

ANNA ACHENBACH

PREDICTIVE ANALYTICS IN AIRLINE OPERATIONS:  
APPLICATION OF MACHINE LEARNING FOR ARRIVAL TIME  
AND FUEL CONSUMPTION PREDICTION

Dissertation in partial fulfillment of the requirements for the degree of  
Doctor of Business and Economics (Dr. rer. pol.)

WHU - Otto Beisheim School of Management  
Vallendar

April 30, 2018

First Advisor:  
Prof. Dr. Stefan Spinler

Second Advisor:  
Prof. Dr. Arnd Huchzermeier

Anna Achenbach: *Predictive Analytics in Airline Operations: Application of Machine Learning for Arrival Time and Fuel Consumption Prediction*,  
© April 30, 2018

*To Klara and Max*



## ABSTRACT

---

This dissertation investigates the application of machine learning to improve decision making in airline operations. In an introductory overview, we<sup>1</sup> discuss the airline industry and the challenges of decision making in airline operations: e. g., complex IT infrastructure, interconnected resources, delay management, fuel price volatility and future environmental regulation.

Machine learning can efficiently integrate a large volume of data from a variety of data sources and formats to generate accurate predictions. To assess the viability of using machine learning models for decision making in airline operations, we develop a model based on linear regression and gradient boosting to predict aircraft arrival time. Furthermore, we integrate cost index optimization to model the impact of aircraft speed on arrival time. While we find that machine learning can improve prediction accuracy by more than 30 %, the optimal cost index varies according to fuel cost, flight distance and delay costs. We propose an overall reduction in cost index for short-haul flights to reduce fuel cost while maintaining punctual operations.

Arrival time predictions serve as input for daily operations planning and control. For network carriers accurate arrival time predictions are key for efficient hub operations. In a next step, we focus on aircraft arrival time prediction for intercontinental flights. We analyze the accuracy of en-route weather data provided by the flight plan, generate features based on en-route weather data and integrate them in our prediction model. We evaluate three machine learning models: linear regression, random forest and gradient boosting. Through our approach, we can assess the impact of en-route weather data. Overall, an increase in prediction accuracy of 25 % is achieved. By including en-route weather data, prediction accuracy is improved by 5 %. Our model outlines the essential features for intercontinental arrival time predictions and assess the value of en-route weather data. Furthermore, we outline organizational challenges in implementing predictive analytics.

Future environmental regulations are a challenge for the airline industry. From 2020 onwards net growth in CO<sub>2</sub> emissions is prohibited. Thus, airlines need to focus on initiatives to limit fuel consumption. We develop a prediction model for fuel consumption considering ten different aircraft types. Our results show that fuel consumption can be improved by more than 30 % for short and long-haul flights resulting in an annual reduction of 10.5 million € in fuel costs and a reduction on green house gas emissions of 66.000 tons. Furthermore, we assess the pilots' willingness to integrate our prediction model in their fuel decision. Our analysis shows that high prediction accuracy, model understanding and a long testing phase are essential for acceptance of the prediction model.

The main implication is that decision making in airline operations can be substantially improved through machine learning. Therefore, more prediction

---

<sup>1</sup>The term "we" refers to the authors of the respective chapters as denoted at the beginning of each chapter. For the abstract, this refers to the authors of Achenbach and Spinler (2018a,b), Achenbach et al. (2017).

models should be developed to enhance data-driven decision making to increase the efficiency in airline operations. Furthermore, more accurate decisions can decrease operating costs and the environmental footprint of airlines. Lastly, organizational challenges when implementing predictive analytics need to be considered to ensure the acceptance and usage of new tools.

## ACKNOWLEDGMENTS

---

This quasi-cumulative dissertation was prepared at the Kühne Institute for Logistics Management at WHU – Otto Beisheim School of Management between 2015 and 2018.

During the past month I have been looking forward to the moment when I would write this brief, albeit important, section of my thesis. First and foremost, I want to thank my professor for his encouragement throughout my doctoral studies. Prof. Dr. Stefan Spinler—in his role as my main advisor at the Otto Beisheim School of Management— has supported me with his continual guidance, patience, and motivation throughout my research project. He not only created an engaging working environment and allowed for tremendous flexibility, but he also encouraged a very preference-based selection of research areas and provided precise methodological advice for each project.

I would also like to thank Prof. Dr. Arnd Huchzermeier for serving in the capacity of my second advisor and for providing helpful perspectives on the topic of this thesis.

I am extremely grateful to the entire team at the Kühne Institute for Logistics Management and the team at the Production Management department. I greatly appreciate a number of helpful discussions on methodological approaches with Maximilian Burkhardt, Andreas Faber, Christian Haehl, Daniel Ringbeck and, especially, Dmitry Smirnov. I would also like to thank my colleagues at Lufthansa, Stanislaw Schmal and Fabian Werner, for their business-centered perspective on my research topics, their ideas and feedback.

Last, but not least, it is unlikely I could have written such a dissertation without the support of my family and friends. My husband, Max, has motivated and encouraged me throughout the whole process. His support in the past years has helped me to overcome any lows and share the highs during the dissertation project. Having a baby while writing my dissertation has certainly been a challenge. I want to thank my parents for all their support in taking care of Klara. I also want to thank Klara for being such a wonderful baby. Finally, I want to thank my American friend Rachel Krebs for all her effort in proof reading my papers and improving my writing.



# CONTENTS

---

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	The Airline Industry . . . . .	1
1.2	Operational Challenges in the Airline Industry . . . . .	2
1.2.1	Decision Making . . . . .	2
1.2.2	Delay management . . . . .	3
1.2.3	Eco-Efficiency . . . . .	4
1.2.4	Fuel Price Volatility . . . . .	5
1.3	Machine Learning . . . . .	5
1.4	Contributions of this Work . . . . .	6
<b>2</b>	<b>ARRIVAL TIME PREDICTION AND COST INDEX OPTIMIZATION FOR SHORT-HAUL FLIGHTS</b>	<b>9</b>
2.1	Introduction . . . . .	9
2.2	Literature review . . . . .	11
2.2.1	Airline cost index . . . . .	12
2.2.2	Arrival time prediction . . . . .	12
2.2.3	Machine learning . . . . .	14
2.3	Model formulation . . . . .	15
2.3.1	Target variable . . . . .	15
2.3.2	Predictor variables . . . . .	16
2.3.3	Predictive model . . . . .	18
2.3.4	Tuning of hyperparameters . . . . .	19
2.4	Aircraft performance data and model . . . . .	20
2.4.1	Cost Index data . . . . .	20
2.4.2	Optimization of Cost Index model . . . . .	22
2.5	Results . . . . .	24
2.5.1	Prediction results . . . . .	24
2.5.2	Optimization results . . . . .	27
2.5.3	Managerial implications . . . . .	31
2.6	Conclusion and Outlook . . . . .	31
2.6.1	Future research . . . . .	32
<b>3</b>	<b>INTERCONTINENTAL ARRIVAL TIME PREDICTION</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Literature review . . . . .	36
3.2.1	Arrival time prediction . . . . .	36
3.2.2	En-route weather data . . . . .	37
3.2.3	Prediction methodology . . . . .	38
3.3	Assessing data accuracy of en-route weather data . . . . .	38
3.4	Model formulation . . . . .	40
3.4.1	Target feature and baseline . . . . .	40
3.4.2	Features for prediction . . . . .	41
3.4.3	Predictive model . . . . .	43

3.5	Results . . . . .	44
3.5.1	Prediction results without weather data . . . . .	44
3.5.2	Prediction results with weather data . . . . .	46
3.5.3	Managerial implications . . . . .	50
3.5.4	Implementation challenges . . . . .	51
3.6	Conclusion and Outlook . . . . .	51
3.6.1	Future research . . . . .	52
<b>4</b>	<b>OPTIMIZING CONTINGENCY FUEL LOAD THROUGH PREDICTIVE ANALYTICS</b>	<b>53</b>
4.1	Introduction . . . . .	53
4.1.1	Background . . . . .	53
4.1.2	Legal fuel requirements and industry practice . . . . .	55
4.2	Literature review . . . . .	56
4.2.1	Fuel burn prediction . . . . .	56
4.2.2	Fuel efficiency . . . . .	57
4.2.3	Implementation challenges of advanced analytics . . . . .	57
4.3	Data . . . . .	58
4.3.1	Dependent variable . . . . .	59
4.3.2	Features . . . . .	59
4.4	Methodology . . . . .	61
4.4.1	Baseline . . . . .	61
4.4.2	Fuel consumption prediction . . . . .	62
4.5	Results . . . . .	63
4.6	Cost to carry estimates . . . . .	67
4.7	Acceptance of prediction model by pilots . . . . .	70
4.8	Conclusion and Outlook . . . . .	72
<b>5</b>	<b>SUMMARY AND OUTLOOK</b>	<b>75</b>
5.1	Summary . . . . .	75
5.2	Outlook . . . . .	78
<b>A</b>	<b>APPENDIX TO CHAPTER 2</b>	<b>79</b>
A.1	Detailed list of features . . . . .	79
A.2	Ensemble model . . . . .	81
A.3	Example results . . . . .	81
A.4	Aircraft Performance Data for OD Pairs . . . . .	82
<b>B</b>	<b>APPENDIX TO CHAPTER 3</b>	<b>85</b>
B.1	Evaluation criteria . . . . .	85
B.2	Dataset structuring . . . . .	85
<b>C</b>	<b>APPENDIX TO CHAPTER 4</b>	<b>87</b>
C.1	Linear Regression . . . . .	87
C.2	Hyperparameter settings . . . . .	87
C.3	Results of SLM model . . . . .	88
C.4	Interview Outline . . . . .	88
	<b>BIBLIOGRAPHY</b>	<b>89</b>

## LIST OF FIGURES

---

Figure 1.1	Global flight passengers and air freight . . . . .	1
Figure 1.2	Schematic overview of IATA 2020 goals . . . . .	4
Figure 1.3	Development of jet fuel price (1990-2025) . . . . .	5
Figure 2.1	Overview of cost index and arrival time prediction . . . . .	10
Figure 2.2	Correlation of distance and total trip time . . . . .	17
Figure 2.3	Correlation of flight time and total trip time . . . . .	17
Figure 2.4	Hyperparameter optimization for gradient boosting model	20
Figure 2.5	Influence of cost index on flight time . . . . .	23
Figure 2.6	Influence of cost index on fuel consumption . . . . .	24
Figure 2.7	Error histogram for arrival time prediction . . . . .	25
Figure 2.8	Variable importance plot for linear regression . . . . .	27
Figure 2.9	Optimal cost index according to distance . . . . .	30
Figure 3.1	Overview of flight stages . . . . .	35
Figure 3.2	Data provided by the operational flight plan . . . . .	39
Figure 3.3	Overview of prediction model and feature subsets . . . . .	43
Figure 3.4	Description of prediction points and prediction models . . . . .	44
Figure 3.5	Reduction in MAE and SD considering prediction points . . . . .	48
Figure 3.6	Variable importance for random forest . . . . .	49
Figure 4.1	Error histogram for short-haul flights . . . . .	65
Figure 4.2	Error histogram for long-haul flights . . . . .	66

## LIST OF TABLES

---

Table 2.1	Influence of cost index on aircraft flight time . . . . .	22
Table 2.2	Overview of Baseline and Prediction model . . . . .	25
Table 2.3	Results of Ensemble model . . . . .	26
Table 2.4	Comparison and evaluation of ensemble model . . . . .	26
Table 2.5	Influence of cost index on flight time and total trip time	28
Table 3.1	Prediction results of the best performing model without en-route weather data . . . . .	46
Table 3.2	Prediction results of the best performing model consid- ering en-route weather data . . . . .	47
Table 3.3	Comparison of best models . . . . .	48
Table 4.1	Overview of aircraft types, average distance and fuel consumption . . . . .	59
Table 4.2	List of features for prediction model . . . . .	60
Table 4.3	Overview of results . . . . .	64
Table 4.4	Comparison of variable importance . . . . .	66
Table 4.5	Cost to carry estimates for each aircraft type . . . . .	69
Table 4.6	Cost to carry factors according to pilot expertise . . . . .	69
Table 4.7	Total fuel and CO <sub>2</sub> savings for the airline . . . . .	70
Table A.1	Features based on time and date . . . . .	79
Table A.2	Weather features . . . . .	79
Table A.3	Features based on aircraft and flight information . . . . .	79
Table A.4	Departure features . . . . .	80
Table A.5	Arrival features . . . . .	80
Table A.6	Example results for 20 flights . . . . .	82
Table A.7	Flight example: HAM - HUB 1 (285 NM) . . . . .	82
Table A.8	Flight example: MXP - HUB 2 (294 NM) . . . . .	83
Table A.9	Flight example: DUS - HUB 2 (307 NM) . . . . .	83
Table A.10	Flight example: MAN - HUB 1 (509 NM) . . . . .	83
Table A.11	Flight example: LHR - HUB 2 (528 NM) . . . . .	83
Table A.12	Flight example: FCO - HUB 1 (609 NM) . . . . .	84
Table A.13	Flight example: ARN - HUB 2 (798 NM) . . . . .	84
Table A.14	Flight example: DME - HUB 1 (1222 NM) . . . . .	84
Table B.1	Hyperparameter settings Gradient Boosting . . . . .	85
Table C.1	Overview of linear regression results . . . . .	87
Table C.2	Hyperparameter settings for each aircraft type . . . . .	88
Table C.3	Overview of results - SLM model . . . . .	88

## ACRONYMS

---

<b>IT</b>	Information Technology
<b>CO<sub>2</sub></b>	Carbon dioxide
<b>IATA</b>	International Air Transport Association
<b>ICAO</b>	International Civil Aviation Organization
<b>ANSP</b>	Air navigation service provider
<b>CART</b>	Classification and regression trees
<b>EU</b>	European Union
<b>DCI</b>	Dynamic cost indexing
<b>CI</b>	Cost Index
<b>NM</b>	Nautical Miles
<b>RMSE</b>	Root mean squared error
<b>TAF</b>	Terminal Aerodrome Forecasts
<b>ntree</b>	Number of trees
<b>OD</b>	Origin - Destination
<b>OFF</b>	Operational flight plan
<b>AMDAR</b>	Aircraft meteorological data relay system
<b>ATC</b>	Air traffic control
<b>GDP</b>	Ground delay program
<b>SD</b>	Standard deviation
<b>PM</b>	Proposed model
<b>AF</b>	Airline forecast
<b>RF</b>	Random Forest
<b>LR</b>	Linear regression
<b>GBM</b>	Gradient Boosting Machine
<b>SESAR</b>	Single European Sky ATM Research Program
<b>GHG</b>	Greenhouse gas
<b>SCF</b>	Statistical contingency fuel

<b>MAE</b>	Mean absolute error
<b>RSS</b>	Residual sum of squares
<b>SLM</b>	Short linear model
<b>NOTAMS</b>	Notices to airmen

## INTRODUCTION

### 1.1 THE AIRLINE INDUSTRY

The first successful flight with a heavier-than-air aircraft was conducted by the Wright brothers in 1903. Twenty years later the first airlines started their service in Europe with aircrafts far more advanced than the Wright's original concept. Airlines operated aircraft between major European cities, such as London and Paris. With the British interest in their colonies, intercontinental services to India, Australia and South Africa began in the 1930s. For many years, flying was a service reserved for the upper class. Airlines were subject to strict governmental oversight. Initially, government regulation focused on ensuring safety of passengers. Later, regulations aimed at controlling and protecting the interests of the respective national carrier. With a desire to lower air travel prices and create a more competitive economic environment, the European Union deregulated air travel in the early 1990s. With deregulation, budget airlines entered the market. By increasing competition, prices fell and air travel was affordable for large parts of the population. Today, more than 100 years after the first commercial flight, air travel has become a commodity for many people. Air freight provides the foundation of global trade by moving 35 % of world trade (by value). In 2017, 36 million flights were conducted transporting four billion passengers and moving 60 million tons of goods. Figure 1.1 showcases how the number of passengers has grown since the deregulation of the industry.

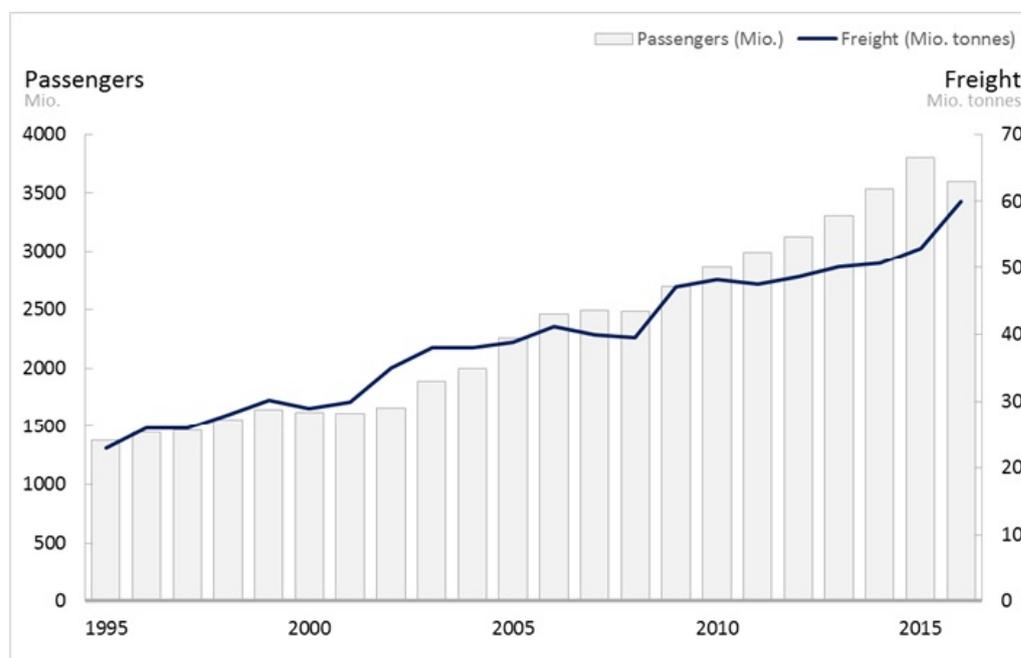


Figure 1.1: Global flight passengers and air freight (1995 - 2016). Source: (ICAO 1995-2016, Worldbank 2017).

The airline industry has been a leading enabler of globalization, moving people and products across continents in a matter of hours. It has grown immensely over the past decades and air travel has become a commodity for many people. Today, airlines face both challenges and opportunities. The European passenger numbers are expected to increase by close to 45 % compared to 2014 (European Transport Commission 2016). Even though this is a positive trend, the increase in demand needs to be accommodated in the schedules. The negative effects, such as delays and GHG emissions, are expected to increase as well. Second, the competition in the airline industry is fierce and expected to intensify. Low cost carriers are aiming for a greater market share which has been held by legacy carriers. Third, airlines must plan and hedge against the volatile jet fuel price. Since the end of 2014, we observe falling jet fuel prices, as low as 0.93 US Dollars per gallon. Soon, the jet fuel prices are expected to rise again (Airbus Market Forecast 2017). Furthermore, airlines will be affected by the upcoming regulations that target the environmental footprint of the industry. The IATA members have agreed to reach a carbon-neutral growth from 2020 onwards. According to IATA (2013b) improvements in air traffic management, airport infrastructure, airline operations and the implementation of known technology will mainly contribute to reach the desired goal. Besides participating in industry-wide initiatives, airlines need to focus on their operations to minimize the undesirable effects of aviation. Considering the complexity of airline operations, decision making is challenging. The implementation of advanced analytics to improve operational decision making is the focus of this dissertation.

## 1.2 OPERATIONAL CHALLENGES IN THE AIRLINE INDUSTRY

Airlines rely on a variety of software systems to operate their aircraft and organize the turnaround process. Not only does the aircraft itself produce large volumes of data, but the operational software does as well. The integration and analysis of the vast amounts of data generated through the operational information technology (IT) systems, will be a critical factor for efficient airline operations in the future. Today, legacy carriers are especially challenged with their legacy IT infrastructure. Oftentimes, data is spread across many systems and those systems lack integration. This results in a lack of data availability and underutilization of the available data for decision making. Due to the complexity of airline operations, advanced models for decision making are required. The following sections aim to give a brief overview of the key challenges in airline operations.

### 1.2.1 *Decision Making*

Airline operations are based on a complex system integrating a large variety of cost intensive resources such as aircraft, crew and ground operations. Due to the interconnection of resources, disruptions quickly spread throughout the whole system. Therefore, airline operations have to consider a large variety of options for action in case of an operational problem. The main entities to

control air and ground operations are operations control and hub control. All information regarding arrivals and departures, ground operations and unforeseen disruptions is centrally handled by those teams. Safety, punctuality and cost sensitivity are considered by operations control in their decision making. Situational awareness is a key aspect of decision making in airline operations. Accurate and reliable information is the foundation of operational awareness. Due to limitations of IT infrastructure and forecasting methods, operations control is provided with information that lacks accuracy and immediate availability. More accurate forecasts and the availability of such information at an earlier point in time can have a substantial impact on operational robustness, safety and punctuality of flight and ground operations.

### 1.2.2 *Delay management*

According to the International Air Transport Association (IATA) the airline industry will face substantial demand increases. Due to a growing demand for global connectivity air travel is predicted to increase by 45 % until 2035 in Europe with a total of 12.8 million passengers (European Transport Commission 2016). With an increase in demand, traffic is expected to grow. This will increase the airlines need for accurate information to ensure punctual operations. Today, airlines are subject to external and internal influences that affect their operations. One of the main external influences is the congested air space across Europe. The European airspace is fragmented as the number of air navigation service provider (ANSP) shows. In total, there are 37 ANSPs operating in the European airspace. Furthermore, each state has military needs that have to be accommodated in airspace management. Compared to the U.S., European flights have an increased delay during the flight phase due to volume and capacity constraints. From an internal perspective, airlines are pressured to use their costly assets, which are aircraft and personnel, as efficiently as possible to stay competitive. This results in tight schedules for flights and the turnaround process. Often, resources are interconnected and a delay of one flight can affect several flights down the line. The airline of our study operates a hub and spoke network. Passengers fly to the hub airport on short-haul flights to connect to intercontinental long-distance flights. Therefore, not only resources managed by the airline can be the cause of delay, but also passengers. Delay management has been a widely researched field in operations research. On a planning level, delay propagation throughout the system and robust scheduling have been assessed to understand and mitigate delay. On an operational level, delay prediction and optimal decision making, in case of a disruption, have been researched. In both cases, models include forecasted arrival times to calculate and predict delays. Therefore, a more accurate prediction of arrival times will increase accuracy of delay prediction models and help enhance disruption management. The early availability of arrival time information can increase decision-making time. More options can be evaluated and an optimal decision can be taken.

### 1.2.3 *Eco-Efficiency*

Today 2 % of global GHG emissions are generated by aviation. Environmental awareness has increased since airlines began commercial travel. Companies and governments initiated various measures to decrease the environmental impact of aviation. Aircraft manufactures focus their research on developing more efficient turbines and light-weight material for aircraft models. Airlines have embarked on a variety of fuel efficiency measures, such as improvements in operations and infrastructure or the research of alternative fuels, to decrease fuel consumption. According to IATA (2018), fuel costs account for 17 % to 25 % of total operating costs and are the single largest cost block. Therefore, airlines are interested in reducing fuel consumption not only to reduce their environmental footprint, but more importantly for economic reasons. The EU supports research focused on the development of alternative fuels and various other initiatives to make aviation more eco-friendly. Nevertheless, the member states of the EU state that they do not expect to decouple CO<sub>2</sub> emissions from the expected growth in the airline sector. Increases in GHG emissions of 45 % are expected until 2035 (European Transport Commission 2016). In the IATA (2013b), airlines agreed to reach a carbon-neutral growth from 2020 onwards. The baseline is defined as the average of the years 2018, 2019 and 2020. Considering that airlines want to increase their service to capture additional demand, measures for an additional reduction in CO<sub>2</sub> emissions needs to be developed. In Figure 1.2 a schematic overview of the measures and estimates for reducing greenhouse gas emission is displayed.

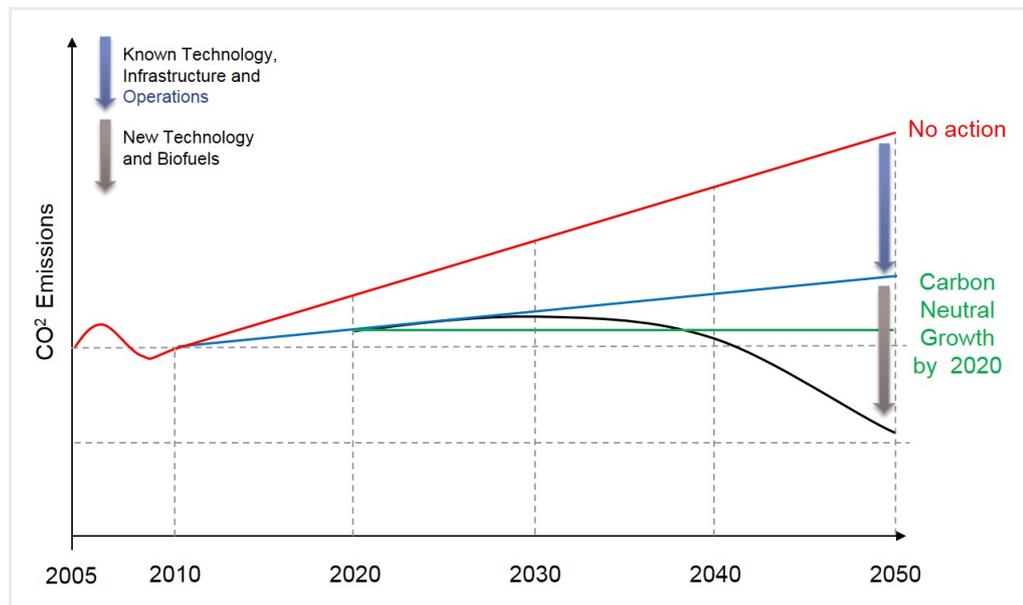


Figure 1.2: Schematic overview of the IATA 2020 goals considering CO<sub>2</sub> Emissions. Source: IATA (2013a).

### 1.2.4 Fuel Price Volatility

Apart from operational challenges such as delay management and data integration, operating costs are volatile, mainly due to the fluctuating oil prices. Most aircrafts today operate on JET A-1 fuel. There are also alternative fuels available based on non petroleum sources, but they constitute less than 1 % of the total jet fuel market (International Air Transport Association 2017). All fuel prices, however, depend on the uncertain oil price and as such complicate cost planning. Airlines can implement various measures to decrease the risk of increasing fuel prices, one example would be hedging. Since 2015, the price for jet fuel is at a historical low. According to Airbus Market Forecast (2017) the oil price and thus the jet fuel price are expected to increase in the next few years. Since jet fuel constitutes the single largest cost block of airline operating costs then even small changes in fuel cost will have a significant impact on airlines. In Figure 1.3 the cost of one gallon jet fuel is displayed also considering future price development until 2025.

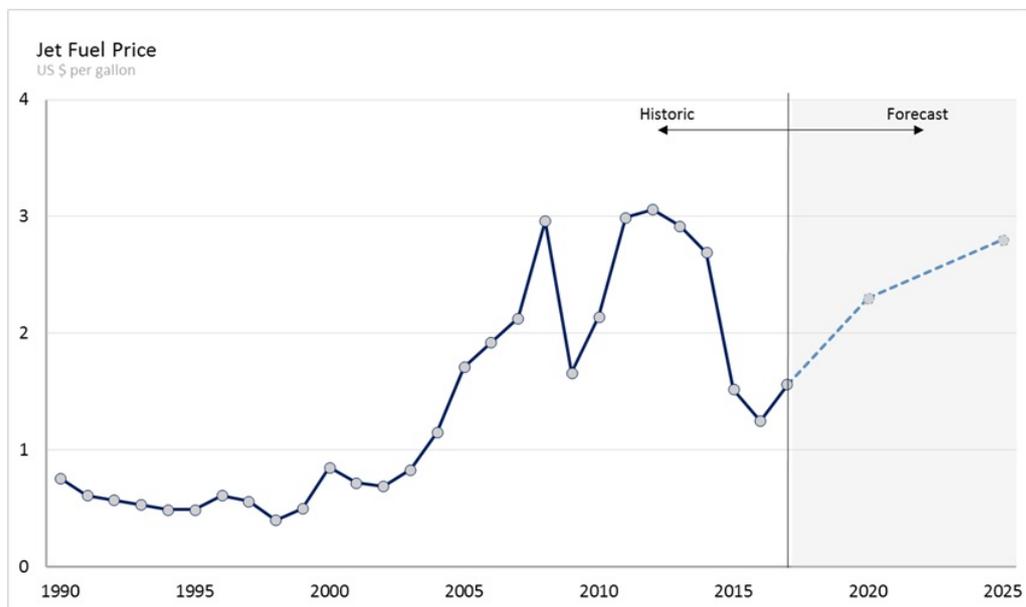


Figure 1.3: Development of jet fuel price considering forecast until 2025 Source: Airbus Market Forecast (2017).

## 1.3 MACHINE LEARNING

In the previous sections we showed that airline operations needs to accommodate a variety of information and data sources for optimal decision making. At the same time, predictions need to be accurate, reliable and available in short time to be able to integrate them in operational decision making. The decision-support models presented in this dissertation are based on machine learning algorithms, which are well suited for integrating and analyzing large volumes of data and generating predictions within a short period of time.

Traditional flight planning software considers variables such as flight route, altitude, aircraft speed, wind conditions to calculate flight time and fuel con-

sumption. Even though these models are accurate for flights traveling under the predicted conditions, these perfect conditions are rarely the reality. Today, fuel and arrival time forecasts lack the integration of additional information. For example, influencing factors like traffic patterns at the arrival airport, seasonal anomalies and crew expertise are not accounted for. Machine learning algorithms are able to integrate large numbers of samples and variables for prediction and, thus, are superior to predictions made by traditional flight planning software.

Machine learning addresses the shortcomings of traditional forecasting methods by considering all available data and recognizing complex patterns. There are a variety of machine learning models that can be considered for a predictive model. Two of the most frequently used algorithms for regression problems are random forest and gradient boosting. Random forest regression builds on the idea of decision trees. The idea was introduced with the CART algorithm (classification and regression trees) by Breiman (1984) in 1984. To reduce overfitting—a common problem for single decision trees—several trees are built and their prediction results are averaged. Furthermore, de-correlation is ensured through bootstrap sampling by randomly selecting predictor variables at each node split. Gradient boosting machines were developed by Friedman and Breiman (2000) in the late 90s. Gradient boosting builds on the idea of combining a set of weak learners to generate a strong predictive model. The model performs sequential operations by assigning higher weights to misclassified outcomes. Wrongly classified outcomes get a higher focus and are modeled more accurately in the next stage. This process is iterated several times until overall prediction accuracy has improved. Another type of prediction method is multiple linear regression. It is a standard tool in prediction and often used as a baseline model for comparison. The regression focuses on minimizing least squares. The predictor variables are assigned regression coefficients based on fitting the model to the data. The models and analyses presented in this dissertation are based on the three described machine learning methods to specifically take into account complex data structures and data sources to increase prediction accuracy. Furthermore, the acceptance of such prediction models is a key success factor. We evaluate the development, implementation and challenges of using machine learning in airline operations in a qualitative assessment <sup>1</sup> that we will apply in Chapter 4.

#### 1.4 CONTRIBUTIONS OF THIS WORK

In the following chapters of this thesis, we<sup>2</sup> aim to analyze options of integrating machine learning models into aircraft operations to improve decision making. It was shown in the previous sections, that the airline industry is facing challenges, that—together with future environmental constraints—require increases in airline operations efficiency. By developing machine learning models based on flight, traffic and weather data, we aim to support decision-making and help increase efficiency. We primarily focus on uncertainty in decision

---

<sup>1</sup>For an overview of our qualitative method, please refer to Cohen and Crabtree (2006).

<sup>2</sup>The term “We” refers to the authors of the respective chapters as noted at the beginning of each chapter.

making during daily operations, as it has the largest potential for improvement. The application of machine learning has been dominant in industries such as health care and marketing. Its use in operations has gained little attention in the past. We base our modeling approaches on a variety of machine learning models such as random forest regression and gradient boosting.

In Chapter 2 we focus on arrival time prediction for short-haul continental flights and quantify the impact of aircraft speed on arrival time. The model is used to show that a slower aircraft speed is more optimal considering cost and environmental aspects. In Chapter 3, we focus on long-haul intercontinental flights to predict gate arrival time. We specifically analyze the effect of including en-route weather data as predictive feature and show the impact on increasing prediction accuracy. In Chapter 4, we develop a fuel prediction model and analyze pilots willingness to integrate our model in their fuel decision. In detail, the three main chapters evolve as follows:

- In Chapter 2, we devise a prediction model for arrival times of short-haul continental flights. Based on a machine learning ensemble model, we are able to improve the arrival time prediction at the point of gate departure by more than 30 %. We integrate arrival time prediction with cost index optimization. The cost index is a tool for optimizing the aircraft's speed. A high cost index implies a faster flight. The cost of time is set in relation to the cost of fuel. We consider three different flight distances to model the impact of cost index changes on gate arrival time. With our model, airlines are able to reduce the cost index without any tangible impact on their overall schedule. We conclude that the optimal cost index level depends on a flight's distance, fuel costs and delay costs. Overall, our model shows that short haul flights should lower cost index. A high cost index has limited impact on reliably decreasing delay during flight. Thus, our model should be taken into account in decision-making.
- In Chapter 3, we present a predictive model to improve the arrival time prediction of intercontinental flights. Arrival time predictions serve as input for daily operations planning and control. For network carriers accurate arrival time predictions are key for efficient hub operations. We analyze the accuracy of en-route weather data provided by the operational flight plan by comparing it to actual weather data. Furthermore, we design and integrate weather features in the arrival time prediction. For the airline we study, arrival time prediction achieves an increase in accuracy of up to 25% over its current practice. 5 % of the accuracy increase can be attributed to en-route weather data. By analyzing the data quality and quantifying its impact our model helps to better understand the value of the data. Our approach and findings can be generalized to other industries for establishing data-driven decision-making.
- In Chapter 4 we present a fuel prediction model based on machine learning algorithms. Our prediction model considers a variety of features not included in the flight planning software. In our model, we estimate fuel values for ten different aircraft types, for both short and long-haul flights. Our results show that the fuel prediction can be improved by 30 % on

average and up to 48 % for specific aircraft types. With our model pilots can load less extra contingency fuel, as we account for more exceptions than the flight planning software. Cost savings of 10.5 million € per year and an annual reduction of 66.000 tons CO<sub>2</sub> emissions can be achieved. To assess the acceptance of our prediction model, we conducted interviews with pilots. The pilots state a general interest in an improved fuel consumption prediction. Nevertheless, the acceptance and use of the prediction model depends on trust, which needs to be built overtime through accurate predictions.

In Chapter 5 we summarize the dissertation and discuss the managerial implications of the three main chapters. Potential areas for future research based on our findings are also outlined.

We contribute to existing airline operations literature and offer decision support models for operations control and flight planning to increase the efficiency of airline operations. The models presented in this thesis help to better understand optimal flight speed considering fuel cost and cost of delay, the impact of weather on arrival punctuality and critical components for an improved fuel loading strategy. The derived findings are relevant for pilots, operations control and the airline strategy. Overall, our findings contribute to the transformation of data-driven decision making in airline operations and showcase several use-cases. Our models help pilots and operations control during their daily work to optimize their decisions with regards to fuel loading, aircraft speed and proactive delay management.

## ARRIVAL TIME PREDICTION AND COST INDEX OPTIMIZATION FOR SHORT-HAUL FLIGHTS<sup>1</sup>

---

### 2.1 INTRODUCTION

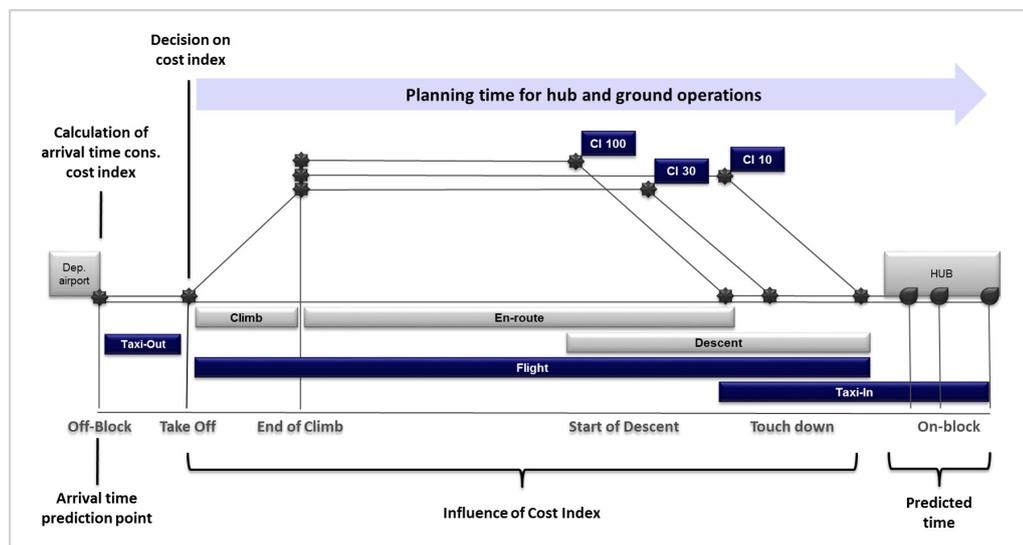
Considering that reactionary delays, which describe delays caused by the late arrival of an aircraft or crew of a previous journey, made up more than 40% of all delays in 2015 ensuring on time arrivals is a critical operational problem for airlines (Walker 2016). Operations control can implement several measures during the flight and on the ground to ensure punctual arrivals. The cost index (CI) is a tool to influence arrival time. It basically describes the speed of the aircraft. The CI defines the trade off between time and fuel cost. An increased CI results in decreasing flight time and increasing fuel cost and vice versa. According to Roberson (2007) airlines can save several millions a year by optimizing their cost index without impairing their schedules. Today operations control are not able to accurately integrate cost index models with their arrival time predictions. This is due to the fact that time related costs are hard to quantify and short term arrival time forecasting is lacking accuracy (Kim 2016, Takacs 2015). Oftentimes delayed flights are flying with a high CI to recover time, even on such short flight routes, where an increase in CI has very limited impact. En-route flight phases are too short to tangibly reduce flight time by flying faster. Time savings are often small and easily lost in the landing and taxiing process.

Significant effort has been made by industry and researchers to ensure the minimization of delays through strategic planning and robust scheduling (Barnhart and Smith 2012). Due to unforeseeable events and tight schedules delays cannot be eliminated. In recent years, predictions focusing on network wide delay propagation as well as on delay predictions for single aircrafts were published (Abdelghany et al. 2008, Rebollo and Balakrishnan 2014). Their work is closely connected to the research field of arrival time prediction, which has entered the focus of researchers and industry in the past years (Kern et al. 2015, Kim 2016, Takacs 2015). Especially the use of machine learning algorithms, for example the ones used by the top five contestants of the GE-Flight Quest Kaggle competition, have shown to outperform linear models (Takacs 2015). They were able to increase the prediction accuracy of a flight's runway and gate arrival time by up to 40%. Similar results were achieved in other studies (Glina et al. 2012). The accurate assessment of the most economic cost index represents a crucial operational problem for airlines. The cost index optimizes the speed of an aircraft to ensure minimum costs of fuel and time. Not only the correct estimation of those different cost blocks is highly dependent on the airline and its aircraft fleet, but even more critical is the correct estimation of change

---

<sup>1</sup>The following chapter is based on the work of Achenbach and Spinler (Achenbach and Spinler 2018b). This manuscript titled "Prescriptive Analytics in Airline Operations: Arrival time prediction and cost index optimization for short-haul flights" was published at Operations Research Perspectives in 2018.

in arrival time considering different cost indices. Work on cost index optimization is rather limited. Cook et al. (2007) provide a generic tool for dynamic cost indexing (DCI). DCI is defined as managing flight delay costs dynamically through trading fuel burn against cost of time. Airline manufacturer maintain cost index guidelines for their fleets (Customer Services AIRBUS 1998, Robertson 2007). Especially network carriers have a strong interest in improving their arrival time predictions and lower their fleet wide cost index. At their hub locations, long distance flights are fed by a large number of local connections. Owing to efficiency and cost pressures tight schedules are implemented. Often times continental flights are sped up to ensure that passengers reach their connections on time. Due to the short cruise phases on continental flights, an increase of the cost index only has a very limited effect on the flight's overall total trip time. In Figure 2.1 we show an example of a flight and the impact of the cost index on the flight's arrival time.



A flight consists of five phases: Taxi out, Climb, En-route, Descent and Taxi in. The en-route flight phase can vary. Very short flights start the descent right after the climb phase. On average the en-route phase lasts for 31 minutes and an average flight for 100 minutes. The cost index only influences the flight time.

Figure 2.1: Overview of cost index and arrival time prediction model

In the current study, a machine learning ensemble is used to predict an aircraft's arrival time. But instead of predicting arrival times once aircraft is airborne, we focus on predicting an aircraft's arrival time at the off-block moment. We implement linear regression and gradient boosting machines as described by Friedman and Breiman (2000). Flight data of a European network carrier from the year 2015 and 2016 is used. The dataset consists of departing flights from over 200 continental European cities, which are arriving at the airline's hub locations. Various predictor variables related to weather data, airport congestion levels and standard flight information are included. We provide accurate arrival time predictions. Furthermore, the arrival times are predicted for all cost index levels of the aircraft. With this information we can calculate change in arrival time due to change in cost index. With our model we can quantify

the delay reduction and cost of decreasing delay during the flight. With this information airlines are able to dynamically adapt the cost index for each flight prior to departure considering current delay and fuel cost.

Our research findings provide several important contributions to the literature. First, to the best of our knowledge, we are the first to combine arrival time prediction and dynamic cost indexing to a prescriptive analytics solution focused on minimizing total costs of a flight. These findings extend existing work of dynamic cost index optimization and arrival time predictions. With this model airlines are able to quantify and weigh fuel costs against cost of time and take more informed decisions in their daily operations. Second, our results show that an ensemble of machine learning algorithms is an adequate tool to predict an aircraft's arrival time. Through accounting for linear and non-linear relationships of influencing factors and considering their interdependencies, predictions reach a high accuracy. Furthermore we show that a high prediction accuracy can already be achieved at the block off moment of an aircraft. With an average block time of 100 minutes operations control have considerable time for analysis and decision making.

Third, our model is the first attempt to predict aircraft arrival time with European airline information. Studies focused on US airspace are numerous (Aktürk et al. 2014, Glina et al. 2012, Kern et al. 2015), but so far no study using European flight data has been published. According to Eurocontrol and FAA (2010) there are several important differences between the European and US airspace, which make arrival time forecasting for European airspace more complex. Even though the number of controlled flights in the US exceeds the number of European flights by more than 65%, European airspace is more fragmented. Instead of one air navigation service provider (ANSP), 37 ANSP are operating for the European airspace. Each ANSP operates their own system and procedures. Furthermore the military needs of all national states need to be accommodated in airspace management. European flights have larger amounts of enroute delay due to volume and capacity constraints. According to Campanelli et al. (2016) a key difference between European and US air space management is the prioritization of flights at the airports. The US follows a first-come first-served basis contrary to the European slot system. Considering those differences, forecasting flight arrival time for European flights needs to be investigated as there are major differences to the US system.

This paper is organized as follows. Section 4.2 reviews the relevant literature. The prediction model is suggested in Section 4.3, wherein the data, variables and the current baseline model are described. In Section 4.4, the cost index model is explained. Section 4.5 reports the results of each model. Section 4.8 provides a brief conclusion and describes implications for theory and practice in addition to limitations and directions for future research.

## 2.2 LITERATURE REVIEW

Several overlapping streams of literature provide the context for studying arrival time prediction and cost index optimization.

### 2.2.1 *Airline cost index*

The ratio of the CI includes fuel and time related costs. Fuel costs fluctuate depending on daily spot rates, location and the hedging strategy of an airline. Time-dependent costs primarily include hourly crew and maintenance costs. Delay costs are typically hard to quantify, because they can vary depending on the magnitude of delay and the airline's policies (Cook et al. 2007, Customer Services AIRBUS 1998, Roberson 2007). The first stream is the literature on airline cost indexing. Both Boeing and Airbus have published cost index strategies for their fleets. In Customer Services AIRBUS (1998) the current industry standard for cost indexing is described. Detailed reports are included for all fleet types detailing fuel increases, speed limitations and operational recommendations. Boeing puts the concept of cost indexing in context with fuel conservation strategies. They state that airlines do not leverage the cost index concept enough in their daily operations and miss out on annual savings of up to \$ 5 million by flying with higher cost indices. According to Boeing lowering the cost index would have a minor impact on schedule. Flight time increases would range from one to three minutes for a 1.000 mile trip (Roberson 2007). Cook et al. (2007) develop a generic tool for dynamic cost indexing, which allows pilots to change the cost index during the flight. Through the use of case studies they identify the key challenges for implementing a dynamic cost index, which is the accurate estimation of time related costs, especially delay costs, and data integration of external and internal sources. Furthermore Cook et al. (2007) have considered the environmental impact of optimized CI levels to reduce CO<sub>2</sub> emissions. According to Cook et al. (2007), two airlines which were interviewed for case studies had differing views on the impact of cost index optimization for short haul flights. Experts of Scandinavian Airlines stated that time recovery for flights under 60 minutes is limited and they should be flown with a low CI regardless. Contrary to that other airlines stated that even small delay recoveries would be beneficial for their operational efficiency. More recent work on the environmental benefits of optimizing CI for different aircraft models was published by Edwards et al. (2016). According to their results, optimal CI values can vary tremendously depending on aircraft type and flight distance, but CI optimization has a higher impact on long haul flights. Closely linked to using cost indices for a delayed flight is the concept of speed optimization to overcome delays. Aktürk et al. (2014) include cruise speed as a decision variable to shorten delays in a network and show that speeding up an aircraft can reduce overall delay costs. Even though all studies consider cost indexing to have an impact on flight times, they do not explain their arrival time calculations in detail. So far no model integrates cost index optimization with detailed arrival time predictions.

### 2.2.2 *Arrival time prediction*

The second stream of literature concerns arrival time estimation. According to Murphy et al. (2003), airlines use aircraft performance models together with parametric or physics-based trajectory models to calculate flight time. Often

these tools are incapable of considering external influences such as weather and airport congestion. Similar conclusions were drawn by Glina et al. (2012). Even though there are more advanced methods to predict an aircraft's arrival time, airlines still heavily rely on simple models due to a lack of data availability and real time data integration (Kletzel and DeWaart 2014, Smith and Kletzel 2015). Proactive delay management is closely linked to arrival time predictions. Researchers have focused on using classification models to predict if an aircraft is delayed. Especially the detection of periodic and reoccurring delays have been investigated. Early work in the field of delay propagation was done by Peterson et al. (1995) and Beatty et al. (1998). In recent years, the use of predictive analytics in delay management has entered the focus of researchers. Abdelghany et al. (2008) developed an airline's integrated recovery approach by predicting schedule disruptions and proposing an integrated recovery plan requiring only one minute computing time. Similar work was done by Rebollo and Balakrishnan (2014) concentrating on network delay states to forecast departure delays 2 - 24 hours in advance. Further work focusing on delay prediction was published by Zonglei et al. (2008) using machine learning to train an unsupervised model on flight delay detection. Xu et al. (2008) focus on delay prediction for 34 major U.S. airports. They predict positive or negative delay. Tu et al. (2008) analyze the distribution of delays for departing flights.

In recent years interest in research focusing on predicting a flight's remaining flight time or arrival time has increased. Glina et al. (2012) aimed at predicting landing times once the aircraft is within a distance of 60 nautical miles (NM) from the airport to optimize aircraft sequencing. A linear relationship between prediction accuracy and decreasing distance was observed. The approach was tested at the Dallas/Fort Worth Int. Airport for a period of five days and delivered satisfying results. Srivastava (2011) focus on taxi out time prediction for John F. Kennedy airport in New York. They use historical data based on airport surface traffic. Kern et al. (2015) used random forest regression for arrival time prediction and compared different sources of information and their predictive power. The inclusion of aircraft and flight information together with weather data and runway utilization data delivered by far the best results than any other combination of information. In the 2013 GE Flight Quest challenge a data science contest was hosted by GE Aviation and Alaska Airlines. By providing extensive historical flight data the companies aimed at getting real time big data analysis solutions. The dataset included over 2.3 million U.S. domestic flights for a period of several months. The winning team achieved a reduction of prediction error by more than 40 %. Especially the feature selection proved to be a critical task in developing a superior predictive model (Takacs 2015). Kim (2016) developed a non-parametric additive model to predict flight arrival times for domestic flights arriving to Denver International Airport. He highlights the importance of departure delay to accurately forecast arrival times and achieve a root mean squared error of twelve minutes.

### 2.2.3 *Machine learning*

In this section we give a brief overview of predictive modeling for regression considering the most relevant machine learning algorithms including gradient boosting machines, random forest and multivariate linear regression. First, the CART algorithm (classification and regression trees) was introduced by Breiman (1984) in 1984. It is the foundation for random forest regression and builds on the idea of decision trees. Random forest regression is a commonly used prediction algorithm in data mining and machine learning. To overcome the drawback of over-fitting, several trees are built and their prediction results are averaged. The trees are de-correlated through bootstrap sampling, which ensures random selection of predictor variables at each node split. Second, Gradient boosting machines were developed by Friedman and Breiman (2000) in the late 90s. The algorithm is an additive ensemble, which subsequently gives more weight to observations that are hard to classify. It is grown sequentially using a modified version of the original dataset by adding a new weak learner to the model at each stage. Through that shortcoming of existing weak learners are compensated and overall prediction accuracy increases. The definition of the loss function as well as parameter tuning are critical for model performance. Third, a standard tool in predictive modeling is multivariate linear regression focused on minimizing least squares. It is based on one response variable being described by a set of predictor variables and their regression coefficients. Continuous as well as categorical variables can be used. The data needs to be linearly separable. For aircraft arrival time prediction all of these methods have been used. Kim (2016) implemented a spline smoothing-based non-parametric additive model to predict aircraft arrival time at the point of departure. Focusing on flights arriving to Denver Int. airport the model achieved a prediction accuracy of root mean squared error (RMSE) 12.2 minutes. Glina et al. (2012) used a quantile regression forest algorithm, which is an extension of random forest providing conditional probability distributions. Overall 4000 flights were used to train and test the model. The algorithm achieved high-fidelity predictions for aircraft landing times. Kern et al. (2015) showed that random forest regression outperforms traditional arrival time predictions based on trajectory models. The random forest was trained on a sample of 20,000 continental US flights and achieved a mean absolute error reduction of over 40 %. The winners of the GE flight quest used a combination of random forest and gradient boosting models to come up with their final solution. Other contestants such as the third team implemented an ensemble of gradient boosting machines and ridge regression trained on the previous stages residuals as part of their final model Kaggle (2012). The contestants of the GE flight quest competition all achieved high prediction accuracy of around 4.2 minutes for gate arrival times.

The three streams of research stress the importance of optimizing a flight's cost index and the potential of machine learning to enhance aircraft arrival time prediction. However, while they offer useful insights into the potential benefits of cost index optimization, they do not provide a detailed model for obtaining a cost minimum given arrival time prediction. Our model is a step towards closing this gap. It can be used to estimate a flight's arrival time for each cost

index level. Based on this information the optimal cost index can be chosen for each flight individually.

In the following sections we first describe the model and then show its application on eight continental European flights. We illustrate how the model serves as a useful tool to determine the feasibility of adapting a flight's cost index and estimate the monetary implications.

## 2.3 MODEL FORMULATION

Contrary to the US data which is published by the U.S. Bureau of Transportation Statistics, European flight data is not made publicly available. Thus, previous studies focused on arrival time prediction for the U.S. airspace. This study is the first to make use of a rich dataset of a European airline. The data includes continental flights of the A320 fleet, which arrive at two hub locations. The dataset is comprised of historical data for the year 2015 and 2016. Prediction accuracy was evaluated on a new dataset. Besides the airline's flight data the dataset was enhanced with weather information and detailed data on the airports' capacity utilization. Incomplete flights with missing or wrong values were excluded from the dataset. Furthermore all airports with less than 20 flights per year were eliminated. In total, over 80,000 flights were used for model training, validation and testing.

We compare our prediction to the airline's arrival time forecast. The airline method is based on a flight planning tool, which considers distance, altitude, temperature and pressure to determine the optimal flight level and route. Additionally headwind and tailwind are included to calculate ground speed and therefore arrival time. The calculated time is transmitted to air traffic control and used by the airline's operations control team. For a detailed description of the calculation please refer to UK Civil Aviation Authority (2006).

### 2.3.1 *Target variable*

The target variable is given by the overall trip time of a flight, rather than the classification and prediction of delay scenarios. We refer to this variable as the actual total trip time. Considering the high impact that taxi times have on a continental flight's total trip time, we focus on predicting gate arrival time rather than runway arrival time. Therefore not only flight time is predicted, but also taxi times at the departure and arrival airport. The total trip time is measured starting at the moment the aircraft leaves the gate (block off) until the aircraft docks on to the gate at the destination (block on). The gate arrival time is more important than the landing time, as this is the most important time for any cost of time related cost index calculation. The target variable (in minutes) was not changed and could be directly extracted from the flight data file provided by the airline. Thus, only positive values are available. Note that scheduled total trip time denotes the airlines flight schedule times and predicted total trip time is the predicted time according to our ensemble model.

### 2.3.2 *Predictor variables*

With our prediction model we aim at generating a more accurate arrival time prediction at the point of gate departure. In this study, we consider the flight data of one airline only. Most of the variables were already included in the original dataset provided by the airline. Nevertheless, great effort was put into feature generation especially for time related variables. As Kern et al. (2015) show in their work, the combination of flight data, weather forecasts and airport congestions levels lead to the highest arrival time prediction accuracy. This is why we also included all of these features in our model. Overall there are 71 features, six categorical and 65 numerical. The features include information related to the departure airport, information related to the arrival airport, general flight and airline data, weather data and time related data. Due to limitations in data types and number of factor levels, data transformation needed to be done for certain variables. These transformations include simple conversions of the data format, such as changing a time format to a numeric variable by converting it into minutes, as well as more complex factor level conversions. The features as well as any data transformation are explained in more detail in the following paragraphs.

Information regarding the departure airport is considered an essential feature for arrival time prediction. According to Kim (2016), departure delay, i.e. the time between scheduled and actual off block, is an important predictor variable. We include it in our model by calculating the difference between the scheduled block off time and the actual block off time. Negative values indicate that the aircraft left the gate earlier than scheduled. Furthermore categorical features such as scheduled departure runway and departure stand are included.

Three additional features were generated for the departure airport. Besides adding the country and city, which were both transformed to numeric features, a variable called destination frequency was added. It represents the total number of flights departing from this airport to the hub locations within one year. Hub locations are the two main locations of the airline, connecting passengers from short-haul to long distance flights. With this variable, we quantify efficiency of crew and ground operations due to frequently working together.

The structure of the arrival information is similar to the departure dataset. It includes arrival airport and scheduled arrival runway. To account for airport congestion detailed calculations for the number of departing and arriving aircraft were constructed. These airport congestion features include flights of all airlines operating at the airport. Taking the scheduled landing time as reference point the number of arriving and departing aircraft is calculated for five minute breaks ranging from 30 minutes prior to landing until 30 minutes after landing. Thus, 24 numeric features are created depicting the number of aircraft departing or arriving within the specific time window. This variable creation process is based on the findings of Kern et al. (2015). The above described approach delivered a higher prediction accuracy than the cumulative score with time breaks including all previous stages.

The weather datasets are based on Terminal Aerodrome Forecasts (TAF). TAF reports are issued according to the standards defined by the International Civil

Aviation Organization (ICAO). National meteorological services publish TAF reports for all major civil airfields. Scheduled departure and landing times are used to match TAF reports for the arrival and departure airports. The report includes numeric features such as cloud base, vertical and horizontal visibility, wind speed, wind gust, wind direction as well as one categorical feature. Due to large amounts of missing values certain variables such as temperature and atmospheric pressure had to be excluded. The categorical feature describes overall weather conditions.

Additionally features related to the flight and the aircraft are incorporated. These range from aircraft and fuel type to total take-off weight. Air distance, great circle distance and maximum altitude are included. Furthermore, data of the airline includes scheduled times for the different flight segments such as taxi times and total trip time. To enhance prediction accuracy we created derived features of the total trip time and taxi times. One feature focuses on the taxi-out time until the landing time. Another one on the scheduled flight time. It was derived by subtracting scheduled taxi times from scheduled total trip time. As shown in Figures 2.2 and 2.3, time and distance related variables have a strong linear relationship with the actual total trip time. The correlation coefficient for the distance and time related variables with actual total trip time ranges from 0.981 to 0.986. Therefore, we include those variables in our model as important features for an accurate total trip time prediction.

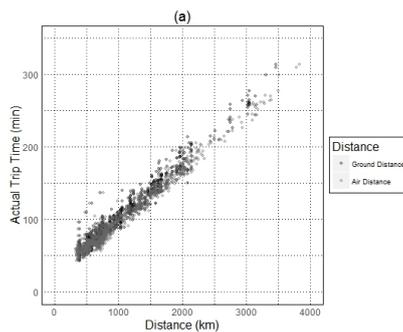


Figure 2.2: Air and ground distance have a correlation coefficient of 0.981 with actual total trip time.

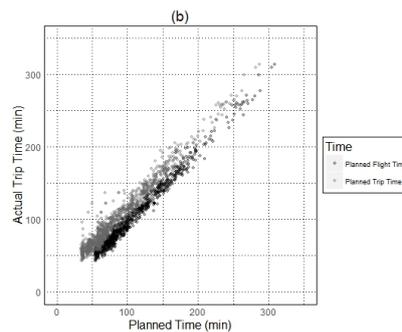


Figure 2.3: Planned flight time and total trip time have a correlation coefficient of 0.986 with actual total trip time.

According to Rebollo and Balakrishnan (2012) and Kim (2016), features describing the time of day are important for estimating delays. In our model, time related features include variables generated from the flight date such as month, season and weekday. Besides, we analyzed peak hours during the day and created the feature "Time of day" to account for peak hours in the morning and evening. Another feature that was created focuses on the number of days until the next national holiday to account for increases in travel before major holidays. Furthermore, time related features include a binary variable to account for weekends and two variables focusing on the hour of departure and arrival. All time and date related features that were generated are numerical. For our optimization model the scheduled flight times are especially important. They will be adapted to generate CI specific arrival time predictions. The

adaptation of features is further explained in the section 4.1 focusing on the optimization model. Please refer to Appendix A for a list of all features and detailed information on data origin and transformation.

### 2.3.3 Predictive model

For our prediction model we implement and test several machine learning models. Our final model is based on an ensemble of a gradient boosting machine and linear regression. Ensembles of different models are frequently used for predictive modeling. We develop an ensemble model that is trained on the residual of the previous stage. Through the use of an ensemble model we are able to combine linear regression and gradient boosting. Furthermore, we include different features in each stage of the ensemble. With the combination of two predictive algorithms, we account for linear and non-linear relationships in the data. The whole system was implemented in R version 3.2.5. Besides "gbm" standard data preprocessing packages such as "dplyr", "stats", "timeseries" and "zoo" were used. The ensemble takes less than three minutes to run.

In the following, each ensemble stage of the model detailing the features and parameter settings is explained. Variable selection was based on literature reviews and importance rankings derived by individual models. The final model was developed through a highly iterative process with the goal of optimizing error rates and standard deviation. Many ensemble combinations as well as single models were run to determine the best combination of features and structure of the ensemble. The ensemble model outperformed any individual model. The number of ensemble stages and the variables used in each stage were organized according to their subject area. Therefore, we include subsets with general information, weather and airport congestion levels. Also the prediction of flight segments was tested, but did not result in a superior prediction accuracy compared to predicting the arrival time at once. Instead of decreasing overall prediction error, the errors of each individual segment (taxi-out time, flight time, taxi-in time) add up to a larger inaccuracy. This is due to the fact that the error of the taxi times is as large as the error of the flight time. Considering that taxi times account for a smaller proportion of total trip time, overall prediction accuracy is less than in the ensemble model focusing on total trip time. A detailed list of the features used for each model stage is included in the Appendix B. The models were trained and tested using data from 2015 and 2016. Root mean squared error, mean absolute error and standard deviation are used for model evaluation.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (c_i - v_i)^2} \quad (2.1)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |c_i - v_i| \quad (2.2)$$

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (c_i - v_i)^2} \quad (2.3)$$

$N$  is the number of predicted values.  $c_i$  and  $v_i$  are the predicted and actual total trip time respectively. After training the ensemble with over 60,000 flights, fine tuning and cross validation was done with a validation dataset of 14,000 flights. The final ensemble was tested on a set of 6,000 flights. The datasets are made up of flight data ranging from January 2015 until April 2016. The datasets are generated through random sampling.

The first stage is a linear regression consisting of a subset of highly informative variables. Most of the variables are numeric and have a strong linear relationship with the actual total trip time. They include scheduled total trip time, scheduled flight time, block off delay, season, time of day, scheduled arrival runway, planned take-off weight as well as planned air and ground distance. This subset of variables is included in all other stages of the ensemble. These include the predicted values in minutes and denote the total trip time from the block off moment until the aircraft reaches the gate at the destination airport (block on).

In the second stage, predictions are made by a gradient boosting machine using a subset of the features not including weather and airport congestion data. The target variable is not actual total trip time, but the residual of the previous stage, which is calculated by subtracting actual total trip time from predicted total trip time. The variables included in this stage are departure stand, scheduled departure runway, departure airport, arrival airport, aircraft type, planned maximum altitude, weekday, month, scheduled taxi out and taxi in times as well as a binary variable for days on the weekend.

In the third stage, weather information such as weather conditions at the arrival and departure airport are used to make a prediction. Weather data includes over 15 different measures.

Stage 4 of the ensemble model combines capacity utilization information for the arrival airport according to the concept proposed by Kern et al. (2015). For every thirty minutes prior and after landing, five minute buckets are created detailing the number of flights to depart and land during these times.

#### 2.3.4 *Tuning of hyperparameters*

Hyperparameters are values that serve as additional information for machine learning algorithms. Different data patterns require different hyperparameters to optimally generalize the data. The hyperparameters for the gradient boosting were determined through a grid search. Random search was also considered for hyperparameter optimization, but due to the low number of dimensions grid search was chosen. It is easily implemented and can be parallelized Bergstra and Bengio (2012).  $N.trees$  describes the number of trees that are grown. Learning rate, also known as shrinkage, reduces the impact of each additional tree by penalizing the importance of each consecutive iteration. The interaction depth describes the number of splits done in each tree. We chose a

large number to allow for complex tree structures. The number of trees (ntree) ranged from 100 to 500, learning rate was set between 0.1 and 0.4 and interaction depth was either 10, 20 or 40. Minimum observations in the final node were kept constant with 100. In Figure 2.4 examples of hyperparameter tuning results for the first gbm stage are displayed. As we can see in the graph, the best RMSE is achieved with a learning rate of 0.1, ntree of 100 and interaction depth of 10. The RMSE for the ntree values of 250 and 500 is almost the same. The RMSE deteriorates with higher learning rate values, even when the number of trees are increasing. Interaction depth did not have a high influence on RMSE. A higher interaction depth lead to a slight increase in RMSE. Therefore the superior GBM settings are as following: n.trees = 100, learning rate = 0.1, interaction depth = 40, minimum number of observations in terminal node = 100. The GBM settings are the same in all stages except for the learning rate. The hyperparameter tuning in those stages showed an optimal learning rate of 0.15 for the third and fourth stage.

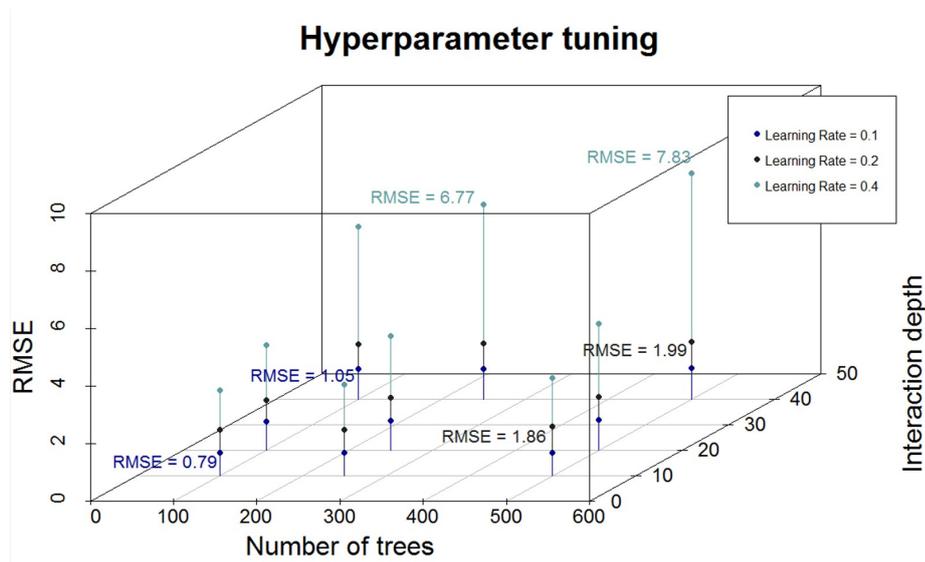


Figure 2.4: Increasing interaction depth leads to a slight increase in RMSE. Overall a low learning rate delivers superior results. The effect of number of trees on RMSE depends on learning rate and interaction depth.

## 2.4 AIRCRAFT PERFORMANCE DATA AND MODEL

### 2.4.1 Cost Index data

The cost index sets cost of time ( $C_{time}$ ) in relation to cost of fuel ( $C_{Fuel}$ ).

$$CI = \frac{C_{time}}{C_{Fuel}} \quad (2.4)$$

To calculate a flight's cost index, several data sources are needed. First of all, the cost of fuel needs to be known. This can be easily derived from daily fuel spot prices. Some airlines implementing fuel hedging might have to adapt

those daily rates according to their current hedging strategy. The cost of jet fuel is subject to daily variations depending on the oil market, location and amount. Cost of time are much harder to calculate. They are made up of marginal costs for each additional minute of flight time. Personnel as well as equipment costs, such as maintenance costs and hourly usage fees for engines, need to be included. Since crews are paid on a monthly fixed salary basis only additional costs for overtime add to overall crew cost. At the end of a month airlines accumulate crew hours that are higher than hours covered by the base salary. These additional hours need to be averaged over the month and year respectively. Different costs and working hours for cabin crew and pilots need to be considered. Thus, an airline can determine marginal cost of their crew for each additional minute. The accuracy of this heuristic depends on an airline's ability to depict crew costs and working hours correctly. The costs might vary depending on the seniority of the crew. Equipment maintenance cost also differ depending on the aircraft's age. Focusing on one aircraft family marginal costs for an additional hour of flight can be derived from the airlines yearly equipment maintenance costs. In case of rented engines hourly usage fees need to be added. On top of that delay costs have to be considered. It is quite difficult to decide on a specific value for delay costs per minute. Studies such as Cook (2004) and Cook and Tanner (2011) have suggested costs of 80 € for every five minutes of delay for the aircraft type A320. Delay costs include hard costs for re-booking the passenger as well as soft costs such as passenger dissatisfaction. The estimation of those values strongly depend on the airline. Ball et al. (2010) provide an innovative review of delay costs also considering passenger delay costs.

Considering that cost of time cannot be calculated comprehensively for the airline we study, we model the cost of time as a function of fuel costs for a given cost index. With the information of the aircraft performance model, describing changes in flight time and fuel consumption for each cost index level, we are able to display the cost of fuel for each cost index level. In Table 2.1 the information of the aircraft performance model is displayed for an example flight. Cost index, total trip time, change in total trip time and change in fuel consumption are the given variables based on the aircraft performance model. This information serves as an input for further calculations and ultimately as a threshold for decision making. Cost of time needs to be equal or higher than fuel costs for the respective cost index multiplied by the overall time saved at the specific CI ( $T_{CI}$ ). An example of this calculation is given in Table 2.1, where the  $T_{CI}$  is specified as minimum cost of time considering fuel costs. Since airline lack the capability to accurately calculate cost of time, the calculation of thresholds serves as decision guideline. If the airline staff estimates  $T_{CI}$  to be larger than fuel costs, an increase in CI is reasonable.

Aircraft performance models are provided by the aircraft manufacturer. It is possible to model a flight considering different cost index levels to understand the effect of a changed cost index on an aircraft's flight time and fuel usage. It is important to note that the models' calculations depend on a specific subtype of an aircraft fleet. For our model the data of the most frequently used aircraft, the Airbus A320, was considered. Furthermore, it is important to note that the distance of a flight can have a significant impact on the aircraft performance

Table 2.1: Example of a flight with time and fuel cost calculation according to aircraft performance model

Cost Index	0	10	20	30	40	50	60	70	80	90	100
Total trip time	82	82	81	80	79	78	77.5	77	77	77	77
Change in total trip time (min.)*	2	2	1	0	-1	-2	-2.5	-3	-3	-3	-3
Change in fuel consumption (kg)*	-60	-50	-30	0	30	60	90	120	135	150	160
Cost of Fuel (0,42€/kg)	-25.20	-21.00	-12.60	0.00	12.60	25.20	37.80	50.40	56.70	63.00	67.20
T <sub>CI</sub> (€/min.)	12.60	10.50	12.60	0.00	-12.60	-12.60	-15.12	-16.80	-18.90	-21.00	-22.40

\* acc. to aircraft performance model

model. To show the validity of our approach the aircraft performance model was developed for different distances. In the aircraft performance model distances are measured in nautical miles. It includes distances of 300 NM up to 1200 NM. Cost index levels from 0 to 100 in steps of ten were generated. It is important to note that the aircraft performance model focuses on the flight time of the aircraft. Fuel and time calculations start at take-off and end at touch-down. Thus, it only shows the percentage change in time and fuel for the flight time of the aircraft and not for the total trip time. The remaining segments of the total trip time, such as taxi-in and out times, are not included. Considering that the cost index will not have an influence on taxi times, it is reasonable to use the same taxi times for all CI levels. Furthermore, normal flight conditions are assumed.

#### 2.4.2 Optimization of Cost Index model

To determine the optimal cost index for each flight we first calculate the change in flight time for each cost index level. Secondly, we use the prediction model to generate arrival time prediction with the altered cost index related features and finally we generate thresholds at which a higher cost index becomes optimal. Our optimization is a full enumeration of all solutions. The optimal solution considers minimum delay and fuel costs.

To show the effect of cost index optimization for various distances, the aircraft performance model was applied to eight different origin-destination (OD) pairs. For each OD pair the following process was used to estimate the fuel and time values. First, we need to define a baseline for our fuel cost and time calculation. According to the airline personnel the standard CI is set to 30. When analyzing the actual flown CI for all OD pairs we use in our study, this was confirmed. The analysis showed that a CI between 29 and 32 was the most commonly used CI on the selected routes. Therefore the cost index level 30 was chosen as a baseline for all OD pairs. This baseline was then used to calculate the estimated change in time according to the aircraft performance model. For example, a flight with a distance of 510 NM has an average flight time of 78 minutes. This value was set as CI 30.

Using the change in percentage given by the aircraft performance model for 500 NM, the expected flight time is calculated for the other CI levels. Keeping all other values the same, only the three variables affected by a change of flight time were altered. These include the scheduled flight time, scheduled time from

block off until landing and the scheduled total trip time. These new values of expected flight time are used in the following to predict a flight's arrival time. Considering the eleven different cost index levels (CI={0, 10, 20, ..., 90, 100}) one flight was predicted eleven times. The results are eleven different arrival time estimates for one flight considering different cost index levels. To ensure the correctness of the model we tested each OD pair with 25 different flights.

According to the aircraft performance model the following time savings can be achieved by varying the cost index level. A short flight of 300 NM and approximately 52 minutes flight time at cost index 0 will be able to decrease the flight time by 3 minutes to a total flight time of 49 minutes at CI 100. This is approximately 5.7%. A flight with 500 NM and a total flight time of 79 minutes at CI 0 can decrease flight time by 5 minutes at CI 100, which is 6.3%. A flight of 1200 NM and 174 minutes total flight time at CI 0 can reduce the flight time to 166 minutes for CI 100, which equals 4.6%. In Figure 2.5 the percentage change in flight time are displayed.

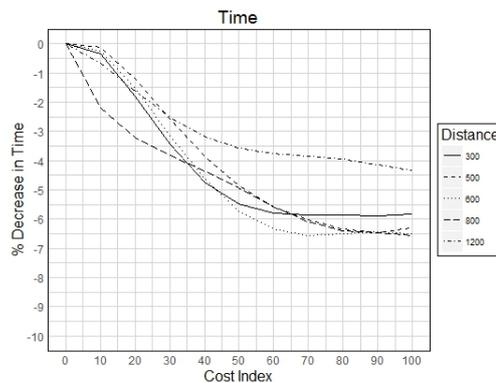


Figure 2.5: Decrease in flight time as a function of cost index (CI:0-100). We observe that the change in flight time varies depending on the distance. For flights of 300 NM the maximum decrease in flight time is 4%. For longer flights (800-1200 NM) a maximum decrease of 6.5% can be achieved.

The eight OD pairs were selected based on their distances. They include short distance flights around 300 NM, medium distance flights of 500 NM to 600 NM and long flights of 800 NM or 1200 NM. For all calculations the aircraft performance model of the A320 fleet was used as they are the most commonly used aircraft type for all OD pairs. Short distance OD pairs include flights from Hamburg to Frankfurt (285 NM), Milan to Munich (294 NM) and Düsseldorf to Munich (307 NM). The aircraft performance model derived for 300 NM was used. The routes Manchester to Frankfurt (509 NM), London to Munich (528 NM) and Rome to Frankfurt (609) make up the OD pairs for medium distance flights. The aircraft performance models for 500 NM and 600 NM were used respectively. The dataset included only a limited number of OD pairs with a distance of more than 800 NM, but for those OD pairs a sufficient number of flights is available. Therefore, only two OD pairs were chosen for long distances. These include Stockholm to Munich (798 NM) and Moscow to Frankfurt (1,222 NM). Aircraft performance models for 800 NM and 1,200 NM were used to estimated changes in time and fuel consumption. Fuel usage can vary up to 9,2% depending on the CI level. With longer distances the variability of fuel

decreases. In Figure 2.6 the percentage increase in fuel usage for each CI level and respective distance is displayed. Please refer to the Appendix D for detailed insights to the aircraft performance model for each OD pair.

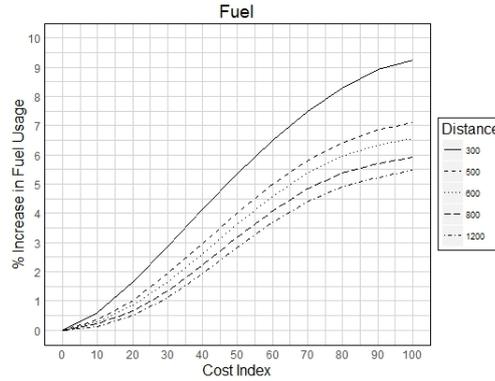


Figure 2.6: Increase in fuel usage as a function of cost index (CI:0-100). The increase in fuel consumption is quite similar for all distances. For shorter flights the fuel usage can increase by more than 9 %. For distances of 1200 NM the maximum increase is 5.5 %.

Due to the fact that planned fuel data for taxi-in is missing, we deducted double the average amount of planned taxi-out fuel from total planned fuel to get an average amount of fuel used during the flight. These values are in line with the fuel amounts shown in the aircraft performance model. Fuel is not a feature in the predictive model, but it needs to be considered for cost calculation. With a total number of eleven arrival time estimates for each flight, we are able to calculate the optimal cost index by fully enumerating all eleven options for CI 0 to CI 100. The change in fuel usage was altered according to the aircraft performance model data provided for fuel. Considering the new estimates of arrival time we can calculate the change in fuel cost depending on the CI.

## 2.5 RESULTS

Our results consist of prediction results for the machine learning ensemble model focused on arrival time prediction and the integration of the cost index in the arrival time prediction. First, we outline the results of the ensemble model. Second, we show how the cost index can be integrated in arrival time prediction and the optimization model in determining the optimal cost index for each flight.

### 2.5.1 Prediction results

We divided our data into three parts consisting of a dataset for training, validation and testing. Our final results are based on the predictions run on the test dataset. The ensemble seems to be the most accurate prediction model considering all evaluation criteria. The baseline model consists of the airline's

Table 2.2: Comparison of results

Method	RMSE	MAE	SD	Dev.>5 min.	Dev.>10 min.
Airline Forecast	8.5	6.80	5.10	60.6 %	33.4 %
Ensemble Model	5.90	4.31	4.03	32.8 %	7.4 %

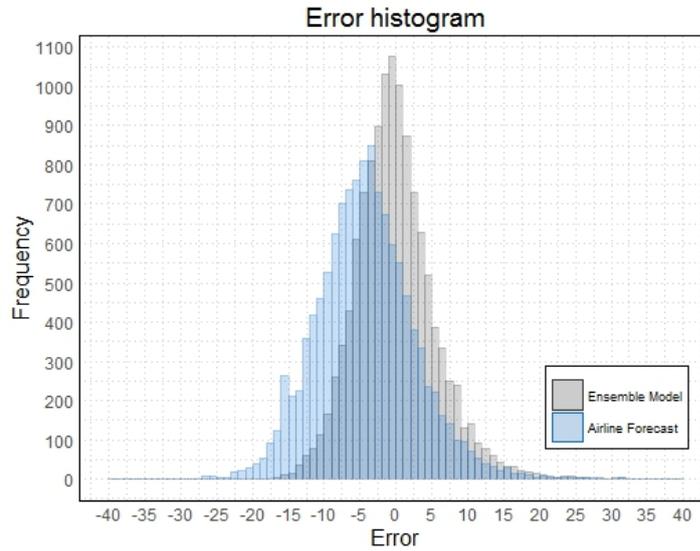


Figure 2.7: Absolute error in minutes: Difference between actual total trip time and predicted/forecasted total trip time. Negative values indicated earlier arrival than predicted/forecasted and positive values indicate later arrival than predicted/forecasted.

scheduled time of arrival calculation provided by operations control. In Table 4.3, the prediction results of the ensemble and the scheduled time of arrival calculation are shown.

In summary, the ensemble of gradient boosting machines and linear regression outperformed the airline's existing model. As expected, the prediction performs well compared to the baseline method of the airline. Prediction accuracy is improved by 30% for RMSE and 40% considering MAE. The GE flight quest challenge as well as Kern et al. (2015) and Glina et al. (2012) have shown that machine learning algorithms, such as random forest, can outperform standard trajectory models in arrival time prediction. We also implemented a random forest model. It was able to achieve comparable accuracy, but only at the cost of significantly larger model size and increased computing time.

In Figure 2.7, the error histogram of our predicted total trip times is shown. The amount of flights with more than five minutes deviation of their actual total trip time amounts to 32.8%. For flights with more than ten minutes deviation this value decreases to 7.4% (Table 4.3). Both of these values are important performance indicators for the airline. Considering that, previously, flights with more than five or ten minutes deviation from their actual trip time accounted for 60.6% and 33.4%, we observe that our model's accuracy has increased total trip time prediction substantially. We attribute the performance

Table 2.3: Results of Ensemble model

Stage	Model	RMSE	MAE	SD
1	Linear Regression	6.54	4.88	4.35
2	GBM (General)	6.15	4.53	4.15
3	GBM (Weather)	5.98	4.39	4.07
4	GBM (Airport Traffic)	5.90	4.31	4.03

Table 2.4: Comparison of single stage models with ensemble model

Method	RMSE	MAE	SD
Linear regression	6,54	4,88	4,35
Gradient boosting	6,16	4,54	4,16
Ensemble model	5,90	4,31	4,03

improvement to the additional data sources, which include information beyond linear aspects.

When considering the improvement in prediction performance from the initial stage up to the 4<sup>th</sup> stage, the model shows an RMSE improvement of 9.78%, MAE improvement of 11.68% and a decrease in SD by 7.35%. Please refer to Table 2.3 to see the improvements of each stage of the ensemble model.

Single stage models did not perform as good as the ensemble model. In Table 2.4 the results of a single stage linear regression and gradient boosting model including all variables at once are compared to the results of the ensemble model. Especially the separate inclusion of weather data (Stage 3) and air traffic information (Stage 4) in the ensemble are important as shown in Table 2.3.

In Figure 2.8 the most important variables for linear regression (absolute t-value) are shown. The variable importance was estimated by running the model on the full dataset including all variables at once. Variable importance is not fully representative for the ensemble model due to variable selection and inclusion of GBM. Nevertheless, the predicted importance measures were used in model generation and testing. As we can see in Figure 2.8 the most important variables for linear regression, such as distance, scheduled trip times, arrival runway and features related to time, are also essential features in our ensemble model.

The sensitivity analysis gave further insight to the prediction accuracy depending on certain data subsets. In general the prediction performs best with short flights. In the dataset flights with a scheduled flight time below 70 minutes, scheduled total trip time below 100 minutes and air distance smaller than 1000 km significantly outperformed flights with increased distances or flight times. Prediction accuracy of national flights are higher than international flights. Surprisingly, the influence of distance does not explain the difference between departure airports. When looking at a national level, we see that flights from Sweden with almost 1.5 times the average distance have a significantly higher prediction accuracy compared to flights from Italy. This might be due to

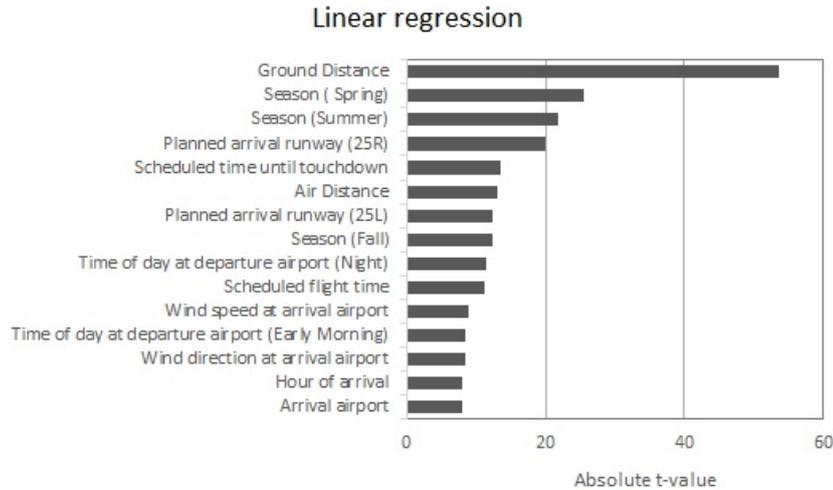


Figure 2.8: Variable importance plot: Linear regression including all variables. Absolute t-value as measurement. Not fully comparable to linear regression in the ensemble model due to variable selection and different ensemble stages.

airport inefficiencies or weather conditions. Not surprising is the effect of several other factors. For example prediction performs worse for flights arriving during peak hours in the morning or evening, during weekdays compared to weekends as well as during the winter months, where adverse weather conditions are more likely to occur. Furthermore, there are other less obvious discoveries. According to the sensitivity analysis the deviation in block off time, either leaving early or late by more than ten minutes, decreases prediction accuracy by more than 15 %. When comparing these results to the airline's scheduled arrival time forecast and applying the same sensitivity analysis the following two discoveries stand out. First of all we observe that in most cases the airline's forecasts reacts similarly to changes in certain variables. For example the decrease in prediction accuracy with increasing flight time and distance is similar to the results of the ensemble model. However, the prediction errors are larger. Contrary to the results of the predictive model, several categorical variables are not important for the airline's scheduled arrival time forecast. These include season, weekday and month. In the Appendix C a comparison of total trip time predictions obtained from the airline's forecast and the ensemble model is shown. The column "Scheduled" refers to the airline's forecast. "Actual" refers to the flight's actual total trip time and "Predicted" to the ensemble model's predicted total trip time.

### 2.5.2 Optimization results

The results show that the prediction is sensitive to changes in scheduled total trip time, scheduled time from block off until landing and scheduled flight time. All three variables are one of the ten most important predictors in our ensemble model.

To ensure the validity of our results, cost index optimization for one route was tested on 25 different flight movements of one OD pair. First of all we can see that the change in CI level has a similar effect on all 25 flights of the same OD pair. Furthermore we compared the results of different flights with similar distances to verify the outcome. Considering different OD pairs with similar distance the effect of flight time changes according to the a change in CI level are also comparable. With these results we are confident that our prediction model is able to accurately depict changes in flight time for total trip time prediction. Table 2.5 shows the impact of a change in CI levels for the short, medium and long distance flights considering all OD pairs of the category.

Table 2.5: Percentage change in Flight time and Total trip time (%)

Cost Index		0	10	20	30	40	50	60	70	80	90	100
Short	Flight Time	0.00	0.35	1.82	3.45	4.73	5.49	5.80	5.86	5.85	5.88	5.82
	Total Trip Time	0.00	0.08	0.38	0.67	0.95	1.15	1.23	1.24	1.24	1.25	1.24
Medium	Flight Time	0.00	0.17	1.37	2.87	4.25	5.29	5.95	6.29	6.42	6.44	6.41
	Total Trip Time	0.00	0.04	0.38	0.79	1.18	1.48	1.67	1.77	1.82	1.84	1.81
Long	Flight Time	0.00	1.44	2.43	3.16	3.76	4.26	4.66	4.97	5.17	5.31	5.46
	Total Trip Time	0.00	0.49	0.79	0.99	1.20	1.38	1.52	1.61	1.67	1.71	1.77

For flights with an average distance of 300 NM the flight time constitutes roughly 72% of total trip time. Taxi times account for the remaining trip time and are comparatively large parts due to the short flight time. Even though the flight time can be reduced by up to three minutes, the total trip time is not impacted strongly. For short haul flights the average decrease of trip time from CI 0 to CI 100 range between half a minute and one minute. These results are consistent for all short flights with an average distance of 300 NM. Comparing the change in minutes to the total trip time and flight time, a change in CI from 0 to 100 can lead 1.24% savings in total trip time (Table 2.5).

The increase in achievable time savings is larger with longer flight times. For flights with an average distance of 500 to 600 NM the flight time constitutes 77% of the total trip time. According to the results a reduction in total trip time of up to two minutes is possible. Comparing the time reductions achieved by an increase in CI, time savings of 1.81% for total trip time and 6.41% for flight time can be reached with the highest CI level (Table 2.5).

This effect is even stronger for the two flights with a distance of 800 NM and 1200NM. Flight time accounts for more than 85% of total trip time. According to the aircraft performance model flight time reductions of up to three minutes can be achieved by increasing the cost index to CI 100. The results of our model show that an increase in CI from 0 to 100 can decrease the total trip time by 2.1% for the 800 NM flight and 1.5% for the 1200 NM flight. This is equal to a reduction of around eight minutes in flight time and three minutes in total trip time.

Through calculating the change in fuel consumption for each CI level we can derive the cost of fuel. We refer to the average commodity price for jet fuel in the year 2015, which is 0.42 €/kg (Index Mundi 2016). Considering the complex cost structures for cost of time, we model cost of time as a function of

fuel cost. In Figure 2.9, we show examples for the flight distances and the effect of different cost of time values on the optimal cost index. The graphs show an overview of the effect of cost of time on optimal CI levels. In the graph three curves are shown. The dotted line depicts fuel cost. The dashed line show cost of time and the solid line is the total cost. All values are in reference to CI 30. Thus, negative values of cost are also possible. The total cost curve greatly varies depending on the distance of the flight and the cost of time. According to the expertise of the airlines' operations control, we generate three different scenarios for cost of time. We set cost of time equal to 10 €/min., 100 €/min. and 500 €/min. to show the effect on total cost. The white sphere shows the global minimum and the solid, black sphere the global maximum. The optimal cost index is equal to the global minimum.

In the following, we describe the step increase in cost of time, at which a new cost index becomes optimal. Short haul flights are the most insensitive to a change in cost of time. With small amounts of cost of time the optimal CI is 0. Only at a cost of time of 50 €/min. CI 0 is no longer optimal. Due to the limited effect of total trip time reduction the optimal cost index does not exceed CI 90. Time savings level off beyond CI 60. CI 60 should not be exceeded, because an increase beyond CI 60 results in a large increase in cost of fuel with tiny cost of time reductions. CI 100 is never optimal and less favorable than a lower CI. As we can see in 2.9 CI 90 is not optimal, even when cost of time reaches 500 €/min. Therefore, CI 90 becomes optimal only at very high costs per minute, which need to be larger than 500 €/min. Flights with an average distance between 500 NM and 600 NM are more sensitive to changes in cost of time. CI 0 is not optimal as soon as cost of time reaches a value of 22 €/min. For cost of time ranging from 22 € to 110 € the optimal CI varies from CI 20 to CI 50. Also for these flights CI 100 is only optimal at very large cost of time. For the two flights with a distance over 800 NM CI 100 becomes optimal at a cost of time of 152 €/min. The amount of minutes that can be saved by an increase in cost index are higher and therefore the impact of increased cost of time is stronger. CI 0 is not optimal once cost of time reach a value greater than 4 € per minute.

Through considering cost of time as a function of fuel cost we are able to show the difference between flights. Distance and flight time have a large impact on the amount of minutes that can be saved with an increase in CI. Comparing the short flights of 300 NM to the longer flights of more than 800 NM, we see that the required cost of time to change from one CI level to the next varies greatly. Therefore we recommend to fly short haul flights with a distance below 500 NM at a cost index of 20. Only at substantial increases in cost of time CI should be adapted. Due to a limited effect on saved time CI 60 should not be exceeded for short haul flights. For flights with a distance of 500 to 600 NM we recommend not to exceed CI 80 as our analysis has shown that a higher CI is only optimal at very high cost of time. For normal conditions CI 20 is recommended. For flights with a distance greater than 800 NM we recommend a CI between 30 and 40. Considerable time savings can be achieved with an increase in CI. Furthermore, the full envelope of CI levels should be used as our analysis has shown that CI 100 is optimal at a cost of time exceeding 152 €/min. With this information airline operations are able to take more differentiated decisions for CI level changes of continental flights.

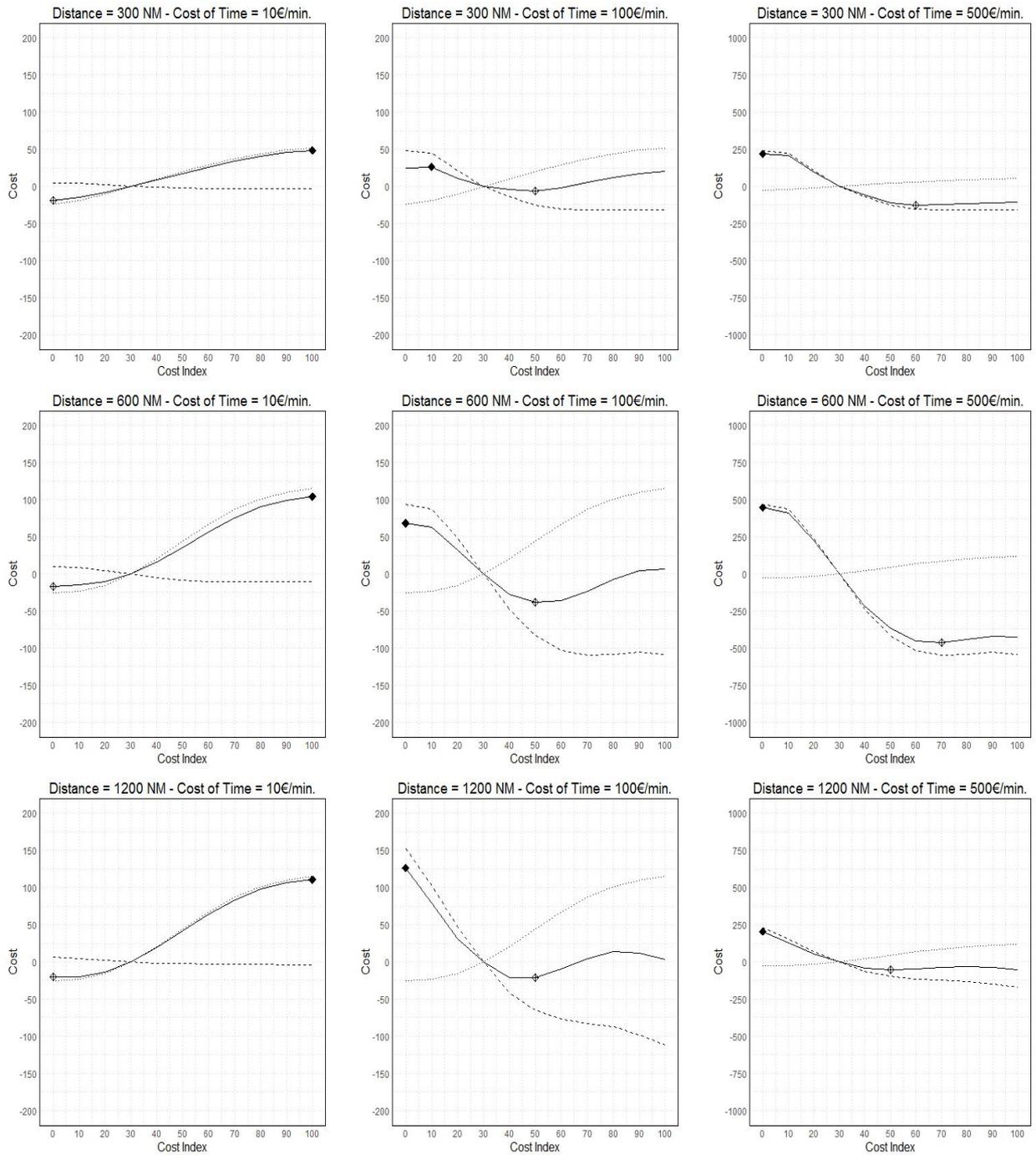


Figure 2.9: Considering fuel cost and different cost of time for the distances 300 NM, 600 NM and 1200 NM we see the variability in cost index. Solid, black square equals maximum cost. White square, equals minimum total cost.

### 2.5.3 *Managerial implications*

Our model is the first model, which is able to give high fidelity arrival time predictions before the flight departs. Through the early and accurate prediction operations control has more time to analyze and decide on measures to ensure smooth operations. Previously such accurate arrival time predictions were only available 30 minutes prior to landing. Considering that continental European flights have an average total trip time of 100 minutes, the time window for decision making increases by more than 60 minutes. Integrating cost index optimization with arrival time predictions has not been done before. With this study we are able to quantify the effect of cost index changes on an aircraft's arrival time. CI 30 is a common CI for most continental European flights. The results presented in this study show that a CI lower than 30 is more economical in case of small cost of time. A decrease of their aircraft's cost index saves fuel without any strong effect on total trip time and therefore arrival time. This is in line with the recommendations published by Aktürk et al. (2014), Cook et al. (2007). Considering that today the standard CI is 30, decreasing it to CI 20 would result in a fuel reduction of up to 100 kg per flight. With fuel costs at an all time low (0.42 €/kg) total cost savings amount to 42 € per flight. Although this is not a large amount, it is multiplied by the large number of flights per year. Furthermore a CI reduction, which will have no considerable impact on arrival times, should be considered from an environmental perspective as it will result in a reduction of greenhouse gas emissions. Nevertheless it is important to note that cost of time can greatly impact the optimal CI. As we have shown with our model increasing the cost index to reduce flight time works, but the overall impact in saved minutes is often not more than 2 %. Nevertheless in case of very high cost of time CI 100 can be optimal. This applies especially for flights with distances over 800 NM.

## 2.6 CONCLUSION AND OUTLOOK

In this paper we apply machine learning to large amounts of flight data. The goal is to model aircraft arrival time more precisely. Several features such as weather forecasts, time related features and airport congestion data, have shown to be important predictor variables. Our prediction is able to improve airline operations, as it provides reliable and accurate arrival time predictions at the point of departure. Furthermore we provide insights to the effects of CI levels on gate arrival time. Through altering the scheduled trip times of a flight according to the time savings associated with the CI level, we can generate arrival time predictions for the different CI levels. With the accurate arrival time prediction and the cost index optimization model, we expect that airlines could reduce CI levels without any significant impact on overall punctuality. Often time savings achieved through a high cost index are lost in the descent and taxi phase. As we show in our model increasing the CI level is in many cases not an efficient option to decrease total trip time. Furthermore, there is a critical trade-off between fuel consumption and minimizing a flight's total flight time. Although cost index optimization is a very popular tool used by airlines to

decrease a flight's trip time, its benefit has not been quantified comprehensively. With our model airlines gain insight in the impact of CI levels on arrival times and through this they are able to lower CI levels, while maintaining efficient and punctual operations. The arrival time prediction results in lower delay costs and the cost index optimization allows for significant fuel savings.

There are several limitations to this research. First our data focuses on one airline. It would be interesting to apply the model to an airline of similar size and set-up to test the results on an unrelated sample. Furthermore we considered cost of time per minute for CI optimization. Often cost of time cannot be calculated per minute for all cost groups. Additional cost for personnel and equipment can be depicted as marginal costs per minute, but delay costs often only occur, if a flight does not reach the airport at a certain time. In the aircraft performance model normal flight conditions are assumed. Weather conditions, especially wind, can influence an aircraft's speed immensely. Therefore we see the lack of en-route weather conditions as a limitation of the aircraft performance model. With a more accurate aircraft performance model the quality of the features for our prediction model would improve and give a more realistic flight time estimates.

#### 2.6.1 *Future research*

To conclude we outline several interesting future research opportunities based on our model. A natural extension of this study would be developing a more accurate and integrated estimate of delay costs so that the cost index optimization can rely on more realistic data. First of all the calculation of all costs related to time need to be investigated further. A model able to combine cost blocks based on the minute of additional flight time as well as cost blocks only occurring, if a flight passes a certain threshold would be optimal. This approach depends on airline's capability to accurately calculate all cost blocks related to an aircraft's arrival time. Secondly this work is focused on continental flights with short en-route phases. Therefore it would be interesting to investigate the influence of cost index changes and resulting cost optimizations for intercontinental flights with longer en-route phases. Third the model can be extended by including additional levers that can have an impact on arrival times. These levers could include permissions to fly direct during the flight, prioritized landing and taxiing or gate swaps.

### 3.1 INTRODUCTION

The International Air Transport Association (IATA) forecasts passenger demand for air travel to almost double until 2035. Even though most of the growth is expected on routes to Asia and within Asia, considerable growth for the European market is also predicted (Oxley 2016). Today European airlines are already struggling with congested air space and tight schedules. This is often the cause for delays. The accommodation of additional passenger traffic will further intensify the problem. Delays are very costly for airlines. Especially delay propagation throughout the network leads to additional costs. Through increasing situational awareness, airlines can anticipate delays and overcome them by adapting their schedules proactively. There are two factors to increase situational awareness. First, the accuracy of the available information has to be improved. Secondly the timely availability of information should be advanced. By providing precise information quickly, this allows for better decision-making. Ensuring an early availability of the more precise information will increase decision-making time. Considering the complexity of flight operations, an increase in decision-making time and information reliability can have a substantial impact on operational robustness and punctuality. Especially the accurate arrival time forecast is valuable. With a more accurate arrival time forecast resources for the turnaround process can be scheduled more efficiently. Airlines need to use their costly assets as efficiently as possible. Maximizing capacity utilization for costly assets such as aircrafts is a key driver for cost efficiency.

The airline we study in our work is a major European legacy carrier. They operate over 350 aircraft including 100 aircraft for long distance flights. In 2016 total sales amounted to approximately 23 billion €. Their network is setup in a hub and spoke system including two main hubs.

Even though airlines were among the early adopters of information technology (IT) for their businesses, older airlines are often challenged by their legacy IT infrastructure. The IT landscape includes many different systems, which often lack integration. This results in a lack of data availability and information gathered can be underutilized. Often times younger airlines have a competitive advantage over legacy carriers with their less fragmented and newer IT infrastructure (Barnhart and Smith 2012). To stay competitive, legacy carriers need to catch up with their more agile competitors. According to Kletzel (2017), Smith and Kletzel (2015) digitalization is at the top-of-mind for many decision makers in the commercial airline industry. Until recently many airlines have fo-

---

<sup>1</sup>The following chapter is based on the work of Achenbach, Schmal and Spinler (Achenbach et al. 2017). This unpublished manuscript titled "Prescriptive Analytics in Airline Operations: Intercontinental arrival time prediction" is currently under revision at Interfaces.

cused their digitalization efforts on marketing and customer relations. But now interest is shifting to operational applications as well. One example is predictive maintenance, which helps in scheduling maintenance work by predicting equipment failure. Improvements in airline operations can be achieved through the application of machine learning for forecasting and prediction. Another focus area is the improvement of decision support tools through data analytics. According to Kletzel (2017) many airlines struggle to move beyond common buzzwords and implement actual digitization projects, which impact their service or cost. They struggle with system integration, data quality and real-time availability making it hard to use information gained through analytics in daily operations.

Several streams of research can be connected to the topic of machine learning in airline operations with the focus on arrival time prediction. Previous studies have focused on predicting arrival times of continental flight connections for the US airspace (Kern et al. 2015, Takacs 2015). With the use of machine learning algorithms such as gradient boosting, generalized linear regression or random forest regression, researchers delivered improvements of up to 40% in prediction accuracy of touchdown and gate arrival times. Other work has focused on predicting landing times (Glina et al. 2012) as well as forecasting gate arrival times for specific airports (Kim 2016). Due to flight data availability in the United States, many studies have focused on continental US flights. Achenbach and Spinler (2018b) focuses on arrival time prediction for continental European flights. As of yet, no work has been published, which focuses on arrival time prediction for intercontinental flight connections.

We use actual flight and weather data of a European legacy carrier for the years 2015 and 2016. The focus of our work is on predicting the aircraft's arrival time not only at the block-off and take-off, but also at the point of entering and leaving the en-route flight phase. The dataset includes intercontinental connections originating from over 58 airports. In total, over 52 features focusing on flight information are used. Furthermore, we include en-route weather data extracted from the operational flight plan (OFP). First, we analyze the data quality of the OFP data by comparing it to actual weather events. With the insights we gain in the data quality analysis, we generate different feature subsets with up to 80 additional features. Using these feature subsets we can quantify the improvement in prediction accuracy by including en-route weather data. Furthermore, the airline gains insights on the data quality and the value of OFP data for arrival time prediction. Additionally, we predict gate arrival time at different times throughout the flight. The different times of prediction allow for a comparison between increasing prediction accuracy and decreasing flight time. In Figure 3.1 the flight process and prediction points are depicted. Furthermore we generate a comparison of predictive models by using linear regression, random forest regression (Breiman 1984) and gradient boosting (Friedman and Breiman 2000) for each prediction. We compare the results to the airline's forecast. We intend to quantify the impact of en-route weather data on overall prediction accuracy, which allows us to answer the following questions:

1. Arrival time prediction: Can we outperform the airline's arrival time forecast by using machine learning models?

2. En-route weather data: How accurate is the weather data of the operational flight plan? Does this weather data further increase prediction accuracy?
3. Trade off between prediction accuracy and remaining flight time: What is the optimal prediction point considering increasing prediction accuracy and decreasing flight time?

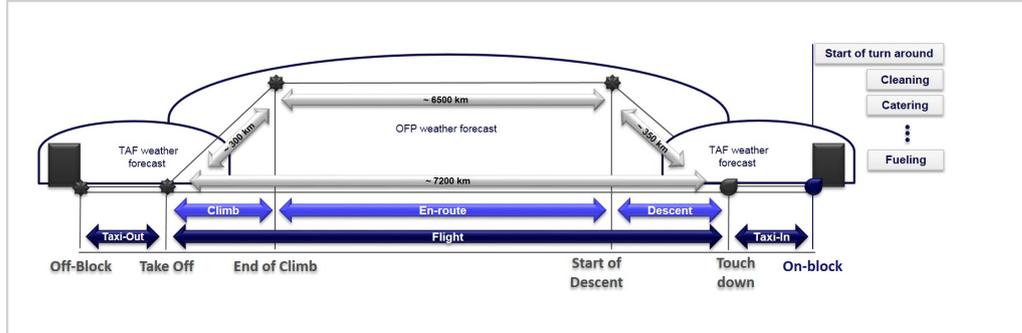


Figure 3.1: The flight process includes five segments: Taxi-out, Climb, En-route, Descent and Taxi-in. For long distance flights the average en-route phase lasts 6,500 km. The OFP weather forecast is available from take-off until touch down. Gate arrival time is connected to a variety of resources and tasks for the turn around process.

This paper discusses the development, implementation and challenges of using machine learning in airline operations. With our work we provide an example of how an airline successfully implements operations research to enhance their operations. Considering that airlines have large amounts of operational data, there are a variety of possible machine learning applications. With our focus on arrival time prediction for intercontinental flights, we focus on an important part of a network carrier's business model. Furthermore, we are able to increase the time of decision-making by delivering a higher prediction accuracy before departure. Thus, operations control has more time to take decisions and implement cost effective solutions to prevent and overcome delays by using our machine learning model. Our results show that the use of machine learning algorithms outperform the airline forecast as well as the multiple linear regression. The machine learning algorithms are able to not only account for the strong linear correlation of distance and flight time, but can also accurately model non-linear factors and their influence on arrival time. Additionally, we are the first to make use of en-route weather data to predict aircraft arrival times. The accuracy of en-route weather data provided by the operational flight plan is compared to actual weather data collected through the aircraft meteorological data relay system (AMDAR). Thus, we can quantify the accuracy of OFP data and adapt our choice of features accordingly. Even though most airlines can access rich datasets of historic flight data, the effort of generating and including additional features such as en-route weather data is hard to quantify. With our work we deliver a subset of tested features that can improve arrival time prediction for intercontinental flights.

This paper is organized as follows. Section 2 reviews the relevant literature. In Section 3 we focus on the verification of the provided en-route weather data in the OFP by matching it with actual observed weather and assess the accuracy.

The prediction model is suggested in Section 4, wherein the data, features and the current baseline model are described. Section 5 reports the results of each model. Section 6 provides a conclusion and describes implications for theory and practice in addition to limitations and directions for future research.

### 3.2 LITERATURE REVIEW

There are several streams of literature which are relevant for our work. First, we outline the current literature on arrival time prediction. We then focus on flight operations and the use of en-route weather data. We conclude with a review of our prediction methods such as machine learning and linear regression and the application of these methods in airline operations.

#### 3.2.1 *Arrival time prediction*

According to Murphy et al. (2003) airlines implement parametric of physics-based trajectory models to calculate a flight's arrival time. There are several common software solutions for airlines to manage their flight operations. Through the use of route optimization, which focus on distance, minimum fuel burn and altitude, flight plans are generated. Additionally, the software considers Air traffic control (ATC) information, weather forecasts and operational aircraft data to generate an arrival time prediction (Lufthansa Systems 2013).

A large amount of research focuses on delay prediction and delay propagation. Mueller et al. (2002) analyzes arrival delay distribution for ten major US airports. Xu et al. (2008) predict positive and negative delays for 34 US airports. Xu et al. (2008) identify Ground delay programs (GDP), carrier, departure demand and previous delays as influencing factors for arrival delays. Tu et al. (2008) develop a delay distribution for flight departures. Rebollo and Balakrishnan (2012) focus on a network-wide delay prediction and predict the network delay state with a two hour horizon. state that airlines operating a hub-and-spoke system have higher costs for arrival delays compared to point-to-point carriers. The work serves as input for feature generation and emphasizes the need for accurate arrival time prediction to minimize delays.

For arrival time prediction, the focus is either on predicting remaining flight time or gate arrival time. Willemain (2001) analyze the systemic variability such as weather and traffic in flight time. Srivastava (2011) develops a model for taxi-out time prediction based on historical data from John F. Kennedy airport. Tobaruela (2014) is one of the view studies focused on the European airspace. They aim at improving collaborative decision-making through increasing air traffic predictability and capacity. Directly focusing on arrival time prediction are the following studies. Glina et al. (2012) uses a quantile regression forest to predict landing times at Dallas/Fort Worth International Airport. The key predictor variables in their model are remaining distance and altitude. Weather data is not included in the model. A similar approach was implemented by Kern et al. (2015). They use random forest regression to predict estimated time

of arrival. In their work, they focus on quantifying the impact of certain variable subgroups on prediction accuracy. Their results show that the inclusion of standard airline and flight information together with weather and airport congestion data delivers the highest accuracy. Most famous is the work done by the different participants of the GE flight quest. GE Aviation and Alaska Airlines challenged data scientist for an improved arrival time prediction model. The flight data focused on U.S. domestic flights including variables such as flight information, flight route, weather reports and air traffic control events. The superior model outperformed the airline's forecast by more than 40% (Takacs 2015). Another important contribution was made by Kim (2016), who developed a non-parametric additive model to predict flight arrival times. The scope was limited to flights arriving to Denver International Airport. The predictor variables do not include any en-route weather data. Only the weather conditions at the departure and arrival airport are included. The best model achieves a prediction accuracy of 8.5 minutes MAE. In their work, Tielrooij et al. (2015) consider the uncertainty of predicted arrival times in an air traffic control setting. They state that the short planning horizon of 20 to 30 minutes limits the ATC choices to ensure efficient flight operations. They consider flights arriving to Amsterdam Schiphol International airport. In the work of Achenbach and Spinler (2018b) gate arrival time predictions for continental European flights are made. Through the use of a machine learning ensemble including linear regression and gradient boosting a prediction accuracy of 4.5 minutes MAE is achieved. The authors do not use en-route weather data except for Terminal Aerodrome forecast (TAF) reports for the departure and arrival airports.

### 3.2.2 *En-route weather data*

The operational flight plan is generated by the flight planning software prior to each flight. It includes information on the aircraft and the planned route. The origin and destination airports as well as the waypoints that will be passed during the flight are listed. Additionally, the flight plan includes a weather forecast for each waypoint. The weather forecast consists of forecasts for temperature, wind speed and wind direction. Furthermore, an outbound track and air speed of the aircraft when passing through the waypoint is listed. The World Meteorological Organization in cooperation with aviation partners initiated the Aircraft Meteorological Data Relay. The AMDAR system serves as a data collection of upper-air data, which is used for weather forecasting, climate monitoring and early warning system for severe weather events. As of May 2017 the global AMDAR system produces more than 700,000 high quality observations per day. The data is collected through sensor-equipped aircraft. Besides the positional and temporal information, data such as air temperature, wind speed, wind direction and humidity are monitored. The data is transmitted to the respective national meteorological services. The national meteorological services ensure quality control and process the data. In total there are 40 participating airlines, many of which are large network carriers such as Air France-KLM and United. Several studies have analyzed the effect of en-route weather on arrival time. Callaham et al. (2001) develop a weather impacted traffic index to make

airspace performance comparable. Closely connected is the work of Chatterji and Sridhar (2005) who identify weather as the most important factor for delays in the US airspace. Abdelghany et al. (2004) focus on delay propagation and state that 75% of system delays can be attributed to bad weather.

### 3.2.3 *Prediction methodology*

Multiple linear regression is a standard tool in prediction. Often it is used as a baseline model for comparison. The regression focuses on minimizing least squares. The predictor variables are assigned regression coefficients based on fitting the model to the data. Random forest regression is one of the most commonly used predictive algorithms in machine learning. The foundations of this algorithm are based on the early work of Breiman (1984). Breiman (1984) developed the CART (classification and regression trees) algorithm in 1984. It extends the ideas of decisions trees and improves the drawbacks of single trees through several functions. First overfitting is decreased by averaging predictions over all trees. In general 500 or more trees are trained for a prediction. Second, trees are de-correlated through bootstrap sampling, which selects only a subset of predictor variables to consider at each node split. Through this process not only the same strong predictors are considered at each node split, but a random subset. Gradient boosting machines came about in the late 90s. They were developed by Friedman and Breiman (2000) and are related to random forests, but instead of the original observation residuals are used for prediction. Gradient boosting builds on the idea of combining a set of weak learners to generate a strong predictive model. The model performs sequential operations by assigning higher weights to misclassified outcomes. Wrongly classified outcomes get a higher focus and are modeled more accurately in the next stage. This process is iterated several times until overall prediction accuracy has improved. Gradient boosting and random forest regression are especially useful for unclean datasets.

The different streams of research show the potential of machine learning to enhance aircraft arrival time predictions. However, while they offer insights on arrival time prediction for continental flights, they do not provide any application to intercontinental flight routes. The inclusion of en-route weather data as predictor variable is very limited in all previous studies. In our model, we aim at closing this gap by providing high-fidelity arrival time predictions for intercontinental flight connections and assessing the value of en-route weather data as features.

## 3.3 ASSESSING DATA ACCURACY OF EN-ROUTE WEATHER DATA

Data quality is a challenge in predictive modeling. In our model we rely on forecasted en-route weather generated by the meteorological services included in the OFP. To ensure that the OFP data is reliable and accurate, we compare the OFP data to actual weather data generated by the AMDAR system. The data

covers the period from 1st of July 2015 until 29th of February 2016, which totals eight months. For this period we are able to include en-route weather data in our prediction model. We quantify the data accuracy of the three weather variables temperature, wind speed and wind direction. In Figure 3.2 we display a random sample of 7,000 OFP data points to illustrate the routes and collection intervals for intercontinental flights.

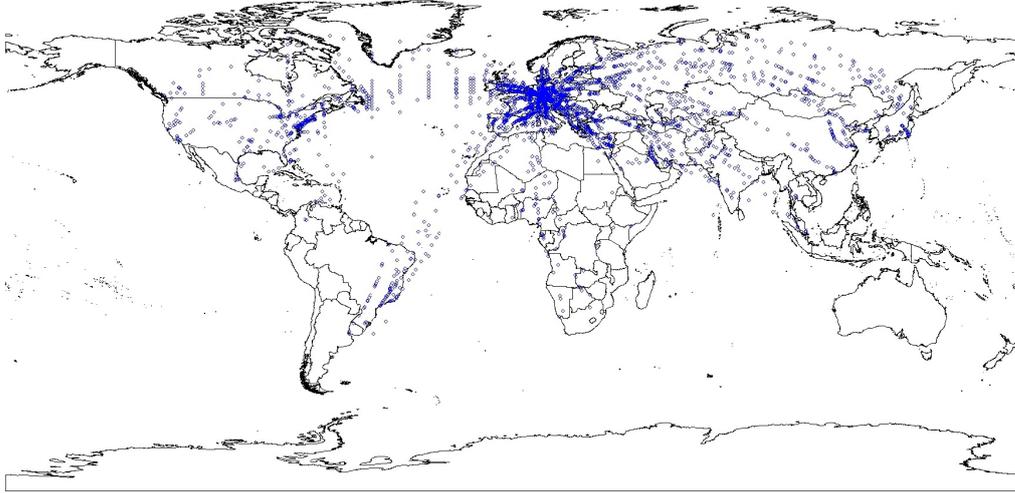


Figure 3.2: This example shows a collection of OFP data points. On average 62 waypoints are passed during a long distance flight. For each waypoint OFP weather data is generated. On average every 7 minutes a new OFP event is produced. Similar maps can be drawn with the AMDAR data.

Our AMDAR dataset consist of data for the year 2016. The overlap of AMDAR and OFP data includes two months: January and February 2016. Considering that those months are in winter and therefore adverse weather conditions commonly occur, we expect that accurately forecasting weather events is challenging for these months. Date, time, latitude, longitude and altitude serve as matching criteria. We allowed for a time window of 30 minutes, a deviation of latitude and longitude of 0.5 degrees and an altitude boundary of 500 meters. In total, we were able to match over 150,000 flight plan data waypoints (en-route) to actual AMDAR weather events. To calculate the accuracy of the three important variables "Temperature", "Wind direction" and "Wind speed", we used mean absolute error. We use MAE instead of RMSE, as it is considered superior evaluation for average error magnitude Willmott and Matsuura (2005). The MAE for wind direction and temperature are 2.8% and 4.4% respectively. For wind speed, there is a higher MAE of 17%. The differences in accuracy were analyzed in more detail by considering remaining ground distance, altitude levels and the region. For temperature, we do not observe a large variation in prediction accuracy. For wind speed we observe larger magnitudes in change of prediction accuracy. For flights not within Europe, the error in prediction accuracy increases up to 36 %. For wind direction it is similar. The farther the aircraft is from its destination, the more inaccurate are the predictions. Nevertheless, the error rates for temperature and wind direction

are quite low and the operational flight plan delivers reliable forecasts for these two features. We only include wind speed in a few features, as the analysis showed that the accuracy is low.

### 3.4 MODEL FORMULATION

Contrary to many previous studies such as Takacs (2015) and Kim (2016) which focus on arrival time prediction for continental routes, this study is the first to make use of a rich dataset of intercontinental flights. The data includes flights originating from all continents with an average distance of 7,000 km. Two major European hubs are the destination airports. We predict the remaining flight time at different critical flight stages such as block-off, take-off, end of climb phase and start of descent phase. The historical flight data ranges from March 2015 to February 2016. To further enhance the prediction accuracy of our model, we include en-route weather data, which was derived from the flight plan. The en-route weather data is comprised of temperature, wind speed, wind direction and flight path forecasts for planned waypoints. For prediction purposes, the vast weather information was aggregated to several informative features, which are explained in detail in Section 3.4.2. In total over 20,000 flights are used for model training, validation and testing. Incomplete flights with missing values were excluded from the dataset.

After training the ensemble with over 15,000 flights, fine tuning and cross validation was done with a validation dataset of 3,100 flights. The final ensemble was tested on a set of 1,500 flights. The datasets are made up of flight data ranging from March 2015 until February 2016. The datasets are generated through random sampling. For the data set including en-route weather data, we have a training data set of over 7,500 flights and a validation data set of 1,600 flights. The final testing was done on a sample of 750 flights.

#### 3.4.1 *Target feature and baseline*

The target feature is remaining flight time including taxi time at the destination airport. We refer to this feature as the actual total trip time. Due to the different points in time when the predictions are made, the target feature changes. Thus, taxi times at the departure airport are only included in the first prediction "block-off". The target feature (in minutes) was not changed and could be directly extracted from the flight data file provided by the airline. Thus, only positive values are available. Note that scheduled total trip time denotes the airlines flight schedule times and predicted total trip time is the predicted time according to our model. The shortest flights have a distance of 3,000 km (ground distance). Therefore, the target feature varies from 224 minutes up to 837 minutes for the longest flight, with a ground distance of 11,600 km. When looking at the distribution of the distances between origin destination pairs, we observe the following: There is a peak at 3,000 km, 6,000 km and 9,000 km.

The average flight covers a ground distance of 7,000 km. The amount of flights landing at the two arrival airports is skewed with 75% of flights arriving at one airport and 25% at the other. On average, we observe a delay at block-off of over 8 minutes. The baseline is derived from the airline's flight planning system. Their arrival time forecast is based on the flight route. Distance, altitude, temperature and pressure are used to calculate the optimal flight path and speed. Head and tail winds are considered to determine ground speed and therefore arrival time. This arrival time forecast is used in the operations control center for all operational decisions. Furthermore, it is transmitted to air traffic control.

### 3.4.2 *Features for prediction*

As features we include some, which were provided by the airline and several, which were developed in the feature generation process. The analysis of previous works, such as Kim (2016) served as input for feature generation. Overall there are 55 features in our prediction model, which does not make use of en-route weather data. In total there are 48 numerical and seven categorical features. The features can be structured in several subgroups and are explained in the following paragraphs.

The first subgroup consists of time related features. These features include the hour of departure and the hour of arrival. Furthermore, we aggregated the hour of arrival into six subgroups to account for busy airport hours. These subgroups are early morning, morning, noon, afternoon, evening and night. Month and weekday are also included. The month are aggregated to the feature "Season" to account for delayed operations during winter times or peak capacity in the summer month. A feature to account for national holidays was created. It counts the days until the next major national holiday to model increase in traffic before holidays.

Through the information provided by the airline, we have several features which describe the expected flight times or taxi times for each flight. These include the planned total trip time for the whole flight and for each flight segment (climb, en-route and descent) and the forecasted taxi out and taxi in times. Additionally we created two features to enhance the information of the different subsections of the flight. One is the planned flight time and the other one the time from block-off until touch down. Furthermore, we include block-off delay. Considering that the first prediction point is taking place at the block-off moment, we can calculate if the block-off moment was delayed, on time or earlier than expected. Early departures are shown by a negative block-off delay. Another important set of features was created by calculating the mean total trip time of previous flights on the same origin destination (OD) pair. First, we calculate the mean total trip time for the last three flights that took place most recently before the flight. If there is only one or two flights before the predicted flight, the average was calculated with those values only. If there are no previous flights, the value is zero. Additionally, we include a feature focusing on the total trip time of the last flight. Lastly, we calculate the

difference in predicted and actual total trip time of the last three flights and include it as feature.

In addition to the information on the departure airport, we include the TAF weather forecast for the departure airport. The TAF weather forecast includes horizontal and vertical visibility, wind direction, wind speed, cloud base and wind gust. The weather data was enhanced by four features. We created binary features for weather events such as snow, rain or fog at the point of departure. This information was extracted from the TAF raw text report. Additionally, we create a binary feature for bad weather to aggregate the information spread across in the different TAF features. If any of the following criteria is fulfilled, we consider it bad weather: wind gust larger than 40 knots, wind speed larger than 20 knots, vertical, horizontal visibility or cloud base are smaller than 1 km. To ensure uniqueness for departure stand and departure runway, we combine them with the departure airport. The features regarding the arrival airport are the same as for departure airports. Only arrival stand is not included as we do not have this information.

Additional information on the aircraft such as expected total take off weight and aircraft type were included as features. Furthermore, we include the fuel type. The number of passengers and the maximum altitude are features as well. Features related to distance are the planned ground distance in km and the expected air distance in km.

According to industry experts, airlines focus on punctual operations especially on competitive routes. To model this focus on punctuality, we include a binary feature for the ten busiest routes according to departure airports. These are Dubai, New York, Newark, Hong Kong, Singapore, Los Angeles, Montreal and Chicago. If the flight is departing from one of these airports the competition feature equals one.

Due to the lack of aggregation in the flight plan data, we had to generate meta features which were calculated for each flight based on the data given in the OFF. As there is no comparable literature, the feature generation process for the en-route weather data followed a trial and error process. In a first step many features were generated and their predictive power was analyzed. If the features improved overall prediction accuracy, they were included in the final model. Due to the different points of prediction, this form of feature evaluation was done for each prediction point. In the following paragraph the different features are described.

We begin calculating the number of waypoints for each flight. Due to the different flight lengths, the number of waypoints can vary from 19 to 138. Next, we determine the average, minimum and maximum for wind speed, altitude, temperature and airspeed. To ensure comparability of those derived features, we generate the features based on en-route weather events, which occur once the aircraft has reached flight level. By limiting the data to only include weather events after end of climb and before start of descent we achieve this. Considering that we include weather forecasts of the arrival and departure airports and considering that climb and descent phase only make up a very small portion of overall flight time (12.5 %) it seems reasonable to exclude the weather data where the flight level is not reached. In a next step, we create a feature for the vector of flight direction and wind direction. We want to understand, if a

flight has favorable wind conditions such as tail wind or less favorable flight conditions such as head wind. Finally, we combine the average wind speed with wind direction into a feature. Furthermore, we calculate the cosine of the difference between outbound track of the aircraft and wind direction. We multiply it by the wind speed to create the feature of average wind component.

To account for the different prediction points, we generate the same features for all remaining distances larger than 2,000 km, smaller than 2,000 km but larger than 500 km and smaller than 500 km. These features are integrated into the prediction model.

### 3.4.3 Predictive model

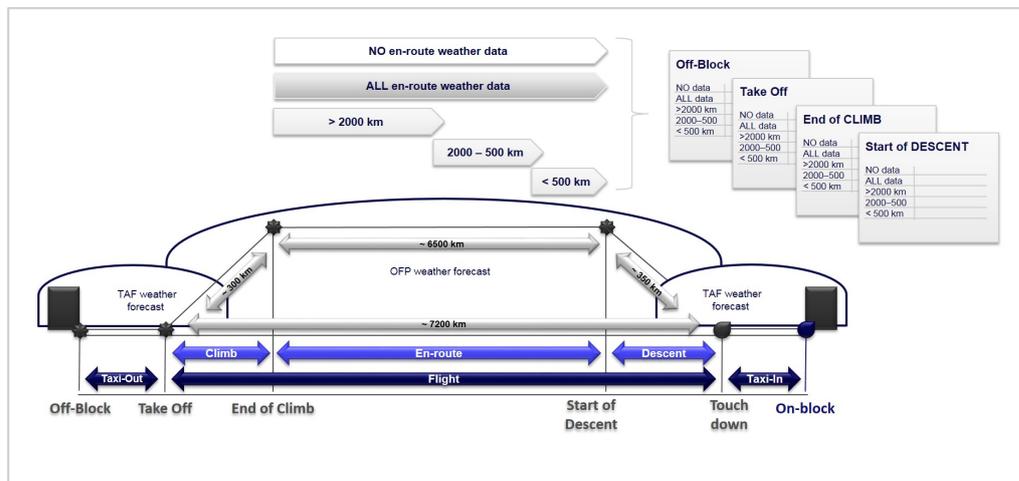


Figure 3.3: In the prediction model we include five different subsets of OFP weather data. They were generated depending on the remaining distance until the start of descent.

We relied on three different methods for prediction: multiple linear regression, gradient boosting and random forest regression. To learn more about the importance of the weather data, we included different subsets of the en-route weather data. First, we make predictions with all three models without including any en-route weather data. Second, we include all features that were calculated with the en-route weather data also including the different subsets. Third, only the features calculated on the whole en-route dataset are used. We call this subset of weather features the "Overall" en-route weather data. In the following the remaining subsets such as weather data including all en-route information occurring 2,000 km before start of descent are included. En-route weather data including all events between 2,000 km and 500 km before descent form a subset, as well as weather data 500 km before the start of descent. All these subsets are used for predictions separately. An overview of the prediction model is shown in Figure 3.3. Considering the four prediction points (block-off, take-off, end of climb, start of descent), the six different data subset combinations and the three

different methods for prediction, we run 72 different predictions to determine the optimal dataset and the superior prediction model.

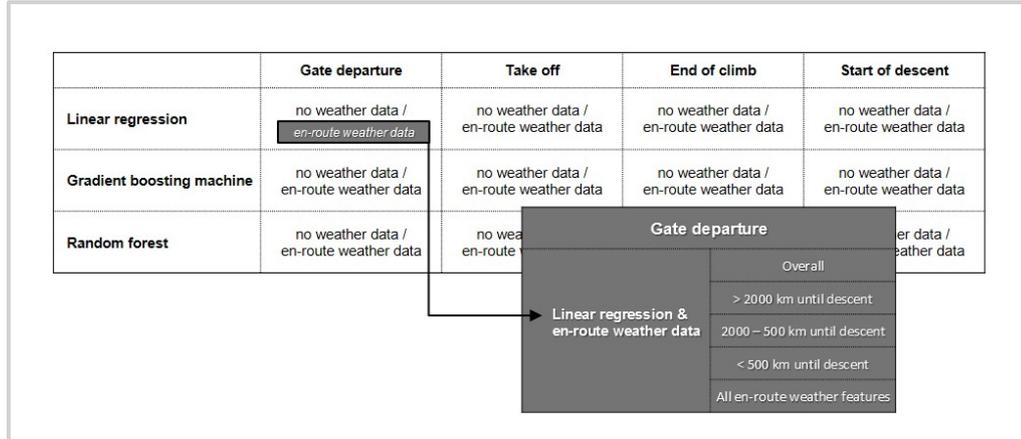


Figure 3.4: For each of the four prediction points (Off block, Take off, End of climb, Start of descent) we generate five feature subsets generated from the OFP weather data and run three prediction models with linear regression, random forest regression and gradient boosting to determine the optimal model.

Results are evaluated using mean absolute error and standard deviation (SD). When evaluating our predictions we try to minimize MAE and SD to achieve the best predictive model. Willmott and Matsuura (2005) state that MAE is a superior evaluation criteria compared to RMSE to measure average error magnitude. Thus, we use MAE and SD for model evaluation. Please refer to Appendix B.1 for detailed formulas of the evaluation criteria and to Appendix C.3 for information on parameter settings for random forest and gradient boosting.

### 3.5 RESULTS

#### 3.5.1 Prediction results without weather data

The prediction results obtained without en-route weather data delivered superior results compared to the airline forecast. The results will be described for each prediction point separately. The predictions were obtained by fitting the data once to the dataset held back for testing. The prediction accuracy and sensitivity was in line with the results of the validation dataset, which was used to fine tune the model. On the validation dataset the predictions were run multiple times to determine the best model. The baseline model consists of the airline's scheduled time of arrival calculation provided by their operations control. The different predictions and models are illustrated in Figure 3.4.

At the point of gate departure (block-off), the linear regression delivered the lowest MAE with 6.53 min closely followed by the gbm model with an MAE of 6.65. The random forest algorithm delivered an MAE of 6.75. The same is observable for the standard deviation. Linear regression is the superior algo-

rithm, closely followed by gradient boosting and then random forest regression. Nevertheless, all models outperform the airline forecast, which has an error of 8.39 minutes for the MAE at the point of gate departure. When we look at the most important variables, we observe the following: Average forecasting error of previous flights on the same route, as well as any delay at the gate, ground and air distance and planned flight times either for the whole flight or parts of the flight are the crucial predictor variables. Less important but still included in the top fifteen predictor variables are the hour of arrival, departure airport and country and the runway at the arrival airport.

Even though the point of gate departure and the point of take-off are quite close (time difference on average 20 minutes) the prediction accuracy at the point of take-off increases substantially. We observe slightly different results as in the block-off prediction with gradient boosting delivering by far superior performance compared to linear regression and random forest regression with an MAE of 5.28. Linear regression and random forest regression both outperform the airline forecast with their MAE of 5.49 and 5.42 compared to an MAE of 6.35. The SD is reduced to 4.41 from 5. Random Forest slightly outperforms linear regression in both metrics, but cannot compare to the results of the gradient boosting. The most important predictor variables vary between linear regression and the other two models. For linear regression forecasted times and planned distances are less important compared to the block off prediction. Instead, variables such as departure airport and stand as well as time of day at departure gain importance. For gradient boosting and random forest regression, forecasted flight times and distances are the most important features. Random forest regression additionally includes features such as arrival airport and runway and planned hour of arrival.

We can only observe a small gain in prediction accuracy when comparing predictions made at the end of the climb phase compared to predictions at take-off. Again, gradient boosting is the superior model. The decrease in MAE from 5.28 at take-off decreases to 5.17 at the end of climb. On average there are 25 minutes between take-off and exiting the climb phase. Linear regression and random forest regression deliver a MAE of 5.34 and 5.33. For SD the results are very similar. Comparing this to the airline forecast with a MAE of 6.22 and SD of 5.32, we see that all models are superior to the airline forecast. The gradient boosting achieves a reduction in prediction error of up to 17%. The analysis of the most important variables is very similar to the prediction at take off. GBM and random forest regression include a large amount of forecasted flight times and distances compared to linear regression, which focuses on other features as well. For example time of day at departure, departure airport and departure stand are important features for linear regression.

Considering the average remaining distance an aircraft has at the start of the descent phase, we can make accurate predictions at this point in time. Linear Regression performs best with a MAE and SD of 3.73 and 3.2 respectively. The MAE is similar for all three models with linear regression having the lowest error. Compared to the airline forecast with an MAE of 4.63, the predictive models outperform the airline forecast by 20%. For the linear regression, important variables are planned en-route flight time, overall planned flight time and planned time for the climb phase. Furthermore, the time of day at arrival is

a critical predictor. For gradient boosting and random forest regression sparse features such as arrival airport and runway, days until the next major holiday as well as weekday and hour of arrival are important. Random forest also includes planned fuel and time values for the descent phase as main predictors.

Table 3.1: Prediction results of the best performing model without en-route weather data

Prediction point	Proposed model (PM)		Airline forecast (AF)		Difference (D=AF-PM)				
	Model	MAE	SD	MAE	SD	% change MAE	% change SD	95% confidence interval	
Block off	Linear regression	6.63	5.76	8.39	7.00	21%	18%	6.34	6.92
Take off	Gradient Boosting	5.28	4.41	6.35	5.41	17%	19%	5.06	5.50
End of Climb	Gradient Boosting	5.17	4.26	6.22	5.32	17%	20%	4.96	5.39
Start of Descent	Linear regression	3.73	3.20	4.63	5.23	20%	39%	3.57	3.89

Overall the use of gradient boosting seems to be the superior prediction method, considering that GBM performed best in two out of four predictions points and only slightly worse in the other two cases. It should be the first choice for all predictions without en-route weather data. Linear regression would be a good choice as well, but in the cases where GBM performs best linear regression is much worse. In Table 3.1 the results of all predictions without en-route weather data are summarized.

### 3.5.2 Prediction results with weather data

The prediction results obtained through the inclusion of different subsets of en-route weather data delivered superior results compared to the predictions without en-route weather data. The increase in prediction accuracy is on average 5% for MAE. Therefore, the inclusion of en-route weather data is beneficial. In the following, the best model for each prediction point will be described.

The superior model for the prediction at the block-off moment is delivered by the random forest algorithm. All subsets of weather data significantly improve prediction accuracy with the subsets of en-route weather data 500 km before descent being the best. The MAE of the linear regression model without weather data is outperformed by 5.6% leading to an MAE of 6.26. Comparing this MAE to the airline forecast, we see an improvement of 25%. SD achieve an improvement of around 17% by using the random forest model. For linear regression, the overall en-route weather data subset delivers the best results. However, when comparing these results to the linear regression without any weather data, we observe that there is almost no change in prediction accuracy for any of the evaluation metrics. In conclusion, including en-route weather data delivers a more accurate total trip time prediction than when not included. Random forest and gradient boosting can make use of the additional features and improve their initial prediction accuracy.

Also, for predictions at the point of take-off, the random forest delivers the best results considering all predictions with en-route weather data. The MAE of 5.04 outperforms the airline model by 21%. SD also improves by 13%, but is not the lowest compared to other models. The random forest makes use of

the en-route weather data larger than 2,000 km. Linear regression results again show there is no significant change in prediction accuracy including en-route weather data. The best performing linear regression model makes use of all en-route weather data, but results are only slightly better than not using any weather data. The GBM performs slightly worse with any en-route weather data subset than without any weather data. The random forest prediction with en-route weather data larger than 2,000 km is comparable in accuracy to the gradient boosting without any en-route weather data. For linear regression and gradient boosting, the ranking of the most important variables does not change significantly. For random forest regression average airspeed and temperature, minimum wind speed and average wind speed are included in the top ten predictor variables.

At the end of the climb phase, the best performing algorithm with en-route weather data is the GBM with the en-route weather data larger than 2,000 km. Compared to the best performing algorithm without any weather data (also GBM) we observe an improvement of 3%. All gradient boosting models outperform any other prediction model. For random forest regression, the inclusion of en-route weather data decreases prediction accuracy. For linear regression the change in prediction accuracy by including en-route weather data is not significant. Comparing the results of the gradient boosting to the airline forecast, we observe an improvement of more than 19 % in SD and MAE. The comparison of variable importance of the two gbm models shows that the top ten variables are similar. The ranking for the top ten to top fifteen changes by including variables such as average wind component and average temperature in the en-route weather data prediction.

Table 3.2: Prediction results of the best performing model considering en-route weather data

Prediction point	Proposed model (PM)		Airline forecast (AF)		Difference (D=AF-PM)			
	Model	MAE	SD	MAE	SD	% change MAE	% change SD	95% confidence interval
Block off	En-route weather data (RF 500 km)	6.26	5.81	8.39	7.00	25%	17%	5.97 6.55
Take off	En-route weather data (RF > 2000 km)	5.04	4.71	6.35	5.41	21%	13%	4.81 5.28
End of Climb	En-route weather data (GBM > 2000 km)	5.02	4.34	6.22	5.32	19%	19%	4.80 5.24
Start of Descent	En-route weather data (RF 2000 - 500 km)	3.49	3.40	4.63	5.23	25%	35%	3.32 3.66

At the start of descent, the superior prediction considering both models with and without en-route weather data is delivered by the random forest model including en-route weather data between 2,000 and 500 km. Comparing the MAE of 4.63 of the airline forecast to the MAE of 3.49 of the random forest model, we see a reduction in prediction error of 25%. This is outperformed by the reduction in SD from 5.23 to 3.40, which equals 35%. The linear regression including en-route weather data performs worse than the linear regression without any weather data, which was the superior model for prediction without weather data. When looking at variable importance, we observe that the ranking has changed significantly. For the random forest algorithm average temperature, average wind speed in combination with wind direction, minimum altitude, maximum airspeed and percentage of positive weather events is included in the top ten. In Table 3.2 the results of all predictions with en-route weather data are summarized.

In Figure 3.5, the change in MAE and SD is shown. For the MAE we observe that the airline forecast is outperformed by the best models including and excluding en-route weather data. The MAE for the first prediction point (block off) is almost as low as the prediction accuracy of the airline forecast at the end of the climb phase. The values are 6.26 and 6.22 respectively. Overall for MAE the predictions including en-route weather data outperform the predictions without weather data considering the MAE. For SD, the opposite is true. Both proposed models outperform the airline forecast by an averaged 20%. However, the SD for all predictions excluding en-route weather data is lower than for all predictions including en-route weather data. In Table 3.3, the results for the proposed models are compared. Considering the focus on MAE error the prediction models including en-route weather data are superior. The decrease in accuracy considering SD depends on the prediction point. Nevertheless, an improved MAE is considered better than the decrease in SD.

Table 3.3: Comparison of best models (excluding and including en-route weather data)

Prediction point	Incl. En-route data		Excl. En-route data		% change MAE	% change SD
	MAE	SD	MAE	SD		
Block off	6.26	5.81	6.63	5.76	5.6%	-0.9%
Take off	5.04	4.71	5.28	4.41	4.5%	-6.8%
End of Climb	5.02	4.34	5.17	4.26	3.0%	-1.8%
Start of Descent	3.49	3.40	3.73	3.20	6.3%	-6.1%

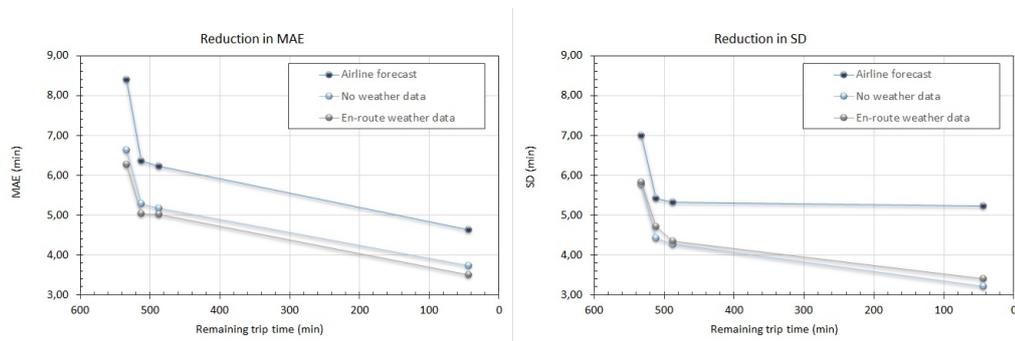


Figure 3.5: Comparison of reduction in MAE and SD including the best prediction models with en-route weather data and without. For MAE inclusion of en-route weather data delivers superior results. Considering SD no weather data is slightly better.

We analyzed the 15 most important features of each predictive model to understand differences in variable importance. Overall the prediction for all models is dependent on the planning values related to time forecasts from the flight plan. These include planned flight time for the different flight segments, overall flight time, delay minutes at gate departure and flight times of previous flights with the same connection. Only for the last prediction point at the start of descent these linear variables lose importance. Other key variables are ground and air distance as well as planned fuel usage (whole flight and certain segments). The hour of arrival, and in a more aggregated form, the time of

day at arrival are also included in the top 15 features. Comparing the random forest model without any weather data to the best performing model at the point of gate departure, the change in variable importance can be attributed to the inclusion of several weather data features. These include the feature of wind speed in connection with wind and flight direction, the percentage of times wind and flight direction are aligned, the average and minimum wind speed as well as the average airspeed. In Figure 6, an overview of the fifteen most important variables for random forest regression without and with en-route weather data are included.

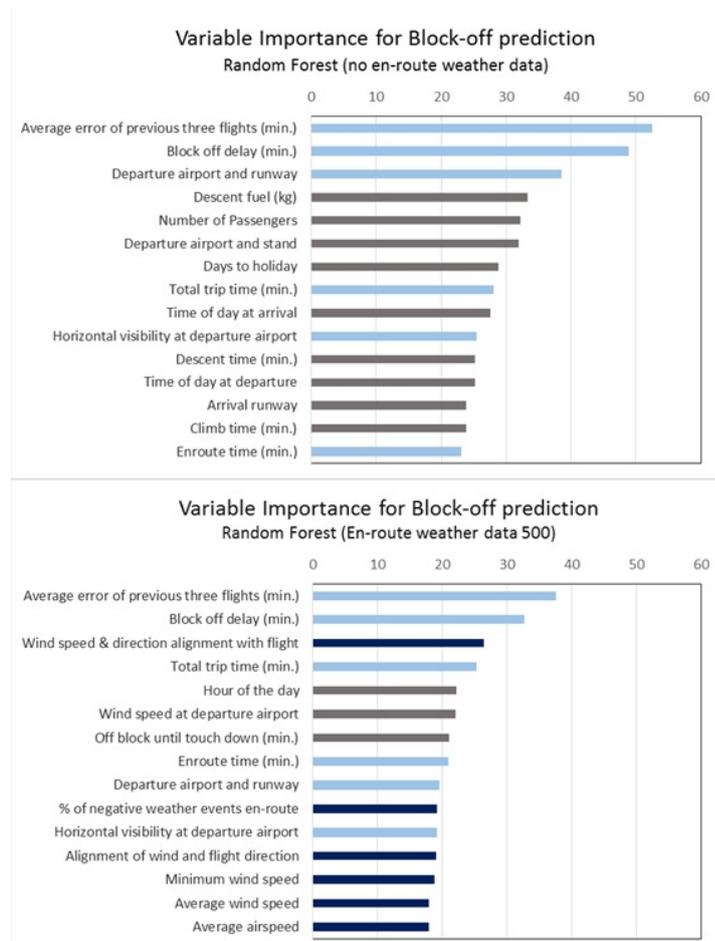


Figure 3.6: Block off: The top two features are the same in both cases. Besides that there are four overlaps of features that we included in both predictions. Prediction including en-route weather data is mostly enhanced with generalized weather features depicting overall weather conditions for the flight.

Overall, the key features for en-route weather data are the alignment of wind speed and direction with flight direction, all measures related to temperature and wind speed. Furthermore, the inclusion of historic flight events as a feature increases prediction accuracy significantly. All three historic features (average total trip time of previous flight/ previous three flights and their prediction error) are frequently included in the top 15 features for all models.

### 3.5.3 *Managerial implications*

Our first main finding focuses on the use of machine learning as a method for total trip time prediction. In all cases, machine learning algorithms such as linear regression, gradient boosting or random forest regression outperformed the standard airline forecast, which is mainly based on flight distance. Improvements in prediction accuracy of 19% to 25% are achieved. Therefore we strongly recommend integrating machine learning algorithms as models for arrival time prediction.

In practice the airline has to weight the benefits of model complexity and accuracy against implementability and ease of use. Additional model tuning such as ensemble models could be implemented to increase prediction accuracy further, but model complexity should be kept at a minimum. Therefore we recommend to use standard machine learning methods as they can be understood by people outside the subject area.

Regarding the value of en-route weather data, our findings are positive. In all cases the prediction accuracy increases when including en-route weather data. Improvements range from 3% (end of climb) up to 6.3% (start of descent). Nevertheless, these improvements in prediction accuracy have to be held against the effort of generating en-route weather features for the prediction model. More data needs to be integrated and monitored, but the feature generation can be automated similar to the other features.

Furthermore, our analysis shows the difference in prediction accuracy considering the time of prediction. At the point of gate departure we can generate an arrival time prediction with an MAE lower than 6.5 minutes. It seems to be reasonable to predict the total trip time at the point of gate departure and at take-off. The prediction at the start of descent shows that there is an increase in prediction accuracy considering decrease in distance. The MAE drops from 6.5 minutes at gate departure to 3.5 minutes. Considering that the average flight distance of an intercontinental flight equals 7,000 km aircraft arrival time prediction is already very accurate at the point of gate departure. The prediction at the start of descent can be used to schedule ground operations such as ramp services, catering and cabin services.

Airlines can streamline their operations and optimize processes with stakeholders, such as service providers for catering or the airport, by having a more accurate arrival time prediction at an earlier point in time. The increase in prediction accuracy contributes to an overall improved information for operations control and strengthens the trust in data-based decision-making. To become a data-driven business many processes need to be adapted based on insights gained through data analytics. Arrival time prediction is one component of that. Its success and the lessons learned serve as example for future projects. In the future the airline aims at improving their operations through other predictive analytics projects such as predicting aircraft weight more accurately.

#### 3.5.4 *Implementation challenges*

In August 2016 the airline implemented a new tool for arrival time prediction based on the insights gained in our study. This includes the use of a GBM model for predicting aircraft arrival times at their hub locations as well as including certain weather features in the model. Furthermore, organizational and technical challenges had to be overcome. The airline created a new analytics department to integrate data-driven decision-making in airline operations. Arrival time prediction was the first project and its success helped to convince all stakeholders, such as crews, operations control and hub control, to expand projects with the focus on data-driven decision-making. Success was measured based on the increase in prediction accuracy and the ease of use once the new method was fully deployed. Nevertheless, it is a challenge to find capable personnel. The airline hired a data scientist responsible for model building and testing, a data engineer responsible for data cleaning and data integration and an analytics ambassador, who can translate between the data scientist and the business. Integrating these new roles with existing structures and processes and ensuring a good communication with process owners was a critical task. On the technical side the airline had to face the following challenges: Integrating data coming from several data warehouses to one database, ensuring real-time data availability and overcoming data quality issues. Verifying and cross-checking for correct values took up a lot of time during the first project phase. After successfully integrating all data sources needed for the new arrival time prediction, the new forecast had to be integrated in existing processes. For example the arrival time prediction is used by hub control for turnaround and passenger transfer preparations. Even though the arrival time prediction outperforms the previous forecast, the new prediction method needed to be explained in detail until it was accepted. It took a lot of effort by the analytics ambassador and management to convince all stakeholders. Stakeholders include pilots and cabin personnel, hub and operations control teams and service providers for ground operations. Initial resistance was due to lack of trust, limited understanding of the new prediction method and fear of change. Today the airline can foresee disturbances earlier and evaluate rescheduling options for crews and aircrafts.

### 3.6 CONCLUSION AND OUTLOOK

In this paper, we show how a major European airline improved its decision-making process through machine learning. In this project, we successfully created a prediction model for aircraft arrival times. Not only challenges during feature generation and model building are considered, but also technical and organizational challenges during implementation. The airline has implemented the model in their daily operations and the new forecast is used by hub and ground operations. Our work serves as guideline for future investments in data and IT infrastructure. With our extensive feature generation and testing process,

we were able to quantify the value of en-route weather data and develop recommendations of additional features to include in the predictive model. Furthermore, our work shows the relationship between increasing prediction accuracy and decreasing flight time. Thus, an optimal point of prediction can be determined. This project was the first analytics project using machine learning. In the coming years the airline plans to implement a variety of machine learning models to enhance forecasts and improve decision-making. The success of our project helped convince staff to establish data-driven decision-making. Arrival time prediction is a critical task for all airlines regardless of which business model they are following. Other airlines can use our approach to establish data-driven decision-making.

Our work is the first to consider the arrival time prediction for intercontinental flights. Previous studies such as Takacs (2015) and Kim (2016) have focused on continental connections with a maximum ground distance of 4,500 km. Data of intercontinental connections including all continents is not freely available and we are the first to publish findings based on such a global dataset. Furthermore, Kim (2016) and Takacs (2015) focus on only one airspace with one set of regulations and traffic control system. The complexity that is added into the arrival time prediction by flying over the ocean as well as having several handovers between different traffic control agencies is not accounted for. Though Kim (2016) includes only data of one airport, the arrival time prediction with the different methods of linear regression are not as accurate as our predictions or the results obtained by Takacs (2015) with the use of a machine learning ensemble model. Considering intercontinental flights and also quantifying the impact of en-route weather data on prediction accuracy is a novel approach.

### 3.6.1 *Future research*

To conclude, we outline several interesting research opportunities based on our work. First, the selection of features can be considered. Either a different subset of features or different features should be generated for prediction. As we have shown in the variable importance analysis, certain features such as airspeed and temperature have a stronger impact on prediction accuracy compared to others. Second, the behavior of pilots and airline operations should be investigated. They might adapt their flight behavior and routes to account for bad weather. If there is a bad weather event on the planned route, an alternative route might be chosen. Furthermore, speed increases can counteract any headwinds or otherwise unfavorable weather conditions.

## OPTIMIZING EXTRA CONTINGENCY FUEL LOAD THROUGH PREDICTIVE ANALYTICS<sup>1</sup>

---

### 4.1 INTRODUCTION

#### 4.1.1 *Background*

Since 1990 the number of flights has grown by 80 % and is expected to grow by another 45 % until 2035. The traffic forecast for Europe expects the number of flights to reach 12.8 million flights per annum. In line with the increase in air travel is the expected increase in greenhouse gas (GHG) emissions. Today the transport sector makes up one quarter of European GHG emissions. Air transport accounts for 13 % (European Transport Commission 2016). The member states of the European Union (EU) coordinate a variety of measures to reduce the environmental impact of aviation. One effort to reduce CO<sub>2</sub> emissions is the EU emission trading scheme. Since 2012 all flights in Europe are required to participate. Through this market-based measure 65 million tonnes of CO<sub>2</sub> emissions were reduced. Furthermore, technological innovations such as aircraft and engine design aim at a reduction of fuel consumption, noise and GHG emissions. Alternative fuels are being researched, but so far production and usage are far behind the set goals. Efficiency gains in air traffic management through the Single European Sky ATM Research Program (SESAR) are expected. SESAR aims at increasing air traffic management performance through automation of systems, pan-European information sharing and standardization of processes. Nevertheless, the SESAR project is behind schedule and first results are only expected in the next years. According to the European aviation environmental report in 2016 experts do not expect to compensate emissions through technological innovations in the near future. Thus, the growth of GHG emissions is expected to reach 44 % until 2035, just like the increase in travel demand (European Transport Commission 2016).

Besides emissions produced during ground operations, the bulk of CO<sub>2</sub> emissions are generated through burning fuel in flight. Reducing fuel consumption has been a key aspect for aircraft manufacturers and airlines not only because of the positive environmental effects, but due to considerable savings in operating costs. As of 2017, fuel is the single largest cost block for airlines with 17 to 25 % of total costs (IATA 2018). Especially for network carriers, fuel costs are considerable. With fuel being the largest single cost block, even a small reduction in fuel consumption would lead to considerable cost savings.

The lifespan of aircraft limits the effect of technical innovations to reduce fuel consumption in the near future. Airlines need to focus on measures, which will

---

<sup>1</sup>The following chapter is based on the work of Achenbach and Spinler (Achenbach and Spinler 2018a). This unpublished manuscript titled "Optimizing extra contingency fuel load through predictive analytics" is currently under revision at Transportation Science.

quickly help to reduce fuel consumption. Aircraft weight, flight speed and distance are the key variables to determine fuel consumption for a flight. Distance cannot be altered, but aircraft weight and flight speed can be optimized. Weight reduction can include reductions of drinking water on board and using lighter seats as the work of the Air Transport Department of Cranfield University (2010) has shown, but fuel itself makes up most of the aircraft's weight. Therefore a reduction in fuel would decrease fuel consumption efficiently. In this work, we focus on the reduction of aircraft take-off weight by reducing the overall fuel load.

For the airline we study, pilots decide on fuel load. They are provided with flight planning information, which includes an estimate for fuel load. However, the prediction accuracy of the flight planning software is limited and pilots tend to distrust the estimates. To account for unforeseen contingencies they load extra fuel on top of the estimated fuel.

We develop a model to improve the prediction accuracy of the fuel estimation model. With a more reliable fuel forecast, pilots will be able to reduce the amount of contingency fuel. Through reducing the aircraft's weight at take-off, less fuel throughout the flight is consumed. This results in substantial savings in fuel costs and CO<sub>2</sub> emissions. The reduction in contingency fuel, however, requires a change in behavior of pilots. Therefore, we also include expert interviews with five pilots to assess their willingness to use our prediction model. Changing fuel loading behavior is easier to implement compared to the outlined alternative measures to reduce fuel consumption. Thus, our prediction model can have an immediate impact on reducing fuel consumption, when implemented and requires little resources.

We implement four different models to predict total fuel consumption. The first model is based on the key factors for fuel consumption such as weight, distance and flight time. The other models include a broad set of information such as weather forecasts for the departure and arrival airport, historic fuel consumption and extended flight information. The data is based on a network carrier and includes short-, medium- and long-haul flights. In total, the fuel consumption of ten different aircraft types is modeled.

Our contribution to the literature is twofold. First, we develop and test four fuel prediction models based on machine learning algorithms to determine the most accurate prediction model. We analyze variable importance and compare the prediction accuracy for ten aircraft types. We analyze short and long-haul flights and outline key differences considering fuel prediction. We show that prediction accuracy can be significantly increased for each aircraft type, for some fleets by up to 48 %. With our prediction model not only increases in accuracy are achieved, but the number of predictions which underestimate the needed fuel amount are decreased by 17 %. The impact of our prediction model is quantified by using cost to carry factors to account for the cost of carrying extra fuel. With estimating cost to carry factors, we are able to translate higher accuracy into cost savings and lower CO<sub>2</sub> emissions. Second, we seek to identify factors that promote adoption of our prediction models. We conduct semi-structured expert interviews with five pilots to understand their willingness to use a new prediction model when making the fuel loading decision. Our results indicate that pilots are willing to use the prediction model to make their

fuel loading decision, but the method and data sources need to be explained thoroughly and an extended testing phase would be needed to build up trust.

This paper is organized in five sections. Section 2 review the relevant literature. Section 3 outlines the feature generation process and the data. In Section 4 the machine learning method is presented, followed by the results in Section 5. In Section 6, the cost to carry factors and CO<sub>2</sub> savings are estimated. Expert Interviews with pilots regarding the acceptance of our prediction model are presented in Section 7. Section 8 concludes with the discussion and outlook. In the following we will outline legal requirements for fuel loading and the industry practice.

#### 4.1.2 *Legal fuel requirements and industry practice*

According to the International Civil Aviation Organization the fuel requirement for operating a jet aircraft for passenger transportation are outlined. This guideline is adopted by the EU Commission (2008). All passenger flights which are operated in an EU member state, Switzerland, Iceland or Norway need to follow the fuel requirements described in EU-OPS 1.255. The new regulation has been effective since 2008 and requires the following fuel to be loaded on an aircraft:

- Trip fuel (according to flight level, aircraft weight and speed)
- Alternate Fuel (flying to an alternate airport after arriving at the destination airport)
- Reserve fuel
  - ◊ 5% contingency fuel based on the trip fuel
  - ◊ Final reserve to fly a 30 minute holding at the destination airport

The airline uses a flight planning tool to calculate the aircraft's fuel usage for the specified flight based on the route and conditions of the operational flight plan. The value that is generated serves as the minimum legal fuel to load for the flight. Extra fuel is added by the pilot. Taxi-in fuel is not considered for the calculation. Under normal conditions the aircraft will land with the reserve and alternate fuel being untouched. This fuel can then be used for taxi-in. There are certain conditions under which the airline is allowed to decrease the reserve or alternate fuel, for example having an alternate airport en-route close to the arrival airport.

For the airline we study, pilots are provided with an overview of under- and overburn fuel rates for the same origin-destination (OD) pair of the last thirty days. Overburn means that the forecasted fuel by the flight planning tool was insufficient. The actual fuel burn was higher than the forecasted fuel burn. Thus, alternate or reserve fuel is used if an insufficient amount of extra fuel was loaded. Alternate and reserve fuel are legally required fuel loads, which are only used in case of emergency for example flying to an alternate

airport. The amount of extra fuel added depends on the pilot's experience and assessment. They estimate the expected additional flight time and add extra fuel based on the standard fuel flow for the aircraft type. It is the airline's goal to carry as little fuel as possible. Nevertheless, the pilot has the final decision on the fuel load and there is no monetary incentive to carry less fuel.

## 4.2 LITERATURE REVIEW

The research field for fuel efficiency is broad. We focus on fuel burn prediction based on machine learning and trajectory models. Studies which compare fuel efficiency of airlines and regions and the impact of weather are also considered. Furthermore, we review the relevant literature for our predictive models.

### 4.2.1 *Fuel burn prediction*

There are several papers focused on predicting the fuel burn during a specific flight phase. Ravizza et al. (2013) analyze fuel consumption during taxi times and propose a model for route optimization at the airport while considering fuel usage for aircraft acceleration. Alligier et al. (2015) indirectly predict fuel burn by focusing on aircraft mass prediction during the climb phase. Patron (2014) aim at optimizing flight trajectories considering winds to reduce fuel burn during the cruise phase. They state that a reduction of 3 % in flight cost can be achieved through optimizing flight trajectories with unfavorable wind conditions. Connected to their work is the approach of Rivas et al. (2016), who develop a methodology to integrate weather conditions in the trajectory planning process. By considering weather uncertainty in the planning phase, contingency fuel loads can be adjusted for the cruise phase. Khardi (2014) focus on the descent phase and suggest an optimal descent trajectory to reduce noise and fuel consumption. Fuel savings of 10% to 20% can be achieved with an optimal flight path during the descent phase. Ryerson et al. (2015) focus on the U.S. airspace and model fuel usage due to excess contingency fuel. According to their data, which is based on a major US network carrier, almost 5 % of fuel consumption is due to carrying unused fuel. They propose to reduce fuel to the statistical contingency fuel level of the 99<sup>th</sup> percentile, which is the extra fuel amount where 99% of the time flights are able to land without using their reserve or alternate fuel. Kang and Hansen (2017) focus on optimizing the method for statistical contingency fuel levels. Instead of estimating the distribution of fuel usage for previous flights on the same route and deriving the 95<sup>th</sup> percentile (SCF 95) of extra fuel, they implement a quantile regression model to estimate the statistical contingency fuel level for the 95<sup>th</sup> percentile. Quantile regression model the relation between certain percentiles of the response variable and predictor variables. This allows to compare how the fuel consumption of the 95<sup>th</sup> percentile of an aircraft is affected by certain predictor variables. Kang and Hansen (2017) state that savings for flights with adverse weather conditions are higher as dispatchers tend to load higher amounts of contingency fuel for such flights. With their method an average

fuel reduction of 110 kg can be achieved. With the SCF 95 method 5.1 % of flights land with reserve fuel being used instead of 3.4 % of flights when using the quantile regression model. Wang et al. (2017) propose a support vector machine to model the fuel consumption of a Boeing 737. Support vector machines create a finite-dimensional space, in which each dimension represents a feature. Through finding an optimal hyperplane the feature space is divided. Support vectors are the data points, which lie closest to the hyperplane and thus have the largest impact on the optimum location of the hyperplane. In their study Wang et al. (2017) use flight manual data to train and validate their model. For testing actual flight data was used, but no details about the region or routes is given. According to their results delivers the SVM-method satisfactory fuel consumption predictions. No information in kilogram or cost savings is given. Horiguchi et al. (2017) combine two problems. They predict fuel consumption and flight delays using the data of a Japanese low cost carrier. They compare their results to the dispatcher's estimates and achieve an improvement of almost 40 %. On average their gradient boosting model performed best and achieved a relative root mean squared error (RMSE) of 8.8 %.

#### 4.2.2 *Fuel efficiency*

Connected to our research is the topic of fuel efficiency. Singh and Sharma (2015) provide a literature review of fuel consumption optimization for the years 1973 until 2014. They consider 277 articles in their review and divide them in four categories: aviation technology, airline operations, socioeconomic factors and alternative fuels. According to their analysis, research in fuel optimization has significantly increased since 2000. In their analysis, Zou et al. (2013) focus on the fuel efficiency of the 15 largest airlines in the U.S. According to their results, airlines operating a hub and spoke system are 20 % less fuel efficient than point-to-point carriers. This is mainly due to newer and more efficient aircraft being operated by low-cost-carrier. The work of Chandra et al. (2014) has a global focus and aims at establishing a benchmark for fuel efficiency that can be referred to by airlines. They show that European airlines are 20 % more fuel efficient than their U.S. competitors. The best performing airlines are two low cost carriers: Ryanair and Southwest. Brueckner and Abreu (2017) provide insights on factors affecting fuel usage. They relate seven variables to an airline's fuel usage. For our work the following insights are important: Longer flights as well as larger aircraft are more fuel efficient. Additionally, aircraft age impacts fuel usage.

#### 4.2.3 *Implementation challenges of advanced analytics*

To evaluate the challenges of implementing advanced analytics we first discuss the literature concerning resources to implement advanced analytics. In a second step we focus on psychological aspects of using information in decision making. Studies such as Davenport (2006) have shown that decisions by business leaders can be improved through the use of analytics. According to Kiron and Ferguson (2012) the change in mindset to a data-driven decision making

is hard to achieve. Several recent studies assess the challenges in implementing advanced analytics in companies. Gupta and George (2016) identify the essential resources to build up a big data analytics capability and state that intangible assets such as data-specific managerial skills and an organizational culture focused on data-driven decision making are key success factors. Côté-Real et al. (2017) analyze the impact of big data analytics on the competitive advantage of European firms. According to their findings big data analytics increases the agility of firms and leads to better operational efficiency. From a psychological perspective there are several cognitive biases that can influence the pilot's decision. Considering the fuel prediction model there are several biases that apply. First, confirmation bias, which describes the tendency to interpret information according to preexisting beliefs Nickerson (1998), implies that the pilots will use the information provided by the fuel prediction model selectively to confirm their personal, initial assessment of the extra fuel load. Second, motivated reasoning, which is used to decrease cognitive dissonance, describes the process of selecting and ignoring information according to one's own belief Kunda (1990). In this case, pilots would tend to estimate extra fuel load based on their previous experiences and not on the information provided by the fuel prediction model. To mitigate the impact of such biases, de-biasing methods can be used. Kunda (1990) show that accuracy goals, implying that people are motivated to be accurate, decrease bias through increasing the persons willingness to process information more carefully. Thus, the pilots' fuel decisions should be evaluated to increase their willingness to be accurate.

#### 4.3 DATA

In this section, we outline the different datasets, which were combined for the prediction model. We collected data from three sources: the first dataset comprises general route and aircraft information for each flight, the second dataset consists of relevant information for fueling and the third dataset includes weather information for arrival and departure airports. A globally operating network carrier provided the flight and fuel information for this research. The Terminal Aerodrome Forecasts (TAF) were collected from the national meteorological service. The data includes continental and intercontinental flights arriving at the hub airport between March 2015 and February 2016. In total there are ten different aircraft types in our dataset including Airbus and Boeing. The dataset contains planned and actual fuel consumption from gate departure until arrival at the destination gate. Additionally, aircraft weight before departure is provided. Flights with fuel loaded for several flight legs, which is called tankering, were excluded.

The dataset can be divided into two subcategories. First, there is the group of short-haul flights including the following aircraft types: A319, A320, A321, B733 and B735. Short-haul flights are defined as aircraft with an average flight distance less than 1,200 km. Short-haul flights make up the bulk of our dataset totaling 146,467 flights. Long-haul flights include the aircraft types A333, A343, A346, B744 and B748. Our dataset includes 20,091 long-haul flights. In Table 4.1

the number of flights, average planned air distance and average planned fuel consumption are shown for each aircraft type.

Table 4.1: Overview of aircraft types, average distance and fuel consumption

Aircraft type	Number of flights*	Ave. Distance (km) *	Ave. Fuel Consumption (kg) *
Short-haul	A319	30,000	840
	A320	50,000	1,000
	A321	50,000	1,200
	B733	7,000	640
	B735	8,000	590
Long-haul	A333	5,000	5,800
	A343	3,000	7,900
	A346	5,000	7,900
	B744	3,200	7,600
	B748	4,000	8,200

\* Rounded planning values

The ranges of average fuel consumption displayed in Table 4.1 are so large due to the variety in destinations. We include all flights of an airline, which arrive at the hub locations. Thus, national flights with a ground distance of 300 kilometers are included as well as intercontinental flights with a ground distance of 11,500 kilometers. Due to data privacy average fuel consumption for each aircraft type is not displayed. After data cleaning the dataset includes 153,728 flights, short and long-haul combined. Flights with missing values as well as wrong values were excluded. We determined wrong values by analyzing outliers for each feature. Our goal is to use the flight and weather information to improve the fuel prediction on an individual flight level.

#### 4.3.1 Dependent variable

The target feature for our prediction is total fuel usage including taxi fuel. Before a flight departs dispatchers and pilots have to decide on the amount of fuel to load. Besides the trip fuel, reserve fuel and alternate fuel is added for additional safety. On top of these legal requirements the pilot is allowed to load additional fuel. In the literature this fuel is termed "Extra contingency fuel". To enhance the pilot's decision capabilities, we predict the total fuel consumed from the block-off moment until reaching the parking position at the destination airport.

#### 4.3.2 Features

We construct several feature sets including flight information, weather data and features based on historic data of similar OD pairs. Features are the input variables for our predictive model. All the features that are used for the fuel prediction need to be available before departure. In total we include 54 features derived from three datasets. In Table 4.2 an overview of the features is shown.

The first part of the data consists of general flight information. We include information on the departure airport such as departure gate and runway, hour of departure and delay at the block-off moment. Furthermore, hour of departure, time of day, weekday, month and season are used. Additionally, we include planned flight times for the flight segments climb, cruise and descent. There are five features focusing on distance. We include the planned ground distance,

Table 4.2: List of features for prediction model

Category	Feature	Type	Number of levels	Example	Unit
Original features	Block time (Total, Climb, En-route, Descent)	Numerical	x	116	min.
	Flight time	Numerical	x	95	
	Block -off until Touchdown time	Numerical	x	110	
	Mission fuel (Total, Climb, En-route, Descent)	Numerical	x	4,500	kg
	Taxi fuel	Numerical	x	400	
	Take-off weight	Numerical	x	40,000	
	Air Distance (Total, Climb, En-route, Descent)	Numerical	x	3,200	km
	Ground Distance	Numerical	x	2,600	
	Maximum altitude	Numerical	x	8,000	
	Departure airport	Categorical	215	HAM	N/A
Arrival airport	Binary	0 / 1	1		
Features generated based on date/ time information	Hour of Arrival	Integer	0 - 23	18	N/A
	Hour of Departure	Integer	0 - 23	10	
	Time of day at Arrival airport	Categorical	6	Evening	
	Weekday	Categorical	7	Monday	
	Month	Categorical	12	March	
Season	Categorical	4	Spring		
Features generated based on other information	Departure airport x Runway	Categorical	x	LIS-13	N/A
	Departure airport x Gate	Categorical	x	LIS- 24	
	Frequency of destination	Integer	1 - 5000 +	30	
	Air distance x Take-off weight	Numerical	x	48,988,085	km x kg
Equal OD pair	Block time x Take-off weight	Numerical	x	4,982,715	min. x kg
	Ave. Fuel previous three flights	Numerical	x	3,275	kg
	Ave. Fuel previous flight	Numerical	x	3,342	
	Ave. Error in Fuel prediction of previous three flights	Numerical	x	231	
	Ave. Block time previous three flights	Numerical	x	84	min.
	Ave. Block time previous flight	Numerical	x	89	
Ave. Error in Block time prediction of previous three flights	Numerical	x	-6		
TAF weather features (Forecasts for departure and arrival airport)	Fog	Binary	0 / 1	1	N/A
	Snow	Binary	0 / 1	0	
	Rain	Binary	0 / 1	1	
	Wind speed	Numerical	x		knots
	Wind gust	Numerical	x	35	
	Wind direction	Integer	1 - 360	52Â°	Degree
	Vertical visibility	Numerical	x	9,999	m
Horizontal visibility	Numerical	x	4,007		

overall air distance and planned air distance for each flight segment. For the arrival airport we include hour of arrival, planned arrival runway and time of day.

National meteorological services publish TAF reports for all major civil airfields. Scheduled departure and landing times are used to match TAF reports for the arrival and departure airports. The report includes numeric features such as cloud base, visibility, wind speed, wind gust and wind direction.

The fuel data includes forecasted fuel usage for the three flight stages climb, cruise and descent. Furthermore, the total planned fuel, take-off weight and taxi out fuel are available features.

#### 4.4 METHODOLOGY

Our goal is to predict aircraft fuel consumption on an individual flight level. Fuel costs are the largest expenditures for airlines ranging from 17% to 25% (IATA 2018). Reducing fleet-wide fuel consumption is one of the most important issues for airlines to reduce operating costs. Furthermore, any decrease in fuel results in considerable reductions of greenhouse gas emissions, which contributes to the International Air Transport Association (IATA) 2020 goal of achieving carbon-neutral growth. In the following we describe the baseline method and the different prediction methods.

##### 4.4.1 *Baseline*

Operations control plans the flight route considering weather, traffic and airline data. With the use of a flight planning software the optimal flight route is determined. Temperature, altitude and pressure are considered in calculating the optimal flight level. Weather conditions influence the flight route. In case of strong winds a different route might be more optimal to either avoid headwinds or use tailwinds. Based on this information, the distance, speed and flight time are calculated. As a result the fuel usage of the aircraft for the specific flight route can be determined. The aircraft's fuel usage also depends on aircraft weight. Distance or flight time multiplied by take-off weight serves as an initial estimate for fuel load. Take-off weight depends on the number of passengers and cargo load. Both values can be determined accurately, but last minute changes in cargo load are common. Nevertheless, average mean error of planned and actual cargo load is on average less than one percent for long-haul flights and slightly more than 1.5 % for short-haul flights.

Based on this a flight plan is generated. The flight plan is used to calculate the legal fuel requirements. The flight dispatcher hands this information over to the pilot. The pilot then makes the final fuel loading decision. This includes adding extra contingency fuel on top of the legal fuel requirements. We use the flight plans fuel recommendation as our baseline and denote it as "LEG-F" (Legal fuel). The pilot estimates the final fuel load shortly before departure. The information available to the pilot is similar to the features in our dataset. Our goal is to improve the information that is available to pilot, when making the final fuel loading decision.

#### 4.4.2 *Fuel consumption prediction*

In this section we outline our prediction method. Fuel prediction may be framed as a regression problem. We implement a prediction model based on three independent features to predict fuel consumption: time, weight and distance. Furthermore, we apply three machine learning models including all described features. These are linear regression, random forest regression and gradient boosting. Each model is explained in detail in the following section. For all models the dependent variable is total fuel load, which describes the actual fuel consumption in kilogram (kg). In all predictions we focus on the reduction of prediction error. The error is measured in root mean squared error (RMSE) and mean absolute error (MAE). All predictions are compared to the baseline provided by the airline. Predictions and evaluation are considered for all aircraft types separately. Due to data privacy and comparability of results prediction error and other information is not reported in kilogram but in percent. Furthermore, we evaluate if the overburn rate can be decreased with our prediction model. Overburn occurs if the forecasted fuel by the flight planning tool was insufficient and more fuel was used during the flight. In this case, alternate or reserve fuel is used, if the pilot did not load sufficient amounts of extra fuel. By decreasing overburn we not only assess the models accuracy, but also the model's increase in safety through decreasing the number of flights with an overburn. The case of an underburn, implying that less fuel is actually used compared to the forecasted fuel, are more favorable in a real life scenario for fuel prediction.

We use linear regression for two prediction models. In the first linear regression model we include planned flight time, planned air distance and planned take-off weight. Additionally, we generate a time and weight feature by multiplying the planned flight time with the planned take-off weight. The same is done with planned distance. With this feature we aim to include the interaction between flight time and distance with the take-off weight. It is possible that the effect of these interaction terms are different on fuel consumption prediction than the variables by themselves. Thus, we intend to model the effect of distance, weight and time on fuel consumption. The model includes five features for prediction. In the following we refer to this model as short linear model (SLM), which relies on the same information as the airlines existing forecast. In the OLS regression the coefficients are selected with the goal of minimizing the residual sum of squares (RSS). In the second linear regression model we include all 54 features, also including the SLM features. We implement both Lasso and ridge regression to minimize the number of features. Both methods assign coefficients to the features. Lasso excludes features by decreasing coefficients to zero, while ridge regression shrinks coefficients to a minimum. Through this only essential features are included. Nevertheless, neither ridge regression nor Lasso deliver superior prediction accuracy than the OLS regression including all 54 features.

Third, random forest regression is implemented to predict fuel usage. Breiman (1984) introduced random forest regression with the CART (classification and regression trees) algorithm in 1984. It eliminates the weaknesses of single decision trees. Overfitting is decreased by growing many trees and averaging pre-

dictions. Additionally, bootstrap sampling, which is the selection of a subset of predictors at each node split, de-correlates the trees. This random sampling ensures that also weak predictors are selected for node splits. The predictor space is divided into high-dimensional regions. For every observation that falls into one region, the same prediction is made. The goal is to minimize the RSS for each region. Similar to random forest regression are gradient boosting machines. They were also developed by Friedman and Breiman (2000). Instead of many individual predictions, like the random forest, an additive model is build. The gradient boosting algorithm performs sequential operations and uses the information of previous stages for later predictions. In case of misclassification weights are assigned and the misclassified instances are granted more importance in the next stage. This process is iterated several times until overall prediction accuracy has improved.

To test the prediction models on an unseen sample, we subdivide our dataset. The test datasets is generated through a random selection. Thus, we can test and verify the prediction accuracy of each model using an unseen sample. The size of the datasets varies depending on the aircraft type. For short distance flights flown with the A320 family we have more than 30,000 flights for training and testing. For the other aircraft models the number of flights ranges between 2,500 and 10,000. For short and long-haul flights we aggregate the specified aircraft types and predict fuel consumption overall for comparison.

We divide our dataset into randomly selected training, validation and test sets. The training data consists of 60 %, validation and test sets of 20 %. Therefore we use the training dataset for model building, validation data for model fine-tuning and the test data is held back for final results evaluation. Gradient boosting and random forest regression require hyperparameter selection. Hyperparameter are additional input values for machine learning algorithms to ensure that the algorithm optimally generalizes the data. They ensure that the error due to bias and variance are decreased to a minimum. According to Bergstra and Bengio (2012) random search is superior to grid search in finding optimal parameter settings. Random search exploits the configuration spaces more effectively using significantly less computation time. For both methods random search was implemented. Detailed parameter tuning results are provided in Appendix B.1.

#### 4.5 RESULTS

To predict fuel usage on an individual flight level, we use the trained models and the test data to generate predictions on unused sample data. We then evaluate prediction accuracy by comparing the flight's actual fuel usage to our prediction models and determine the best model. Overall the prediction based on any of the machine learning models delivered superior results compared to the airline's baseline model.

For each aircraft type we compare prediction accuracy to the baseline values provided by the airline. We do not report the differences in kilogram of fuel. Instead we show the error in MAE and RMSE as percentage based on average fuel. Thus, we can compare prediction accuracy across different aircraft types

as there are significant differences in average fuel amounts. The improvement in prediction accuracy was validated through statistical testing with different data subsets.

For the A320 family we observe the following: All prediction models also including the SLM model based only on time, distance and weight, outperform the baseline model. Linear regression, gradient boosting and random forest regression even further outperform the SLM model. The best performing model considering RMSE is gradient boosting. For MAE it is random forest. The number of flights with overburn is reduced by 20 % compared to the baseline forecast. Most accurate are the predictions for the aircraft A321 with only 3.2 % MAE and 4.6 % RMSE. For aircraft types B733 and B735 we observe the highest inaccuracy in the baseline model for all aircraft types. In both cases RMSE is above 10% and close to 9 % for MAE. The SLM model outperforms the baseline, but the other three models are superior. We observe the largest improvements in RMSE and MAE for B733 and B735. With gradient boosting RMSE and MAE are reduced by more than 40 % for both aircraft types. The error is not due to the aircraft type, but due to the routes. On average, B733 and B735 fly the shortest distances, for which prediction error is highest.

For long-haul flights we consider Airbus aircraft types A333, A343, A346 and Boeing aircraft B744 and B748. The dataset for long-haul flights is small compared to the short haul flights, but our testing showed that the predictions are robust even with training and validation datasets of only 2,000 flights. For long-haul flights linear regression performs best considering RMSE and random forest performs best considering MAE. There is no statistical significant difference between prediction accuracy of those models. Again, all machine learning models outperform the baseline and SLM model. The SLM model performs better than the baseline model in all cases except for the aircraft type B748. Even though the linear regression outperforms the baseline model in all cases, there is a large variation in prediction accuracy. For aircraft type B748 it is as low as 8 % considering MAE and as high as 46 % for aircraft A333. Similar variations are shown in the RMSE reduction. On average the MAE and RMSE are reduced by 31 % and 25 % respectively. The achieved prediction accuracy is best for aircraft types A343 with MAE of 1.0 % and RMSE of 1.4 %. The overburn rate was reduced by 4 %. Please refer to Table 4.3 to see the detailed results for each aircraft type.

Table 4.3: Overview of results

Aircraft type		% Mean absolute error				% Root mean squared error			
		Baseline	Prediction	Best Model	Improvement	Baseline	Prediction	Best Model	Improvement
Short-haul	A319	5.3	3.8	RF	28%	7.1	5.2	GBM	27%
	A320	5.2	3.5	GBM	33%	6.8	4.7	GBM	31%
	A321	4.5	3.2	RF	29%	6.0	4.6	GBM	23%
	B733	9.0	4.7	GBM	48%	10.9	6.3	RF	42%
	B735	8.7	5.0	GBM	43%	10.8	6.6	GBM	39%
	All SH flights	5.1	3.6	GBM	29%	6.6	4.8	GBM	27%
Long-haul	A333	2.8	1.5	LR	46%	3.4	1.9	LR	44%
	A343	1.8	1.0	LR	44%	2.2	1.4	LR	36%
	A346	1.3	1.1	LR	15%	1.8	1.5	LR	17%
	B744	1.8	1.2	LR	33%	2.2	1.6	LR	27%
	B748	1.2	1.1	LR	8%	1.6	1.4	LR	13%
	All LH flights	1.6	1.1	LR	31%	2.0	1.5	LR	25%

Considering prediction performance for short and long-haul flights we observe the following: As shown in Table 4.3 the prediction error of the baseline

model for long-haul flights is considerably lower compared to short-haul flights. The extra fuel loaded to account for contingencies is less compared to the absolute amount of fuel loaded for long-haul flights. The prediction error of the baseline model for short-haul flights is on average three times larger than the prediction error for long-haul flights. The reductions in MAE and RMSE for long-haul flights are subject to larger variations considering the improvement compared to short-haul flights. For example, we achieve an improvement of 8 % MAE for aircraft type B748 and an improvement of 46 % MAE for aircraft type A333. When considering the number of flights averaged over the increase in prediction accuracy the reduction in MAE and RMSE are similar for short and long-haul flights. Nevertheless, fuel savings for long-haul flights are larger in absolute terms. In Figures 4.1 and 4.2 we show the distribution of prediction error for the baseline model and the respective prediction model. The goal is to provide accurate predictions by neither predicting overburn or underburn. Thus, predictions should be centered around zero. Nevertheless, in real life predicting slightly more fuel is more favorable than predicting too little fuel. In this example negative values describe an underburn and positive values overburn. For the short-haul flights the average error is decreased and fewer flights are subject to an overburn. we build several decision trees to analyse the root cause for overburn. The first split for short-haul flights with an overburn depends on the average fuel error of previous flights. The second node split is based on planned take off weight. For long-haul flights the node splits are dependent on the features block time and planned take off weight as well as departure airport.

The average is more accurate and moved towards the left (underburn) area. Furthermore, the two lines show the limits of the 90<sup>th</sup> percentile. We see that the area is narrowed down compared to the baseline model. For long-haul flights the shift to the left and peak is not as striking. Nevertheless, the error reduction is achieved and the limits of the 90<sup>th</sup> percentile are also narrower compared to the baseline forecasts.

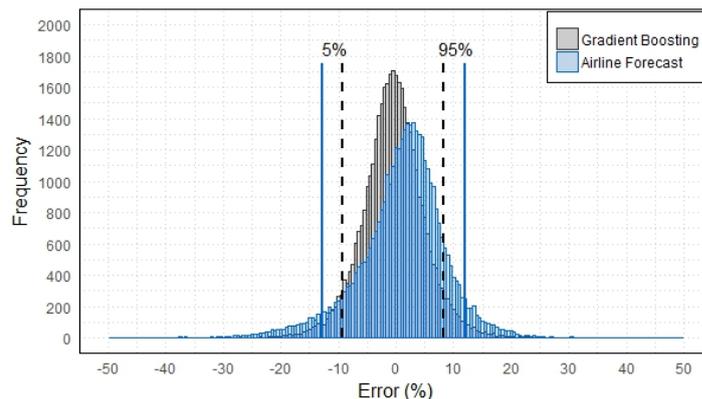


Figure 4.1: Gradient Boosting improves RMSE by 27 % for short-haul flights. Prediction accuracy is increased as shown by the peak in predictions around 0. Additionally, overburn is reduced by 20 %.

When comparing variable importance for the gradient boosting model for all short-haul flights and the linear regression for long-haul flights there are

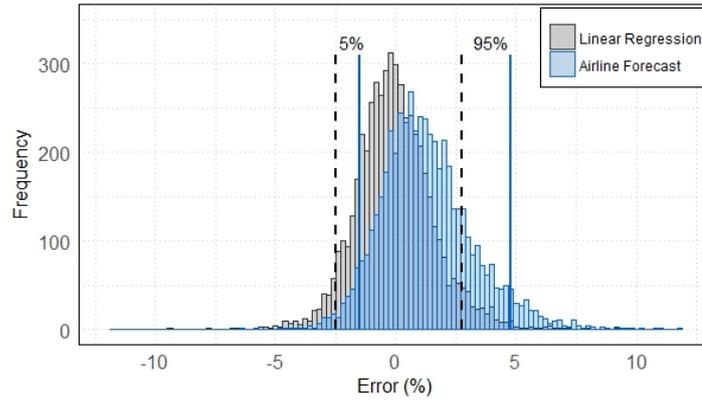


Figure 4.2: Linear regression improves RMSE by 25 % for long-haul flights. Prediction space of 90 % of flights is narrowed down and moved to the left. Thus, overburn rates are reduced as well.

the following findings: Planned fuel value calculated by the flight planning software is the most important feature in both models. Planned take-off weight, ground distance and the historic fuel error are features included in the top fifteen of both models. Except for wind direction at arrival and aircraft type, all features for the gradient boosting model focus on numeric estimates such as distance, fuel and time. The linear regression for long-haul flights includes a variety of non-numeric features such as season, departure airport and weather conditions at arrival. Bad weather conditions at the arrival airport can lead to multiple landing attempts or increased flight time due to holdings. Especially in winter, de-icing has to be performed before take-off leading to longer waiting time on the ground.

Table 4.4: Variable Importance

Linear Regression		Gradient Boosting		
Variable	t-value	Variable	rel. Inf.	
Long-haul	<b>Planned Fuel</b>	87.64	<b>Planned Fuel</b>	62.43
	<b>Ground distance</b>	-13.62	Feature of Time x Weight	22.01
	Departure airport	-8.69	Feature of Distance x Weight	8.18
	Wind gust at arrival	6.16	Planned Enroute Fuel	6.38
	Off-block until touch down time	-5.49	<b>Fuel error (historic)</b>	0.09
	Fuel type	-5.30	<b>Ground distance</b>	0.08
	Flight time	5.13	Ave. Fuel (historic)	0.05
	Planned fuel for descent	-5.00	Block-off delay	0.05
	Rain at arrival	4.81	Aircraft type	0.05
	Hour of Departure	4.69	<b>Planned Take-off weight</b>	0.04
	<b>Planned Take-off weight</b>	4.32	Air distance Descent zone	0.04
	Planned Block time	4.17	Wind direction at arrival	0.04
	Block time error (historic)	3.57	Planned fuel for climb	0.04
	<b>Fuel error (historic)</b>	-3.46	Air distance En-route zone	0.03
	Season	-3.51	Air distance Climb zone	0.03
Short-haul				

The difference in prediction accuracy between aircraft types in the respective category is considerable. For example the two short-haul aircraft A321 and B753 have a difference in RMSE of 2 % with 4.6 % and 6.6 % respectively. Even larger is the difference when considering long-haul flights, which reach a RMSE of 1.4 %. This can be attributed to flight distance. The average distance of A321 aircraft is 1,200 km, which is double the average of B735 flights. Overall,

we observe that the prediction error of the baseline model and our proposed method decreases with increasing flight distance.

When considering the superior prediction model we see that the machine learning models are superior in all examples. The SLM model outperforms the baseline, but cannot compare to the results of the machine learning models. For short-haul flights gradient boosting is the best prediction method. Even though linear regression and random forest are close. For long-haul flights linear regression performs best. Gradient boosting and random forest also deliver accurate predictions, but on average they are 2 % worse than the linear regression model. We expected that the SLM model would perform better for long-haul flights. Our results do not support this view. There are large variances in prediction accuracy for the SLM model ranging from an improvement in RMSE of 35 % (A333) to a decrease in RMSE of -25 % (B748). Please refer to Appendix C.3 for a detailed summary of the SLM model results.

According to Kang and Hansen (2017) there was a significant variation in predicted fuel load between weather and non-weather impacted flights. Weather-impacted flights are described as flights, which have to land under certain conditions such as the presence of a thunderstorm, low ceiling, snow or low visibility. We tested the sensitivity of our prediction model focusing on weather, time and aircraft speed. Based on the insights we gained in the literature and through the expert interviews, these factors have the largest impact on fuel consumption. With this we want to determine, if one of these factors significantly increases the prediction error. Due to differences between long and short-haul flights, we determine the influence of the effects separately. Considering the season and month we observe a slight increase in prediction error for winter due to increased delays and de-icing. For short-haul flights this is mainly due to an increase in prediction error in January. For long-haul flights we see an increase in prediction error for January, February and March. For both, the summer months have a positive impact on prediction error. Time of day and Weekday influence prediction error as well. They indirectly display the traffic volumes at the arrival and departure airports. For short-haul flights prediction error is lowest on Thursday and Fridays and highest on Mondays, when traffic volumes are the highest. Furthermore, arrival early in the morning and at night show an increase in error. For long-haul flights Sundays and Mondays are the worst weekdays. Arrivals in the evening hours show an increase in prediction error. The largest effect on prediction error has aircraft speed. For both short and long-haul flights we see a significant increase in prediction error, when the aircraft speed is increased above regular values (Cost Index larger than 50). Furthermore, weather at the arrival airport affects prediction error considerably. Low visibility, strong wind gust or wind speed lead to an increase in fuel consumption. Weather effects at the departure airport have no effect.

#### 4.6 COST TO CARRY ESTIMATES

To quantify the benefit of our prediction model we estimate the amount of fuel saved compared to the baseline fuel prediction. We assume that pilots will decrease the extra contingency fuel load according to the increase in prediction

accuracy of our model. Therefore, we calculate the fuel saving according to the reduction in RMSE (in kg) for each aircraft type. For example the RMSE for aircraft type A320 was reduced by 31 % from an error of 6.8 % to 4.7 %. The difference in fuel load is 75 kg. Together with the cost to carry factor for this additional fuel we estimate the total amount of possible fuel saving.

The study airline follows a conservative fueling strategy. Thus, there is room for improvement and the fuel load can be optimized without increasing risk. Our model will help to safely adapt fuel load by integrating a variety of information in the prediction. Through this both underburn and overburn will be limited. Our prediction model increases the accuracy considering the underburn prediction. Furthermore, our model decreases the error in fuel prediction with overburn from 63% to 45 % compared to the baseline model. Our prediction models are trained with the conservative fuel loading data. Thus, total fuel burn values of the training data tend to be high due the inclusion of large amounts of extra fuel. The cost to carry this extra fuel further increases the total fuel consumed during the flight. Through the elevated fuel amounts in the training data all models use a dependent variable for training, which tends to be larger than the optimum.

We use the cost to carry factors estimated by Ryerson et al. (2015) and estimates provided by the pilots. In the following the cost to carry factor according to Ryerson et al. (2015) is explained. Cost to carry factors calculate the quantity of fuel consumed to carry one kg of extra weight over one kilometer. Ryerson et al. (2015) estimated the cost to carry factors for a variety of aircraft types and distances. Block fuel (b) can be calculated considering take-off weight (m) and distance (d). They are set into relation with the block fuel consumed for flight (i) with aircraft type (a)

$$b_{i,a} = \beta_{1,a} m_{i,a} + \beta_{2,a} m_{i,a} d_i + \beta_{3,a} d_i. \quad (4.1)$$

From this equation the coefficients are used to estimate cost to carry factors, which are based on weight and weight-distance factors shown in

$$\gamma_{i,a} = \frac{\beta_{1,a}}{d_i} + \beta_{2,a}. \quad (4.2)$$

Ryerson et al. (2015) based cost to carry calculations on the fuel planning tool Piano 5. In our data different aircraft types are included. Therefore, we adapt cost to carry factors of similar aircraft types. Please refer to Table 4.5 to see the matching of aircraft types and cost to carry factors.

Due to the difference in aircraft type we observe large variations in fuel savings and cost to carry estimates. For short-haul flights the average decrease in RMSE is 32 %, which is equivalent to a fuel saving of 82 kg. Additionally, 4 kg cost to carry fuel is added. Based on the number of flights per year the total fuel saving ranges between 700 t to more than 4,000 t depending on the aircraft type. For long-haul flights the variance in RMSE is much larger. It ranges from 13 % up to 44 %. On average the fuel overload accumulates to nearly half a ton with 423 kg extra contingency fuel and cost to carry fuel of an additional 76 kg. Fleet-wide fuel savings range from 800 t to more than 3,000 t. Detailed cost to carry and fuel saving estimates are shown in Table 4.5.

Table 4.5: Cost to carry estimates for each aircraft type

Aircraft type	Variable	Coefficient Estimates	Ave. Fuel Overload (kg)	CTC (kg)	Total Fuel Saving / flight (kg)
Short-haul	A319	Mass Distance $2.854 \times 10^{-5}$ 0.713	57	3	60
	A320	Mass Distance $2.879 \times 10^{-5}$ 0.887	75	4	79
	A321	Mass Distance $2.879 \times 10^{-5}$ 0.887	79	4	83
	B733	Mass Distance $3.395 \times 10^{-5}$ 0.521	98	4	102
	B735	Mass Distance $3.395 \times 10^{-5}$ 0.521	84	3	87
	A333	Mass Distance $2.149 \times 10^{-5}$ 1.280	584	86	670
Long-haul	A343	Mass Distance $2.149 \times 10^{-5}$ 1.280	536	104	640
	A346	Mass Distance $2.149 \times 10^{-5}$ 1.280	239	46	285
	B744	Mass Distance $2.13 \times 10^{-5}$ 2.655	582	110	692
	B748	Mass Distance $2.13 \times 10^{-5}$ 2.655	175	36	211

The pilots that were interviewed used different estimates to calculate cost to carry values. In their logic the extra fuel needs to be available at the start of the descent phase. They estimate the cost to carry extra fuel until the destination airport with one third of the extra fuel amount. Thus, extra fuel multiplied by 1.3 is added to the aircraft. Even though the pilots stated that this is a conservative approach, it is the standard procedure. In Table 4.6 we show the cost to carry estimates based on this approach. Compared to the cost to carry estimates based on the work of Ryerson et al. (2015) we observe that the standard approach by pilots has larger deviations for short-haul flights. The average increase in fuel overload for short-haul flights is close to 20 %. For long-haul the value averages out at 9 %. We estimate cost savings and reductions in CO<sub>2</sub> emissions for both calculations.

Table 4.6: Cost to carry factors according to pilot expertise

Aircraft type	CTC Estimate	Ave. Fuel Overload (kg)	CTC (kg)	Total Fuel Saving / flight (kg)	Ave. total increase	
Short-haul	A319	1.3	57	17	74	20%
	A320	1.3	75	23	98	19%
	A321	1.3	79	24	103	19%
	B733	1.3	98	29	127	20%
	B735	1.3	84	25	109	20%
	A333	1.3	584	175	759	12%
Long-haul	A343	1.3	536	161	697	8%
	A346	1.3	239	72	311	8%
	B744	1.3	582	175	757	9%
	B748	1.3	175	53	228	7%

With our prediction model the airline can achieve considerable cost savings and reductions in CO<sub>2</sub> emissions. We calculate the fuel savings at 0.50 euro per liter based on the average jet fuel price for 2017 (Index Mundi 2016). One liter kerosene is equivalent to 0.82 kilograms kerosene. In total the airline could save 10.5 million € per annum in fuel costs by implementing our prediction model based on the CTC factors of Ryerson et al. (2015). Considering the CTC factors assumed by pilots annual savings in fuel cost amount to 12.3 million €. Furthermore, savings in CO<sub>2</sub> emissions amount to 66,000 tons up to 78,000 tons considering the respective CTC factors by Ryerson et al. (2015) and based

on the pilots' estimates. The airline industry aims at a carbon-neutral growth from 2020 onwards. Thus, a variety of fuel efficiency initiatives are needed. According to IATA (2013b) the baseline for the IATA 2020 goals are calculated based on the average emissions of the individual carrier for the years 2018 to 2020. The airline we study emits 29 million tons of direct CO<sub>2</sub> emissions per year and aims at an annual reduction of 600,000 tons CO<sub>2</sub>. With our improved prediction model, we would contribute between 10% and 12.5% to the airline's annual goal of reaching the IATA 2020 goals.

Table 4.7: Total fuel and CO<sub>2</sub> savings for the airline

	Aircraft type	Number of flights	Total fuel savings (t)	Total fuel savings (€)	CO <sub>2</sub> saving per flight (kg)	Total CO <sub>2</sub> saving (t)
Short-haul	A319	30,000	1,785	892,500	187	5,623
	A320	50,000	3,931	1,965,250	248	12,381
	A321	50,000	4,159	2,079,250	262	13,099
	B733	7,000	715	357,595	322	2,253
	B735	8,000	699	349,720	275	2,203
	Long-haul	A333	5,000	3,351	1,675,250	2,111
	A343	3,000	1,919	959,445	2,015	6,045
	A346	5,000	1,425	712,625	898	4,490
	B744	3,200	2,213	1,106,688	2,179	6,972
	B748	4,000	842	421,000	663	2,652
				10,519,323		66,272

#### 4.7 ACCEPTANCE OF PREDICTION MODEL BY PILOTS

To understand the pilot's willingness to use the prediction model for their fuel loading decision, we conducted semi-structured interviews with five pilots. The group of pilots is diverse including some with more than twenty years experience and others starting their career. Furthermore, the interviewed pilots fly short, medium and long-haul flights to a large variety of destinations. We conduct semi-structured expert interviews based on Cohen and Crabtree (2006). The interviews are transcribed and paraphrased followed by a thematic comparison and conceptualization. The detailed interview outline is included in Appendix C.4. The insights gained are explained in the following. In the interviews, we first focus on the current fuel loading procedure and the importance of the fuel estimate provided by the flight planning software. We ask the pilots to describe their fuel loading decision and the information that they use to estimate the fuel load. Furthermore, we ask them to assess the accuracy of the information that is provided. All pilots stated that the minimum take off fuel, which is provided by the flight planning software, serves as an initial estimate for fuel load. The extra fuel load is determined by considering NOTAMS (Notices to airmen) and weather data. NOTAMS mainly consist of runway closures. Especially the weather at the destination airport is critical for determining extra fuel. Pilots conducting long-haul flights further stated that the zero fuel weight, which is the aircraft's total weight including all content except for fuel, expected flight level and route are assessed to determine fuel load. Furthermore, three pilots stated that traffic at the destination airport is important to evaluate extra fuel due to holdings. The pilots stated that the initial fuel estimate is accurate, if the flight proceeds as described by the flight plan. En-route weather and trip fuel calculation were cited as being very accurate. Weather forecasts at the destination airport and zero fuel weight was considered to be rather inaccurate especially for long-haul flights. Additionally, the aircraft systems that monitor fuel load lack accuracy. Differences of 100 kg between actual and monitored

fuel are common. Overall, unforeseeable conditions at the destination airport lead to an increase in extra fuel. In the next question we ask pilots if they think that a more accurate prediction model is needed. All pilots stated that a more accurate fuel prediction would be useful. More accurate raw data was also named as useful addition. Furthermore, one pilot stated that the impact of a fuel prediction model depends on the OD pair. For some airports fuel usage is very predictable due to constant weather conditions and low traffic volumes. For other airports high traffic volumes lead to considerable uncertainty, which increases extra fuel added by the pilot. In the last part of the first section, we assess how much the pilot relies on the provided data to make the fuel decision. Most pilots stated that they rely on their experience and intuition to decide on extra fuel. The provided information is used to validate the fuel decision. Furthermore, co-pilots and captains employ a combined decision making process. Both separately calculate the total fuel amount based on the provided information and the larger number is used as final fuel load. Furthermore, the fuel usage of previous flights is assessed. For short-haul flights pilots stated that they lack time to look at all the additional information.

Second, we investigate the pilot's acceptance of a prediction based on machine learning algorithms. We intend to assess if the pilot would use our prediction model to enhance the fuel loading decision. Furthermore, the pilot's acceptance of "black-box" model is investigated. First, we ask the pilots if they are familiar with machine learning or analytics. Three out of five pilots are not familiar with the concept. Two pilots have worked on projects related to analytics. They have basic knowledge of the concept and models of machine learning. In a second step we ask the pilots about their trust in a prediction, that is generated through a "black-box" model. A variety of answers was given to this question. All pilots stated that a long testing phase for such a model would be needed. According to their assessment, trust needs to be build up over several month. The accuracy of the prediction model and the explanation of the model to pilots was also named as critical success factor. Furthermore, it was stated that pilots are generally inclined to assess and use information in their decision making. Next, we ask what additional information would be needed to increase the trust in the prediction model. Part of this question was answered in the previous question, where a thorough explanation of the model was named. This question seemed to be hard to answer. Two out of the five pilots did not name anything. One pilot stated that a real-time prediction in flight would be useful. Two pilots said that the model should provide information why extra fuel needs to be added or decreased considering the root cause. For example the likelihood of flying a holding is high and thus 300 kg of extra fuel should added or the weather conditions at the destination airport are favorable implying a decrease in extra fuel of 250 kg.

At the end of the interview the pilots were asked to add any missing information or aspects considering the interview. Several pilots elaborated on the potential of cost savings by decreasing fuel load. According to their assessment savings especially for long-haul flights are substantial. On average flights land with an additional 30 % extra fuel compared to the planned amount of arrival fuel.

We hypothesize that with a more accurate prediction, pilots will load less extra contingency fuel as we account for situations in our model that are not included in the calculation by the flight planning software. These include all time related features, which model the traffic at departure and arrival airports, weather conditions at take-off and landing as well as historic fuel information. Considering the insights we gained in the expert interviews, there are several important factors for the success of such a fuel prediction model. All pilots stated a general interest in a more accurate fuel prediction model. Nevertheless, some pilots said, that it will be hard to convince them to load less fuel than they estimated. The communication and implementation strategy of the fuel prediction model will be critical. The model and data need to be explained thoroughly. This will increase trust in data quality and in the methodology. Furthermore, pilots stated that the model will only be appreciated if it delivers a high prediction accuracy. Considering the increases in prediction accuracy that the machine learning model deliver, we are confident that the increase in accuracy is sufficient. A testing phase of half a year will be needed to slowly build up trust. Ideally, pilots would provide feedback on prediction accuracy through evaluating the fuel prediction and actual fuel consumption after each flight. Considering the impact of motivated reasoning the airline should implement some evaluation measure to increase the pilots' willingness to assess the additional information provided through the fuel prediction model and incorporate it in their fuel decision. Otherwise, the fuel prediction information will be disregarded by the majority by only focusing on their experience.

#### 4.8 CONCLUSION AND OUTLOOK

In this paper we propose a new fuel estimation model for flight planning based on machine learning algorithms. Due to unforeseeable events and safety measures estimated fuel for a flight includes various safety buffers, such as alternate fuel and reserve fuel. On top of that extra contingency fuel is added based on the pilot's assessment. To improve the pilot's information when making the final fueling decision for the aircraft, we use machine learning to predict fuel consumption for a specific flight. We include a variety of features in our model, such as weather at departure and arrival airport and historic fuel values. Our machine learning models substantially outperform the current airline fuel forecast. Especially gradient boosting delivers a superior fuel prediction. With a more accurate prediction pilots will feel comfortable in reducing the amount of extra contingency fuel. This will result in a cost reduction for the airline and substantial reductions in CO<sub>2</sub> emissions.

An extensive dataset of one airline was used to train and validate the different prediction models. We achieve an improvement of prediction accuracy of 30 % on average of all aircraft types. Especially fuel reduction for long-haul flights will result in considerable savings. For certain aircraft types such as A333, A343 and B744 reductions of 700 kg are possible. The features we use in the prediction model are available, when the pilot makes the final fuel loading decision and can therefore be easily implemented.

We calculate the cost of carrying extra contingency fuel based on flight data of a network carrier for the year 2015 and 2016. We develop fuel prediction models for all ten aircraft types. With the use of the CTC factors, we are able to estimate the extra fuel burn due to carrying extra contingency fuel. We estimate the fuel saving achieved with our prediction model compared to actual fuel burn. The extra fuel burn can then be translated to costs. For the whole airline fuel savings between 10.5 million € and 12.3 million € can be achieved through implementing the improved prediction model. This would result in savings of CO<sub>2</sub> emissions of 66,272 tons up to 78,024 tons, which would contribute 10.5 to 12.6 % to the annual goal of reducing GHG emissions by 600,000 tons.

In the expert interviews conducted with pilots we have analyzed the decision making process of estimating extra contingency fuel for a flight. Pilots consider a variety of information to make the fuel decision. With our prediction model we are able to integrate the scattered information provided to the pilots into one model. According to the experts, pilots are inclined to use a variety of information for decision making and consider a more accurate fuel prediction model useful. Furthermore, experts stated that the success of such a model will depend on the pilots' trust in the model. Trust needs to be created through high prediction accuracy and a testing phase. Our prediction model delivers an increase in accuracy of 30 %. Thus, we are confident that it will gain the pilots' trust. Nevertheless, a successful implementation will be challenging. To overcome organizational barriers a change in mindset towards a more data-driven decision making needs to be achieved. Confirmation bias and motivated reasoning can be overcome by setting accuracy goals for example through evaluating the pilots' decision on extra fuel.

There are several interesting research opportunities, which we outline in the following. First, additional features for prediction can be considered to further improve accuracy. Additional information, for example bad weather conditions during the flight, could be included as variables. Second, we have focused our prediction model on one airline. Comparing and integrating our model with an airline of similar size and network structure would deliver valuable insights on the generalizability of our model. Furthermore, en-route traffic data of the Single European Sky Air Traffic Management Research (SESAR) project could be included to enhance prediction accuracy.



## SUMMARY AND OUTLOOK

---

### 5.1 SUMMARY

The airline industry has been facing considerable changes in the past decades. Since 1990 the number of flights has grown by 80 % and competition has been intensifying. The emergence of low cost carrier, volatile oil prices and demand volatility due to the global financial crisis in 2009 have altered the industry. Large consolidation of network carriers have been the result <sup>1</sup>. The International Air Transport Association (IATA) forecasts passenger demand for air travel to almost double until 2035. Most growth is expected on routes to Asia and within Asia. Nevertheless, considerable growth for the European market is also predicted. The traffic forecast for Europe expects the number of flights to increase by 3 million and reach 12.8 million per annum. Today, airlines operating in Europe are already struggling with congested air space, tight schedules and delays. Furthermore, an increasing focus is on reducing the industry's environmental footprint. With a growing number of flights, today's challenges will further intensify. As a result, measures need to be taken to increase the efficiency of air traffic operations, while decreasing greenhouse gas emissions. On an industry level, the European Union initiated the Single European Sky ATM Research Program (SESAR). SESAR aims at increasing air traffic management performance through automation of systems, pan-European information sharing and standardization of processes. Furthermore, research on alternative fuels and technical innovations is supported. On a business level, airlines focus their efforts on strategic and operational measures to minimize delays and GHG emissions. Strategic initiatives are long-term projects such as fleet renewal and infrastructure investments. On an operational level, measures focus on efficiency improvements in airline operations that have an instantaneous, positive impact. Therefore, the application of machine learning and analytics to improve decision support in airline operations needs to be investigated.

This dissertation aims to contribute to improve decision making in airline operations. To this end, we<sup>2</sup> develop machine learning models that are used for prediction of key metrics in airline operations. We first design a model to predict gate arrival time for short-haul flights. This model is further enhanced with a prescriptive analytics solution focused on cost index optimization before departure (Chapter 2). Second, we focus on arrival time prediction for intercontinental flights and assess the impact of en-route weather data on prediction accuracy (Chapter 3). As part of the paper, we also assess the technical and organizational challenges of implementing a predictive analytics solution. Lastly,

---

<sup>1</sup>See, for example, the consolidation of KLM / Air France and Iberia / British Airways (Canelas and Ramos 2014)

<sup>2</sup>The term "we" refers to the authors of the respective chapters as denoted at the beginning of each chapter. For the conclusion, this refers to the authors of Achenbach and Spinler (2018a,b), Achenbach et al. (2017).

we shift the focus to aircraft fuel prediction (Chapter 4). We generate fuel prediction models for ten different aircraft types and analyze the potential of cost savings and reduction in GHG emissions. Furthermore, the pilot's willingness to use our model for the fuel loading decision is analyzed. Findings help the airline to understand data quality issues, the capability of predictive modeling in the context of airline operations and implementation challenges.

In Chapter 2, we explore the application of arrival time prediction and aircraft speed optimization for short-haul continental flights. In a first step, we develop an ensemble model of linear regression and gradient boosting for predicting gate arrival time. Predictions are made at the block-off moment at the departure airport. They are based on more than 80.000 short-haul flights conducted by a European network carrier. Our arrival time forecasts outperforms the airline model by more than 30 %. We further integrate the aircraft speed also known as cost index in our prediction model. The cost index defines the trade off between time and fuel cost. An increase in cost index implies a faster more fuel intensive flight. Although cost index optimization is a very popular tool used by airlines to decrease flight time, its benefit has not been quantified comprehensively. We show that the impact of a high cost index is very limited for short-haul flights. Oftentimes, flying with a high cost index is only optimal in case of longer flights and considerable delay costs. With our model, airlines gain insight on the impact of cost index changes on arrival times. Through this model, they are able to lower cost index levels, while maintaining punctual operations. The arrival time prediction results in lower delay costs and the cost index optimization allows for significant fuel savings.

In Chapter 3, we focus on arrival time prediction for intercontinental flights. Due to the long en-route flight phase we include en-route weather data in our prediction model. First, we assess the accuracy of the forecasted weather data available in the operational flight plan by comparing it to actual weather data collected through the aircraft meteorological data relay system (AMDAR). Based on the results, we can assess the data quality of each feature provided in the operational flight plan and generate en-route weather features accordingly. The prediction accuracy of different feature subsets and machine learning models is tested to determine the best combination. In total, we predict aircraft arrival time at four different times throughout the flight and generate 72 predictions to determine the optimal feature subset and prediction method. With our best model, we are able to outperform the airline forecast by 25 %. The increase in prediction accuracy that is due to including en-route weather data is 5 %. Furthermore, our results show the difference in accuracy for each prediction point throughout the flight. Thus, an optimal point of prediction can be determined. We contribute to existing literature by providing the first study considering arrival time prediction for intercontinental flights. The airline has implemented the model in their daily operations and the new forecast is used by hub and ground operations. Today, the airline can foresee disturbances earlier and evaluate rescheduling options for crews and aircrafts. Our work serves as guideline for future investments in data and IT infrastructure and assess data quality. By considering technical and organizational challenges during implementation, we provide insights for key features of decision-support tools and processes.

In Chapter 4, we shift focus to fuel prediction for ten different aircraft types. We propose a new machine learning model to predict fuel consumption for short and long-haul flights. Our model aims at providing additional decision support to pilots, when making the fuel loading decision. We include a variety of additional features in our fuel prediction model compared to the airline's flight planning software. Our machine learning models substantially outperform the current airline fuel forecast. Depending on aircraft type, prediction accuracy is increased by more than 45 %. With the use of cost to carry factors, we estimate the additional fuel burned to carry extra contingency fuel. Assuming that the pilots will reduce fuel load with our more accurate prediction model, we calculate cost savings and reductions in CO<sub>2</sub> emissions. For the whole airline, fuel savings of 10.5 million € and 66.000 tons of greenhouse gas emissions can be achieved. By analyzing the pilot's willingness to rely on our prediction model, when making the fuel decision, we add a qualitative aspect to the quantitative results. In total, we interviewed five pilots and asked for their trust in using our prediction model. The answers showed that there is a considerable interest in a more accurate fuel prediction. Furthermore, the pilots stated that the acceptance and the use of the prediction model will depend on its accuracy. Our findings are applicable to all airlines, since we include different aircraft types and distances in our prediction model. With the pressure to reduce emissions, more precise prediction models become important. For the airline we study, our model contributes 10 % to the airline's annual goal of reaching the IATA 2020 goals.

There are three main limitations to the findings presented in this dissertation. First, we have focused our work on one airline. With diverse business models resulting in different network structures and operational procedures, prediction accuracy may vary. Considering the sensible information included in the flight data, for example fuel amounts and take-off weight, it was not possible to compare our results with a dataset of a competitor. As other research focused on the U.S. airspace has found the same improvement potential for arrival time and fuel prediction (see Kang and Hansen (2017), Takacs (2015)), we are confident that our focus on one airline can be generalized to other airlines. The second limitation affects only findings presented in Chapter 2, where we include cost index optimization for short-haul flights. The quantification of cost of time for a specific flight requires the integration of various data sources. We lack the data to include a cost per minute for each cost index calculation separately. The last limitation concerns the findings presented in Chapter 3. While the feature generation process and testing of subsets was a complex and highly iterative process, it was not exhaustive. Constructing other subsets or including additional features could further enhance prediction accuracy.

With this dissertation, we contribute to machine learning literature by providing examples of applying machine learning models to operational data. With our assessment of implementation challenges and acceptance of machine learning models, we also contribute to the qualitative aspects of research. All models presented can be modified and calibrated to other industries than airlines as well. We further contribute to airline operations literature and present relevant insights both for airlines and environmental agencies. Operators can benefit from an advanced understanding of accuracy increases by implementing ma-

chine learning models for flight arrival time prediction and fuel prediction. Our detailed assessment of variable importance serves as guideline for future research. Furthermore, our results show how airlines can efficiently decrease greenhouse gas emissions without major investment by either limiting flight speed for short-haul flights or by optimizing fuel prediction.

## 5.2 OUTLOOK

This dissertation focused on the application of machine learning in airline operations. While the models proposed deliver many interesting and valuable findings, such research also leads to new questions and avenues for future research to be addressed.

The models presented in this dissertation were calibrated to a European network carrier by using typical operational data. It would be interesting to test the application of our models on low cost carriers or on network carriers operating on other continents. Furthermore, applying arrival time prediction to other industries such as shipping is an interesting avenue for future research.

Our prediction models can be further extended. In our models, we focus on arrival time prediction for flights arriving at the airline's hub locations. Fuel and arrival time prediction for outgoing flights could be an extension. Furthermore, forecasting arrival times for flight legs and integrating the information in a delay propagation model could enhance operational robustness. Second, apart from the features we include in our prediction models, there is a multitude of other information that can be included. In our modeling, we have focused on flight and weather data of one airline. For short-haul flights, the inclusion of air traffic data and airport ground movements holds a great potential. For long-haul flights actual weather data during flight could be integrated in our model. Overall, developing a model for continuous flight arrival time prediction with the most recent information could be developed and evaluated.

We have discussed in Sections 3 and 4, that arrival time prediction and fuel consumption also depends on the behavior of the pilot. For fuel prediction, we assess the acceptance of a machine learning-based model. A qualitative assessment of the pilot's decision making process and its effect on arrival time and fuel loading decisions could enhance our prediction model.

Lastly, we have focused our analyses on a specific aircraft. Future research could add value by integrating our prediction models into the SESAR project. Air traffic management could be substantially increased, if the accuracy of flight times is increased in a similar magnitude as the results of our models have shown. At the same time, integrating real-time SESAR traffic data into our prediction model could substantially enhance prediction accuracy for flight arrival time and fuel consumption.

In conclusion, the application of machine learning in an operational context holds great potential. There are plenty of aspects remaining for further research, of which the ones presented above are the ones closest related to this dissertation. We therefore believe that there is a promising research agenda at the crossing point of machine learning and airline operations.



## APPENDIX TO CHAPTER 2

### A.1 DETAILED LIST OF FEATURES

Table A.1: Features based on time and date

Time dependent	Description	Structure	Transformation	Source
Hour	Block on/ block off-Date and Time	integer (1-23)	none	derived
Time of day	Block on- Date and Time & clustering of hours for one day	factor (6 levels)	none	derived
Month	Block on- Date and Time	numeric (1-12)	none	derived
Season	Block on-Date and Time & clustering of month	factor (4 levels)	none	derived
Weekday	Block on- Date and Time	numeric (1-7)	none	derived
Days to holiday	minimum days until the next major holiday	numeric	none	derived
Weekend	Weekend yes or no variable (yes=1)	binary (0/1)	none	derived

Table A.2: Weather features

Weather	Description	Structure	Transformation	Source
Cloud base	TAF	numeric	none	existing
Vertical visibility	TAF	numeric	none	existing
Horizontal visibility	TAF	numeric	none	existing
Wind speed	TAF	numeric	none	existing
Wind gust	TAF	numeric	none	existing
Wind direction	TAF	numeric	none	existing
Weather description	TAF	numeric	Factor level conversion	existing

Table A.3: Features based on aircraft and flight information

Flight	Description	Structure	Transformation	Source
Aircraft type	combination of letters and number	factor (3 levels)	none	existing
Fuel type	numbers	factor (3 levels)	none	existing
Air distance	Total flight air distance	numeric (km)	none	existing
Great circle distance	Total flight ground distance	numeric (km)	none	existing
Maximum altitude	Maximum altitude reached during the flight	numeric	none	existing
Total weight	Take-off weight of aircraft	numeric (kg)	none	existing
Actual total trip time	Actual block off time until actual block on time	numeric (min)	none	existing
Scheduled total trip time	Scheduled block off time until scheduled block on time	numeric (min)	none	existing
Taxi-in Time	Touchdown until block on time	numeric (min)	none	existing
Taxi-out Time	Block off until take-off time	numeric (min)	none	existing
ETA Time	Scheduled block off time until scheduled touchdown time	numeric (min)	none	derived
Flight Time	Scheduled take-off time until scheduled touch down time	numeric (min)	none	derived

Table A.4: Departure features

Departure Airport	Description	Structure	Transformation	Source
Departure airport	IATA code	numeric	Factor level conversion	existing
Departure runway	combination of letters and number	numeric	Factor level conversion	existing
Departure stand	combination of letters and number	numeric	Factor level conversion	existing
Country	Name	numeric	Factor level conversion	derived
City	Name	numeric	Factor level conversion	derived
Block off delay	Scheduled block off Time - Actual block off time	numeric (min)	none	derived
Frequency of destination	number of times destination is visited within a year	numeric	none	derived

Table A.5: Arrival features

Arrival Airport	Description	Structure	Transformation	Source
Number of arriving flights	Planned landing time used as the reference to calculate number of arriving flights Steps of 5 min (from 30 min prior to landing until 30 min after landing)	numeric	none	derived
Number of departing flights	Planned landing time used as the reference to calculate number of arriving/ departing flights Steps of 5 min (from 30 min prior to landing until 30 min after landing)	numeric	none	derived
Arrival airport	IATA code	factor (2 levels)	none	existing
Arrival runway	combination of letters and number	factor (12 levels)	none	existing

## A.2 ENSEMBLE MODEL

$S_0$  = new variable subset

Stage 1 - Linear regression

$S_1$  =  $S_0 \cup$  (Actual total trip time, scheduled total trip time, block off delay, ETA time, flight time, hour, time of day, season, days to holiday, arrival runway, air distance, great circle distance, total weight)

No parameter settings

Stage 2 - Gradient boosting machine

$S_2$  =  $S_0 \cup$  ( $S_1$ , Departure stand, departure runway, departure airport, arrival airport, aircraft type, fuel type, maximum altitude, month, weekday, taxi-out time, taxi-in time, frequency of destination, weekend)

Parameter settings: target variable = residual of previous stage, n.trees = 100, shrinkage = 0.1, interaction depth = 40, n.minobsinnode = 100

Stage 3 - Gradient boosting machine

$S_3$  =  $S_0 \cup$  ( $S_1$ , Weather description, cloud base, vertical visibility, horizontal visibility, wind speed, wind gust, wind direction)

Parameter settings: target variable = residual of previous stage, n.trees = 100, shrinkage = 0.15, interaction depth = 40, n.minobsinnode = 100

Stage 4 - Gradient boosting machine

$S_4$  =  $S_0 \cup$  ( $S_1$ , Number of arriving flights, number of departing flights)

Parameter settings: target variable = residual of previous stage, n.trees = 100, shrinkage = 0.15, interaction depth = 40, n.minobsinnode = 100

## A.3 EXAMPLE RESULTS

Table A.6: Example results for 20 flights

Route		Total trip time (min.)		
Departure	Arrival	Scheduled	Actual	Predicted
MUC	HUB	63	58.95	57
OSL	HUB	132	128.79	126
KBP	HUB	154	140.44	137
BIO	HUB	137	126.68	133
BCN	HUB	130	119.65	117
MAN	HUB	98	93.13	97
TLV	HUB	267	256.46	262
TXL	HUB	70	63.97	65
DUS	HUB	69	61.72	64
CPH	HUB	89	84.59	86
GOT	HUB	97	97.85	95
LED	HUB	168	167.33	169
MAD	HUB	143	142.85	132
MXP	HUB	86	77.02	92
ARN	HUB	138	122.09	122
HAM	HUB	66	64.04	72
HAJ	HUB	70	63.36	59
FCO	HUB	98	90.43	87
KBP	HUB	186	176.19	168
MUC	HUB	58	51.90	47
TUN	HUB	142	137.05	134
OPO	HUB	151	145.96	138

## A.4 AIRCRAFT PERFORMANCE DATA FOR OD PAIRS

Table A.7: Flight example: HAM - HUB<sub>1</sub> (285 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	49.66	1.66	0.00%	1950.51	-57.49
10	0.35%	49.48	1.48	0.61%	1962.46	-45.54
20	1.82%	48.75	0.75	1.64%	1982.50	-25.50
30	3.45%	48.00	0.00	2.86%	2008.00	0.00
40	4.73%	47.31	-0.69	4.13%	2031.15	23.15
50	5.49%	46.93	-1.07	5.36%	2055.08	47.08
60	5.80%	46.78	-1.22	6.49%	2077.06	69.06
70	5.86%	46.75	-1.25	7.48%	2096.40	88.40
80	5.85%	46.75	-1.25	8.30%	2112.50	104.50
90	5.88%	46.74	-1.26	8.92%	2124.48	116.48
99	5.84%	46.75	-1.25	9.23%	2130.55	122.55
100	5.82%	46.76	-1.24	9.25%	2130.88	122.88

Table A.8: Flight example: MXP - HUB 2 (294 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	51.21	1.71	0.00%	1950.51	-57.49
10	0.35%	51.03	1.53	0.61%	1962.46	-45.54
20	1.82%	50.27	0.77	1.64%	1982.50	-25.50
30	3.45%	49.50	0.00	2.86%	2008.00	0.00
40	4.73%	48.79	-0.71	4.13%	2031.15	23.15
50	5.49%	48.40	-1.10	5.36%	2055.08	47.08
60	5.80%	48.24	-1.26	6.49%	2077.06	69.06
70	5.86%	48.21	-1.29	7.48%	2096.40	88.40
80	5.85%	48.21	-1.29	8.30%	2112.50	104.50
90	5.88%	48.20	-1.30	8.92%	2124.48	116.48
99	5.84%	48.22	-1.28	9.23%	2130.55	122.55
100	5.82%	48.23	-1.27	9.25%	2130.88	122.88

Table A.9: Flight example: DUS - HUB 2 (307 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	52.55	1.75	0.00%	2056.39	-60.61
10	0.35%	52.37	1.57	0.35%	2068.99	-48.01
20	1.82%	51.59	0.79	1.02%	2090.11	-26.89
30	3.45%	50.80	0.00	1.93%	2117.00	0.00
40	4.73%	50.07	-0.73	2.97%	2141.41	24.41
50	5.49%	49.67	-1.13	4.02%	2166.64	49.64
60	5.80%	49.50	-1.30	5.00%	2189.81	72.81
70	5.86%	49.48	-1.32	5.81%	2210.20	93.20
80	5.85%	49.48	-1.32	6.42%	2227.17	110.17
90	5.88%	49.46	-1.34	6.84%	2239.80	122.80
99	5.84%	49.48	-1.32	7.10%	2246.20	129.20
100	5.82%	49.49	-1.31	7.12%	2246.55	129.55

Table A.10: Flight example: MAN - HUB 1 (509 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	79.50	2.00	0.00%	3123.50	-61.50
10	0.11%	79.41	1.91	0.61%	3134.41	-50.59
20	1.20%	78.55	1.05	1.64%	3155.44	-29.56
30	2.58%	77.50	0.00	2.86%	3185.00	0.00
40	3.86%	76.44	-1.06	4.13%	3216.21	31.21
50	4.85%	75.64	-1.86	5.36%	3249.21	64.21
60	5.56%	75.08	-2.42	6.49%	3279.64	94.64
70	6.02%	74.71	-2.79	7.48%	3305.07	120.07
80	6.33%	74.47	-3.03	8.30%	3324.16	139.16
90	6.47%	74.36	-3.14	8.92%	3337.09	152.09
99	6.35%	74.45	-3.05	9.23%	3345.18	160.18
100	6.31%	74.48	-3.02	9.25%	3345.99	160.99

Table A.11: Flight example: LHR - HUB 2 (528 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	82.07	2.07	0.00%	3070.06	-60.44
10	0.11%	81.98	1.98	0.35%	3080.78	-49.72
20	1.20%	81.08	1.08	1.02%	3101.45	-29.05
30	2.58%	80.00	0.00	1.93%	3130.50	0.00
40	3.86%	78.90	-1.10	2.97%	3161.18	30.68
50	4.85%	78.08	-1.92	4.02%	3193.61	63.11
60	5.56%	77.51	-2.49	5.00%	3223.52	93.02
70	6.02%	77.12	-2.88	5.81%	3248.52	118.02
80	6.33%	76.87	-3.13	6.42%	3267.28	136.78
90	6.47%	76.75	-3.25	6.84%	3279.99	149.49
99	6.35%	76.85	-3.15	7.10%	3287.94	157.44
100	6.31%	76.88	-3.12	7.12%	3288.74	158.24

Table A.12: Flight example: FCO - HUB 1 (609 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	93.35	2.85	0.00%	3788.68	-63.82
10	0.24%	93.13	2.63	0.28%	3799.44	-53.06
20	1.53%	91.92	1.42	0.85%	3820.85	-31.65
30	3.15%	90.50	0.00	1.66%	3852.50	0.00
40	4.63%	89.03	-1.47	2.62%	3888.01	35.51
50	5.72%	88.01	-2.49	3.64%	3926.44	73.94
60	6.34%	87.43	-3.07	4.59%	3962.47	109.97
70	6.55%	87.24	-3.26	5.38%	3992.54	140.04
80	6.51%	87.28	-3.22	5.96%	4014.51	162.01
90	6.42%	87.36	-3.14	6.33%	4028.51	176.01
99	6.47%	87.31	-3.19	6.55%	4036.84	184.34
100	6.50%	87.29	-3.21	6.57%	4037.71	185.21

Table A.13: Flight example: ARN - HUB 2 (798 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	120.42	4.42	0.00%	4762.16	-64.84
10	2.20%	117.77	1.77	0.21%	4772.04	-54.96
20	3.22%	116.54	0.54	0.65%	4793.19	-33.81
30	3.81%	116.00	0.00	1.34%	4827.00	0.00
40	4.35%	115.18	-0.82	2.22%	4867.88	40.88
50	4.95%	114.45	-1.55	3.17%	4913.33	86.33
60	5.58%	113.70	-2.30	4.08%	4956.65	129.65
70	6.09%	113.08	-2.92	4.84%	4992.66	165.66
80	6.39%	112.72	-3.28	5.38%	5018.22	191.22
90	6.48%	112.62	-3.38	5.70%	5033.62	206.62
99	6.55%	112.53	-3.47	5.90%	5042.92	215.92
100	6.57%	112.50	-3.50	5.92%	5044.00	217.00

Table A.14: Flight example: DME - HUB 1 (1222 NM)

CI	Time Diff in %	Time (min.)	Deviation	Fuel Diff in %	Fuel (kg)	Deviation
0	0.00%	177.87	4.37	0.00%	6549.52	-61.98
10	0.67%	176.68	3.18	0.08%	6554.82	-56.68
20	1.63%	174.97	1.47	0.39%	6574.76	-36.74
30	2.52%	173.50	0.00	0.94%	6611.50	0.00
40	3.17%	172.23	-1.27	1.69%	6660.09	48.59
50	3.56%	171.54	-1.96	2.54%	6716.02	104.52
60	3.74%	171.21	-2.29	3.38%	6771.14	159.64
70	3.84%	171.04	-2.46	4.10%	6818.36	206.86
80	3.95%	170.84	-2.66	4.63%	6852.76	241.26
90	4.14%	170.51	-2.99	4.95%	6873.39	261.89
99	4.32%	170.18	-3.32	5.11%	6883.98	272.48
100	4.34%	170.15	-3.35	5.12%	6885.02	273.52

## APPENDIX TO CHAPTER 3

## B.1 EVALUATION CRITERIA

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |c_i - v_i|$$

$$\text{SD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (c_i - v_i)^2}$$

$N$  is the number of predicted values.  $c_i$  and  $v_i$  are the predicted and actual total trip time respectively.

## B.2 DATASET STRUCTURING

For random forest regression we observed improvement in setting the number of trees value (`ntree`) to 1,000 compared to the default. All other parameters are kept at the default values. For the gradient boosting machine we use a grid search to determine parameter settings. The interaction depth and number of observations in the final node was kept constant for all predictions and was set to 30 and 10 respectively. Values for `ntree` ranged from 100 to 700. The shrinkage parameter was set to a value between 0.1 and 0.2.

Table B.1: Hyperparameter settings Gradient Boosting

Prediction point	Excl. En-route data		Incl. En-route data	
	<code>ntree</code>	<code>shrinkage</code>	<code>ntree</code>	<code>shrinkage</code>
Block off	700	0.15	500	0.1
Take off	700	0.15	700	0.1
End of Climb	700	0.15	700	0.15
Start of Descent	100	0.2	700	0.15



## APPENDIX TO CHAPTER 4

## C.1 LINEAR REGRESSION

Residual standard error and adjusted R-squared are provided for each aircraft type in the following table.

Table C.1: Overview of linear regression results

Linear Regression		
Aircraft type	Residual standard error	Adj. R-squared
A319	165	0.9871
A320	185	0.9883
A321	223	0.9924
B733	171	0.9532
B735	155	0.9235
SH	190	0.9915
A333	755	0.9915
A343	868	0.9975
A346	1073	0.9976
B744	1307	0.9953
B748	1305	0.9971
LH	1080	0.9984

## C.2 HYPERPARAMETER SETTINGS

Random Search for gradient boosting was performed with the following configuration space:

- Number of trees: 0 to 2000 with an increment of 100
- Minimum number of observations in final node: 1, 5, 10, 20, 50, or 100
- Learning rate: 0.08 to 0.2 with an increment of 0.001
- Interaction depth: 1 to 20

Random Search for random forest was performed with the following configuration space:

- Number of trees: 50 to 2000 with an increment of 100 until 1000 then 500
- Interaction depth: 1 to 35

Table C.2: Hyperparameter settings for each aircraft type

Aircraft type	Gradient Boosting				Random Forest	
	Learning rate	Num. of trees	Interaction	Min. obs. in final node	Num. of trees	Num. of variables each split
A319	0.103	1000	19	10	1000	19
A320	0.157	1000	20	5	500	17
A321	0.103	1000	19	10	500	16
B733	0.199	1000	17	5	1000	20
B735	0.199	1000	17	5	1000	20
Short-haul	0.103	1000	19	10	1000	20
A333	0.199	1000	17	5	1000	18
A343	0.157	1000	17	5	2000	20
A346	0.157	1000	20	5	2000	20
B744	0.199	1000	17	5	1000	20
B748	0.199	1000	17	5	500	17
Long-haul	0.199	1000	19	5	500	17

### C.3 RESULTS OF SLM MODEL

Table C.3: Overview of results - SLM model

Aircraft type	% Mean absolute error			% Root mean squared error			
	Baseline	Prediction	Improvement	Baseline	Prediction	Improvement	
Short-haul	A319	5.3	4.8	9%	7.1	6.4	10%
	A320	5.2	4.5	13%	6.8	6.0	12%
	A321	4.5	4.0	11%	6.0	5.8	3%
	B733	9.0	5.7	37%	10.9	8.0	27%
	B735	8.7	5.9	32%	10.8	8.0	26%
	All SH flights	5.1	4.5	12%	6.6	6.1	8%
Long-haul	A333	2.8	1.7	39%	3.4	2.2	35%
	A343	1.8	1.2	33%	2.2	1.6	27%
	A346	1.3	1.3	0%	1.8	1.7	6%
	B744	1.8	1.4	22%	2.2	1.8	18%
	B748	1.2	1.6	-33%	1.6	2.0	-25%
	All LH flights	1.6	3.8	-138%	2.0	5.0	-150%

### C.4 INTERVIEW OUTLINE

Questions:

- 1) What is the essential information for you to determine fuel load ?
- 2) How do you rate the prediction accuracy of the fuel estimate provided by the flight planning software?
- 3) Do you think a more accurate prediction model is needed ?
- 4) How much do you rely on the provided data when making your fuel loading decision?
- 5) Do you know what advanced analytics and machine learning is ?
- 6) Would you trust the new prediction for your fuel loading decision ? ("Black-box" model)
- 7) What additional information would be needed to increase your trust in the prediction method ?

## BIBLIOGRAPHY

---

- Abdelghany, K., S. Shah, S., Raina, S., 2004. A model for projecting flight delays during irregular operation conditions. *Journal of Air Transport Management* 10 (6), 385–394.
- Abdelghany, K. F., Abdelghany, A. F., Ekollu, G., 2008. An integrated decision support tool for airlines schedule recovery during irregular operations. *European Journal of Operational Research* 185 (2), 825–848.
- Achenbach, A., Spinler, S., 2018a. Optimizing extra contingency fuel load through predictive analytics, unpublished working paper, WHU-Otto Beisheim School of Management.
- Achenbach, A., Spinler, S., 2018b. Prescriptive Analytics in Airline Operations: Arrival time prediction and cost index optimization for short-haul flights. *Operations Research Perspectives* 2018 - 5, 265–279.
- Achenbach, A., Spinler, S., Schmal, S., 2017. Prescriptive Analytics in Airline Operations: Intercontinental arrival time prediction, unpublished working paper, WHU-Otto Beisheim School of Management.
- Air Transport Department of Cranfield University, 2010. Fuel and air transport - A report for the European Commission. Tech. rep.
- Airbus Market Forecast, 2017. Growing Horizons 2017/2036. Tech. Rep. 4, Airbus Market Forecast.  
URL <http://www.airbus.com/content/dam/corporate-topics/publications/backgrounders/Airbus{ }Global{ }Market{ }Forecast{ }2017-2036{ }Growing{ }Horizons{ }full{ }book.pdf>
- Aktürk, M. S., Atamtürk, A., Gürel, S., 2014. Aircraft Rescheduling with Cruise Speed Control. *Operations Research* 62 (4), 829–845.
- Alligier, R., Gianazza, D., Durand, N., 2015. Machine Learning and Mass Estimation Methods for Ground-Based Aircraft Climb Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 1–12.
- Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., Zou, B., 2010. Total delay impact study: A comprehensive assessment of the costs and impacts of flight delay in the united states. Tech. rep.
- Barnhart, C., Smith, B., 2012. *Quantitative Problem Solving Methods in the Airline Industry*, 1st Edition. Springer US.
- Beatty, R., Hsu, R., Berry, L., Rome, J., 1998. Preliminary evaluation of flight delay propagation through and airline schedule. 2nd USA/Europe Air Traffic Management R&D Seminar (December), 1–9.
- Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research* 13, 281–305.
- Breiman, L., 1984. *Classification and regression trees*. Chapman & Hall/CRC.

- Brueckner, J. K., Abreu, C., 2017. Airline fuel usage and carbon emissions: Determining factors. *Journal of Air Transport Management* 62, 10–17.  
URL <http://dx.doi.org/10.1016/j.jairtraman.2017.01.004>
- Callaham, M. B., Sc, D., Dearmon, J. S., Cooper, M., Goodfriend, J. H., Ph, D., 2001. Assessing NAS Performance : Normalizing for the Effects of Weather. In: 4th USA/Europe Air Traffic Management R&D Symposium. pp. 3–7.
- Campanelli, B., Fleurquin, P., Arranz, A., Etxebarria, I., Ciruelos, C., Egu luz, V. M., Ramasco, J. J., 2016. Comparing the modeling of delay propagation in the us and european air traffic networks. *Journal of Air Transport Management* 56, 12–18.
- Canelas, B. H., Ramos, P., 2014. Consolidation in Europe’s Airline Industry. Tech. rep.
- Chandra, S., Chitgopeker, C. K., Crawford, B., Dwyer, J., Gao, Y., 2014. Establishing a Benchmark of Fuel Efficiency for Commercial Airline Operations. *Journal of Aviation Technology and Engineering* 1, 32–39.
- Chatterji, G. B., Sridhar, B., 2005. National Airspace System Delay Estimation Using Weather Weighted Traffic Counts. In: AIAA Guidance, Navigation, and Control conference and Exhibit. No. March 2004. pp. 1–17.
- Cohen, D., Crabtree, B., 2006. Qualitative Research Guidelines Project. Tech. rep.  
URL <http://www.qualres.org/HomeSemi-3629.html>
- Cook, A., 2004. Evaluating the true cost to airlines of one minute of airborne or ground delay; Final report. Review Literature And Arts Of The Americas.
- Cook, A., Tanner, G., Williams, V., Meise, G., 2007. Dynamic Cost Indexing. 6th EUROCONTROL Innovative Research Workshops & Exhibition (December), 4–6.
- Cook, A. J., Tanner, G., 2011. European airline delay cost reference values. Tech. Rep. March.
- C rte-Real, N., Oliveira, T., Ruivo, P., 2017. Assessing business value of Big Data Analytics in European firms. *Journal of Business Research* 70, 379–390.  
URL <http://dx.doi.org/10.1016/j.jbusres.2016.08.011>
- Customer Services AIRBUS, 1998. Cost Index. Tech. Rep. II.
- Davenport, T. H., 2006. Competing on analytics. *Harvard Business Review* 84 (1), 98–107.  
URL <http://libproxy.trinity.edu:80/login?url=http://search.ebscohost.com/login.aspx?direct=true{&}db=a9h{&}AN=31316546{&}site=ehost-live>
- Edwards, H. A., Dixon-Hardy, D., Wadud, Z., feb 2016. Aircraft cost index and the future of carbon emissions from air travel. *Applied Energy* 164, 553–562.
- EU Comission, 2008. Amtsblatt der Europ ischen Union VO Nr. 859/2008.
- European Transport Comission, 2016. European Aviation Environmental Report 2016. Tech. rep.

- Eurocontrol, FAA, 2010. 2010 U.S./Europe Comparison of ATM-Related Operational Performance. Tech. rep.
- Friedman, J. H., Breiman, L., 2000. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 29 (5), 1189–1232.
- Glina, Y., Jordan, R., Ishutkina, M., 2012. A Tree-Based Ensemble Method for the Prediction and Uncertainty Quantification of Aircraft Landing Times. In: American Meteorological Society - 10th Conference on Artificial Intelligence Applications to Environmental Science.
- Gupta, M., George, J. F., 2016. Toward the development of a big data analytics capability. *Information and Management* 53 (8), 1049–1064.  
URL <http://dx.doi.org/10.1016/j.im.2016.07.004>
- Horiguchi, Y., Baba, Y., Kashima, H., 2017. Predicting Fuel Consumption and Flight Delays for Low-Cost Airlines. *Proceedings of the 31th Conference on Artificial Intelligence (AAAI 2017)*, 4686–4693.
- IATA, 2013a. IATA Technology Roadmap. Tech. Rep. June.  
URL <https://www.iata.org/whatwedo/environment/Documents/technology-roadmap-2013.pdf>
- IATA, 2013b. Resolution on the implementation of the aviation "CNG2020" strategy. Tech. rep.
- IATA, 2018. Fact Sheet – Fuel. Tech. Rep. December 2017.  
URL <https://www.iata.org/pressroom/factsfigures/factsheets/Documents/fact-sheet-fuel.pdf>
- ICAO, 1995-2016. Annual Reports of the Council - ICAO.  
URL <https://www.icao.int/publications/Pages/annual-reports.aspx>
- Index Mundi, 2016. Jet Fuel Monthly Price - IndexMundi.  
URL <http://www.indexmundi.com/commodities/?commodity=jet-fuel&months=60&currency=eur>
- International Air Transport Association, 2017. Fact Sheet: Alternative Fuels. Tech. Rep. December.
- Kaggle, 2012. Winners - GE Flight Quest.  
URL <http://www.gequest.com/c/flight/details/winners>
- Kang, L., Hansen, M., 2017. Quantile Regression Based Estimation of Statistical Contingency Fuel. Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017).
- Kern, C. S., de Medeiros, I. P., Yoneyama, T., April 2015. Data-driven aircraft estimated time of arrival prediction. In: *Systems Conference (SysCon), 2015 9th Annual IEEE International*. pp. 727–733.
- Khaldi, S., 2014. Environmental impact reduction of commercial aircraft around airports . Less noise and less fuel consumption. *European Transport Research Review* 6, 71–84.
- Kim, M. S., 2016. Analysis of short-term forecasting for flight arrival time. *Journal of Air Transport Management* 52, 35–41.  
URL <http://dx.doi.org/10.1016/j.jairtraman.2015.12.002>
- Kiron, D., Ferguson, R. B., 2012. Innovating With Analytics. *MIT Sloan Management Review* 54 (1), 1–8.

- Kletzel, J., 2017. 2017 Commercial Aviation Trends - Digitize and reassess your competitive position. Tech. rep.
- Kletzel, J., DeWaart, D., 2014. Tailwinds - 2014 airline industry trends. Tech. rep.
- Kunda, Z., 1990. The case of motivated reasoning. *Psychological Bulletin* 108 (3), 480–498.
- Lufthansa Systems, 2013. Rely on our efficient flight-planning solution and dispatch support. Tech. rep.  
URL <https://www.lhsystems.com/solutions-services/airline-solutions-services/flight-deck-solutions/lidoflightplanning/lidoflight.html>
- Mueller, E. R., Chatterji, G. B., Field, M., 2002. Analysis of aircraft arrival and departure delay. *AIAA's Aircraft Technology, Integration, and Operations (ATIO) 5866* (October), 1–14.
- Murphy, J. R., Reisman, R. J., Clayton, J., Wright, R., Volpe, J. A., 2003. Physics-Based and Parametric Trajectory Prediction Performance Comparison for Traffic Flow Management. *AIAA Guidance, Navigation and Control Conference and Exhibit* (August), 1–11.
- Nickerson, R. S., 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2 (2), 175–220.  
URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/1089-2680.2.2.175>
- Oxley, G., 2016. Air Passenger Forecasts Global Report UPDATED. Tech. rep.
- Patron, R. S. F., 2014. Horizontal Flight Trajectories Optimization for Commercial Aircraft through a Flight Management System. *Aeronautical Journal* (January 2015), 1–36.
- Peterson, M. D., Bertsimas, D. J., Odoni, A. R., Rooml, S., 1995. Models and Algorithms Congestion for at Transient Airports Queueing. *Management Science* 41 (8), 1279–1295.
- Ravizza, S., Chen, J., Atkin, J. a. D., Burke, E. K., Stewart, P., 2013. The trade-off between taxi time and fuel consumption in airport ground movement. *Public Transport* 5 (1-2), 25–40.
- Rebollo, J., Balakrishnan, H., 2012. A Network-Based Model for Predicting Air Traffic Delays. In: *International Conference on Research in Air Transportation(ICRAT 2012)*. No. ICRAT.  
URL <http://web.mit.edu/hamsa/www/pubs/RebolloBalakrishnanICRAT2012.pdf>
- Rebollo, J. J., Balakrishnan, H., 2014. Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies* 44, 231–241.
- Rivas, D., Vazquez, R., Franco, A., 2016. Probabilistic Analysis of Aircraft Fuel Consumption Using Ensemble Weather Forecasts. *7th International Conference on Research in Air Transportation (ICRAT)*, 1–8.
- Roberson, B., 2007. Cost Index Explained. Tech. Rep. 2.
- Ryerson, M. S., Hansen, M., Hao, L., Seelhorst, M., 2015. Landing on empty : estimating the benefits from reducing fuel uplift in US Civil Aviation. *Environ* 10, 1–11.

- Singh, V., Sharma, S. K., 2015. Fuel consumption optimization in air transport : a review , classification , critique , simple meta-analysis , and future research implications. *European Transport Research Review* 7 (12), 1–24. URL <http://dx.doi.org/10.1007/s12544-015-0160-x>
- Smith, J., Kletzel, J., 2015. Getting clear of the clouds - Will the upward trajectory continue? Tech. rep.
- Srivastava, A., 2011. Improving departure taxi time predictions using ASDE-X surveillance data. In: 2011 IEEE/AIAA 30th Digital Avionics Systems Conference. pp. 2B5-1-2B5-14.
- Takacs, G., 2015. Predicting flight arrival times with a multistage model. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014* 2, 78–84.
- Tielrooij, M., Borst, C., Van Paassen, M. M., Mulder, M., 2015. Predicting Arrival Time Uncertainty from Actual Flight Information. *Europe Air Traffic Management Research and Development Seminar* (November).
- Tobaruela, G. & Fransen, P. . S. W. . Y. O. W. . M. A., 2014. Air traffic predictability framework - Development, performance evaluation and application. *Journal of Air Transport Management* 39 (6), 48–58.
- Tu, Y., Ball, M. O., Jank, W. S., 2008. Estimating Flight Departure Delay Distributions-A Statistical Approach With Long-Term Trend and Short-Term Pattern. *Journal of the American Statistical Association* 103 (481), 112–125.
- UK Civil Aviation Authority, jul 2006. Flight planning manual. Tech. rep.
- Walker, C., 2016. CODA DIGEST 2015. Tech. rep., Eurocontrol.
- Wang, X., Chen, X., Bi, Z., 2017. Support vector machine and ROC curves for modeling of aircraft fuel consumption. *Journal of Management Analytics* 2 (1), 22–34. URL <http://dx.doi.org/10.1080/23270012.2015.1019582>
- Willemain, T. R., 2001. Estimating Components of Variation in Flight Times. *NEXTOR WP-01-2*, 1–48.
- Willmott, C. J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30 (1), 79–82.
- Worldbank, 2017. Air transport - passengers carried. URL <https://tcddata360.worldbank.org/indicators/IS.AIR.PSGR?country=BRA{&}indicator=1733{&}viz=line{&}chart{&}years=1970,2016>
- Xu, N., Sherry, L., Laskey, K. B., 2008. Multi-factor model for predicting delays at U.S. airports. *Transportation Research Record: Journal of the Transportation Research Board* 2052, 62–71.
- Zonglei, L., Jiandong, W., Guansheng, Z., dec 2008. A New Method to Alarm Large Scale of Flights Delay Based on Machine Learning. In: 2008 International Symposium on Knowledge Acquisition and Modeling. pp. 589–592.

- Zou, B., Elke, M., Hansen, M., 2013. Evaluating Air Carrier Fuel Efficiency in the U . S . Airline Industry. Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013).

## DECLARATION

---

**Last Name:** Achenbach

**First Name:** Anna

**Affirmation – Statutory Declaration**  
**According to § 12 part 1 no. 6 of the Doctoral Program Regulations**  
**(dated 20th May 2015)**

I hereby declare, that the

**Dissertation**

submitted to Wissenschaftliche Hochschule für Unternehmensführung (WHU) - Otto-Beisheim-Hochschule - was produced independently and without the aid of sources other than those which have been indicated. All ideas and thoughts coming both directly and indirectly from outside sources have been noted as such. This work has previously not been presented in any similar form to any other board of examiners.

Sentences or text phrases, taken out of other sources either literally or as regards contents, have been marked accordingly. Without notion of its origin, including sources which are available via internet, those phrases or sentences are to be considered as plagiarisms. It is the WHU's right to check submitted dissertations with the aid of software that is able to identify plagiarisms in order to make sure that those dissertations have been rightfully composed. I agree to that kind of checking, and I will upload an electronic version of my dissertation on the according website to enable the automatic identification of plagiarisms.

The following persons helped me gratuitous / non-gratuitous in the indicated way in selecting and evaluating the used materials: Further persons have not

---

<b>Last name</b>	<b>First name</b>	<b>Kind of Support</b>	<b>gratuitous / non-gratuitous</b>
Krebs	Rachel	Proofreading (Chapters 3 & 4)	non-gratuitous

---

been involved in the preparation of the presented dissertation as regards contents or in substance. In particular, I have not drawn on the non-gratuitous help of placement or advisory services (doctoral counsels / PhD advisors or other persons). Nobody has received direct or indirect monetary benefits for services that are in connection with the contents of the presented dissertation.

The dissertation does not contain texts or (parts of) chapters that are subject of current or completed dissertation projects.

*Vallendar, April 30, 2018*

---

Anna Achenbach