forall s \in \varphi.$$

3.3.1 Combinatorial properties

To follow we discuss the combinatorial properties of this formulation. The set packing model can be solved by solving the associated stable set problem on the intersection graph of $\{0, 1\}$ matrix A [40]. This means to find a set of pairwise nonadjacent vertices of maximum weight in a graph $G(V, E)$. If all weights of vertices V are equal to one, the problem reduces to maximum cardinality stable set problem. Each decision variable x_{is} corresponds to a pair $(i, s) \in (\mathcal{F} \times \varphi)$ and can be represented as a node $v \in V_I$, where $\mathrm{G}_I = (V_I, E_I)$ is an intersection graph (more explanation provided in [40]).

Let us illustrate the behavior of this model formulation on an example. We consider a small airport k and 4 flights, which are supposed to land during next 45 minutes. Through first 15 minutes, the airport is able to handle 2 arriving flights and therefore for $t = 1$ we have two capacity slots S^1_{k1}, S^2_{k1}. However, the airport is able to handle only one arriving flight per every 15 minutes afterwards. Due to this we create only 1 capacity slot for t_2 and t_3, those are S^1_{k2} and S^1_{k3}. The problem is now to optimally assign each flight to land on one of the possible landing slots, such that the ground holding costs are minimized. Due to the constraints of our model, the sum of all rows has to be less or equal to one and the same has to hold for columns. More of the properties of the graph induced by our model can be found in [40].

$$\begin{array}{cc} & \begin{array}{cccc} V(S^1_{k1}) & V(S^2_{k1}) & V(S^1_{k2}) & V(S^1_{k3}) \end{array} \\ \begin{array}{c} V(1) \\ V(2) \\ V(3) \\ V(4) \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

The set packing model can be extended to more complicated versions of GHP (see in extensions of [40]). It gives an advantage of solving a MIP problem by LP and under some assumptions assures the integrality of decision variables. It was shown, that the rank inequalities reduce the integrality gap also in large instances, however, for classical scenarios, the set packing model remains efficient. The set packing model was applied and solved for instances of the size of real-world problems in [40]. It outperformed other previously known GHP algorithms and reduced the infeasibility of solutions (cancellation of flights).

3.4 The ATFM under capacity uncertainty

In the previously mentioned ATFM models we consider the landing capacities of airports as functions of time only. We suppose that for every time period $t \in T$ and every airport $k \in \mathcal{K}$ there exists a deterministically given value $R_k(t)$ representing the capacity of a given airport at time t. However, in reality this not always the case. Usually, when the schedule is being made, one can not exactly estimate the future capacity values and therefore the values $R_k(t)$ are only estimates. If the estimates are too optimistic, that means considering higher landing capacities than the real ones, we would have to face airborne delays of flights close on arrival. Costs of such delays are usually much higher than for delays caused by ground-holds. In contrast, a pessimistic estimate results in avoidable ground delays for some flights and the subsequent propagated delays to connecting flights [27]. To follow we will study the ATFM problem with uncertain capacities. We model the weather-front induced landing capacities and propose a robust optimization formulation of ATFM problem.

3.4.1 Delays

According to Aviation System Performance Metrics (ASPM), main factors causing delays of flights are carrier delays, weather delays, NAS delays, security delays and late aircraft delays [26]. The data on on-time performance of flights in the American National Airspace (ANA) are provided by Federal Aviation Administration (FAA) and are freely accessible at [3]. Carrier delay is usually caused by the control of the air carrier. Security delays are caused for example by the evacuation of the terminal or re-boarding of the aircraft. NAS delays are delays that are within the control of the American National Airspace System. Late aircraft delays are delays caused by late arrival of the same aircraft at previous airport.

In the U.S., most of the delays in the airspace during the last 15 years were caused by extreme or hazardous weather conditions (Fig. 3.1). Let us remind, that more than 60% of yearly delays are those caused by weather. Furthermore, an evident seasonal pattern can be seen between the time of the year and the number of delays (Fig. 3.2). The amount of delays increases during the summer months due to the higher number of flights in this period, which causes bigger congestions on the airports. Also, the changes in weather occur more often in summer and are generally harder to predict than during the rest of the year.

Following the previous facts, we consider necessary to study the ATFM problem with impact of unpredictable weather conditions. We believe that this can be done by

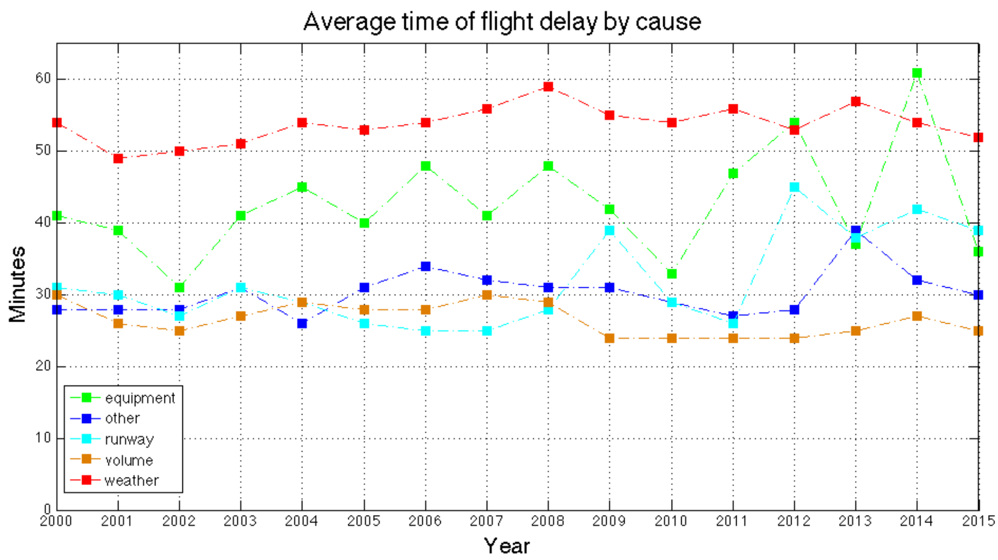

Figure 3.1: Average time of flight delay by cause in the U.S. (Data Source: OPSNET [3]).

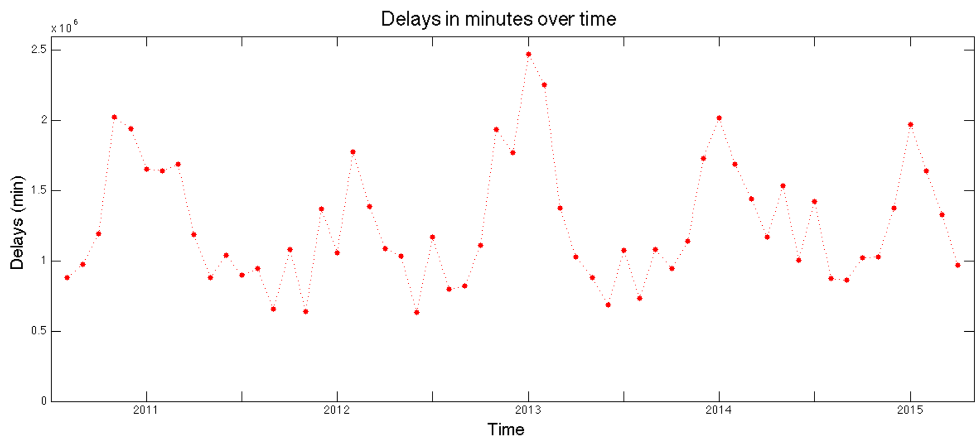

Figure 3.2: Seasonal pattern of delays in the U.S. (Data Source: OPSNET [3]). The red dots represent monthwise average delays over the past 5 years.

implementing robust optimization frameworks described in Chapter 2 into the discrete models, which have been proposed previously in this chapter.

3.5 Modelling the capacity uncertainty

Following the previous findings, the main influence on the uncertainty in capacities of the landing slots comes from the uncertainty in weather. For practical reasons, it is important to model this uncertainty by only small number of parameters. We will model now the weather uncertainty and describe its relation with capacity reductions.

3.5.1 Defining parameters

Imagine that it is a typical summer day and the storm is expected to occur in the area of an airport $\mathcal{K}$ at some time in the afternoon. Let us suppose that in the morning, when the schedule for the day is being produced, we get the information that the storm will arrive between 17:00 and 19:00. For us this information would mean that the capacity slots of airport K would be reduced by some number at some time between 17:00 and 19:00. Naturally thinking we assume, that the stronger the storm is, the more reduction of capacities we can expect. This means, that in order to better predict the capacity reductions, we would like to know how strong the storm would be. Also, it is important for us to know how long will the storm be. If it is a short one, let's say 8 min, we can assume that the capacities will be violated only at one of the time slots (since each $t_i \epsilon T$ represents 15 min). But if the storm takes longer time, i.e. 40 min, the reduction will affect more slots (in this case 40 min correspond to 3 time slots).

Mathematically speaking, after previous assumptions we can derive the key parameters affecting the weather induced capacity uncertainty:

1. Time of arrival T_a of weather front
2. Duration d of weather front
3. Capacity reduction α.

Thus, $T_a \in \{T_{min}, ..., T_{max}\}$, where T_{min} represents the first possible time when the weather front could occure and T_{max} stands for the last possible time slot, when the weather-front can start acting. Note, that if the weather-front arrives at T_{max}, the capacity reduction will occure until $T_e = T_{max} + d$, where $d \in \{d_{min}, ..., d_{max}\}$ and

$$T_e \in \{T_{min} + d_{min}, ..., T_{max} + d_{max}\} .$$

Capacity reduction α represents the portion of reduced slots. Now, we can rewrite the capacity vector of airport k with the weather front realization at timeslot t_a, duration d and capacity reduction α followingly (for easier notation we skip the index k):

$$\boldsymbol{R}(t) = \{R(t_1), ..., R(t_{a-1}), \alpha R(t_a), ..., \alpha R(t_{a+d}), R(t_{a+d+1}), ..., R(t_n)\},$$

that is

$$R(t_i) = \begin{cases} R(t_i), & i \in T\text{-}\{T_a, ..., T_b\} \\ \alpha R(t_i), & i \in \{T_a, ..., T_b\}. \end{cases}$$

Since we are in the discrete setting, we assume that α is always given such that $\alpha R(t_i)$ is an integer.

Example: Suppose that the capacity $R_k(t)$ of airport k is during the whole time equal to 25, that is $R_k(t) = 25\ \forall t$. Considering $T = 4$, $T_a \in \{2, 3\}$, $d \in \{1, 2\}$ and $\alpha \in \{0,2,\ 1\}$ we get the following set of possible realizations of uncertainty:
$\Omega = \{(25, 20, 25, 25), (25, 20, 20, 25), (25, 0, 25, 25), (25, 0, 0, 25), (25, 25, 20, 25),$
$(25, 25, 20, 20), (25, 25, 0, 25), (25, 25, 0, 0)\}$.

3.5.2 Characterization of the uncertainty set

Fom Chapter 2 we know, that in order to computationally solve the problem, it is necessary to characterize the uncertainty set given by described parameters. We present the characterization of uncertainty set coming from [27]. He firstly derives the polyhedral description of conv(Ω_α), showing that $\mathcal{P}_\alpha = \text{conv}(\Omega_{\alpha_{min}})$. Next, he proves the relation of conv(Ω)=conv(Ω_α) and concludes with the statement $\mathcal{P}_{\alpha_{min}} = \text{conv}(\Omega)$. These findings are extremely important for the derivation of our model and therefore next we present a deeper explanation.

He begins with introducing auxiliary variables for every $t \in \{T_{min}, ..., T_{max} + d_{min}\}$:

$$y_t = \begin{cases} 1, & \text{if capacity drops by time } t \\ 0 & \text{otherwise.} \end{cases}$$

$$z_t = \begin{cases} 1, & \text{if capacity revives by time } t \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, $y_t = 1$ if $t \in \{T_{max}, ..., T_{max} + d_{max}\}$, $z_t = 0$ if $t \in \{T_{min}, ..., T_{min} + d_{min} - 1\}$ and $z_t = 1$ for $t \in \{T_{max} + d_{max}\}$.

Furthermore, Gupta provides a mathematical description of an uncertainty set for particular realization of α followingly [27]:

$$\Omega_\alpha = \{\mathbf{b} \in \boldsymbol{Z}_+^{\mathbf{m}} | \, b_t = C(1-y_t) + \alpha C y_t + (1-\alpha) C z_t, \ \forall t \in \{T_{min}, ..., T_{max} + d_{max}\} \, ;$$

$$b_t = C, \ \forall t \in T - \{T_{min}, ..., T_{max} + d_{max}\} \, ;$$

$$y_t \leq y_{t+1}; \ z_t \leq z_{t+1}; \ z_t \leq y_t;$$

$$y_{T_{max}} = 1; \ z_{T_{min}+d_{min}-1} = 0; \ z_{T_{max}+d_{max}} = 1; \ y_{t,} \, z_t \in \{0,1\}\}.$$

Additionally, he claims that after replacing $y_{t,} \, z_t \in \{0,1\}$ by $0 \leq y_t, z_t \leq 1$ we obtain the description of $\text{conv}(\Omega_\alpha)$ and in the case where α is chosen in a way that αC is an integer, then $\text{conv}(\Omega_\alpha)$ is exactly a polyhedron;

$$\text{conv}(\Omega_\alpha) = \mathcal{P}_\alpha.$$

The proof can be seen in [27].

Let $\alpha \in \{\alpha_{min}, ..., \alpha_{max}\}$. We can describe the whole uncertainty set as:

$$\Omega = \left(\cup_{\alpha \in \{\alpha_{min}, ..., \alpha_{max}\}} \Omega_\alpha\right).$$

Moreover, since

$$\cup_{\alpha \in \{\alpha_{min}, ..., \alpha_{max}\}} \text{conv}(\Omega_\alpha) \subseteq \text{conv}\Big(\cup_{\alpha \in \{\alpha_{min}, ..., \alpha_{max}\}} (\Omega_\alpha)\Big) = \text{conv}(\Omega),$$

and due to additional assumptions (see the proof of Theorem 4. in [27]),

$$\text{conv}(\Omega) = \text{conv}(\Omega_{\alpha_{min}}).$$

Under assumptions mentioned above, the author concludes with following theorem:

Theorem: ([27])

$$\mathcal{P}_{\alpha_{min}} = \text{conv}(\Omega).$$

3.6 Robust formulation of the ATFM problem

In this section, we propose formulations of the air traffic flow management problem under capacity uncertainty. We firstly derive a single-stage robust model from the nominal formulation. Using nominal formulation allows us to formulate the uncertainty by the right hand side robust approach. After that we propose a two-stage robust model. The main difference of these two models is in the number of decisions which are being made. In the single-stage robust model, we suppose that the decision maker makes the decision only once, in the beginning and this is in the morning when the schedule is produced. This model does not allow any further changes in the schedule. On the contrary, in the two-stage robust model the decision maker is expected to get an updated information of the landing capacities and he is allowed to change a schedule accordingly. This will be explained in details later in this section. Finally, we propose a single stage robust model of the set-packing ATFM formulation, where the capacity uncertainty is moved from the right-hand side vector inside the matrix of constraints. We show that it is equivalent to the deterministic formulation if assuming the worst-case capacity reductions.

3.6.1 Single-stage robust ATFM model

In the section 3.2 we proposed the nominal formulation of the ATFM problem in which several constraint functions are presented, including capacity constraints. In this model, the capacities for each airport $k \in \mathcal{K}$ at each time period $t \in T$ are considered to be deterministically given. However, as we have already described before, this is not always in reality the case. We will now apply the knowledge from Chapter 2 and reformulate this model into single-stage robust optimization problem.

Note, that we can understand the deterministic capacities $R_k(t)$ from the nominal formulation as a vector defining capacities in the space of airports and time, thus $R_k(t) \in \boldsymbol{R}^{|\mathcal{T}|\times|\mathcal{K}|}$. This vector appears on the right-hand side of the capacity constraints. Following the framework of previous section, we want to introduce the uncertainty in this vector of capacities, and therefore we can claim that the uncertainty of this model will occure only on the right-hand side vector of the capacity constraints of the nominal formulation.

In the end of section 3.5.1 we showed on a small example how an uncertainty set would look like for one airport when a capacity reduction is expected during four time periods, but without the knowledge of the exact time of arrival, duration and strength of weather realization. Following this example, we can similarly compose the uncertainty set Ω for the whole vector $R_k(t) \in \boldsymbol{R}^{|\mathcal{T}|\times|\mathcal{K}|}$. Setting $n = |T_a| \times |\alpha| \times |d|$ and $m = |\mathcal{K}| \times |T|$ the uncertainty set Ω will be composed as follows:

$$\Omega = \left\{(\omega_{1,1}, ..., \omega_{1,m}),\, (\omega_{2,1}, ..., \omega_{2,m}), ...,\, (\omega_{n,1}, ..., \omega_{n,m})\right\}.$$

In the case where only one airport is taken into account, $\omega_{1,1}$ would represent the first realization of uncertainty at first time period, $\omega_{2,m}$ the second realization of uncertainty at $m-$th time period and so on.

The main purpose of the robust optimization is to find a solution, which stays feasible for any realization of uncertainty. In our model, we are looking for a solution, which would satisfy the constraint

$$\sum_{i \in \mathcal{F}(k)} y_{it} \leq R_k(t) \qquad \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, R_k(t) \in \Omega \qquad . \tag{3.1}$$

Due to the worst case nature of robust optimization, we define the vector $R_k^*(t)$ as:

$$R_k^*(t) = \min \left\{ R_k(t) : R_k(t) \in \Omega \right\}.$$

Naturally it follows then, that if $\sum_{i \in \mathcal{F}(k)} y_{it} \leq R_k^*(t)$ is satisfied for $\forall k \in \mathcal{K}, \forall t \in \mathcal{T}$, then (3.1) is fulfilled. To follow, for any uncertainty set Ω describing the right hand side capacity uncertainty from the nominal formulation of ATFM, the single-stage robust ATFM formulation becomes:

$$\begin{array}{ll}
\min \sum_{i \in \mathcal{F}} \left(\left(1 - \sum_{t \in \mathcal{T}_i} y_{it}\right) (c_i + w_i r_i) + w_i \left(\sum_{t \in \mathcal{T}_i} t y_{it} - r_i \right) \right) & \\
\qquad \sum_{i \in \mathcal{F}(k)} y_{it} \leq R_k^*(t) & \forall k \in \mathcal{K}, \forall t \in \mathcal{T} \\
\qquad \sum_{t \in \mathcal{T}_i} y_{it} \leq 1 & \forall i \in \mathcal{F} \\
\qquad \sum_{t \geq t'} y_{it} + \sum_{t \leq t''} y_{jt} \leq 1 & \forall i \in \mathcal{F}', j = succ(i);\ t', t'' \text{coupled} \\
\qquad y_{it} \in \{0, 1\} & \forall i \in \mathcal{F},\ \forall t \in \mathcal{T}.
\end{array}$$

As a consequence, the proposed single-stage robust ATFM model is expected to be efficiently solvable in practice. However, the trade-off between the feasibility and value of the cost function is expected to be high due to the worst-case nature of the proposed formulation.

3.6.2 Single-stage robust ATFM set packing model

Following the framework of previous section, we can implement the uncertainty of capacities into the ATFM set packing formulation and solve this problem with the robust approach. This means finding the solution, which would stay feasible regardless the outcome of the uncertainty set. We propose now three approaches how to impose uncertainty in the set packing formulation.

Initializing the size of the set of capacity slots

First approach is to define the uncertainty set of capacities $R_k(t) \in \Omega$ in the beginning and finding $R_k^*(t) = \min\{R_k(t) : R_k(t) \in \Omega\}$. After $R_k^*(t)$ is known for every $k \in \mathcal{K}$ and $t \in T$, we are able to construct the set of landing slots such that $S_{kt} = \left\{S_{k,t}^1, \ldots, S_{k,t}^{R_k^*(t)}\right\}$. Now, following the assumptions from section 3.3, we formulate the set-packing model followingly:

$$\begin{array}{cc}
\min \sum_{i\in\mathcal{F}} \left(1 - \sum_{t\in\mathcal{T}_i, s\in S_{k_i t}} x_{is}\right)(c_i + w_i r_i) + & w_i\left(\sum_{t\in\mathcal{T}_i, s\in S_{k_i t}} t x_{is} - r_i\right) \\
\sum_{i\in\mathcal{F}} x_{is} \leq 1 & \forall s \in \varphi \\
\sum_{t\in\mathcal{T}_i, s\in S_{k_i t}} x_{is} \leq 1 & \forall i \in \mathcal{F} \\
\sum_{t\geq t', s\in S_{k_i t}} x_{is} + \sum_{t\leq t'', q\in S_{k_j t}} x_{jq} \leq 1 & \forall i \in \mathcal{F}', j = succ(i);\ t', t''\text{coupled} \\
x_{is} \in \{0,1\} & \forall i \in \mathcal{F},\ \forall s \in \varphi.
\end{array} \tag{3.2}$$

Approach by imposing an additional variable

Let us investigate now the case where we firstly consider the set packing formulation of deterministically given capacities and we embed robust approach by cancelling the landing slots in it. This is possible due to the fact that we do not expect the increase of capacities, the uncertainty is effecting only in decreasing the number of landing slots. Let us explain this on a small example.

Example: Suppose that there are three flights (f_1, f_2, f_3) landing at the same airport k. All of them should land at time t_1 and only flights f_2, f_3 are allowed to be delayed, meaning that the flight f_1 has to land at t_1. However, the capacity of airport k at t_1 is only 2, from which we conclude, that at least one of the flights will be assigned to land at t_2 at some cost. Formulating this with a set packing approach gives us a $\{0,1\}$ matrix $\boldsymbol{A}$, where $a_{i,j}$ is set to 1 if flight i can land at time t and slot s:

$$a_{i,j} = 1 \,\text{if}\, j \in (\sum_{l=1}^{t-1} s(l) + 1, \sum_{l=1}^{t} s(l) + \sum_{l=t+1}^{Di} (s(l)),$$

where D_i represents the maximum allowed delay for flight i. We get the max stable set problem:

$$
\begin{aligned}
\max x_{1,1}^1 + x_{2,1}^1 + x_{3,1}^1 + x_{1,1}^2 + x_{2,1}^2 + x_{3,1}^2 + x_{1,2}^1 + x_{2,2}^1 + x_{3,2}^1 + x_{1,2}^2 + x_{2,2}^2 + x_{3,2}^2 & \\
x_{2,1}^1 + x_{2,1}^2 + x_{2,2}^1 + x_{2,2}^2 & \leq 1 \\
x_{3,1}^1 + x_{3,1}^2 + x_{3,2}^1 + x_{3,2}^2 & \leq 1 \\
x_{1,1}^1 + x_{2,1}^1 + x_{3,1}^1 & \leq 1 \\
x_{1,1}^2 + x_{2,1}^2 + x_{3,1}^2 & \leq 1 \\
x_{2,2}^1 + x_{3,2}^1 & \leq 1 \\
x_{2,2}^2 + x_{3,2}^2 & \leq 1 \\
x_{1,1}^1 + x_{1,1}^2 & \leq 1.
\end{aligned}
$$

Considering the case where some of the landing slots may be cancelled due to unfavorable weather conditions, we propose to introduce for every airport k, time slot t and landing slot s a new variable $y_{k,t}^s$ followingly:

$$
y_{k,t}^s = \begin{cases} 0 & \text{if the slot } s \text{ of airport } k \text{ at time } t \text{ is cancelled} \\ 1 & \text{otherwise} \end{cases} .
$$

Note, that if the capacity of an airport is reduced by a scalar h, it means that h slots become not available anymore, and we consider the set of those h slots to be the same for every flight (the set of the same slots). We have to do so in order to prevent the assumption on the size of capacity. If we would consider different groups of slots for each flight to be cancelled, it could happen that the actual capacity of an airport at some time is lower than the one used in our formulation. Therefore without loss of generality, if h slots are to be cancelled at time t, we always consider the last h slots of time t. For example, if one of the two slots at time t_1 would be canceled in our previous example, it would be the slot t_1^2. And if so, by introducing $y_{k,t}^s$ as previously described, we would set $y_{k,t_1}^2 = 0$. Since this landing spot does not exist anymore, none of the flights can be assigned to land on it. We introduce additional constraint:

$$
x_{i,t_1}^2 \leq y_{k,t_1}^2 \quad \forall i \in \mathcal{F} \text{ assigned to land at } k,
$$

This shrinks our set of feasible solutions, making the first flight to land at t_1 and second and third at t_2. Moreover, if one more of the slots would be cancelled, one of the flights would have to be cancelled as well. Therefore this problem would have a feasible solution only if the cancellation of flights is allowed.

Let us derive this formulation for a general setting. With this approach we transfer the uncertainty from the vector of capacities into a new variable $y^s_{k,t}$. For doing so, we construct a mapping from $R_k(t) \in \Omega$ to $y^s_{k,t}(R_{k,t})$ and add following constraints to the nominal set packing formulation:

$$\begin{aligned} x^s_{i,t} \leq y^s_{k,t}(R_k(t)) \quad & \forall i \in \mathcal{F} \text{ assigned to land at } k,\ \forall t \in \mathcal{T},\ \forall s \in \varphi \text{ and } \forall R_k(t) \in \Omega \\ y^s_{kt} \in \{0, 1\} \quad & \qquad\qquad\qquad \forall k \in \mathcal{K},\ \forall t \in T,\ \forall s \in \varphi. \end{aligned}$$

There is a direct dependence between the size of $S_{k,t}$ and the capacity $R_k(t)$. In the weather induced uncertainty case, when the exact capacity is not known, we construct the set of capacity slots of our set packing formulation by using the deterministic capacity $R_k(t)$ and we expect some of them to be cancelled according to the behavior of the uncertainty set. This cancellation is explained by the variable $y^s_{k,t}(R_k(t)) \in \{0, 1\}$. This means, that if we expect a violation on capacity in airport k at time t, we create a set of possible realizations on capacity reduction. If for example at most 5 slots are expected to be violated then at most last 5 variables $y^s_{k,t}$ will be set to 0 and previous $y^s_{k,t}$ will be set to 1. If we want to be protected in our formulation against all the possible realizations, we have to consider the worst case realization of weather front. In this case it means to consider the cancellation of 5 slots and therefore

$$\left(y^s_{k,t},\, y^{s-1}_{k,t},\, y^{s-2}_{k,t},\, y^{s-3}_{k,t},\, y^{s-4}_{k,t}\right) = (0,\, 0,\, 0,\, 0,\, 0)\,.$$

Applying the worst-case concept of robust optimization, we can rewrite the first constraint as:

$$x^s_{i,t} \leq \min_{R_k(t) \in \Omega} y^s_{k,t}(R_k(t)) \quad \forall i \in \mathcal{F} \text{ assigned to land at } k,\ \forall t \in \mathcal{T},\ \forall s \in \varphi.$$

Therefore using again the notation from 3.6.1, where $R^*_k(t) = \min\{R_k(t) : R_k(t) \in \Omega\}$ we characterize:

$$y^s_{k,t} = \begin{cases} 1 & \text{if } s \in \left[1,\, R^*_{k,t}\right] \\ 0 & \text{if } s \in (R^*_{k,t},\, R_{k,t}]. \end{cases}$$

Since $x^s_{i,t} \leq y^s_{k,t}$, the decision variables for which $x^s_{i,t}$, $s \in (R^*_{k,t},\, R_{k,t}]$ are forced to be 0 and therefore none of the flights can be assigned to land at $x^s_{i,t}$, $s \in (R^*_{k,t},\, R_{k,t}]$. Due to this and other assumptions stated in this section we conclude that this approach yields the same formulation as (3.2).

Budget of uncertainty approach

In this section we apply the budget of uncertainty approach of Bertsimas and Sim (see more in 2.2.5) to the set packing ATFM model and we show, that this approach yields again the same formulation as (3.2). Following the framework of the authors we introduce for every row i of matrix $\boldsymbol{A}$ a new parameter Γ_i which controls the number of entries which are expected to be violated as follows:

$$\begin{array}{cc} \Gamma_i(t) = \sum_{\mathcal{T}_i}(R_k(t) - R_k(t,\omega)) & \forall\, i \in \mathcal{F},\, \forall\, \omega \in \Omega \\ \Gamma_i \in \boldsymbol{N} & \end{array}.$$

Let us remind that $\mathcal{T}_i = \{t \in \mathcal{T} \colon r_i \le t \le r_i + D_i\}$ represents a set of time slots in which a flight i can be assigned to land (with some possible costs of delay). Applying again the worst-case nature of robust optimization we get:

$$\begin{array}{cc} \Gamma_i(t) = \sum_{\mathcal{T}_i}(R_k(t) - R_k^*(t)) & \forall\, i \in \mathcal{F} \\ \Gamma_i \in \boldsymbol{N} & \end{array}.$$

From this clearly follows that for every flight i we expect exactly Γ_i landing slots to be violated. Due to the fact that the decision maker is able to find $R_k^*(t)$ and without loss of generality he cancels last Γ_i landing slots for every t (for explanation see previous section) we expect that there exists a deterministic formulation of this approach. We introduce a set of possible landing slots for a flight i which is supposed to land at the airport k as

$$\mathcal{S}(\mathcal{T}_i) = \left\{ s : \, |s| = \sum_{t \in \mathcal{T}_i} |s_k(t) - \Gamma_i(t)|, \, \forall\, t \in \mathcal{T} \colon \, r_i \le t \le r_i + D_i \right\}.$$

We characterize each entry of matrix $\boldsymbol{A}$ as follows:

$$\begin{array}{ll} a_{i,j} = 1 & \text{for}\, i \in \mathcal{F},\, j \in S(\mathcal{T}_i), \\ a_{i,j} = 0 & \text{otherwise} \end{array}.$$

Now clearly, $a_{i,j}$ stays 1 only for those slots, which are not cancelled and becomes 0 for those which are expected to be cancelled in the worst-case scenario of predicted weather-front. This leads the decision variables, which are representing cancelled landing slots to be set to 0. Following from previous assumptions we claim that this reformulation of the set packing ATFM under capacity uncertainty leads again to the exact same formulation as 3.2).

To conclude this section, we see that the only difference between the deterministic and the robust single stage formulations is in the evaluation of the uncertainty set and finding the minimal value for every airport at every time slot. Since the deterministic versions of the models are efficiently solvable, we expect the single-stage robust models to be efficiently solvable as well. Moreover, we expect similar computational times, since the only extra time has to be dedicated for finding $R_k^*(t)$, which is generally not hard. However, since we are optimizing against the worst case scenario, we expect the value of the objective function to be higher than for deterministic cases. In general, the trade-off between the optimality and robustness in the single-stage robust approach is very high. Therefore, we consider reasonable to introduce a model where this trade-off would become more promising. This may be achieved by investigating a robust model of ATFM in which the schedule can be updated according to the current weather conditions. Hence we introduce next a two-stage robust ATFM formulation.

3.6.3 Two-stage robust ATFM model

Let us suppose that the schedule is produced in the morning and it can be updated later during the day. In reality, the closer we are to the realization of weather, the more better we are able to predict its behavior. This gives an advantage of more precise future knowledge of the uncertain parameters T_a, d, α, which would most probably result into a smaller size of uncertainty set in later stage. As well, if we didn't know the exact time of the capacity reduction and in the time of the second-stage decision we know, that it has already happened, the decision maker can cancel the uncertainty vectors concerning this particular reduction from the future uncertainty set. Mathematically we can handle this by second-stage robust optimization approach.

The weather uncertainty will generate the capacity uncertainty set, where the resulting capacities will be highly correlated. This comes from the following observations. If the predicted weather front appears, but moves more quickly or more slowly than forecasted, the modification of the whole capacity vector is highly correlated, since it will cause a change in capacities in several subsequent time slots and airports. The decision maker is then able in the second stage to update the information according to the current situation and to choose the optimal schedule.

Let us divide decision variables into two vectors depending on the time. In reality, we are able to predict the weather with quite good accuracy for next couple of hours. Therefore, for producing a schedule for one day we can without loss of generality divide the decision variables followingly:

$$\begin{aligned} \boldsymbol{u} &= y_{i,t} \qquad t \in \{1, ..., 39\} \\ \boldsymbol{v}(\boldsymbol{\omega}) &= y_{i,t} \quad t \in \{40, ..., 96\}\,. \end{aligned}$$

We consider decision variables $\boldsymbol{u}$ as non-adjustable variables describing the schedule for next 10 hours. Variables from $\boldsymbol{v}$ are adjustable variables and they produce the schedule for the later stage. We can reformulate then the objective function as follows:

$$\begin{aligned} Obj_{adapt} = \quad & \min_{\boldsymbol{u},\boldsymbol{v}(\omega)} \textstyle\sum_{i\in\mathcal{F}} \left(\left(1 - \sum_{t\in\mathcal{T}_i} \boldsymbol{u}\right)(c_i + w_i r_i) + w_i \left(\sum_{t\in\mathcal{T}_i} t\boldsymbol{u} - r_i\right)\right) \\ & + \max_{\omega\in\Omega} \textstyle\sum_{i\in\mathcal{F}} \left(\left(1 - \sum_{t\in\mathcal{T}_i} \boldsymbol{v(\omega)}\right)(c_i + w_i r_i) + w_i \left(\sum_{t\in\mathcal{T}_i} t\boldsymbol{v(\omega)} - r_i\right)\right). \end{aligned}$$

The constraints of the two-stage adaptive ATFM model will become then:

$$\begin{array}{cc} \sum_{i\in\mathcal{F}(k)}(\boldsymbol{u} + \boldsymbol{v(\omega)}) \le R_k^*(t) & \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, \forall \omega \in \Omega \\ \sum_{t\in\mathcal{T}_i}(\boldsymbol{u} + \boldsymbol{v(\omega)}) \le 1 & \forall i \in \mathcal{F}, \forall \omega \in \Omega \\ \sum_{t\ge t'}(\boldsymbol{u} + \boldsymbol{v(\omega)}) + \sum_{t\le t''}(\boldsymbol{u} + \boldsymbol{v(\omega)}) \le 1 & \forall i \in \mathcal{F}', j = succ(i);\, t', t''\text{coupled}, \forall \omega \in \Omega \\ \boldsymbol{u},\ \boldsymbol{v(\omega)} \in \{0,1\} & \forall i \in \mathcal{F},\ \forall t \in \mathcal{T}, \forall \omega \in \Omega. \end{array}$$

Chapter 4

Implementation in Practice

In this chapter, numerical results of the solution approaches introduced previously will be presented. Firstly, the setup of data will be explained. Twelve data sets have been generated in a way that each of them describes the real situation, which can occur in the European airspace during a typical summer day. The nominal ATFM optimization problem was implemented and solved using the open source software GLPK. In order to optimize the computational time and effort, the data was preprocessed accordingly using MATLAB. Later on, we show the obtained results of single-stage and two-stage robust optimization ATFM problems and we compare their performance towards the deterministic solutions. The main focus is taken on evaluating the profitability of two-stage robust ATFM model. We compare the costs of second-stage optimal solutions, where one of the weather scenarios was fixed in the second stage, and the costs of deterministic solutions under the realization of the same weather scenario. To conclude this study, we show that in cases of high congestion and very unfavorable weather realization, the results of the second-stage robust optimization formulation outperforms those coming from deterministic and single-stage robust formulations.

4.1 Data setup

We utilize twelve data sets of daily flight schedules to present proof-of-concept of the usefulness of the optimization methodologies proposed in the previous chapter. The created sets of data follow the distribution of the real flight delays situations of European airspace. Generated data are describing days when high congestion occurs on the airports and follow the collection of real facts. On a typical July day, there are around 30 000 flights across European airspace [25]. There are currently 153 large airports in Europe [1]. In order to represent a typical summer situation on a smaller scale, we create a set of 10 airports and to each of them we assign a number of flights which are supposed to land on it that day. For datasets 1 to 4, 1820 flights are considered, for datasets 5 to 9 it is 1800 flights and 1780 flights for the rest. This gives us an advantage to study the problem under different realizations of congestion.

The capacity values used for all the airports are chosen to be at the "infeasibility border" during peak hours, i.e. their small perturbation on the conservative side may lead to infeasibility of the overall problem. Firstly, we consider the landing capacities as deterministically given values which differ only as a function of airports. The process of obtaining weather-front induced capacities for the robust formulation of the problem will be explained in the next section. Each of the airports has different size and thus different amount of flights should be assigned to land on each of them. Airports 1 and 2 are the biggest ones and hence approximately 15% of flights are assigned to both of them. Next, airports 3, 4 and 5 are middle sized, and consequently 12%, 11% and 9% of flights are scheduled to land there. Finally, 8% of flights are landing at airports 6, 7, 8, and 7% are assigned to last two airports 9 and 10. The average distribution of flights as a function of airport is represented by histogram on the left side of Figure 4.1.

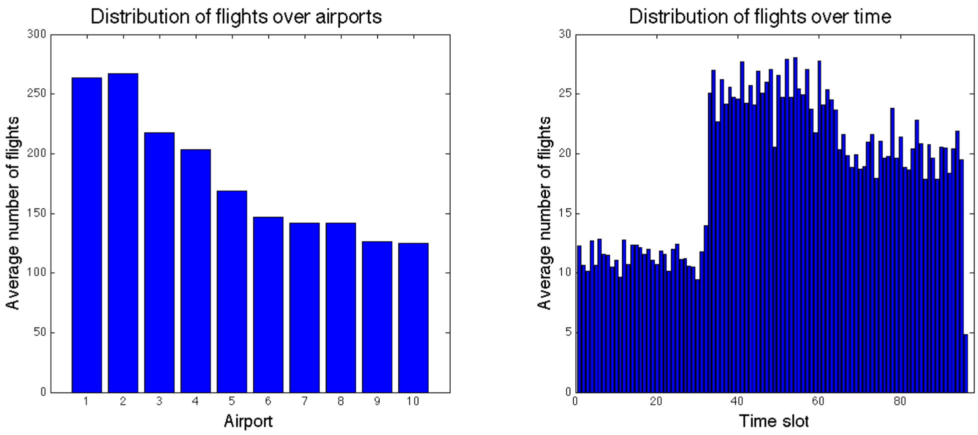

Figure 4.1: Average flight traffic of all the data sets.
Left: Flight traffic as a function of airport; **Right:** and a function of time.

The average distribution of flights as a function of landing time for our data sets can be seen on Fig. 4.1 (Right). We assume, that between midnight and 8 AM only a little part of the flights are scheduled to land and the congestion starts to occur afterwards [29]. The actual distribution of flights over time for each data set is shown on Figure 4.2.

The flights are not allowed to be delayed for more than three hours and the maximum allowed delay times are randomly distributed between 8 and 12 time slots. The average

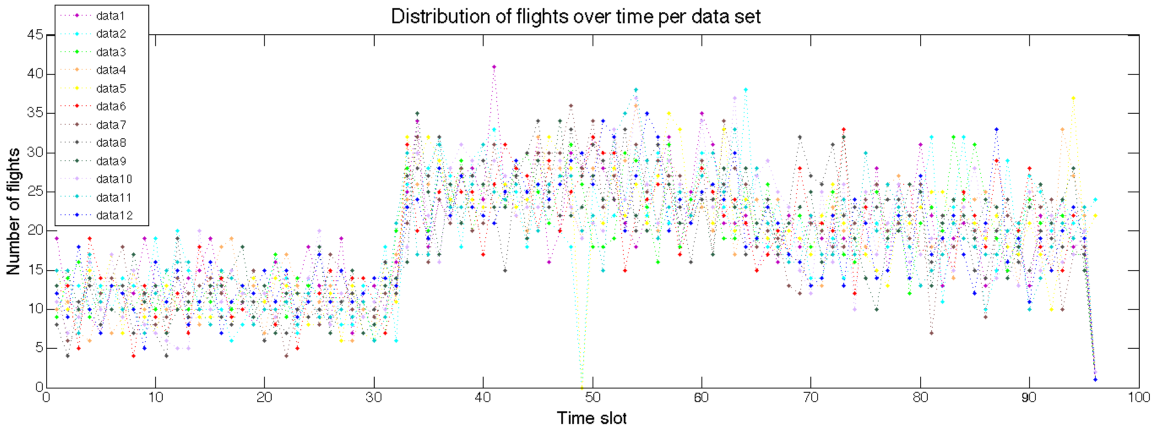

Figure 4.2: Distribution of flights over time for each data set.

price of a delay per one time slot is set to be 480. Since cancelation cost is usually much higher than delay cost [40], without the loss of generality we set the cancelation cost for each flight $i \epsilon F$ to be $(D_i + 2) * dCost_i * 5$, where D represents maximum allowed delay and $dCost$ is cost of delay per time slot. Moreover, for every data set we assume 20% of all the flights to have a successor and we generate the set of pairs of successive flights as well as the set of time-coupled combinations.

To solve the robust and deterministic models, we use an open source software GLPK, the GNU Linear Programming Kit. It is a software package intended for solving large-scale linear and mixed integer programming problems. GLPK uses the branch-and-bound algorithm together with Gomory's mixed integer cuts for (mixed) integer problems [2]. In order to optimize the computational time and effort, we preprocessed the data accordingly in MATLAB. The outline of the code solving the nominal ATFM problem can be find in the appendix.

After performing a set of computational experiments, we wish to answer following questions:

1. Are the running times convenient for practical implementation?

2. What is the price of robustness of the single-stage formulation?

3. How different are the robust single-stage and two-stage solutions?

4. Is the performance of second-stage solution better than the deterministic case with incorporating actual weather scenario?

4.2 Scenario configuration

In this section, we explain how the data to test different scenarios and approaches were created and implemented in the GLPK optimization model. The experiment was run on Mac Book Air 2013, OS X 10.9.3., with processor 1.3 GHz Intel Core i5 and RAM 4 GB 1600 MHz DDR3. GLPK was called from the terminal.

4.2.1 Weather scenarios

In order to test the behavior of robust models, two weather scenarios were generated for each data set and studied separately. Each of them can have different realizations, meaning that the arrival time, duration or strength can have different values. For example, considering that we know, that there will be a storm affecting the airspace around airports 1 and 4 at some time between 3 PM and 5 PM, we create a set of all possible realizations of this storm. The time of arrival, duration and the strength of each weather-front are randomly generated from appropriate intervals and represent the situations, which can possibly occur on a typical summer day. Following the theory developed in Chapter 3, each weather-front results in the reduction of landing capacities during its realization. Hence, the uncertainty set of weather-front induced landing capacities was created for each data set and each weather scenario.

4.2.2 Single-stage robust optimization

The purpose of robust optimization is to find an optimal solution, which will stay feasible under any realization of uncertainty set (see Chapter 2). To fulfill this requirement, the minimum landing capacity for every airport and every time slot is found among all the weather realizations (for explanation see Chapter 3). Consequently, a set of "worst-case" landing capacities is obtained and considered as a deterministic input for the optimization model. For the weather scenario 1 the worst-case capacity reduction corresponds to 7% of deterministic capacities and 9,74% for weather scenario 2. The results of this approach are shown and discussed in the Section 4.3.

4.2.3 Two-stage robust optimization

While implementing the two-stage robust problem, we allow the possibility of updating the schedule during the day. We want to study the impact of this approach on the overall cost of the problem and we desire to show, that the total cost decreases, if further changes in schedule are allowed to be implemented accordingly during the day. We allow the decision maker to change the schedule at 10 AM under assumption, that he is provided by a new, updated information about possible weather realizations occurring later that day. However, while changing the schedule, he is not allowed to reschedule (hold on the ground) those flights, which are currently already in the air. This is protected by imposing an additional constraint to the problem formulation.

The adaptive solutions were implemented on a rolling horizon basis. We firstly compute the single-stage decisions of the robust problem. Subsequently, we fix the scenario from the uncertainty set and use this in the second stage as the actual capacity profile. These scenarios are then assumed as capacity profiles which certainly materialized. Namely, for weather scenario 1 we fix the capacity profiles corresponding to the reduction of 5,4% and 3,7% of deterministic cases. For weather scenario 2 we fix in the second stage capacity profiles corresponding to 7,8% and 4,7% reductions of deterministic capacities. We then re-optimize the second stage decisions under the input of this fixed capacity profile for all the scenarios and data sets. The analysis of results is presented in Section 4.3.

4.2.4 Computing the cost of deterministic solution with fixed weather realization

In order to illustrate the situation in the airspace when one of the weather scenarios arise but the schedule was produced with deterministic capacities, we perform the following experiment. We produce a schedule for a given data set with deterministic capacities. Subsequently, we fix the scenario, which actually materialized and we compute the cost of the deviation of the deterministic schedule. Since the capacities have decreased, following the deterministic schedule, some of the flights have to be held in the air until the next landing spot becomes available. This is computed by comparing the vectors of assigned flights and actual (decreased) capacities. If the capacity of an airport k at some time t became lower than the amount of flights which were assigned to land on it, the algorithm finds next possible empty landing spot and assign the flight to it. As a result, it returns the number of deviations (number of flights which couldn't land as scheduled). For each of those, the amount of time which it had to be hold in the air as well as the airborne costs were computed. This was done in two ways.

Firstly, we assume, that the controllers are not re-assigning the landing spots to flights optimally, but randomly. This can happen in reality, when a big congestion occurs unpredictably or when the controllers are not allowed to change the planned landing slots, for example due to the competition of the air industry. In this case, the

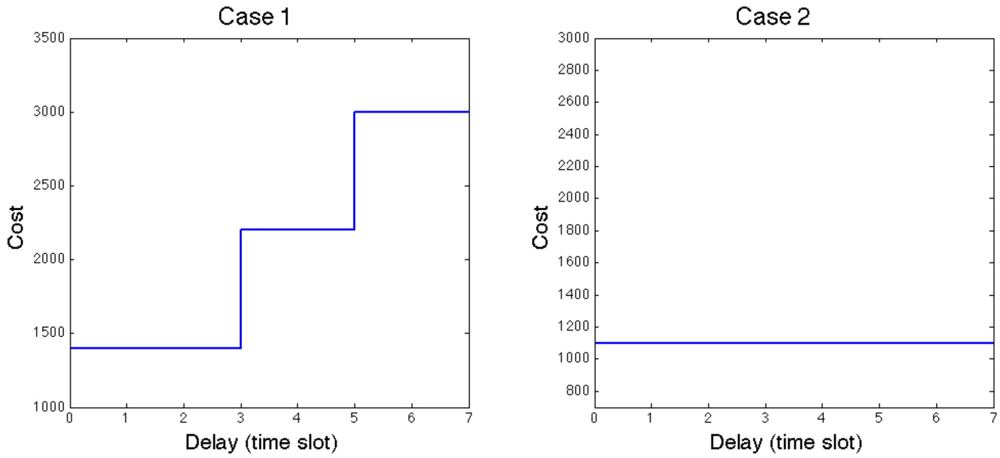

Figure 4.3: Cost of delay as a function of time.
Left: stepwise approximation of exponential behavior; **Right:** constant cost.

airborne costs increase as a function of time, meaning the longer is an aircraft hold in the air, the more costly becomes an additional delay [17]. For simpler computation we approximate this exponential behavior by a step function (Fig. 4.3 Left). In the second case (Fig. 4.3 Right), more sophisticated re-assigning of flights by controllers is taken into account. If a flight can not land as scheduled and the waiting time for the next available landing spot is too long, controller would optimally choose another flight landing later on and hold that one in the air so the one scheduled previously would land on its spot. By doing so on a rolling horizon basis, the air holding costs decrease (see next section) while naturally, the safety stays preserved.

4.3 Analysis of results

In this section, analysis of results coming from our computational experiments is presented. Table 4.1 shows the performance of single-stage robust model for two different weather scenarios. To clarify proposed theoretical expectations, robust solutions have higher costs in all of the cases. Moreover, comparing the values of scenario 1 and scenario 2, the higher the capacity reduction, higher becomes the cost of optimal solution. Furthermore, even while considering the fact that our instances are of much smaller size than real situations, the running times are still promising for practical implementation.

In order to study the characteristics of robust schedules, we define a new quantity-schedule deviation. It is computed for each solution as an amount of time slots, which were delayed altogether for all the flights. This quantity governs the price of robustness in that the higher the schedule deviation, higher is the value of optimal cost for the

	Deterministic		Robust- scenario 1		Robust- scenario 2	
data set (# of flights)	Cost	Time (sec.)	Cost	Time (sec.)	Cost	Time (sec.)
data1 (1820)	142740	3,3	718650	3,9	1139360	4,0
data2 (1820)	89870	3,0	355540	3,2	929050	3,2
data3 (1820)	180850	3,2	750960	3,8	1032810	3,6
data4 (1820)	84500	3,1	747950	3,8	901480	3,5
data5 (1800)	72680	3,5	607970	3,6	921150	3,4
data6 1800)	94830	3,4	600510	3,8	1155480	4,5
data7 (1800)	69560	3,5	543660	3,4	1219760	3,8
data8 (1800)	218080	3,4	871430	4,0	1426010	3,7
data9 (1800)	67200	3,4	387320	3,8	756560	3,2
data10 (1780)	94980	3,0	516060	3,7	946920	3,4
data11 (1780)	156960	3,5	977530	3,7	1108840	3,2
data12 (1780)	114320	3,5	586790	3,3	1149130	3,3

Table 4.1: Computational experience with single-stage robust model comparing to deterministic cases.
Weather scenario 1 corresponds to 7% capacity reduction of deterministic case, scenario 2 corresponds to 9,7% reduction.

robust solutions (Figure 4.4 Left). Also, an approximately linear relationship can be seen between the schedule deviation and values of objective costs (represented by the green best linear fit line). The value of correlation coefficient between these two quantities is for our experiments 0,87, which implies statistical significance and proves theoretical results developed earlier in our work.

To provide a better picture of the price of robustness, we fix a particular realization of uncertainty set and use this to compute new deterministic costs. This represents in reality the situation when schedule was produced with deterministic approach whereas weather-front occurred during the day. Since not all of the flights were able to land as scheduled, some of them had to be hold in the air. We propose two ways of computing new deterministic costs describing this situation in Section 4.2.4. The realizations which have been fixed in our experiment correspond to 5,4% capacity reduction for costs coming from scenario 1, and 7,8% for scenario 2. Meanwhile the first way gives higher deterministic costs than robust objective values in most of the cases, the second way behave oppositely (Fig. 4.4 Right). The green line plots the new deterministic costs and divides the graph into two sections. The results on the left side of this line have higher robust cost than new deterministic cost and therefore in those cases, it is more

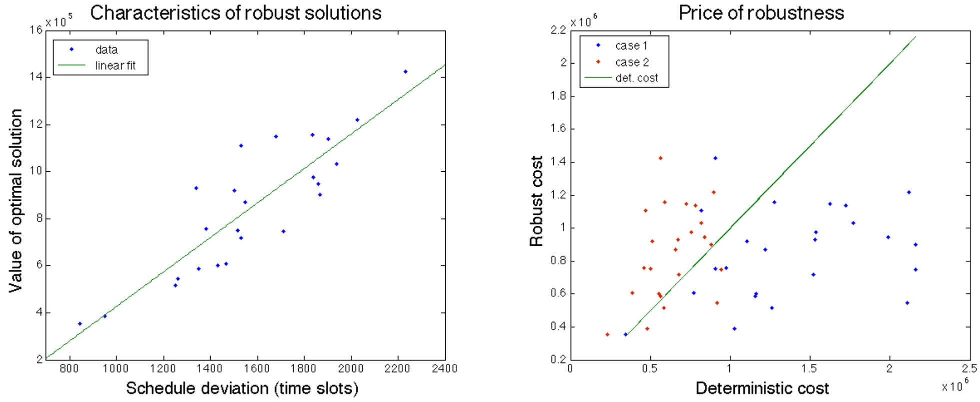

Figure 4.4: Characteristics of robust solutions.
Left: Price of robustness as a function of schedule deviation. **Right:** Price of robustness as a function of new deterministic cost with fixed weather realization for both cases.

favorable (cost wise) to produce schedules with deterministic approach. For the results on the right side, the opposite is accurate. If α is the capacity reduction for a particular realization of weather uncertainty, then the fixed realizations correspond in both cases to approximately $\frac{\alpha_{average}-\alpha_{min}}{2}$. Since these realizations are on the conservative side of the uncertainty set and the price of robustness is high in a lot of cases, we conclude that the single-stage robust approach is costly in general. However, in the cases of high congestion and unfavorable weather realization it produces better results and therefore leads to reduction of costs (Figure 4.4 Right).

As expected from theoretical perspective, second-stage solutions give better results in terms of cost value of objective (Table 4.2). The adaptive (second-stage) solutions were implemented on a rolling horizon basis. Firstly, the robust worst-case solution was computed. Afterwards, two realizations of uncertainty set were fixed (those corresponding approximately to $\frac{\alpha_{average}-\alpha_{min}}{2}$ and $\alpha_{average}$) and used together with first-stage decision variables as an input for second-stage re-optimization (read more in Section 4.2.3). Here we want to remark, that the decision maker is in reality able to choose the uncertainty realization according to an updated information. We chose those two to represent the reality in a following way: a) the first one is on the conservative side, b) the second one is the average capacity reduction. In the case of a), the realization was fairly close to the worst-case and the solution remains optimal for approximately 75% of all the possible realizations. For the case b), the solution stays feasible under the "better" half of uncertainty set. This experiment was done for both of the weather scenarios and all the data sets (Table 4.2). Since it is unlikely that the exact worst-case realization reveals in reality, we consider this approach to be more realistic for practical implementation.

To begin with analysis of second-stage solutions, firstly the objective costs were compared to single-stage robust costs. The values of objective were improved and in some of the cases by over 50% (Fig. 4.5). As expected, solutions for second realization produce lower values of objective and therefore are less costly. Also, for the more congested situations we get a higher value of cost reduction. Furthermore, there is again an approximately linear relationship between the schedule deviation and objective costs, with coefficient of correlation equal to 0,93 (Fig. 4.6 Left). In terms of running times, the second-stage model is approximately two times more expensive than the single-stage (Fig. 4.6 Right). This is due to the fact, that in order to produce adaptive solution, we firstly compute the single-stage solution and re-optimize this later on.

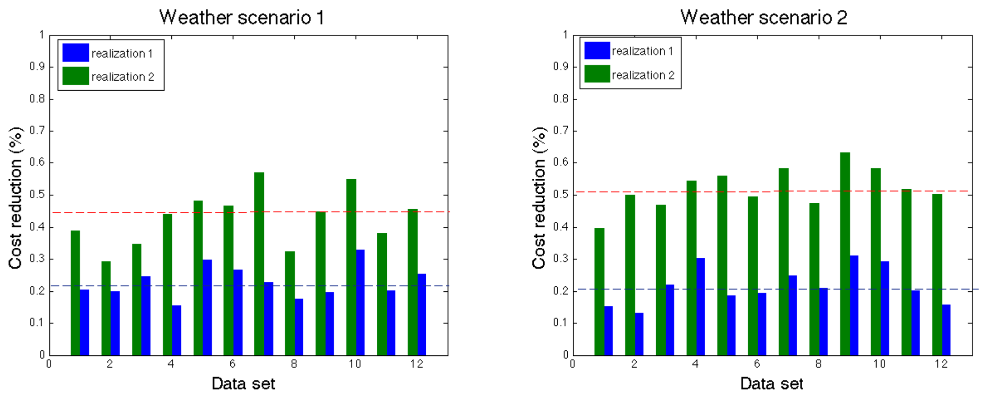

Figure 4.5: Percentual cost reduction.
Percentual cost reduction of second-stage solutions compared to first-stage solutions; **Left:** for weather scenario 1; and **Right:** weather scenario 2. Dashed lines plot the values of median (blue for realization 1 and red for realization 2).

In order to provide a better understanding of the real situation, we compare the price of adaptive solutions to the price of the deterministic ones. The new deterministic cost is obtained by fixing the same weather realization and following the previously mentioned approach (see Section 4.2.4). Figure 4.7 depicts the different natures of two proposed approaches. While there is an approximate linear relation between number of flights, which couldn't land as scheduled and new deterministic costs given by approach two, approach one follows an exponential trend (Fig. 4.7).

Let us remind the consequences of using the deterministic schedule in situations, when unfavorable weather-front occurs. Following the regulations, the lower the visibility or the higher the wind, distances of aircrafts in the airspace have to increase (due to safety reasons). This naturally results into lower sector and landing capacities, for every time slot concerning such a setting. Consequently, it happens, that some of the flights

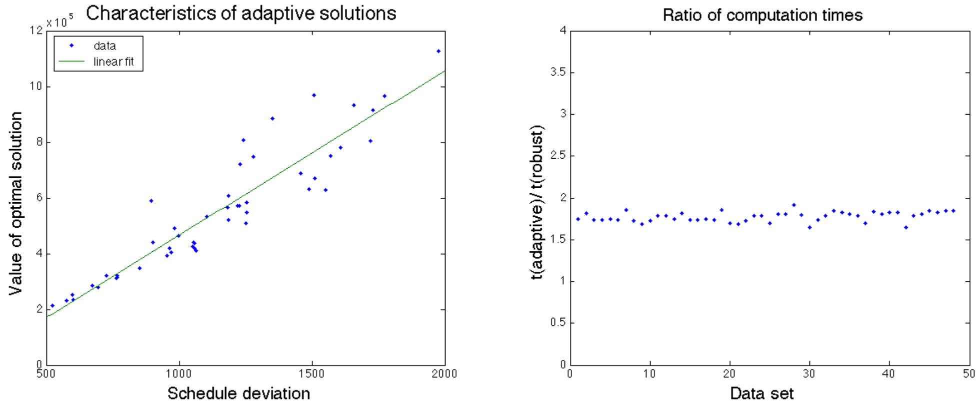

Figure 4.6: Characteristics of adaptive solutions.
Left: Price of robustness as a function of schedule deviation for adaptive problems. **Right:** Ratio of adaptive and robust computation times.

can not land at their scheduled times, because either the airport, or the sectors around the airport are not able to accommodate the same amount of aircrafts as regurarly. Therefore, those planes have to be hold in the air and wait, until the next available landing spot occurs. Since the costs of air holding of aircrafts are significantly higher than ground-holding [8], this may result into extremely high price. Especially in the cases of high congestion, when there are a lot of flights scheduled to land during the same time and the capacity of airport is violated by a big factor. We have previously proposed two ways of computing the costs of such a situation (see Section 4.2.4). The first approach has an exponential behavior, meaning that the air-holding costs increase exponentially by time. This may in reality happen when the controllers don't have time (or tools) to re-optimize the congested scenario and they deal with such a situation "as it comes". On the other hand, the second approach takes into account that the controllers may be able to handle the congested scenario "optimally".

To follow the theoretical background of our work, the second-stage solutions are protected against all the possible realizations of weather, which are better than the realization used in re-optimization. Therefore in reality, using the second-stage robust optimization approach for producing schedules should avoid the described situation. However, since robust methodologies give overly pessimistic results in general, we want to study the profitability of this approach. To do so, the objective costs of second-stage robust solutions are compared to deterministic costs with the same fixed weather realization (referred as "new deterministic costs"). This is done for all the data sets, by fixing two weather realizations for every scenario (one corresponding to $\frac{\alpha_{average}-\alpha_{min}}{2}$ and second one to $\alpha_{average}$) and computed for both new deterministic cost approaches.

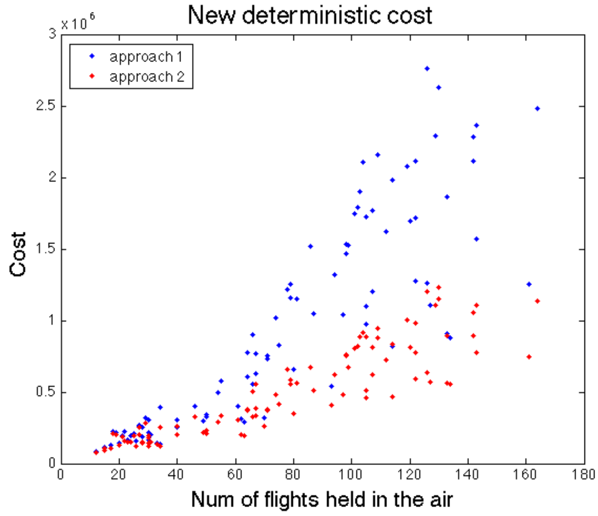

Figure 4.7: New deterministic cost.
Cost of deterministic solutions with fixed weather scenario as a function of number of flights, which couldn't land as scheduled. Blue points represent the values for which the cost was computed by the first approach, red by second approach.

The main result of our experiment is the observation, that for most of the cases, second-stage robust solutions give better results than the new deterministic costs computed by approach one (Fig. 4.8 Left). Mainly, if the actual weather realization belongs to the conservative side of the uncertainty set, the second-stage solutions produce lower costs. Additionally, if the worst-case weather scenario causes high capacity reduction, then it is more likely, that even if the average capacity reduction materialized, second-stage model will still produce results of lower costs than the new deterministic (pink points in Fig. 4.8 Left). However, if the re-optimization is taken in the congested scenario, it is in most of the cases of lower price to use the deterministic approach than second-stage robust (Fig. 4.8 Right). This can be explained followingly. While computing the second stage solution, we re-schedule the flights only once (in the second-stage) and we stay protected against all the possible realizations of new uncertainty set. On the contrary, approach two of computing the new deterministic costs allows controllers to re-optimize flights constantly. The longer is the aircraft hold in the air, the higher priority to land it gets in the re-optimization. That means, that in order to keep the overall costs low, controller is allowed to let the flights with higher priority to land and hold the new ones in the air. By this approach, the number of delayed flights increases, but the air-holding costs won't capture the exponential growth.

To conclude the experiment, we analyze the percentual reduction of costs, which is given by second-stage solutions. Since in reality controllers are usually not able to

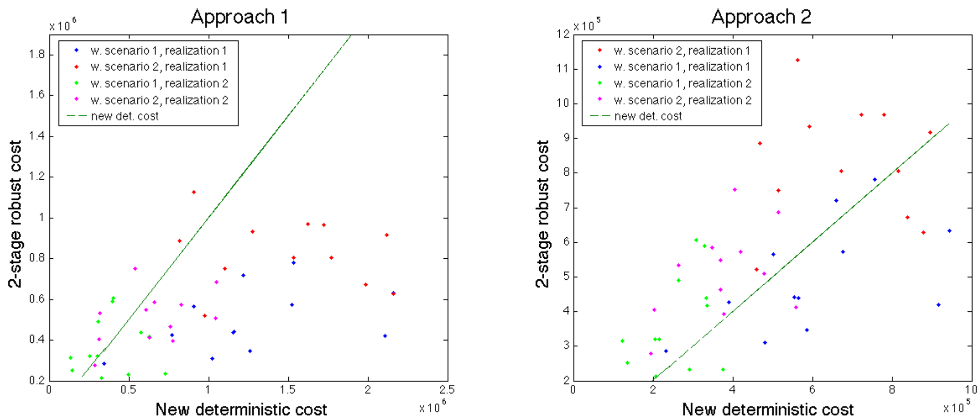

Figure 4.8: Price of second-stage solutions.
Left: Price of second-stage adaptive solutions compared to new deterministic costs computed by approach 1. **Right:** Price of second-stage adaptive solutions compared to new deterministic costs computed by approach 2.

assign flights to land optimally in the congested scenario (due to the competition of air industry), the analysis was done by comparison of results towards the new deterministic costs computed by approach one. The results were computed for all the data sets, weather scenarios and both weather realizations. The following findings were observed:

- The cost of objective improved in average by 22% (Fig. 4.9)
- The total cost decreased in 75% of all the cases (Fig. 4.9)
- Moreover, if the actual weather realization comes from the conservative side of uncertainty set, there is a significant cost reduction (Fig. 4.10):
  - by 55,7% in average for weather scenario one
  - by 37,8% in average for weather scenario two.

While choosing the fixed weather realization as the one corresponding to the average value of capacity reduction, the value of cost reduction decreased. Namely, for the weather scenario 1, the cost is in average 15,4% worse than the new deterministic. This in reality means, that in this case it would be cheaper to hold the aircrafts in the air. However, by doing so, the protection over other possible realizations of uncertainty set is lost. In the case of higher capacity reduction (weather scenario 2), the second-stage solution remains profitable even for the fixed realization being the average one. While still staying protected under the optimistic half of the uncertainty set, it reduces the costs by 10% in average.

Extensive empirical results were presented and discussed in this chapter. They proved theoretical expectations and showed the possibility of their implementation in practice. The results highlight the utility of robust and adaptive solutions. We showed, that under some assumptions, the second-stage robust model leads to significant cost reduction in our computational experiments. Using this knowledge in practice could benefit to profitability of the air traffic industry.

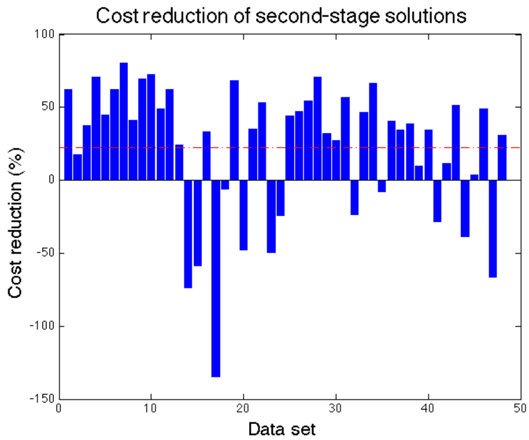

Figure 4.9: Percentual cost reduction of second-stage solutions.
Dashed red line showes the mean value.

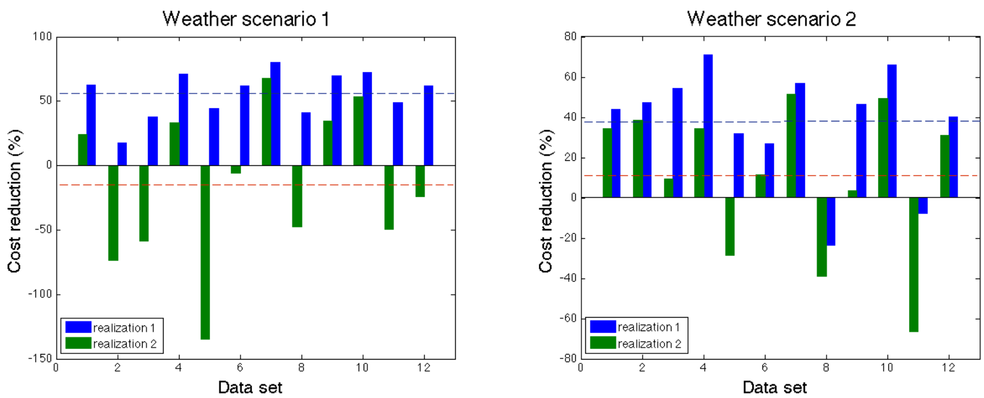

Figure 4.10: Percentual cost reduction of second-stage solutions.
Cost reduction of second-stage solutions compared to new deterministic cost computed by first approach; **Left:** for weather scenario 1; and **Right:** for weather scenario 2.

Scenario 1	Robust		Two-stage, realization 1		Two-stage, realization 2	
data set (# of flights)	Cost	Time (sec.)	Cost	Time (sec.)	Cost	Time (sec.)
data1 (1820)	718650	3,9	572040	6,8	438970	7
data2 (1820)	355540	3,2	285020	5,8	251680	6,2
data3 (1820)	750960	3,8	565840	6,6	490400	6,5
data4 (1820)	747950	3,8	632350	6,6	418270	6,5
data5 (1800)	607970	3,6	426810	6,3	315090	6,3
data6 1800)	600510	3,8	441310	6,6	321040	6,6
data7 (1800)	543660	3,4	420110	6,3	233370	6,4
data8 (1800)	871430	4,0	719630	6,8	589900	6,8
data9 (1800)	387320	3,8	310880	6,4	213940	6,5
data10 (1780)	516060	3,7	346690	6,4	232160	6,3
data11 (1780)	977530	3,7	781270	6,6	606690	6,3
data12 (1780)	586790	3,3	438140	5,9	319750	5,9

Scenario 2	Robust		Two-stage, realization 1		Two-stage, realization 2	
data set (# of flights)	Cost	Time (sec)	Cost	Time (sec.)	Cost	Time (sec.)
data1 (1820)	1139360	4,0	967120	6,8	686590	6,8
data2 (1820)	929050	3,2	806040	5,8	464450	5,9
data3 (1820)	1032810	3,6	804840	6,5	548780	6,5
data4 (1820)	901480	3,5	627840	6,7	411740	6,4
data5 (1800)	921150	3,4	749090	6,1	404570	6,2
data6 1800)	1155480	4,5	933260	7,4	584400	7,4
data7 (1800)	1219760	3,8	916240	6,6	509240	6,8
data8 (1800)	1426010	3,7	1126900	6,6	751610	6,7
data9 (1800)	756560	3,2	521310	5,9	278060	5,9
data10 (1780)	946920	3,4	671060	6,2	394020	6,2
data11 (1780)	1108840	3,2	885490	5,8	533360	5,9
data12 (1780)	1149130	3,3	968050	5,9	572250	6,1

Table 4.2: Computational experience with second-stage robust model.
Single-stage solutions for weather scenario 1 (upper) and scenario 2 (lower) compared to second-stage solutions. For scenario 1, realization 1 corresponds to 5,4% capacity reduction of deterministic case and realization 2 corresponds to 3,7% reduction. For scenario 2, realization 1 corresponds to 7,8% capacity reduction of deterministic case and realization 2 corresponds to 4,7% reduction.

Chapter 5

Conclusion

The problem of assigning optimal and accurate schedules to flights has been long studied by research community. Hence development of integrated, safe and reliable models has become an ambitious and challenging task. Due to enormous impact of the air traffic industry on social welfare and worldwide economy, this issue has to be addressed. We summarize now the overall contributions of our work and point out directions for future research.

5.1 Thesis summary

Our aim in this thesis was to address the problem of the Air Traffic Flow Management, specifically the task of developing optimal schedules, managing flight delays due to dynamic weather conditions and minimizing overall costs. In doing so, we applied the robust optimization framework into the deterministic formulation of the ground holding problem. Our overall aspiration was to show and analyze the profitability of the second-stage robust ATFM model, which allows the changes in schedule on a rolling horizon basis. Our work consists of three key parts: presenting the mathematical theory, incorporating weather-front induced capacities in the formulation of the problem, and computational implementation.

- **Mathematical theory**

As a crucial starting point of our work, we have presented mathematical theory of robust optimization. Robust optimization is a framework of addressing uncertainty in optimization problems in a tractable way. One way of dealing with its overly pessimistic solutions is a multi-stage decision-making approach in which the decisions are produced over time and can be adapted to capture the behavior of uncertainty. We introduced the notion of two-stage robust optimization with right hand side uncertainty, which we later implemented in our model.

- **Problem formulation**

We have introduced two formulations of the ATFM problem, which consider deterministic capacities of the airports. In order to deal with capacity uncertainty, we proposed a way to model an uncertain weather-front propagation. Afterwards, a robust approach is utilized in the deterministic models and a two-stage robust formulation of the ATFM problem with uncertain capacities of arrival airports is proposed. Also, we have showed that the resulting model is just another deterministic instance.

- **Computational implementation**

The proposed methodologies were implemented and solved efficiently using MATLAB and GLPK. In order to present proof-of-concept of the usefulness of the optimization methodologies, twelve data sets of daily flight schedules were created accordingly. Reported empirical results from all the experiments based on the proposed models showed the possible profitability and benefits of two-stage robust approach.

5.2 Directions for future research

We conclude our work by providing some directions for future research, which follow the research done in this thesis.

1. **Implementation of the set packing robust model.**

   In our work, we embedded the robust optimization framework into the set packing formulation of the ATFM problem. This formulation gives an advantage of solving a mixed-integer programming problem by LP and under some assumptions assures the integrality of decision variables. Hence computational implementation could possibly yield improvements of the solutions.

2. **Extensions of the formulations.**

   In this thesis, we investigated and implemented a nominal formulation of the ATFM problem. This formulation does not take into account the capacities of departure airports, neither sector capacities nor possible re-routings of flights. In particular, there is a significant tractability challenge in considering rerouting in the presence of capacity uncertainty and from the practical perspectives of ATFM algorithms, this complication should be addressed.

3. **Computational experiments with real data sets.**

   Due to a notable difficulty of accessing an actual air traffic data, all of the computational experiments in this thesis were done on self-generated data sets. Moreover, in order to prevent the computational possibility of our resources, data sets represented much smaller instances than those in reality. Although all the data were generated in a way that they describe real situations, in order to prove the adequacy of proposed methodologies a natural need of running the experiments on the actual air traffic data arises. This task would involve the analysis of historical flight data and implementation of the proposed robust and adaptive models on large-scale instances.

References

[1] Airportdatabase.net.

[2] https://en.wikibooks.org/w/index.php?title=glpkoldid=3009280.

[3] The operations network (opsnet) database.

[4] L.F. Escudero A. Agustín, A. Alonso-Ayuso and C. Pizarro. Mathematical optimizationg models for air traffic flow management: A review 1.

[5] A. Nemirovski A. Ben-Tal. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[6] A. Nemirovski A. Ben-Tal, L. El Ghaoui. *Robust Optimization*. Princeton University Press, 2009.

[7] E. Guslitzer A. Nemirovski A. Ben-Tal, A. Goryashko. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99:351–376, 2004.

[8] D. Bertsimas B. Vranas and A. Odoni. The multi-airport ground-holding problem in air traffic control. *Massachusetts Institute of Technology, Cambridge, Massachusetts*, 1993.

[9] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

[10] D. Bertsimas and C. Maes. On the complexity of constructing adaptive policies in two-stage linear optimization. 2011.

[11] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. 2002.

[12] D. Bertsimas and S. Stock. The multi airport flow management problem with en route capacities. Technical Report 3736-94 MSA, Massachusetts Institute of Technology, 1994.

[13] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[14] L. Bianco. Trends in transportation system and relation with air transportation demand. In *Advanced Workshop in Air Traffic Control.*, pages 2–6, 1995.

[15] Constantine Caramanis. *Adaptable Optimization: Theory and Algorithms*. PhD thesis, Massachusetts Institute of Technology, 2006.

[16] Walker Chris. All-cause delay and cancellations to air transport in europe- 2014. Technical report, The European Organisation for the Safety of Air Navigation (EUROCONTROL), 2015.

[17] Performance Review Commission. Performance review report : An assessment of air traffic management in europe during the calendar year 2014. Technical report, The European Organisation for the Safety of Air Navigation (EUROCONTROL), May 2015.

[18] A. Odoni D. Bertsimas, G. Lulli. The air traffic flow management problem: An integer optimization approach. *Springer-Verlag*, pages 34–46, 2008.

[19] C. Caramanis D. Bertsimas, D. B. Brown. Theory and applications of robust optimization. *Society for Industrial and Applied Mathematics Review*, 53:464–501, 2011.

[20] D. Pachamanova D. Bertsimas and M. Sim. Robust linear optimization under general norms. *Operations Research Letters 32*, pages 510–516, 2004.

[21] G. Lulli D. Bertsimas and A. Odoni. An integer optimization approach to large-scale air traffic flow management. *Operations Research*, 59(1):211–227, January-February 2011.

[22] S. Stock Patterson D. Bertsimas. The traffic flow management rerouting problem in air traffic control: A dynamic network flow approach. *Transportation Science*, 34(3):239–255, 2000.

[23] V. Goyal D. Bertsimas. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, 2011.

[24] G. Dunn. Heathrow, gatwick expansion costs under-estimated: Davies.

[25] EUROCONTROL. Eurocontrol flight movement forecasts to 2019: go east, young man.

[26] Federal Aviation Administration (FAA). Aviation system performance metrics (aspm).

[27] Shubham Gupta. *A Tractable Optimization Framework for Air Traffic Flow Management Addressing Fairness, Collaboration and Stochasticity*. PhD thesis, Massachusetts Institute of Technology, 2012.

[28] R. Hoffman and M.O. Ball. A comparison of formulations for the single-airport ground-holding problem with banking constraints. *Operations Research*, 48(4):578–590, 2000.

[29] A. Jacquillat and A. R. Odoni. An integrated scheduling and operations approach to airport congestion mitigation. Massachusetts Institute of Technology.

[30] P. Kouvelis and G. Yu. Robust discrete optimization and its applications. *Kluwer Academic Publishers*, 1997.

[31] P. Dell'Olmo L. Bianco and S. Giordani. Scheduling models and algorithms for air traffic management. In *Modeling and Simulation in Air Traffic Management*, pages 139–167. Springer-Verlag, 1997.

[32] M. Laguna. Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. *Management Science*, 44(11):101–110, 1998.

[33] G. Rinaldi L.Bianco and A. Sassano. *A combinatorial optimization approach to aircraft sequencing problem*, pages 323–339. Springer-Verlag, 1987.

[34] G. Lulli and A. Odoni. The european air traffic flow management problem. *Transportation Science*, 41(4):431–443, 2007.

[35] M. Dresner M. Hansen K. Neels A. Odoni E. Peterson L. Sherry A. Trani M. Ball, C. Barnhart and B. Zou. Total delay impact study. Technical report, National Center of Excellence for Aviation Operations Research, 2010.

[36] A. Schöbel M. Goerigk. Algorithm engineering in robust optimization.

[37] Michel Minoux. On 2-stage robust lp with rhs uncertainty: complexity results and applications. *Journal of Global Optimization*, 49:521–537, 2011.

[38] A. R. Odoni. The flow management problem in air traffic control. *Springer-Verlag*, pages 269–288, 1987.

[39] Bureau of Transportation Statistics U.S. Dept. of Transportation. Airline on- time statistics.

[40] F. Rossi and S. Smriglio. A set packing model for the ground holding problem in congested networks. *European Journal of Operational Research*, 131:400–416, 2001.

[41] D. Bertsimas S. Gupta. Multistage air traffic flow management under capacity uncertainty: A robust and adaptive optimization approach. 2011.

[42] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

[43] M. Terrab and A.R. Odoni. Strategic flow control on an air traffic network. *Operations Research*, 41:138–152, 1993.

[44] U. Völkers and D. Böhme. Dynamic planning for airport surface traffic management. In *Advanced Workshop in Air Traffic Control.*, pages 2–6, 1995.

Appendix A

Appendix

For practical computations, the following GLPK code was implemented:

```
% defining parameters
param nTimes;
param nFlights;
param nSuccFlights;
param nAirports;
param nCoupledCombinations;
param maxNumArrs;
param capacity {k in 1..nAirports, t in 1..nTimes};
param cCost {i in 1..nFlights};
param dCost {i in 1..nFlights};
param aTime {i in 1..nFlights };
param maxDelay {i in 1..nFlights };
param arrFlight {k in 1..nAirports, i in 1..maxNumArrs};
param preFlight {i in 1..nSuccFlights };
param succFlight {i in 1..nSuccFlights };
param delta {i in 1..nSuccFlights };
param cBreak {i in 1..nSuccFlights };
param tPrime {m in 1..nCoupledCombinations };
param tPrimePrime {m in 1..nCoupledCombinations };
param numArr {k in 1..nAirports};
```

```
% desicion variables
var y {i in 1..nFlights, t in 1..nTimes }, binary;

% objective function
minimize obj: sum {i in 1..nFlights } ((1-sum { t in aTime[i]..
(aTime[i] + maxDelay[i] ) } y[i,t] ) * ( cCost[i] +
dCost[i]* aTime[i] ) + dCost[i]* ( sum { t in aTime[i]..
(aTime[i] + maxDelay[i]) } y[i,t] *t - aTime[i] ) );

% constraints

s.t. first {i in 1..nFlights }:
sum { t in aTime[i]..(aTime[i] + maxDelay[i] ) } y[i,t] <= 1;

s.t. second {k in 1..nAirports, t in 1..nTimes }:
sum { i in 1..numArr[k] }y[arrFlight[k,i],t] <= capacity[k,t];

s.t. third {m in 1..(nSuccFlights-1), ind in cBreak[m]..
(cBreak[m+1]-1) }:
sum{t in tPrime[ind]..nTimes} y[preFlight[m],t] +
sum{t in 1..tPrimePrime[ind]} y[succFlight[m],t] <= 1;

s.t. thirdLast {ind in cBreak[nSuccFlights]..
nCoupledCombinations}:
sum{t in tPrime[ind]..nTimes} y[preFlight[nSuccFlights],t] +
sum{t in 1..tPrimePrime[ind]} y[succFlight[nSuccFlights],t] <= 1;

s.t. fourth {i in 1..nFlights }:
sum {t in 1..(aTime[i]-1) } y[i,t] = 0;

s.t. fifth {i in 1..nFlights }:
sum {t in (aTime[i] + maxDelay[i]+1)..nTimes } y[i,t] = 0;

solve;
end;
```