Operations Research Perspectives 5 (2018) 265–279

Contents lists available at ScienceDirect

Operations Research Perspectives

journal homepage: www.elsevier.com/locate/orp

Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights

Anna Achenbach*, Stefan Spinler

WHU - Otto Beisheim School of Management, Kuehne Foundation Endowed Chair in Logistics Management, Burgplatz 2, Vallendar 56179, Germany

ARTICLE INFO

Keywords:
OR in airlines
Analytics
Aircraft arrival time prediction
Prescriptive analytics
Cost index optimization

ABSTRACT

In this paper, we provide arrival time prediction combined with a cost index optimization model for short haul flights. Our work is based on flight data of a European network carrier. We focus on predicting the arrival time for incoming flights at two hub locations. Airlines focus on two aspects in their operations: Minimizing cost while ensuring on-time arrivals. Especially network carriers with hub connections need to ensure that incoming flights are on time for passenger, crew and aircraft transfer. 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. Today there is no model for arrival time prediction and integrated cost index optimization. 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 heavily depends on a flight's distance, fuel costs and delay costs. Especially for short haul flights we recommend lowering the cost index as a high cost index has limited impact on gate arrival time.

1. Introduction

Considering that reactionary delays made up more than 40% of all delays in 2015 ensuring on time arrivals is a critical operational problem for airlines [33]. 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 [27] 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 [21,30]. 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 [4]. 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 [1,26]. 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 [20,21,30]. 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 [30]. 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 [17]. 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 in arrival time considering different cost indices. Work on cost index optimization is rather limited. Cook et al. [10] provide a generic tool for dynamic cost indexing (DCI). DCI is defined as managing flight delay costs dynamicly through trading fuel burn against cost of time. Airline manufacturer maintain cost index guidelines for their fleets [12,27]. 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

* Corresponding author.
E-mail addresses: anna.achenbach@whu.edu (A. Achenbach), stefan.spinler@whu.edu (S. Spinler).

https://doi.org/10.1016/j.orp.2018.08.004
Received 8 March 2018; Received in revised form 13 July 2018; Accepted 21 August 2018
Available online 25 August 2018

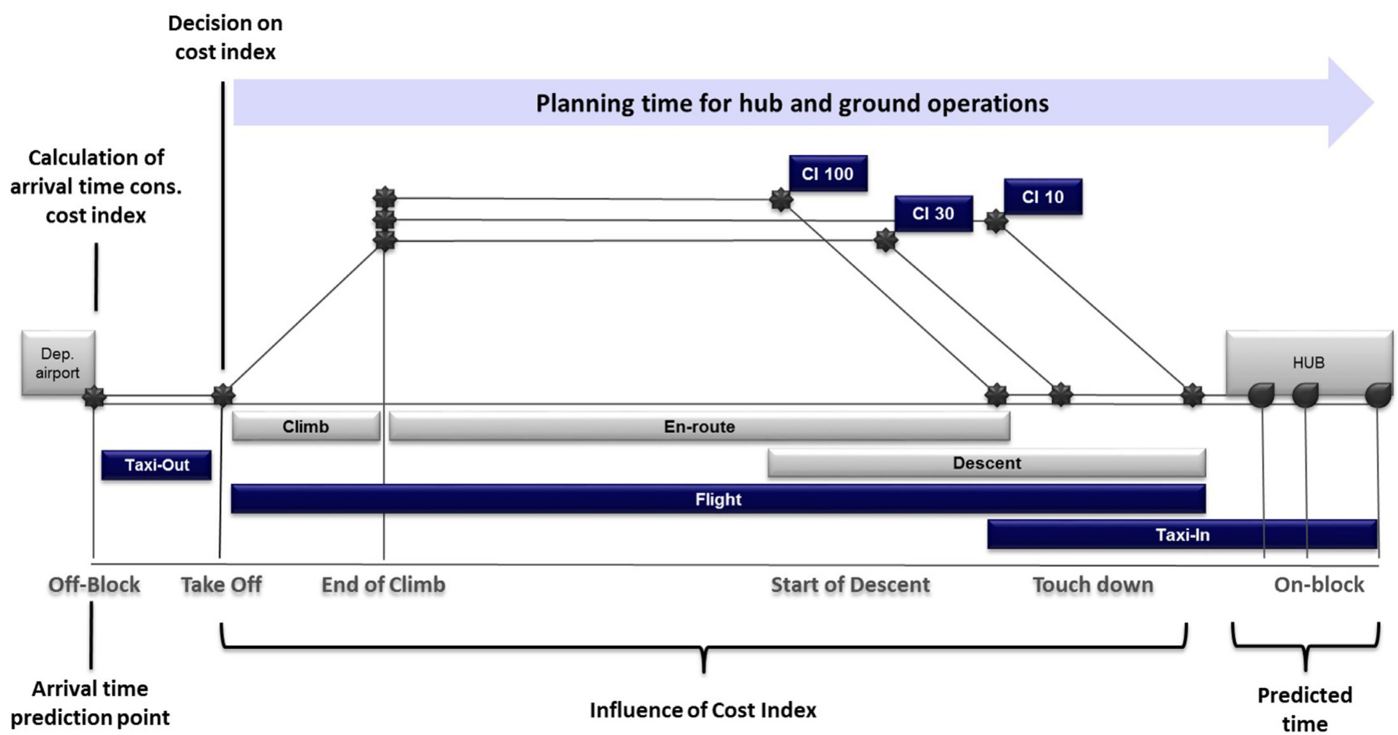

Fig. 1. Overview of cost index and arrival time prediction model. We adopt five flight phases for our model: Taxi out, Climb, En-route, Descent and Taxi in. The en-route flight phase can vary according to distance. Very short flights start the descent right after the climb phase. On average the en-route phase lasts for 31 min and an average flight for 100 min. The cost index only influences the flight time.

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 Fig. 1 we show an example of a flight and the impact of the cost index on the flight's arrival time.

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 block-off moment. We implement linear regression and gradient boosting machines as described by Friedman and Breiman [16]. 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 min 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 [2,17,20], but so far no study using European flight data has been published. According to [15] 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 [8] 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 fights needs to be investigated as there are major differences to the US system.

This paper is organized as follows. Section 2 reviews the relevant literature. The prediction model is suggested in Section 3, wherein the data, variables and the current baseline model are described. In Section 4, the cost index model is explained. Section 5 reports the results of each model. Section 6 provides a brief conclusion and describes implications for theory and practice in addition to limitations and directions for future research.

2. Literature review

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

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 [10,12,27]. The first stream is the literature on airline cost indexing. As early as 1991, Honeywell Inc. has developed patented technology to integrate a cost index parameter into arrival time forecasting [13]. Both Boeing and Airbus have published cost index strategies for their fleets. Furthermore, In [12] 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 [27]. Cook et al. [10] 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 [10] have considered the environmental impact of optimized CI levels to reduce CO_2 emissions. According to [10], 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 min 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. [14]. 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. [2] 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. Arrival time prediction

The second stream of literature concerns arrival time estimation. According to [23], 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. [17]. 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 [22,28]. 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. [5,24]. In recent years, the use of predictive analytics in delay management has entered the focus of researchers. Abdelghany et al. [1] 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 [26] concentrating on network delay states to forecast departure delays 2–24 h in advance. Further work focusing on delay prediction was published by Zonglei et al. [35] using machine learning to train an unsupervised model on flight delay detection. Xu et al. [34] focus on delay prediction for 34 major U.S. airports. They predict positive or negative delay. Tu et al. [31] 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. [17] 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.[29] focus on taxi out time prediction for John F. Kennedy airport in New York. They use historical data based on airport surface traffic. Srivastava [20] 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 month. 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 [30]. Kim [21] 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.

2.3. 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. Fist, the CART algorithm (classification and regression trees) was introduced by Breiman [7] 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 overfitting, 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 [16] 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 [21] 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 min. Glina et al. [17] 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. [20] 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 [19]. The contestants of the GE flight quest competition all achieved high prediction accuracy of around 4.2 min 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.

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 [32].

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 the total trip including taxi times at the departure and arrival airport is predicted. The total trip time is

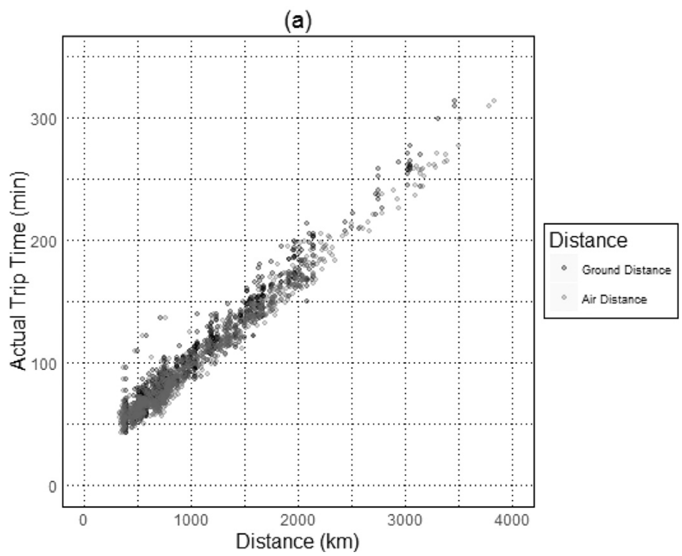

Fig. 2. Air and ground distance have a correlation coefficient of 0.981 with actual total trip time.

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.

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 [20] 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 [21], 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 due to algorithmic limitations of a maximum of 34 classes for categorical features, a variable called destination frequency was added. The destination frequency represents the total number of flights between two airports within one year. With this variable, we quantify the effect of tacit knowledge of the crew and ground personal that is shown in an efficient routine, 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 min prior to landing until 30 min 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 [20]. 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 Figs. 2 and 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

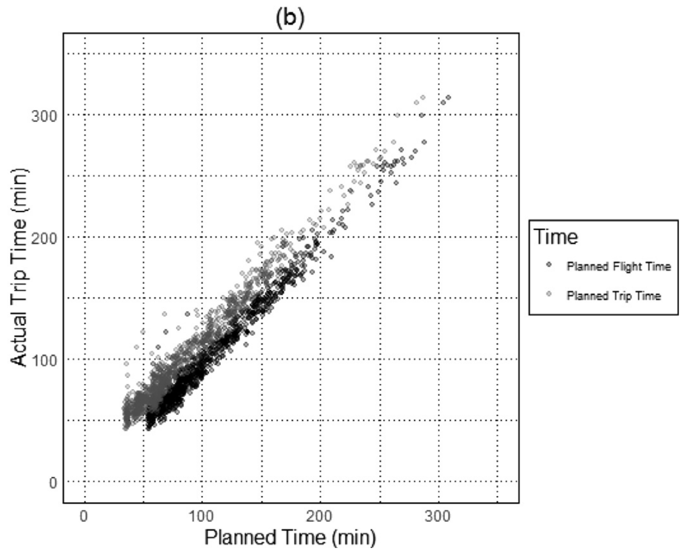

Fig. 3. Planned flight time and total trip time have a correlation coefficient of 0.986 with actual total trip time.

trip time prediction.

According to [21,25], 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 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.

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 such as [17,19,20]. Furthermore, importance rankings were derived by individual models for variable selection. 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 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). The aim of the feature selection for the first stage favors features, which increase overall prediction accuracy. In the first stage we aim for an initial accurate arrival time prediction, which is further enhanced in the next stages of the ensemble model. Thus, features that considerably increase overall prediction accuracy were included in the first stage.

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. The feature selection in the second stage focuses on features, which further fine-tune the initial prediction. Those features did not have a prominent effect in the linear regression, but showed an improved prediction accuracy in the gradient boosting model.

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. None of the weather features has a very high variable importance score compared to the

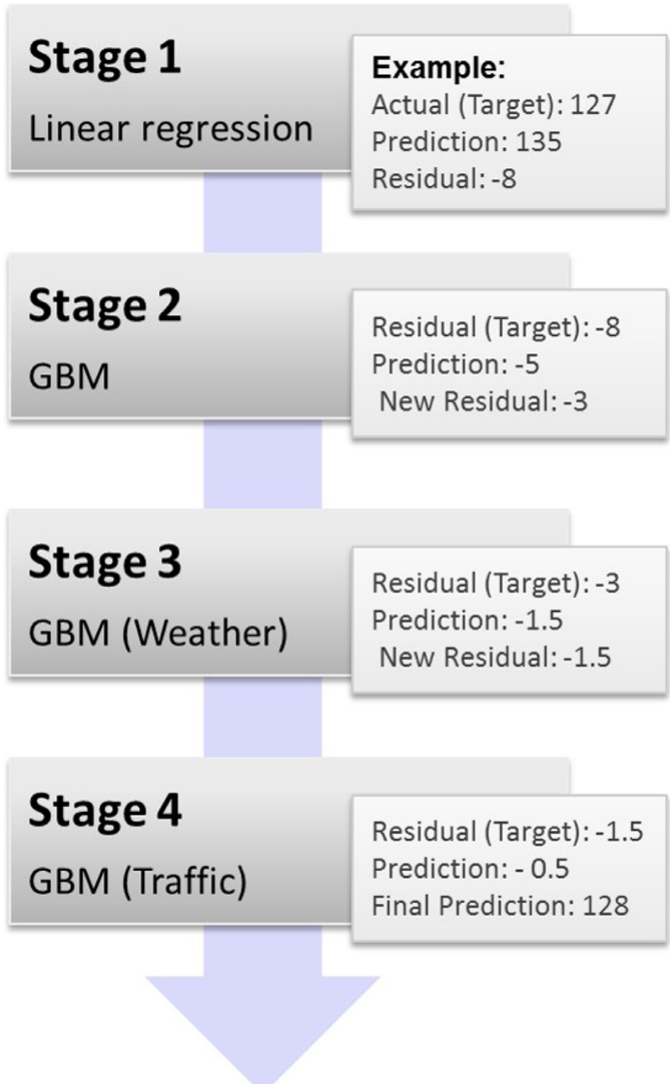

Fig. 4. Overview of ensemble model.

other weather features. Thus, all weather features were included without any further reduction.

Stage 4 of the ensemble model combines capacity utilization information for the arrival airport according to the concept proposed by [20]. 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. Also in this stage, the initial prediction is further fine-tuned with additional information. Depending on the time of day different traffic patterns are important for prediction accuracy. For every flight each traffic related feature is calculated and included in the prediction of the forth stage. Please refer to Fig. 4 to see a graphic display of the ensemble model.

3.4. Performance evaluation

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.

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|c_i - v_i| \tag{2}$$

$$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(c_i - v_i\right)^2} \tag{3}$$

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 6000 flights. The datasets are made up of flight data ranging from January 2015 until April 2016. The datasets are generated through random sampling.

3.5. Tuning of hyperparameters

Hyperparamters are values that set up machine learning algorithms operators. 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. Grid search can be easily implemented and computations can be parallelized [6]. N.trees

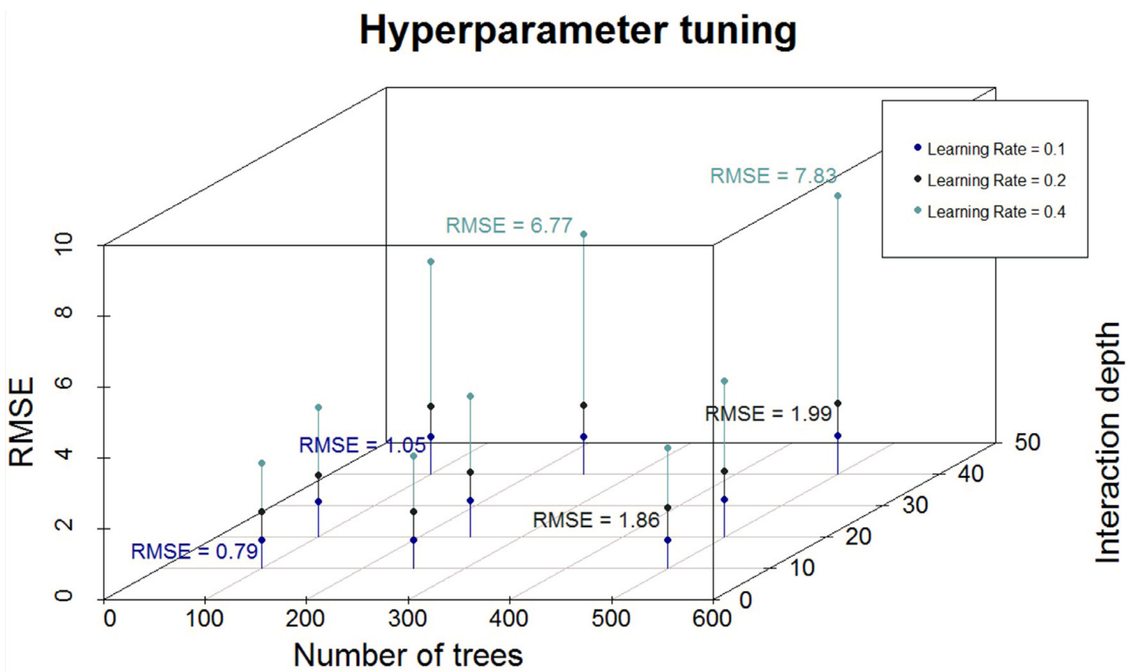

Fig. 5. Increasing interaction depth leads to a slight increase in RMSE. Overall a low learning rate delivers superior results. The effect of number of trees on RMSE depends on learning rate and interaction depth.

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 Fig. 5 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.

4. Aircraft performance data and model

4.1. Cost index data

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

$$CI = \frac{C_{time}}{C_{Fuel}} \tag{4}$$

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 [9,11] 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. [3] provide an innovative review of delay costs also considering passenger delay costs.

Because of the limited information of time related costs we model the cost of time as a function of fuel costs for a given cost index. Thus, we are able to display the cost of fuel for each cost index level. This information serves 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 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

Table 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.)[a]	2	2	1	0	−1	−2	−2.5	−3	−3	−3	−3
Change in fuel consumption (kg)[a]	−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_{CI} (€ /min.)	12.60	10.50	12.60	0.00	−12.60	−12.60	−15.12	−16.80	−18.90	−21.00	−22.40

[a] acc. to aircraft performance model

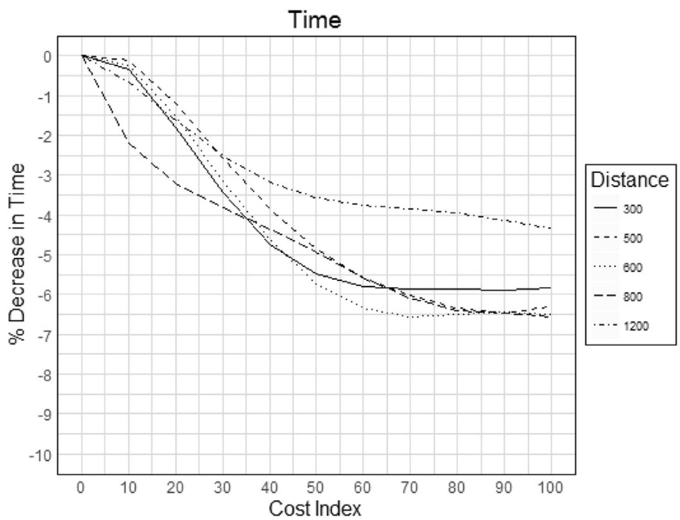

Fig. 6. Decrease in flight time as a function of cost index (CI:0-100). We observe that the change in flight time varies depending on the distance. For flights of 300 NM the maximum decrease in flight time is 4%. For longer flights (800–1200 NM) a maximum decrease of 6.5% can be achieved.

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 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 (NM). 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 touchdown. 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.

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 run 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 min. 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 min flight time at cost index 0 will be able to decrease the flight time by 3 min to a total flight time of 49 min at CI 100. This is approximately 5.7%. A flight with 500 NM and a total flight time of 79 min at CI 0 can decrease flight time by 5 min at CI 100 , which is 6.3%. A flight of 1200 NM and 174 min total flight time at CI 0 can reduce the flight time to 166 min for CI 100, which equals 4.6%. In Fig. 6 the percentage change in flight time are displayed.

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),

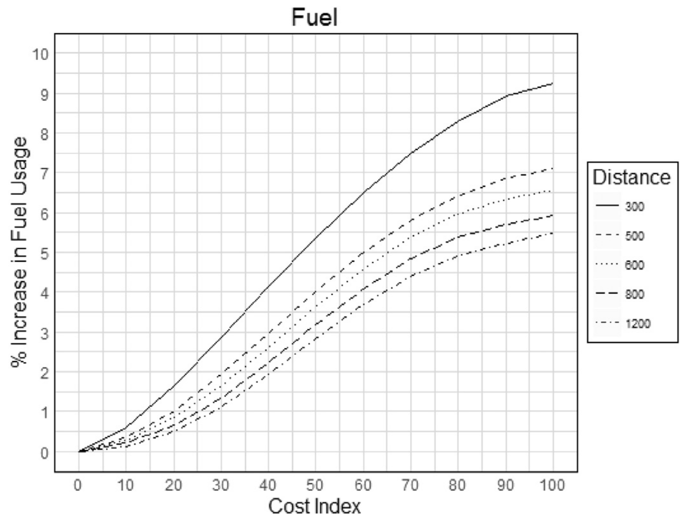

Fig. 7. Increase in fuel usage as a function of cost index (CI:0-100). The increase in fuel consumption is quite similar for all distances. For shorter flights the fuel usage can increase by more than 9%. For distances of 1200 NM the maximum increase is 5.5%.

Table 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%

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 1200 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 Fig. 7 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.

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.

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.

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 scheduled time of arrival calculation provided by operations control. In Table 2, 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. Such high improvements in prediction accuracy are comparable to the results of the GE flight quest challenge. Furthermore [17,20] 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 Fig. 8, 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 2). 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 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 4th 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 4 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 3 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 4.

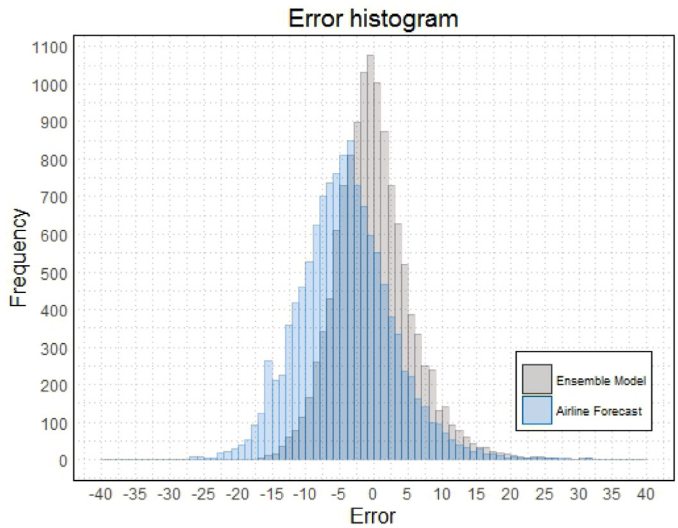

Fig. 8. Absolute error in minutes: Difference between actual total trip time and predicted/forecasted total trip time. Negative values indicated earlier arrival than predicted/forecasted and positive values indicate later arrival than predicted/forecasted.

Table 3
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

Table 4
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

In Fig. 9 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 Fig. 9 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 min, scheduled total trip time below 100 min 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

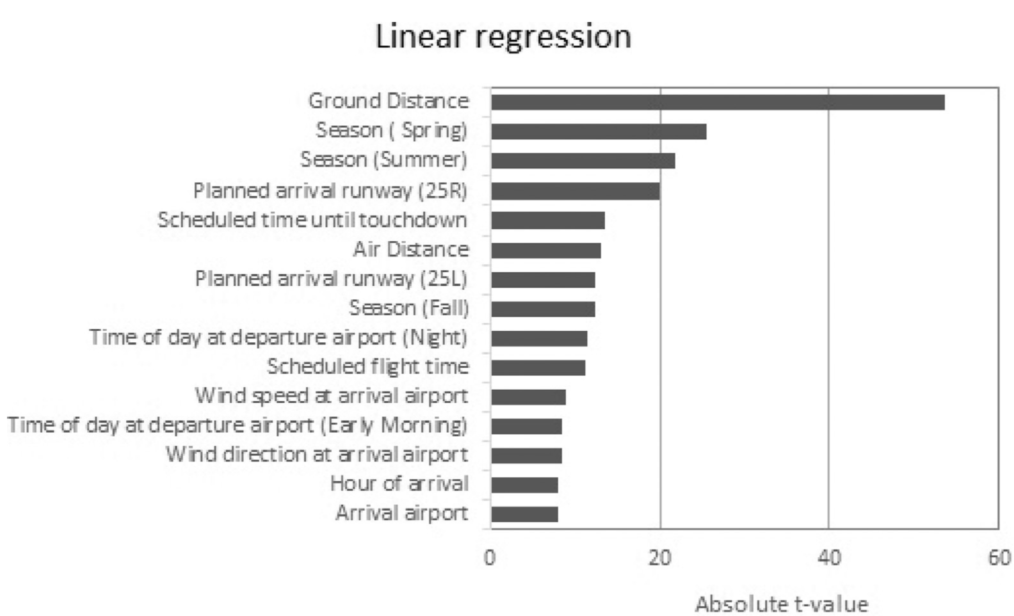

Fig. 9. Variable importance plot: Linear regression including all variables. Absolute *t*-value as measurement. Not fully comparable to linear regression in the ensemble model due to variable selection and different ensemble stages.

almost 1.5 times the average distance have a significantly higher prediction accuracy compared to flights from Italy. This might be due to 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 month, 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.

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 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.

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 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 5).

This effect is even stronger for the two flights with a distance of 800 NM and 1200 NM. 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 [18]. Considering the complex cost structures for cost of time, we model cost of time as a function of fuel cost. In Fig. 10, 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. 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 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 10 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

Table 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

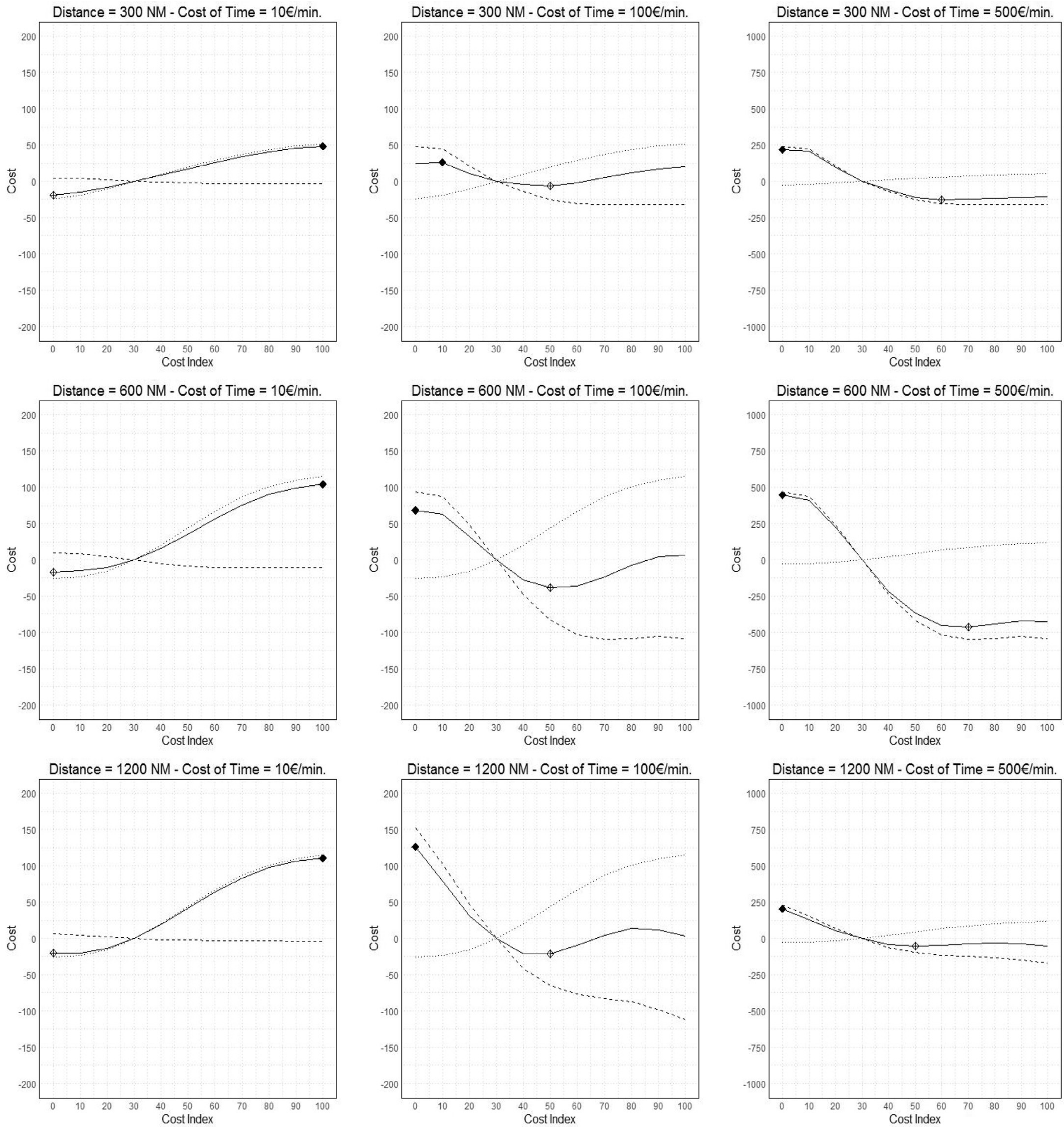

Fig. 10. Optimal cost index: Considering fuel cost and different cost of time for the distances 300 NM, 600 NM and 1200 NM we see the variability in cost index. Black square equals maximum cost. Grey square, equals minimum total cost.

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.

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 min prior to landing. Considering that continental European flights have an average total trip time of 100 min, the time window for decision making increases by more than 60 min. 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 Cook et al. [2,10]. 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.

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, airlines can reduce CI levels without any tangible 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.

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.

7. Declarations of interest and funding

Declarations of interest: none.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.orp.2018.08.004

References

[1] Abdelghany KF, Abdelghany AF, Ekollu G. An integrated decision support tool for airlines schedule recovery during irregular operations. Eur J Oper Res 2008;185(2):825–48. https://doi.org/10.1016/j.ejor.2006.12.045.

[2] Aktürk MS, Atamtürk A, Gürel S. Aircraft rescheduling with cruise speed control. Oper Res 2014;62(4):829–45. https://doi.org/10.1287/opre.2014.1279.

[3] Ball M, Barnhart C, Dresner M, Hansen M, Neels K, Odoni A, et al. Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the united states. Technical Report. 2010.

[4] Barnhart C, Smith B. Quantitative problem solving methods in the airline industry. first ed. US: Springer9781461416074; 2012. https://doi.org/10.1007/978-1-4614-1608-1.

[5] Beatty R., Hsu R., Berry L., Rome J.. Preliminary evaluation of flight delay propagation through and airline schedule. In: Proceedings of the 2nd USA Europe Air Traffic Management R&D Semin1998;(December):1–9.

[6] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13:281–305. https://doi.org/10.1162/153244303322533223. 1504.05070.

[7] Breiman L. Classification and regression trees. Chapman & Hall/CRC; 1984.

[8] Campanelli B, Fleurquin P, Arranz A, Etxebarria I, Ciruelos C, EguÃluz VM, et al. Comparing the modeling of delay propagation in the us and european air traffic networks. J Air Transp Manag 2016;56:12–8. https://doi.org/10.1016/j.jairtraman.2016.03.017.

[9] Cook A. Evaluating the true cost to airlines of one minute of airborne or ground delay; final report. Rev Literat Arts Am 2004.

[10] Cook A., Tanner G., Williams V., Meise G.. Dynamic cost indexing. Proceedings of the 6th EUROCONTROL Innovative Research Workshops & Exhibition2007;(December):4–6.

[11] Cook AJ, Tanner G. European airline delay cost reference values. Technical Report. 2011.

[12] Customer Services AIRBUS. Cost Index. Technical Report. 1998.

[13] DeJonge M.K.. Required time of arrival (RTA) control system. 1992. US Patent 5,121,325.

[14] Edwards HA, Dixon-Hardy D, Wadud Z. Aircraft cost index and the future of carbon emissions from air travel. Appl Energy 2016;164:553–62. https://doi.org/10.1016/j.apenergy.2015.11.058.

[15] Eurocontrol, FAA. 2010 U.S./Europe Comparison of ATM-Related Operational Performance. Technical Report. 2010.

[16] Friedman JH, Breiman L. Greedy function approximation: a gradient boosting machine. Ann Stat 2000;29(5):1189–232. https://doi.org/10.1214/aos/1013203451.

[17] Glina Y, Jordan R, Ishutkina M. A tree-based ensemble method for the prediction and uncertainty quantification of aircraft landing times. Proceedings of the american meteorological society - 10th conference on aritificial intelligence applications to environmental science. 2012.
[18] Index Mundi. Jet Fuel Monthly Price - IndexMundi. 2016. URL http://www.indexmundi.com/commodities/?commodity = jet-fuel&months = 60&currency = eur.
[19] Kaggle. Winners - GE Flight Quest. 2012. URL http://www.gequest.com/c/flight/details/winners.
[20] Kern CS, de Medeiros IP, Yoneyama T. Data-driven aircraft estimated time of arrival prediction. Proceedings of the 9th annual ieee international systems conference (SysCon). 2015. p. 727–33. https://doi.org/10.1109/SYSCON.2015.7116837.
[21] Kim MS. Analysis of short-term forecasting for flight arrival time. J Air Transp Manag 2016;52:35–41. https://doi.org/10.1016/j.jairtraman.2015.12.002.
[22] Kletzel J, DeWaart D. Tailwinds - 2014 airline industry trends. Technical Report. 2014.
[23] Murphy J.R., Reisman R.J., Clayton J., Wright R., Volpe J.A.. Physics-Based and parametric trajectory prediction performance comparison for traffic flow management. Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit2003;(August):1–11.
[24] Peterson MD, Bertsimas DJ, Odoni AR, Rooml S. Models and algorithms congestion for at transient airports queueing. Manage Sci 1995;41(8):1279–95.
[25] Rebollo J, Balakrishnan H. A network-Based model for predicting air traffic delays. Proceedings of the International Conference on Research in Air Transportation (ICRAT 2012). 2012.
[26] Rebollo JJ, Balakrishnan H. Characterization and prediction of air traffic delays. Transp Res Part C Emerg Technol 2014;44:231–41. https://doi.org/10.1016/j.trc.2014.04.007.
[27] Roberson B. Cost Index Explained. Technical Report. 2007.
[28] Smith J, Kletzel J. Getting clear of the clouds - Will the upward trajectory continue? Technical Report. 2015.
[29] Srivastava A. Improving departure taxi time predictions using ASDE-X surveillance data. Proceedings of the IEEE/AIAA 30th Digital Avionics Systems Conference. 2011.
[30] Takacs G.. Predicting flight arrival times with a multistage model. In: Proceedings of the ieee international conference on big data2015; 2:78–84. 10.1109/BigData.2014.7004435.
[31] Tu Y, Ball MO, Jank WS. Estimating flight departure delay distributions a statistical approach with long-Term trend and short-Term pattern. J Am Stat Assoc 2008;103(481):112–25. https://doi.org/10.1198/016214507000000257.
[32] UK Civil Aviation Authority. Flight planning manual. Technical Report. 2006.
[33] Walker C. CODA DIGEST 2015. Technical Report. Eurocontrol; 2016.
[34] Xu N, Sherry L, Laskey KB. Multi-factor model for predicting delays at u.s. airports. Transp Res Record J Transp Res Board 2008;2052:62–71.
[35] Zonglei L, Jiandong W, Guansheng Z. A New Method to alarm large scale of flights delay based on machine learning. Proceedings of the international symposium on knowledge acquisition and modeling. 2008978-0-7695-3488-6. p. 589–92. https://doi.org/10.1109/KAM.2008.18.