Aircraft taxi time prediction: Comparisons and insights

Stefan Ravizza [a,*], Jun Chen [b], Jason A.D. Atkin [a], Paul Stewart [b], Edmund K. Burke [c]

[a] School of Computer Science, University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, UK
[b] School of Engineering, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK
[c] Department of Computing and Mathematics, University of Stirling, Cottrell Building, Stirling FK9 4LA, UK

ARTICLE INFO

ABSTRACT

The predicted growth in air transportation and the ambitious goal of the European Commission to have on-time performance of flights within 1 min makes efficient and predictable ground operations at airports indispensable. Accurately predicting taxi times of arrivals and departures serves as an important key task for runway sequencing, gate assignment and ground movement itself. This research tests different statistical regression approaches and also various regression methods which fall into the realm of soft computing to more accurately predict taxi times. Historic data from two major European airports is utilised for cross-validation. Detailed comparisons show that a TSK fuzzy rule-based system outperformed the other approaches in terms of prediction accuracy. Insights from this approach are then presented, focusing on the analysis of taxi-in times, which is rarely discussed in literature. The aim of this research is to unleash the power of soft computing methods, in particular fuzzy rule-based systems, for taxi time prediction problems. Moreover, we aim to show that, although these methods have only been recently applied to airport problems, they present promising and potential features for such problems.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The latest vision for air transportation in Europe is predicting marked growth in this sector. The European Commission [1] assumes an increase in the global volume of air traffic from 2.5 billion passengers in 2011 to 16 billion passengers in 2050. Thus, the number of commercial flights in Europe per year is expected to increase from 9.4 million to 25 million during the same time period. Nevertheless, one of the formulated goals by 2050 is also that on-time performance of flights is within 1 min.

Efficient ground movement operations are key to successful operations of air transportation networks [2]. Especially since airport ground movement operations link runway sequencing and gate assignment problems and accurate taxi time predictions support various different stakeholders at airports. For example, the benefits for the take-off sequence of having accurate taxi times was shown in [3] and recent developments for Heathrow [4] require accurate taxi times both for take-off sequencing and for allocating pushback times to aircraft, at which they should leave the stands. A significant proportion of the actual travel time can be spent at the airport, especially with short-haul flights. To achieve the stated on-time performance from the European Commission, it is crucial to more accurately predict taxi times at European airports. This will enable runway planning with less uncertainty and will provide important predictions for all of the parties which are involved in the handling of aircraft. Furthermore, it will also be increasingly important to have accurate control of the aircraft on the ground using an advanced decision support system [5].

Idris et al. [6] published the first paper on taxi-out time estimation based on multiple linear regression, with data from Boston Logan International Airport (BOS). The importance of taking into account the current amount of traffic on the surface, when predicting taxi-out times, was highlighted for the first time. The take-off queue size was found to be an especially important explanatory variable for the problem. With the introduction of Collaborative Decision Making (CDM) systems at airports within the last few years, practitioners at airports realised the need for having more accurate taxi times and, driven by that, more researchers have analysed the problem of taxi time prediction. Several authors have published their results about taxi-out time prediction at US airports [7–12]. Balakrishna et al. [7–9,11] used a reinforcement learning algorithm which showed good results for data from Detroit International Airport (DTW) and Tampa International Airport (TPA), but the results were not very consistent for data from John F. Kennedy International Airport (JFK). However, this approach cannot provide the same insights into the problem as some other approaches. Clewlow et al. [10] highlighted that the number of arrivals does affect the taxi-out times, which was not sufficiently taken into account before that. Their multiple regression approach was based on John F. Kennedy International Airport and Boston Logan International Airport. Jordan et al. [13] developed a sequential forward floating subset selection method with the aim of selecting the most

influential explanatory variables from a set. It seems to be one of the few sources which analysed not only taxi-out times, but also taxi-in times. The analysis was performed with data from Dallas/Fort Worth International Airport (DFW). Kistler and Gupta [14] developed a multiple linear regression approach, for the same airport, with several different explanatory variables, to predict taxi-in and taxi-out times.

All of the aforementioned publications were based on data from US airports. One problem of adopting these findings for Europe is that US airports are usually structurally different from European airports. For example, they distinguish between gate-ramps which are operated by airlines, and taxiways, which are controlled by tower ground controllers. In addition, it seems that the problem of taxi time prediction in the US is dominated by the runway queue size and is less related to the actual distance that an aircraft has to taxi [15]. Furthermore, since no cross-validation details were often given in the papers, and the assumptions for multiple linear regression were not discussed, the importance of some of the findings was not clear.

More recently, Ravizza et al. [15] identified which explanatory variables affect the taxi time the most at two major European airports, Stockholm-Arlanda Airport (ARN) and Zurich Airport (ZRH). The utilised multiple linear regression approach incorporated explanatory variables based on the airport layout and not only fitted historic data well, but it also predicted taxi times accurately. The assumptions for multiple linear regression were also tested, making the findings more reliable. Chen et al. [16] further improved the accuracy of the prediction by using a Mamdani fuzzy rule-based system based on the same explanatory variables which had been identified for Zurich Airport.

This paper uses the same explanatory variables as in the research by Ravizza et al. [15] on datasets from the same airports, but with considerably longer time periods now being considered. The aim is to test different statistical regression approaches, along with regression methods which fall into the realm of soft computing, to more accurately predict taxi times, to demonstrate the advantages and disadvantages of these approaches and to give further insights into the problem, especially for taxi-in times. The aim of this research is also to unleash the power of soft computing methods, and fuzzy rule-based systems in particular, for taxi time prediction, and to show that, although these methods have only been recently applied to airport problems, they present promising potential features and deserve more attention from practitioners.

The remainder of the paper discusses the airport ground movement problem and the utilised datasets from Stockholm-Arlanda Airport and Zurich Airport in Section 2. Section 3 introduces six different regression approaches, which are tested in Section 4. This section also presents insights from the best performing approach, before Section 5 ends with the conclusions.

2. Problem description

The airport ground movement problem is discussed first in this section, before the dataset details for the two European airports are introduced.

2.1. The airport ground movement problem

Airport operations are getting more and more complicated with increasing numbers of flights. A major part of having efficient handling of aircraft at an airport is the ground movement problem. Atkin et al. [2] highlighted how the airport ground movement problem links other important airport operations such as runway sequencing [17] and gate assignment [18]. In general, research into the airport ground movement problem aims at helping ground

Table 1
Overview of datasets from Stockholm-Arlanda Airport and Zurich Airport.

Taxi time	Range for ARN [1,30]	Range for ZRH [0.2,34.0]	Type Ordinal for ARN, scale/ordinal for ZRH
ARR	{0,1}	{0,1}	Nominal
Distance	–	[119,4320]	Scale
$\log_{10}$(Distance)	[2.8,3.7]	[2.1,3.6]	Scale
$\log_{10}$(Angle)	–	[1.9,3.0]	Scale
Distance2	[0,3373]	–	Scale
$Q_{DEP,\#DEP}$	[0,16]	[0,23]	Ordinal
$Q_{DEP,\#ARR}$	[0,13]	[0,26]	Ordinal
$Q_{ARR,\#DEP}$	[0,9]	[0,8]	Ordinal
$Q_{ARR,\#ARR}$	[0,9]	[0,7]	Ordinal
$N_{DEP,\#DEP}$	[0,12]	[0,15]	Ordinal
$N_{DEP,\#ARR}$	[0,6]	[0,6]	Ordinal
$N_{ARR,\#DEP}$	[0,12]	[0,16]	Ordinal
$N_{ARR,\#ARR}$	[0,6]	[0,6]	Ordinal
Mode 1	–	{0,1}	Nominal
Mode 3	–	{0,1}	Nominal
Mode 5	–	{0,1}	Nominal

controllers in airport towers, to support them in making better decisions. These consist of routing and scheduling aircraft on the taxiways, from the gate or runway exit to the runway entrance or back to the gate. Thereby, total taxi time is aimed to be reduced or, more broadly, it is aimed that the operations serve the other linked airport operations as smoothly as possible, while simultaneously reducing the environmental issues. Several operational constraints have to be considered. Aircraft have time windows when they can start taxiing and when they have to reach the target. For a departing aircraft, this consists of an earliest pushback time and either a specific time or a time window at the runway where the aircraft has to take-off. An arriving aircraft has to leave the runway immediately after landing and taxi to the gate, which should be free when the aircraft is expected to be there. Depending upon the layout of an airport, not all aircraft can taxi on all taxiways due to limitations of size or weight. This means that the route may be aircraft type specific. In addition, aircraft cannot cross or overtake on a single taxiway, which makes the routing more complex.

This research supports the aforementioned problem by understanding and more accurately predicting taxi times. Such predictions can be used to make better overall decisions at airports and also to improve the quality of decision support systems for the ground movement problem at airports, by applying the findings and integrating the different aircraft speeds into such models.

2.2. Considered airport data

The authors had access to historic data from two European airports. All available data from each airport was combined into one dataset each and they were tested separately. This approach was used, since, as discussed by Demšar [19], no statistical test exists which could compare different prediction methods based on different datasets where each prediction method is utilised for several repetitions of 10-fold cross-validation, due to the overlaps of the training data in different random samples.

This research aims to compare various prediction methods and to find further insights into taxi time prediction at airports. Thereby, it extends the research by Ravizza et al. [15] which highlighted the statistically significant explanatory variables of this problem. The same explanatory variables were used in this study, which is based on more data from the same airports. All of the explanatory variables and their ranges for both datasets are shown in Table 1. The remainder of this section presents the two airports together with the explanatory variables to predict taxi times.
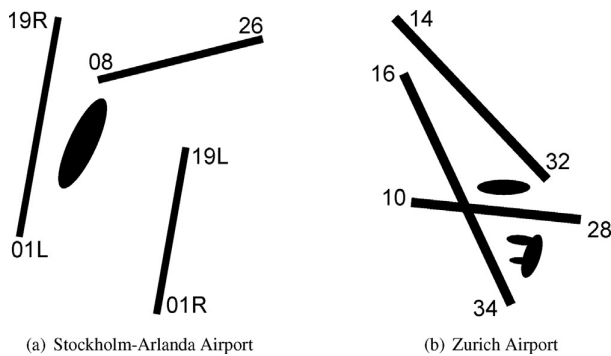
(a) Stockholm-Arlanda Airport   (b) Zurich Airport

Fig. 1. Sketch of both airport layouts.

2.2.1. Stockholm-Arlanda Airport

Stockholm-Arlanda Airport (ARN) is the largest airport in Sweden, with over 19 million passengers in 2011. Scandinavian Airlines is the main hub carrier at this airport. Three runways are arranged around the terminals with two parallel runways (with the directions 19R/01L and 19L/01R) and a third runway in the north (with the direction 08/26), see Fig. 1(a). Data from two entire days' operations was used within this analysis: the 7th of September 2010 (661 movements) and the 14th of October 2010 (656 movements).

Besides the reported taxi times of the aircraft, the dataset for Stockholm-Arlanda Airport consists of the following 11 explanatory variables which were found to be significant in earlier work: the indication of whether an aircraft is arriving or departing (ARR), the logarithmic transformation of the total taxi distance, the total distance which an aircraft is taxiing on longer straight taxiways (Distance2 in meters), and four Q and four N values which indicate the amount of traffic of different types which is on the airport's surface while the considered aircraft is taxiing. The Q value counts the number of other aircraft on the airport's surface which cease taxiing during the time of taxiing of the particular aircraft. Similarly, the N value counts how many other aircraft are already taxiing when the particular aircraft starts taxiing. These values were differentiated into eight integer values to have each combination of arrival and departure for both sets of values for the aircraft under consideration and the aircraft which were actually counted.

2.2.2. Zurich Airport

Zurich Airport (ZRH) is the largest airport in Switzerland, with over 24 million passengers in 2011. This airport consists of the following three runways: 10/28, 14/32 and 16/34, where the first and the last runway intersect (see Fig. 1(b)). The main terminals are located to the south of runway 10/28 and a remote terminal is located to its north. Swiss International Air Lines are based in Zurich as the main hub carrier there. The dataset for Zurich Airport consists of an entire day's operations from the 19th of October 2007 (679 movements) and an entire week's operation between the 27th of June and the 3rd of July 2011 (5611 movements).

Fifteen different explanatory variables were found to be relevant for the dataset for Zurich Airport, in addition to the reported taxi times. Again the indication of arrival or departure was used, together with the eight Q and N values. For this dataset, the total taxi distance in meters and its logarithmic transformation is employed. In addition, the logarithmic transformation of the total angle was taken into account. The angle is calculated as the total angular deviation between adjacent arcs on the shortest path between the start and target location of the aircraft. As a last set of explanatory variables, three dummy variables where introduced to model four different

operational modes. Such operational modes determine which runways can be used to take-off and to land and are affected by the time of the day and heavy winds.

The reported taxi time information is only to the minute rather than to the second. The only exception is the information about landing times on the runway at Zurich Airport, where detailed times have been recorded.

3. Statistical and soft computing regression approaches to predict taxi time

The aim of this research is to compare a wide range of different regression approaches for the problem of predicting taxi times at airports. WEKA [20] is an open source collection of machine learning algorithms for data mining tasks. It was used to explore which regression models were promising for delivering good taxi time predictions, as this software contains a large selection of approaches to compare against our existing approaches. During an initial selection analysis, several regression approaches were not showing promising results and thus are not included in the rest of this study. These regression approaches included, among others: a decision tree learner using reduced-error pruning, nearest neighbourhood methods, Pace regression, multilayer perceptron (back propagation neural network) and Gaussian processes [21].

Four regression techniques from WEKA which are showing promising results are explained in the rest of this section. These are: multiple linear regression, least median squared linear regression, support vector regression and M5 model trees. In addition, the approach published by Chen et al. [16] with a modified and enhanced Mamdani fuzzy rule-based system is also used for comparison. Finally, a very promising extended version of another fuzzy rule-based system was developed for this research and is explained.

3.1. Multiple linear regression

Multiple linear regression is a very widely used regression methodology. It is not only very well studied, but it also has the advantage of determining which explanatory variables of a model are significant. In this study, this technique is mainly used to act as a baseline for comparison with the other approaches. Multiple linear regression is a statistical method attempting to model the dependent variable as a linear weighted function of the explanatory variables. The weights, or regression coefficients, can be estimated by using the least square approach. More in-depth coverage of multiple linear regression can be found in the book by Montgomery et al. [22] and the paper by Ravizza et al. [15].

3.2. Least median squared linear regression

Least median squared linear regression is a more robust linear regression approach than multiple linear regression. Instead of minimising the mean of the squares of the errors, this approach aims to minimise the median of these squares. Standard linear regression is applied iteratively to subsamples of the data and the solution with the smallest median of the squared errors is output [23]. The advantage of being robust against outliers comes with the disadvantage of higher computational costs.

3.3. Support vector regression

Support vector machines are supervised learning methods and can be used for classification and regression analysis. Support vector regression ignores training data within a specified threshold of the model prediction. The objective is to minimise the norm of the weights of the explanatory variables together with the error for the training data which is further away from the prediction than the

set threshold. A value C is normally defined to weight the trade-off between the two objective functions. The dual formulation of this optimisation model is often solved by preference. Support vector regression can be extended to non-linear models by incorporating a kernel function which transforms the original training data into a higher dimensional space. The best kernel found for this particular problem and these datasets turned out to be a normalised polynomial kernel. A tutorial for this approach can be found in [24].

3.4. M5 model trees

Another way of predicting numeric values is by using decision trees which store linear regression models on their leaves. Such trees are called model trees and are similar to piecewise linear functions for the entire model. Model trees are usually smaller and more accurate than regression trees which have only an average value on their leaves [25]. The tree is constructed by the divide-and-conquer method, where the splitting criterion determines the best explanatory variables to split on, based on the expected error reduction. The splitting process finishes when the standard deviation of the subset of the training data is below a certain threshold or the size of this subset is too small. Afterwards, a linear regression model is calculated for each leaf. Pruning can be applied in a second stage: all non-leaves are tested for whether it is better to keep the subtree or whether a linear model could replace the subtree. Additionally, a smoothing stage can be added to reduce the discontinuities between the linear models for different leaves. More details of the M5 model tree can be found in the paper by Quinlan [25].

3.5. Mamdani fuzzy rule-based systems

Fuzzy Rule-Based Systems (FRBSs) are a way of modeling processes which have the ability to be interpreted with linguistic statements. First introduced by Zadeh [26], they give the possibility of combining human expertise together with mathematical models. The input is first mapped with a fuzzification interface then decisions can be made, before these are mapped to a single crisp output using a defuzzification interface.

The concept behind a modified and enhanced Mamdani FRBSs is introduced first, before discussing another concept based on a different type of FRBS in the next section. The general 'rule-base' of an FRBS has the following form, with a number of fuzzy if-then rules R_i:

$$R_i : \text{If } x_1 \text{ is } A_i^1 \text{ and } x_2 \text{ is } A_i^2, \ldots, \text{ and } x_j \text{ is } A_i^j \quad \text{Then} \quad y_i = Z_i. \tag{1}$$

The values x_l for all $l = 1, \ldots, j$ are the explanatory variables and the A_i^l are the ith linguistic values (fuzzy sets) which are each represented by a membership function $\mu_{A_i^l}(x_l)$. They take the form of Gaussian functions of the following form in this work in order to give smooth predictions and to ensure its partial differentiability:

$$\mu_{A_i^l}(x_l) = \exp\left[-\frac{1}{2} \cdot \frac{(x_l - c_i^l)^2}{(\sigma_i^l)^2}\right], \tag{2}$$

where c_i^l denotes the centre of the bell-shape curve and σ_i^l denotes the standard deviation. The consequence part Z_i is a fuzzy set for a Mamdani FRBS, which is here modelled as a bell-shaped membership function. The utilisation of a bell-shape membership function ensures again the function's partial differentiability. The chosen membership types along with the centre of gravity defuzzification ensure that gradient based optimization method, such as the Back-Error-Propagation technique, can be utilised to fine tune the obtained Mamdani FRBS. More details of the approach, the tuning

for the problem of estimating aircraft taxi times, and preliminary results, can be found in the work by Chen et al. [16].

Some of the key features of Mamdani FRBSs highlighted in [16] are (1) the ability to approximate complex non-linear systems, (2) the ability for rules to differ in different regions, (3) the ability to integrate human expertise, and (4) the ability to interpret the underlying system.

3.6. TSK fuzzy rule-based systems

Another form of fuzzy inference system, originally proposed by Takagi and Sugeno [27], has fuzzy sets involved only in the premise part. By using Takagi and Sugeno's fuzzy inference scheme (TSK), one can describe the fuzzy if-then rules as follows:

$$R_i : \text{If } x_1 \text{ is } A_i^1 \text{ and } x_2 \text{ is } A_i^2, \ldots, \text{ and } x_j \text{ is } A_i^j$$

$$\text{Then} \quad y_i = g_i(x_1, x_2, \ldots, x_j). \tag{3}$$

R_i denotes the ith rule to be considered. $g_i()$ is any function, and could, for example, be linear or quadratic. Normally, using a linear function for $g_i()$ is enough, since the fuzzy if-then rule has already embedded non-linearity inherently. When a linear model structure is assumed then a rule base with k rules takes the following format:

$$R_1 : \text{If } x_1 \text{ is } A_1^1 \text{ and } x_2 \text{ is } A_1^2, \ldots, \text{ and } x_j \text{ is } A_1^j \quad \text{Then} \quad y_1 = b_1^0 + b_1^1 \cdot x_1 + \ldots + b_1^j \cdot x_j$$

$$\ldots \tag{4}$$

$$R_k : \text{If } x_1 \text{ is } A_k^1 \text{ and } x_2 \text{ is } A_k^2, \ldots, \text{ and } x_j \text{ is } A_k^j \quad \text{Then} \quad y_k = b_k^0 + b_k^1 \cdot x_1 + \ldots + b_k^j \cdot x_j.$$

The crisp output from the input $(x_1, x_2, \ldots, x_j)$ is obtained as the weighted sum of the consequences of the k rules:

$$y = \sum_{i=1}^{k} \beta_i \cdot y_i = \sum_{i=1}^{k} \beta_i \cdot (b_i^0 + b_i^1 \cdot x_1 + \ldots + b_i^j \cdot x_j), \tag{5}$$

where β_i actually represents the certainty of each rule contributed by the premise of the corresponding rule:

$$\beta_i = \frac{\mu_{A_i^1}(x_1) * \mu_{A_i^2}(x_2) * \ldots * \mu_{A_i^j}(x_j)}{\sum_{l=1}^{k} \mu_{A_l^1}(x_1) * \mu_{A_l^2}(x_2) * \ldots * \mu_{A_l^j}(x_j)}. \tag{6}$$

The membership function $\mu_{A_i^l}(x_l)$ of the premise part is again a Gaussian function as in Eq. (2).

The above presented Mamdani FRBS utilised a combined k-means algorithm [16] and genetic algorithm [28], to automatically identify the initial values of the parameters both in the premise and consequent parts. The same method is used to determine the parameters associated with the premise part for the TSK FRBS. The least square approach is then used to determine the initial values of $b_i^0, b_i^1, \ldots, b_i^j$ for all $i = 1, \ldots, k$ for the consequent part. To further refine the initial fuzzy system which is obtained, a genetic algorithm, namely G3PCX [29], is incorporated into TSK to fine-tune the premise part, followed by a least square approach to obtain the consequent part. This process continues iteratively until a pre-specified condition is met in order to reach a more accurate fuzzy system. G3PCX is a real-parameter genetic algorithm using a parent-centric recombination operator (PCX) and an elite-preserving, computationally fast evolutionary model (G3).

As mentioned in [16], in comparison to the Mamdani FRBS, the following distinctive features associated with the TSK FRBS can be identified:

- The TSK FRBS could in some ways be viewed as an extension of multiple linear regression. Each rule in the rule base resembles a multiple linear regression model for a decomposed explanatory

variable region. Hence, the explanatory ability associated with multiple linear regression automatically applies to the TSK FRBS.

- These rules work cooperatively to produce estimations, which may result in more accurate estimations.
- Although one could lose certain linguistic meanings in the consequent part in comparison to the Mamdani FRBS, due to the function form of the TSK consequent part, such a form should be able to approximate the sub region more accurately than a fuzzy set.

4. Comparisons and insights

This section analyses the different regression approaches for predicting taxi times at airports and shows comparisons between approaches and insights from the best performing approach. First, it is necessary to specify the experimental setup and the performance measures used.

4.1. Experiment setup

All of the experiments were executed on a standard desktop PC (Intel Core 2 Duo, 3 GHz, 2 GB RAM). WEKA was used to perform all of the experiments apart from the last two models, which are related to the fuzzy rule-based systems [20,21]. Analysis with the support vector regression method identified that the best parameter for these experiments was to use the value C equal to 2 and to employ a normalised polynomial kernel with exponent 3. Both of the fuzzy rule-based systems were implemented and tested in MATLAB R2010a. Mamdani FRBS was based on 12 rules, as published in [16]. TSK FRBS was analysed in detail and preliminary results showed that 4 rules for Stockholm-Arlanda Airport and 8 rules for Zurich Airport were the most promising settings. All experiments were based on 10-fold cross-validation if not otherwise stated. This is suggested to be the recommended setting and leads to relative low bias and variance [30,31]. Furthermore, 15 repetitions were done for each experiment, as recommended by Nadeau and Bengio [31]. They also recommend using the corrected resample t-test to test whether the difference between the two prediction models is significant. The corrected resample t-test should be preferred over a normal paired t-test, because the suggested test adjusts the variance in relation to the overlaps between subsets of the data [19]. To compare the different models the same seeds were used to generate the subsets for cross-validation for the different repetitions for both utilised software packages. The significance level α was set to 0.05.

4.2. Performance measures

This research aims to compare the different prediction methods using various performance measures to provide as much insight as possible and to enable further comparisons. The utilised measures are discussed below.

4.2.1. Root mean-squared error

The root mean-squared error (RMSE) is a very commonly used measure and was preferred over the mean-squared error in this paper since it gives values in the same dimension as the predicted values. The formula is

$$\text{RMSE} = \sqrt{\frac{(y_1 - \hat{y}_1)^2 + \ldots + (y_n - \hat{y}_n)^2}{n}}, \tag{7}$$

where y_i is the actual taxi time of aircraft i and $\hat{y}_i$ is the corresponding predicted taxi time. The value n represents the number of aircraft in the dataset.

4.2.2. Mean-absolute error

A second measure, which is also in the same dimension as the predicted value, is the mean-absolute error (MAE). This performance measure averages the individual errors by neglecting their sign. It is defined as follows:

$$\text{MAE} = \frac{|y_1 - \hat{y}_1| + \ldots + |y_n - \hat{y}_n|}{n}. \tag{8}$$

4.2.3. Root relative-squared error

In the root relative-squared error (RRSE) the total squared errors are divided by the total squared errors when using the simplest prediction model (which just outputs the average value $\bar{y}$) as can be seen below:

$$\text{RRSE} = \sqrt{\frac{(y_1 - \hat{y}_1)^2 + \ldots + (y_n - \hat{y}_n)^2}{(y_1 - \bar{y})^2 + \ldots + (y_n - \bar{y})^2}}. \tag{9}$$

4.2.4. Relative-absolute error

The relative-absolute error (RAE) is the total absolute error, which is normalised as in the root relative-squared error by using a simple average predictor:

$$\text{RAE} = \frac{|y_1 - \hat{y}_1| + \ldots + |y_n - \hat{y}_n|}{|y_1 - \bar{y}| + \ldots + |y_n - \bar{y}|}. \tag{10}$$

4.2.5. Coefficient of determination

A commonly used performance measure related to linear regression is the coefficient of determination R^2. It can be determined from the root relative-squared error and takes values between 0 and 1 for linear regression models, with values closer to 1 indicating a better fit. R^2 is defined as follows:

$$R^2 = 1 - \frac{(y_1 - \hat{y}_1)^2 + \ldots + (y_n - \hat{y}_n)^2}{(y_1 - \bar{y})^2 + \ldots + (y_n - \bar{y})^2}. \tag{11}$$

Sometimes an adjusted coefficient of determination is used, which penalises models with many explanatory variables. In this study, the explanatory variables are fixed and the size of the datasets is much larger than the number of explanatory variables, making such a correction term unnecessary.

4.2.6. Prediction accuracy

The last set of performance measures is used to show practitioners the accuracy of the models and is of a form that they will be familiar with. The percentage of the prediction accuracy measure indicates what percentage of the flights in the dataset are predicted within ± 1, 2, 3, 5 or 10 min.

4.3. Visual comparisons

Fig. 2 shows the predication accuracy of the 6 different regression approaches for Zurich Airport. The x-axis represents the aircraft, which were sorted from underestimated to overestimated taxi times within each approach. The analysis is based on 15 repetitions of 10-fold cross-validation. The range on the y-axis is only shown within the interval of ± 5 min. To halve the height of the graph and make the differences more obvious, the absolute error is shown, so the left hand side shows under predictions and the right hand side over predictions. The solid black line visualises the multiple linear regression approach (LinReg) which is used as a baseline analysis. It is clear that least median square linear regression (LMS) performs poorly for predictions which underestimate the actual taxi time. Support vector regression (SMOreg), Mamdani FRBS and TSK FRBS seem to perform the best, but it is hard to distinguish clearly based on this figure, so a numerical comparison will now be presented.
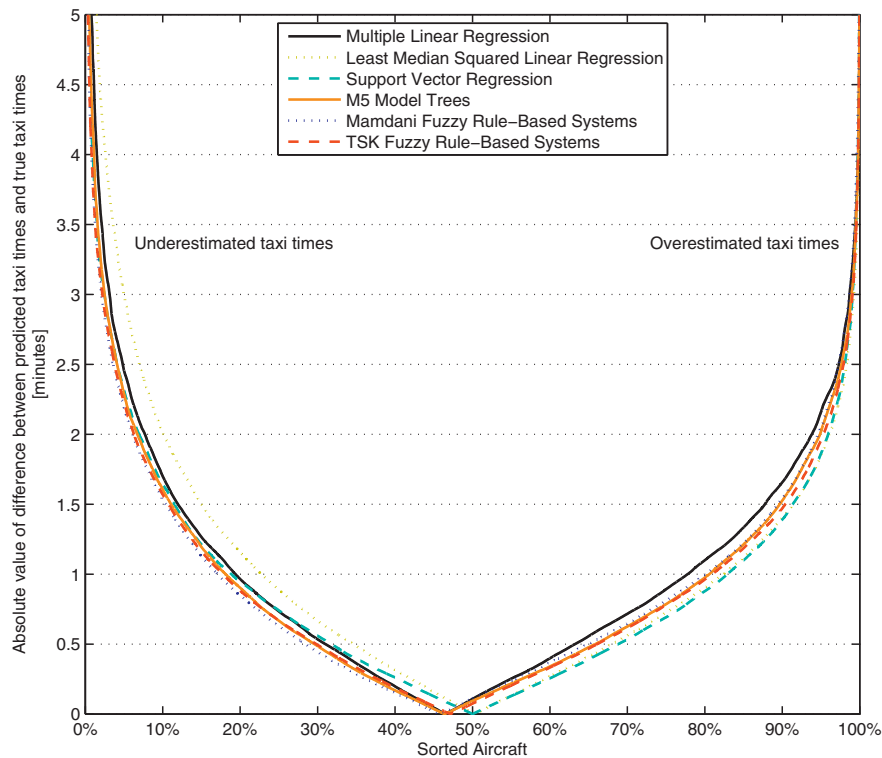
Fig. 2. Taxi time prediction accuracy at Zurich Airport.

4.4. Numeric comparisons

Table 2 shows the first four performance measures for both airport datasets. Bold numbers highlight the best (smallest) result for each performance measure at each airport. The newly introduced TSK FRBS outperforms the other approaches in almost all cases. Only in the case of Zurich Airport does support vector regression have the same result for the mean-absolute error and is slightly better in terms of the relative-absolute error. Tests with the corrected resample *t*-test showed that there is always a significant improvement between the multiple linear regression approach and TSK FRBS at Zurich Airport. For this dataset, TSK FRBS also significantly outperformed least median square linear regression and, apart from the relative-absolute error, also outperform the M5 model trees. Although the numeric results are better for the TSK FRBS, it only significantly outperformed the Mamdani approach in terms of the root mean-square error and the root relative-squared error and did not outperform the support vector regression. The results for Stockholm-Arlanda Airport are very similar, but fewer tests identify significant differences. The best values found for the coefficient of determination R^2 were 80.85% and 93.25% for Stockholm-Arlanda Airport and Zurich Airport, respectively, and were both obtained using the TSK FRBS.

The prediction accuracy within ±1, 2, 3, 5 and 10 min can be found in Table 3. Again, TSK FRBS outperformed, in most cases, the other regression approaches in both datasets. Mamdani FRBS also did very well in comparison to the others. Our findings are in line with the analysis by Wu et al. [32] where TSK FRBSs with fewer rules were more successful than Mamdani FRBSs. Although support vector regression often seemed a good alternative, the ±2 and 3 min accuracy in the case of Stockholm-Arlanda Airport reported the worst values.

4.5. Insights from prediction models

As the comparisons in the last section showed, TSK FRBS seems to perform the best for the analysed prediction problem. Therefore, this section focuses on the results of that particular approach. Earlier results and analysis for multiple linear regression and Mamdani FRBS can be found in the papers by Ravizza et al. [15] and by Chen et al. [16], respectively.

Fig. 3 illustrates the membership functions of two explanatory variables for the TSK FRBS at Stockholm-Arlanda Airport. The first figure is related to the indication of whether an aircraft is arriving or departing and the second figure shows the final model related to the logarithmic transformation of the total taxi distance. Input values are scaled to the range [−1,1] using a linear scaling from their range (see Table 1). The other explanatory variables had less distinct membership functions for the different rules and were omitted due to space limitations. This model is based upon using the approach with the entire dataset as training data. The four rules are represented with different lines, showing that rules 1 and 2 are more focusing on departures and rules 3 and 4 on arrivals. Furthermore, the rules cover different total taxi distances starting with rule 3 for smaller distances, following by rules 2, 4 and 1. The consequence part of the four rules can be seen in Table 4. Each rule has the form of a multiple linear regression approach with the coefficients stated in the table.

TSK FRBS has the ability to model non-linear relationships, which gives such a method an advantage over multiple linear regression, least median squared linear regression or M5 model trees, where automatic transformations of linear functions with a polynomial transformation are only possible to a certain extent [33]. Fig. 4 shows how the functions look for the TSK FRBS and multiple linear regression on the example of the total taxi distance

Table 2
Comparisons of performance measures for Stockholm-Arlanda Airport and Zurich Airport.

Performance measure	Airport	LinReg	LMS	SMOreg	M5P	Mamdani	TSK
Root mean-squared error	ARN	1.52	1.57	1.50	1.51	1.46	**1.44**
	ZRH	1.47	1.60	1.32	1.36	1.33	**1.30**
Mean-absolute error	ARN	1.14	1.14	1.09	1.13	1.07	**1.06**
	ZRH	1.08	1.10	**0.96**	0.99	0.97	**0.96**
Root relative-squared error	ARN	45.70%	47.47%	45.27%	45.54%	44.19%	**43.53%**
	ZRH	29.29%	31.76%	26.28%	27.00%	26.41%	**25.89%**
Relative-absolute error	ARN	45.80%	45.92%	43.98%	45.61%	43.28%	**42.83%**
	ZRH	26.85%	27.52%	**23.87%**	24.55%	24.30%	23.93%

Table 3
Comparisons of accuracies for Stockholm-Arlanda Airport and Zurich Airport.

Accuracy	Airport	LinReg	LMS	SMOreg	M5P	Mamdani	TSK
Accuracy within ± 1 min	ARN	54.28%	56.02%	57.86%	54.61%	58.21%	**58.80%**
	ZRH	58.38%	59.66%	**64.05%**	62.49%	62.97%	63.33%
Accuracy within ± 2 min	ARN	85.30%	85.18%	84.91%	85.19%	86.73%	**86.81%**
	ZRH	86.12%	85.99%	88.98%	88.15%	88.55%	**89.07%**
Accuracy within ± 3 min	ARN	95.40%	94.80%	94.32%	95.43%	95.72%	**96.16%**
	ZRH	95.55%	94.26%	96.60%	96.46%	96.54%	**96.89%**
Accuracy within ± 5 min	ARN	99.16%	98.81%	99.16%	**99.18%**	98.97%	99.08%
	ZRH	99.21%	98.56%	99.45%	99.46%	99.53%	**99.62%**
Accuracy within ± 10 min	ARN	**99.92%**	**99.92%**	**99.92%**	**99.92%**	**99.92%**	**99.92%**
	ZRH	99.92%	99.87%	99.97%	99.97%	**99.98%**	99.97%

Table 4
Consequence part of the TSK FRBS analysis for Stockholm-Arlanda Airport.

	Rule 1	Rule 2	Rule 3	Rule 4
(Constant)	0.793	−0.146	−0.964	1.666
ARR	0.772	0.339	−1.452	−1.080
Distance2	0.082	−0.208	−0.587	−0.224
$\log_{10}(Distance)$	−0.200	0.091	−0.871	−0.368
$Q_{DEP,\#DEP}$	0.594	2.891	−1.554	−2.144
$Q_{DEP,\#ARR}$	0.550	−6.788	10.588	5.890
$Q_{ARR,\#DEP}$	−0.415	−0.569	1.118	1.042
$Q_{ARR,\#ARR}$	0.041	3.828	−1.002	−0.080
$N_{DEP,\#DEP}$	−0.019	−0.273	−1.767	0.575
$N_{DEP,\#ARR}$	−0.127	5.007	−7.334	−4.586
$N_{ARR,\#DEP}$	−0.169	0.455	−0.472	−0.038
$N_{ARR,\#ARR}$	−0.139	−2.314	0.906	0.089

at Stockholm-Arlanda Airport. Two functions are plotted for each approach, one showing the predicted taxi times for departing aircraft (solid lines) and one for arriving aircraft (dashed lines). In the case of multiple linear regression, the two functions always differ by the same taxi time, which is equal to the coefficient of the explanatory variable indicating the difference between arrivals and departures of the multiple linear regression model. The two

functions have a logarithmic shape due to the transformation of the explanatory variable related to distance. On the other hand, the shapes of the two functions of the TSK FRBS can differ and can model non-linear behaviour. This is demonstrated in Fig. 4, where the model only consists of four fuzzy rules.

Fig. 5 shows the predicted taxi times of arriving aircraft at Stockholm-Arlanda Airport with the TSK FRBS. One axis again represents the different total taxi distances and the other axis shows the amount of traffic at the airport. For this analysis, the Q and N values counting other arrivals are indicated on the axis while the Q and N values counting other departures were set to 1 for the entire analysis. An increase in the total taxi distance increases the predicted taxi time and the curve flattens for longer distances. Analysing the amount of traffic on the surface shows that the slope of the curve grows with higher Q and N values. In addition, the influence of the total taxi distance decreases with more traffic on the surface. Again, it should be highlighted that such behaviours cannot be modelled with a linear regression approach, suggesting that one reason for TSK FRBS outperforming the other approaches is due to its ability to model complex non-linear systems.

The model for the TSK FRBS for Zurich Airport is shown in Fig. 6 in the same way as in Fig. 3. As mentioned earlier, the best results
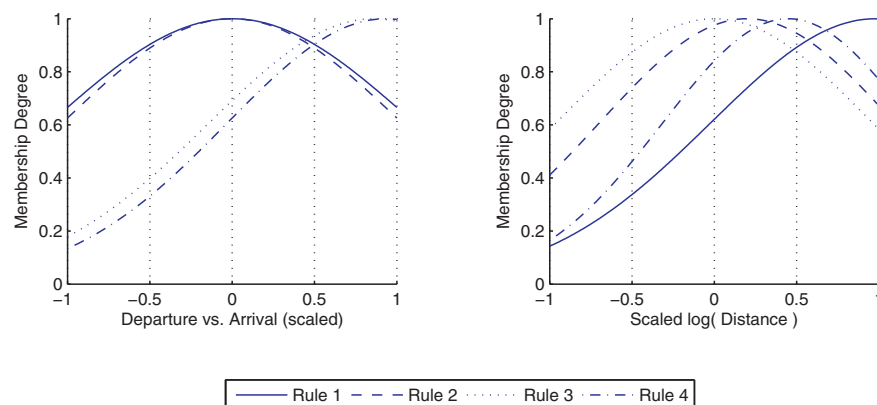

Fig. 3. Four fuzzy rules extracted from the TSK FRBS analysis for Stockholm-Arlanda Airport.
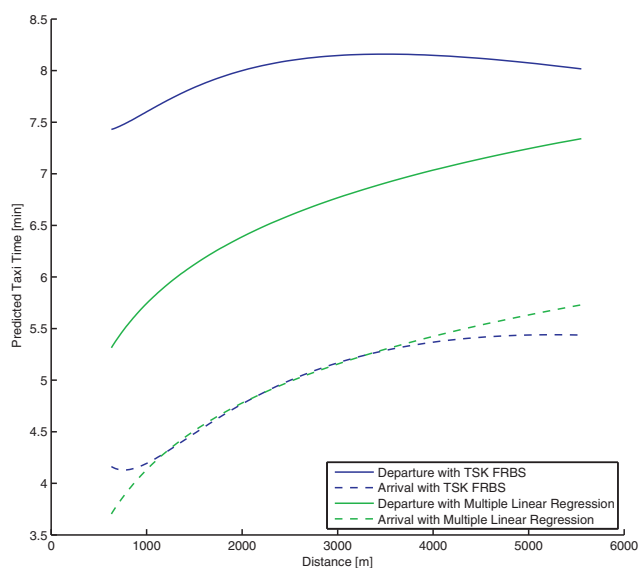
Fig. 4. Non-linearity of TSK FRBS models shown on data from Stockholm-Arlanda Airport.

Fig. 5. Analysis of predicted taxi-in times at Stockholm-Arlanda Airport with the TSK FRBS.

at Zurich Airport were found with eight fuzzy rules. The most interesting membership functions were for the indication of arrivals and departures and the logarithmic transformation of the total taxi
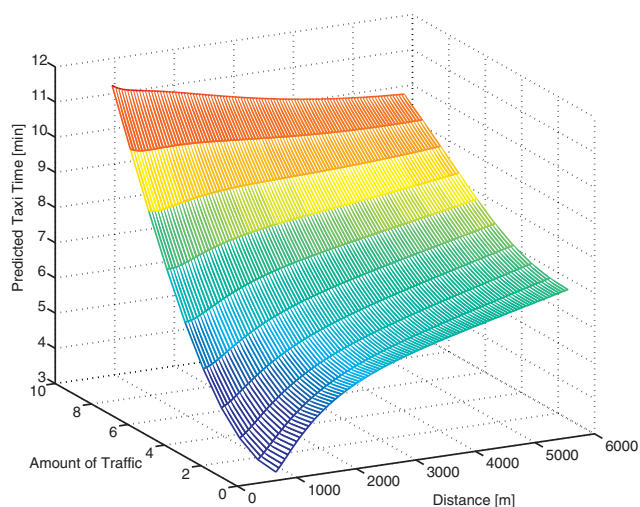
distance as it was before for Stockholm-Arlanda Airport. In addition, the logarithmic transformation of the total angles and the three operational modes also have distinct functions. The corresponding multiple linear regression models for the eight different fuzzy rules can be found in Table 5.
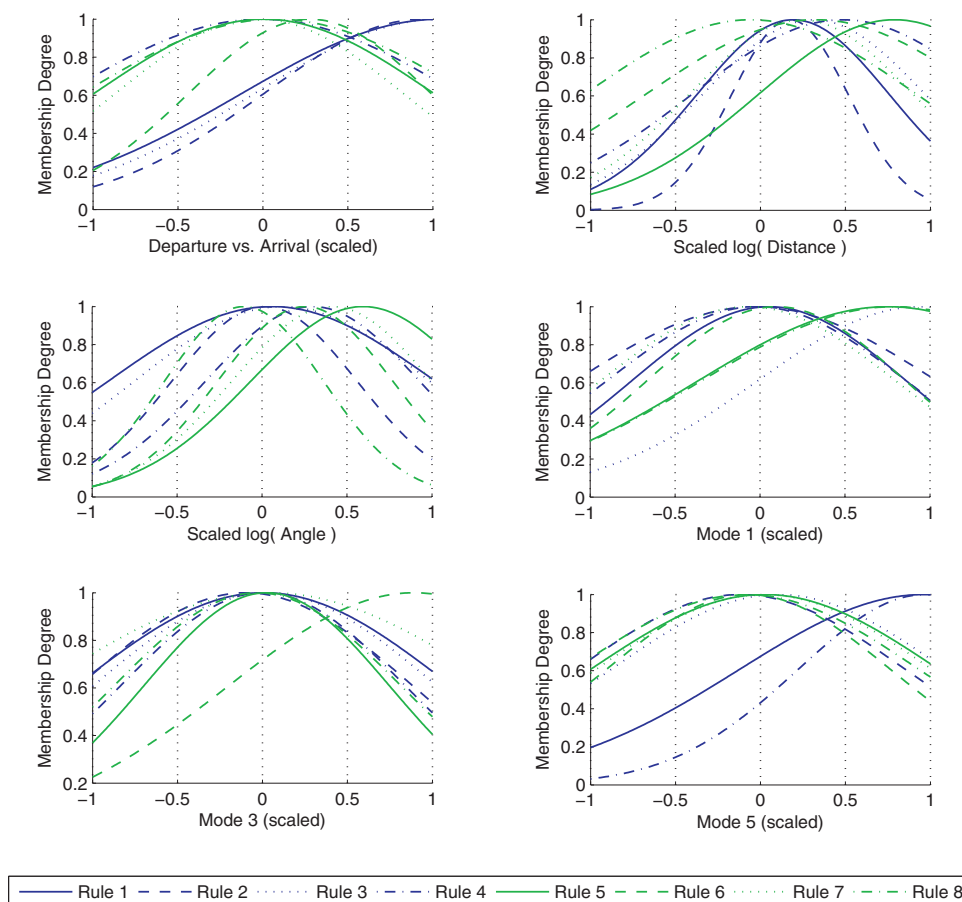

Fig. 6. Eight fuzzy rules extracted from the TSK FRBS analysis for Zurich Airport.

Table 5
Consequence part of the TSK FRBS analysis for Zurich Airport.

	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8
(Constant)	4.983	−4.299	0.424	−19.232	0.923	−0.203	3.008	2.233
ARR	2.608	0.398	−0.273	−2.787	−0.037	−0.062	−0.334	−0.169
Distance	−1.294	−2.751	0.466	2.004	0.954	−2.952	−0.342	1.694
$\log_{10}$(Distance)	0.374	2.778	−0.321	−1.031	−2.494	3.096	0.280	−0.582
$\log_{10}$(Angle)	1.598	−0.443	0.060	−0.579	0.688	0.155	−0.603	−0.270
$Q_{DEP,\#DEP}$	−0.215	−1.108	0.825	1.096	0.294	−2.289	1.952	0.936
$Q_{DEP,\#ARR}$	7.583	0.063	0.232	−1.360	0.039	0.528	1.633	0.827
$Q_{ARR,\#DEP}$	−0.028	−0.247	0.168	−1.639	0.025	−0.941	0.545	0.887
$Q_{ARR,\#ARR}$	1.422	0.351	0.009	−0.333	1.023	1.772	−0.448	1.069
$N_{DEP,\#DEP}$	1.727	1.229	−0.707	−0.176	0.283	−0.763	−0.367	0.194
$N_{DEP,\#ARR}$	0.764	−0.213	−0.240	−0.377	0.074	1.022	−0.379	−0.077
$N_{ARR,\#DEP}$	0.296	0.001	0.061	1.111	−0.955	−0.818	0.542	−0.369
$N_{ARR,\#ARR}$	−0.643	−0.810	0.088	−1.055	−1.326	−0.508	0.568	−1.157
Mode 1	1.072	0.214	−0.034	5.916	−0.797	−2.149	0.201	−0.166
Mode 3	−2.264	−0.317	0.154	2.814	0.122	−2.580	−0.099	0.074
Mode 5	1.342	−0.035	0.056	21.321	0.251	1.665	0.086	0.127

5. Conclusions

Airports commonly have many objectives and most will already be automating some of their operations or decision support, or be planning to do so in the future. Some of the aims are to improve predictability, improve on-time performance, reduce ground movement costs, enhance the use of ground handling resources, stands, gates and terminals and reduce apron and taxiway congestion [34]. Key to such improvements are better taxi time predictions, which can help many different decision support systems. This research should also help the development of new and more accurate decision support systems for the ground movement problem.

A recent study [15] focused on finding the explanatory variables for taxi time prediction for both arrivals and departures, using multiple linear regression to highlight their statistical significance. This paper uses the same explanatory variables and shows an extensive analysis of different regression approaches for predicting taxi times at airports to demonstrate the performance of each. Six different approaches were analysed in detail: multiple linear regression, least median squared linear regression, support vector regression, M5 model trees, Mamdani fuzzy rule-based systems and TSK fuzzy rule-based systems. The latter outperformed the other approaches on datasets from two European hub airports. TSK fuzzy rule-based systems use fuzzy membership functions to subdivide the input space in the premise part and a weighted sum of multiple linear regression approaches in the consequent part. As the different fuzzy rules work cooperatively, in contrast to approaches such as M5 model trees, the approach may potentially give more accurate estimates and can also model non-linear patterns in the data. Furthermore, this paper gave insights into the different rules found by the TSK fuzzy rule-based system and considered taxi-in times, which seems to be a less understood problem in this field.

It would be interesting to also compare these regression approaches for other airports to see whether these findings can be extended to the busiest airports in the world, and particularly to US airports, where the airport operations are often managed differently or are operated under differing constraints. In addition, this research could be integrated into decision support systems which help controllers in the towers, followed by a fine-tuning phase of the models and the decision support systems to provide more valuable decision-making aids.

References

[1] European Commission Flightpath 2050, Europe's vision for aviation. Report of the High Level Group on Aviation Research; 2011. http://dx.doi.org/10.2777/50266

[2] J.A.D. Atkin, E.K. Burke, S. Ravizza, The airport ground movement problem: past and current research and future directions, in: Proceedings of the 4th International Conference on Research in Air Transportation (ICRAT), Budapest, Hungary, 2010, pp. 131–138.

[3] J.A.D. Atkin, E.K. Burke, J.S. Greenwood, D. Reeson, On-line decision support for take-off runway scheduling with uncertain taxi times at London Heathrow Airport, Journal of Scheduling 11 (5) (2008) 323–346, http://dx.doi.org/10.1007/s10951-008-0065-9.

[4] J.A.D. Atkin, G. De Maere, E.K. Burke, J.S. Greenwood, Addressing the pushback time allocation problem at Heathrow Airport, Transportation Science (2012), available first online, http://dx.doi.org/10.1287/trsc.1120.0446.

[5] J.A.D. Atkin, E.K. Burke, S. Ravizza, A more realistic approach for airport ground movement optimisation with stand holding, in: Proceedings of the 5th Multidisciplinary International Scheduling Conference (MISTA), Phoenix, Arizona, USA, 2011.

[6] H.R. Idris, J.P. Clarke, R. Bhuva, L. Kang, Queuing model for taxi-out time estimation, Air Traffic Control Quarterly 10 (1) (2002) 1–22.

[7] P. Balakrishna, R. Ganesan, L. Sherry, B. Levy, Estimating taxi-out times with a reinforcement learning algorithm, in: Proceedings of the IEEE/AIAA 27th Digital Avionics Systems Conference (DASC), 2008, http://dx.doi.org/10.1109/DASC.2008.4702812.

[8] P. Balakrishna, R. Ganesan, L. Sherry, Application of reinforcement learning algorithms for predicting taxi-out times, in: Proceedings of the 8th ATM R&D Seminars, Napa, USA, 2009.

[9] P. Balakrishna, R. Ganesan, L. Sherry, Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: a case-study of Tampa Bay departures, Transportation Research Part C: Emerging Technologies 18 (6) (2010) 950–962, http://dx.doi.org/10.1016/j.trc.2010.03.003.

[10] R. Clewlow, I. Simaiakis, H. Balakrishnan, Impact of arrivals on departure taxi operations at airports, in: Proceedings of the AIAA Guidance, Navigation, and Control Conference, 2010.

[11] R. Ganesan, P. Balakrishna, L. Sherry, Improving quality of prediction in highly dynamic environments using approximate dynamic programming, Quality and Reliability Engineering International 26 (7) (2010) 717–732, http://dx.doi.org/10.1002/qre.1127.

[12] Y. Zhang, A. Chauhan, X. Chen, Modeling and predicting taxi out times, in: Proceedings of the 4th International Conference on Research in Air Transportation (ICRAT), Budapest, Hungary, 2010, pp. 31–35.

[13] R. Jordan, M. Ishutkina, T. Reynolds, A statistical learning approach to the modeling of aircraft taxi time, in: Proceedings of the IEEE/AIAA 29th Digital Avionics Systems Conference (DASC), 2010, http://dx.doi.org/10.1109/DASC.2010.5655532.

[14] M.S. Kistler, G. Gupta, Relationship between airport efficiency and surface traffic, in: Proceedings of 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), Hilton Head, SC, USA, 2009.

[15] S. Ravizza, J.A.D. Atkin, M.H. Maathuis, E.K. Burke, A statistical approach for improving taxi time estimations at airports, Journal of the Operational Research Society 64 (2013) 1347–1360, http://dx.doi.org/10.1057/jors.2012.123.

[16] J. Chen, S. Ravizza, J.A.D. Atkin, P. Stewart, On the utilisation of fuzzy rule-based systems for taxi time estimations at airports, in: Proceedings of the 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS), Saarbrücken, Germany, 2011, vol. 20

of OpenAccess Series in Informatics (OASIcs). ISBN 978-3-939897-33-0; pp. 134–145. doi:10.4230/OASIcs.ATMOS.2011.134.

[17] J.A.D. Atkin, E.K. Burke, J.S. Greenwood, D. Reeson, Hybrid metaheuristics to aid runway scheduling at London Heathrow Airport, Transportation Science 41 (1) (2007) 90–106, http://dx.doi.org/10.1287/trsc.1060.0163.

[18] U. Dorndorf, A. Drexl, Y. Nikulin, E. Pesch, Flight gate scheduling: state-of-the-art and recent developments, Omega 35 (3) (2007) 326–334, http://dx.doi.org/10.1016/j.omega.2005.07.001.

[19] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10–18, http://dx.doi.org/10.1145/1656274.1656278.

[21] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques. The Morgan Kaufmann Series in Data Management Systems, 3rd ed., Elsevier Science, 2011, ISBN: 9780123748560.

[22] D.C. Montgomery, E.A. Peck, G.G. Vining, Introduction to Linear Regression Analysis, 3rd ed., John Wiley & Sons, Inc, 2001.

[23] P. Rousseeuw, A. Leroy, Robust Regression and Outlier Detection. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, Wiley, 1987, ISBN: 9780471852339.

[24] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing 14 (2004) 199–222, http://dx.doi.org/10.1023/B:STCO.0000035301.49549.88.

[25] J.R. Quinlan, Learning with continuous classes, in: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, 1992.

[26] L. Zadeh, Fuzzy sets, Information and Control 8 (3) (1965) 338–353, http://dx.doi.org/10.1016/S0019-9958(65)90241-X.

[27] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Transactions on Systems, Man and Cybernetics 15 (1) (1985) 116–132.

[28] J. Chen, Biological inspired optimisation algorithms for transparent knowledge extraction allied to engineering materials process, Department of Automatic Control and Systems Engineering, The University of Sheffield, 2009 (Ph.D. thesis).

[29] K. Deb, A. Anand, D. Joshi, A computationally efficient evolutionary algorithm for real-parameter optimization, Evolutionary computation 10 (4) (2002) 371–395, http://dx.doi.org/10.1162/106365602760972767.

[30] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems, 3rd ed., Elsevier Science, 2011, ISBN: 9780123814791.

[31] C. Nadeau, Y. Bengio, Inference for the generalization error, Machine Learning 52 (2003) 239–281, http://dx.doi.org/10.1023/A:1024068626366.

[32] Y. Wu, B. Zhang, J. Lu, K.L. Du, Fuzzy logic and neuro-fuzzy systems: a systematic introduction, International Journal of Artificial Intelligence and Expert Systems 2 (2) (2011) 47–80.

[33] G.E.P. Box, D.R. Cox, An analysis of transformations, Journal of the Royal Statistical Society Series B (Methodological) 26 (2) (1964) 211–252.

[34] Eurocontrol. Airport CDM implementation manual. Tech. Rep. version 4; Eurocontrol; 2012.