

 Open access • Journal Article • DOI:10.1057/JORS.2012.123

A combined statistical approach and ground movement model for improving taxi time estimations at airports — [Source link](#)

Stefan Ravizza, Jason A. D. Atkin, Marloes H. Maathuis, Edmund K. Burke

Institutions: University of Nottingham, ETH Zurich, University of Stirling

Published on: 01 Sep 2013 - Journal of the Operational Research Society (Palgrave Macmillan UK)

Related papers:

- [Queuing Model for Taxi-Out Time Estimation](#)
- [Aircraft taxi time prediction: Comparisons and insights](#)
- [Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures](#)
- [Optimization of Taxiway Routing and Runway Scheduling](#)
- [The Airport Ground Movement Problem: Past and Current Research and Future Directions](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/a-combined-statistical-approach-and-ground-movement-model-3slisfuvpm>



A combined statistical approach and ground movement model for improving taxi time estimations at airports

S Ravizza^{1*}, JAD Atkin¹, MH Maathuis² and EK Burke³

¹University of Nottingham, Jubilee Campus, Nottingham, UK; ²ETH Zurich, Zurich, Switzerland; and

³University of Stirling, Cottrell Building, Stirling, UK

With the expected continued increases in air transportation, the mitigation of the consequent delays and environmental effects is becoming more and more important, requiring increasingly sophisticated approaches for airside airport operations. Improved on-stand time predictions (for improved resource allocation at the stands) and take-off time predictions (for improved airport-airspace coordination) both require more accurate taxi time predictions, as do the increasingly sophisticated ground movement models which are being developed. Calibrating such models requires historic data showing how long aircraft will actually take to move around the airport, but recorded data usually includes significant delays due to contention between aircraft. This research was motivated by the need to both predict taxi times and to quantify and eliminate the effects of airport load from historic taxi time data, since delays and re-routing are usually explicitly considered in ground movement models. A prediction model is presented here that combines both airport layout and historic taxi time information within a multiple linear regression analysis, identifying the most relevant factors affecting the variability of taxi times for both arrivals and departures. The promising results for two different European hub airports are compared against previous results for US airports.

Journal of the Operational Research Society (2013) 64, 1347–1360. doi:10.1057/jors.2012.123

Published online 3 October 2012

Keywords: airport operations; ground movement; taxi time prediction; real-world scheduling support; statistical analysis; multiple linear regression

1. Introduction

Airlines and airports face several key challenges in the near future. Firstly, the number of flights is predicted to increase in the next few years (SESAR, 2006). Secondly, there is an increasing focus upon environmental considerations, and this is likely to increase in importance. Thirdly, the use of computerised tools is enabling increased aircraft utilisation, reduced idle times, and increased passenger connection options, leading to ever more complex and interlinked flight schedules. The on-time performance of flights at each airport and the earlier visibility of any delays (allowing corrective measures to be put into place) are becoming increasingly important, since many downstream flights can be affected by delays to single aircraft. Consequently, the operations at busy hub airports are experiencing an increased focus of attention, and this is likely to increase in the face of future challenges.

Total taxi times from stand/gate to runway are needed if advance predictions of take-off times are required, for use by en-route controllers (or decision support systems to help them) or for improving arrival time predictions for the destination airports, allowing the effects of any predicted delays to be mitigated. Taxi times are already needed by several existing search algorithms for take-off time prediction and take-off sequencing (Atkin *et al*, 2007; Eurocontrol, 2010) and for allocating appropriate stand holds to aircraft to absorb ground delay at the gate/stand, decreasing the fuel burn and environmental effects (Burgain *et al*, 2009; Atkin *et al*, 2010a). Although the effects have been less well studied, taxi times are also useful for arrivals, being necessary for predicting stand/gate arrival times, to ensure that adequate resources are available at the correct time (Eurocontrol, 2010). Taxi time predictions will become even more important if the efficiency of stand resource utilisation is to be improved in future. Current common practice is to use standard mean taxi times for each taxi source/destination pairs. A better understanding of the influencing factors, and a model to estimate such taxi times to a higher level of accuracy,

*Correspondence: S Ravizza, School of Computer Science, The University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, Nottinghamshire NG8 1BB, UK.

would have positive effects for both the published approaches and the systems which are currently in use.

The importance of the ground movement problem was explained in Atkin *et al* (2010b), highlighting how it links several other airport operations such as runway sequencing and gate assignment. Improved ground movement can increase on-time performance at airports, so ground movement simulations and optimisers are extremely useful. These usually explicitly model the interaction between aircraft (modelling delays due to other aircraft and any necessary re-routing on longer paths to avoid conflicts) and, thus, require predictions for taxi times which do not already include these elements (Gotteland and Durand, 2003; Smeltink *et al*, 2004; Balakrishnan and Jung, 2007; Roling and Visser, 2008; Lesire, 2010). The use of historic data would be preferable for calibrating models. However, such recorded data usually includes significant delays due to the interactions between aircraft. There are obvious benefits from being able to quantify the effects of this interaction and the model which is considered in this paper aims to provide this facility. Although average speeds have often had to be used in the past due to the lack of reliable predictions, it is important to understand aircraft speed in more detail if more realistic ground movement decision support systems are desired.

The causes and effects of taxi time variability are both often neglected. However, some elements have been considered in the past. Rappaport *et al* (2009) analysed the effect on taxi times of having to reduce speed for turns and it was shown that aircraft travelling straight forward reached higher average speeds than those with upcoming turns. In addition, Idris *et al* (2002) performed a statistical analysis of departing aircraft at Boston Logan International Airport with the conclusion that the taxi-out time for each airline/runway configuration combination was highly dependent upon the take-off queue size. However, the analysis by Idris *et al* (2002) only covered taxi times for departing aircraft. The problem also seems to differ between North American and European airports, with much shorter take-off queues usually being observed at European hubs. More recently, two further estimation approaches were published for North American airports. Simaiakis and Balakrishnan (2009) presented a queuing model and potential impact on emissions reduction. The statistical analysis exclusively used the size of the take-off queue to estimate the taxi-out time. Balakrishna *et al* (2009) presented a model for taxi-out time prediction based on reinforcement learning algorithms. In other work, Tu *et al* (2008) analysed push-back delays at Denver International Airport with seasonal trends and daily propagation patterns.

The aim of this paper is to extensively study the variation of taxi times not only for departing aircraft, but also for arriving aircraft. In contrast to earlier studies, we focus on European hub airports in this paper where the

taxi process is less dominated by queuing and hence other factors have a proportionately greater effect upon taxi times. The use of the airport layout is essential for this research and was not considered in the past. The outcomes will enable researchers to make increasingly accurate taxi time predictions and to develop more realistic decision support systems for ground movement, potentially resulting in smoother airport operations, emission reductions for the taxi process and better on-time performance at airports.

The remainder of this paper is structured as follows: Section 2 provides a description of the problem and the available data. The statistical taxi time prediction method is then detailed in Section 3, where the influence of the ground movement model will be observed. The results and their applications are discussed in Sections 4 and 5, respectively, and the paper ends by drawing important conclusions from this work in Section 6.

2. Problem description

The problem considered in this paper involves the identification of a function to estimate taxi times for both arriving and departing aircraft, which can then be used in an airport decision support system. The problem description in this section has two parts. Firstly, we summarise the airport ground movement problem, explaining why accurate taxi times are very important. Secondly, we discuss the data which we can expect to be available from an airport for use in calibrating ground speed models.

2.1. The airport ground movement problem

This research was motivated by our work on the airport ground movement problem (Atkin *et al*, 2011a; Ravizza and Atkin, 2011), which is basically a routing and scheduling problem. It involves directing aircraft on the surface of an airport to their destinations in a timely manner, with the aim usually being to reduce the overall travel time, to meet some target time windows and/or to absorb the delay at a preferred time, such as when the engines are not running. It is crucial, for reasons of safety, that two aircraft never conflict with each other throughout the ground movement.

For larger airports, especially during peak hours, decision support systems are advantageous to deal with the complexity of the problem (Gotteland and Durand, 2003; Smeltink *et al*, 2004; Balakrishnan and Jung, 2007; Roling and Visser, 2008; Lesire, 2010). Sophisticated algorithms are needed to route and schedule all the aircraft simultaneously on the surface. In doing so, some aircraft might be allocated to a longer route and/or waiting times might need to be added to some schedules to handle conflicts, aiming for a globally better solution.

A detailed survey was recently published showing the state-of-the-art in this research area (Atkin *et al.*, 2010b). For the purpose of this paper, the important feature of this problem is that decision support systems need taxi time predictions for aircraft in isolation, ignoring the presence of other aircraft, but historic data is rarely able to provide this information. However, it is clear that the use of historic data is vital in order to ensure that results are realistic and can be compared with the *status quo* at an airport, in order to quantify any potential improvements from new airport ground movement decision support systems, without running expensive trials.

2.2. Available airport data

This analysis utilised data from two hub airports in Europe: Stockholm-Arlanda Airport (ARN), the largest airport in Sweden and Zurich Airport (ZRH), the largest airport in Switzerland. Both airports have a main hub carrier, Scandinavian Airlines at Stockholm-Arlanda and Swiss International Air Lines at Zurich. Sketches of the two airport layouts are provided in Figure 1.

In collaboration with colleagues at both of the airports, we had access to the data for an entire day's operation: for the 7th of September 2010 at Stockholm-Arlanda (661 movements) and the 19th of October 2007 at Zurich (679 movements). Both data sets represent days with no extraordinary occurrences to be taken into account. The main elements of the supplied data consisted of information about each aircraft, detailing the stand, the runway, the start and end time of taxiing, the aircraft type and whether the aircraft was an arrival or a departure.

In visually analysing the average taxi speeds, it was obvious that there were major differences between different groups of aircraft. A boxplot is presented in Figure 2, showing the general variability in the average speed of the

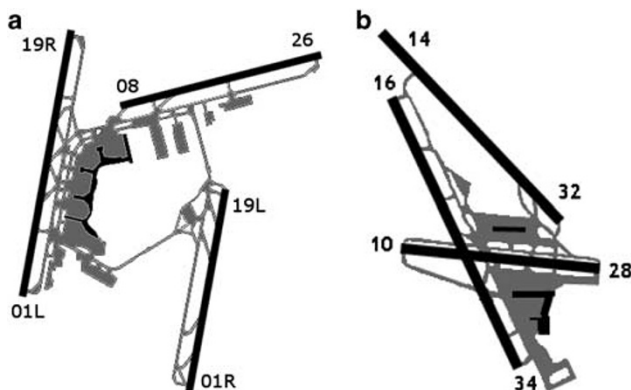


Figure 1 Sketch of airport layouts where both airports operate with three runways: (a) Stockholm-Arlanda Airport; (b) Zurich Airport.

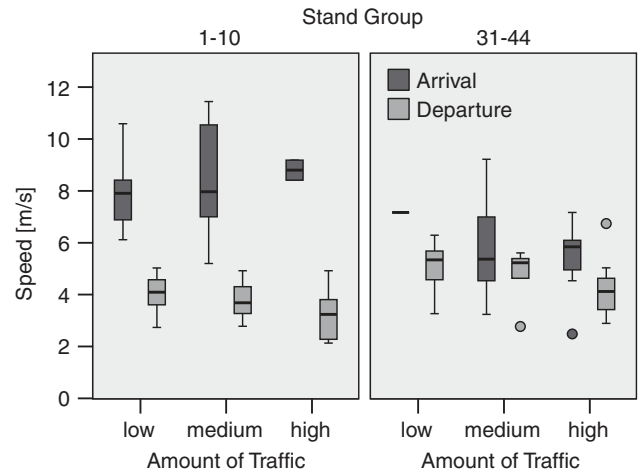


Figure 2 Average speed at Stockholm-Arlanda Airport from two different stand groups to the runway 19R.

aircraft for two stand groups at Stockholm-Arlanda Airport. Major differences are apparent between arriving and departing aircraft as well as between low, medium and high traffic situations at the airport.

3. Approach for estimating taxi speed

The aim of this research is to estimate a function which can more accurately predict taxi times for aircraft or, equivalently, better predict their average speeds. It is not obvious which factors are important for calculating such taxi times and which factors can be ignored. Discussions with practitioners can help in understanding the problem and identifying potential factors but this has its limits for mathematically determining the importance of factors. Multiple linear regression was able not only to answer this question, but also to estimate a function that could predict the taxi speed and was easy to interpret. Of course, the accuracy of the estimation has to be verified, but given such a function, the aim is to eliminate the effects of factors that represent the actual amount of traffic at the airport, by setting the respective variables to 0. Our aim is to be able to predict the taxi times for independent aircraft, for use in a more advanced ground movement decision support system. This would provide the opportunity to compare scenarios with the way in which an airport is currently operating.

3.1. Summary of multiple linear regression

A brief summary of multiple linear regression is given here for reasons of completeness, before providing the details of how it has been applied to the problem of estimating taxi times by incorporating details of the airport layout. The interested reader is directed to the book by Montgomery *et al.* (2001) for more in-depth coverage.

Multiple linear regression is a statistical approach that attempts to model the i th dependent variable y_i as a linear weighted function of other explanatory variables x_{i1}, \dots, x_{ip} and an error term ε_i . The random error terms $\varepsilon_1, \dots, \varepsilon_n$ are assumed to be uncorrelated and to have a normal distribution with mean zero and constant variance σ^2 . The regression coefficients can be estimated using least squares regression, yielding estimated coefficients $\hat{\beta}_1, \dots, \hat{\beta}_p$. The predicted y value for the i th observation is then given by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}. \quad (1)$$

The difference between y_i and \hat{y}_i is called the residual, $e_i = y_i - \hat{y}_i$.

The adjusted coefficient of determination R_{Adj}^2 can be used to measure how well the model fits the data. It is defined as follows:

$$R_{\text{Adj}}^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2 / (n - p - 1)}{\sum_i (y_i - \bar{y})^2 / (n - 1)}, \quad (2)$$

where \bar{y} is the mean of y_1, \dots, y_n . R_{Adj}^2 takes values between 0 and 1, with values closer to 1 indicating a better fit. The measure incorporates a trade-off between goodness of fit and the complexity of the model, favouring simpler models when possible.

3.2. Analysis of the dependent variable

It was discovered that estimations of taxi speeds (in m/s) better fit the linear requirements of the models than direct estimates of the taxi times of aircraft. Furthermore, it was also discovered that a logarithmic transformation of the dependent variable (Equation (3)) was required in order to fulfil the stated assumptions of multiple linear regression and such a transformation is used throughout the following sections:

$$y := \log_{10}(\text{Speed}). \quad (3)$$

A good estimate for $\log_{10}(\text{Speed})$ can then be used for the calculation of a good estimate of the taxi time.

3.3. Analysis with only one explanatory factor

Different individual factors are analysed in this section. The analysed factors were derived from a combination of previously published work in this area, discussions with practitioners and data-driven transformations. The factors that appeared to be statistically relevant were then included together in a combined model. For reasons of simplicity, we focus within this section only on the settings for Stockholm-Arlanda Airport, although many

results are similar for both airports, as can be observed in Section 3.4.

3.3.1. Distances. The first factor that was analysed considered the distance (in meters) that an aircraft was taxiing. To determine such distances, it was useful to model the airport ground layout as a graph, where the arcs represented the taxiways and the nodes represented the junctions or intermediate points (see Figure 3). Based on this underlying graph, it was then assumed that aircraft were travelling on their shortest path and Dijkstra's algorithm (see Cormen *et al* (2001) for more details) was used to determine, for each aircraft, the taxi distance from the stand to the runway or back again. The incorporation of the actual airport layout was essential for the approach as will be seen later. We note that further improvements may be possible from using the actual route taken, but that information was not available at the time. Further research will consider this.

Regressing $\log_{10}(\text{Speed})$ on 'Distance' yielded an adjusted coefficient of determination $R_{\text{Adj}}^2 = 0.473$, with a p -value smaller than $2.2\text{e-}16$ (the p -value comes from the F-test that compares the given model to a model with only an intercept). Figure 4(a) shows a plot of the observed values, y , against the explanatory variables, x .

The nonlinear shape in Figure 4(a) encouraged the application of a logarithmic transformation to the distance. The resulting fit can be seen in Figure 4(b), and has a better linear shape. Regressing $\log_{10}(\text{Speed})$ on $\log_{10}(\text{Distance})$ yielded an R_{Adj}^2 value of 0.479 (p -value $< 2.2\text{e-}16$), which is only marginally better, but it will be observed later that it leads to significant improvements in the final model for both airports.

The R_{Adj}^2 value indicates that almost half of the variance can be explained by this factor, showing the importance of this indicator. Therefore, additional time was invested in analysing it. Instead of only using the entire distance of an aircraft as a variable, it was divided into three different components based upon the known behaviour of aircraft as they taxi around the airport. 'Distance0' represented the length of the path directly around the gates, 'Distance2' represented the length of the path which was comprised of long sub-paths without any junctions and 'Distance1' represented the remaining distance (where all values were in meters). These distances were determined using the directed graph model of Stockholm-Arlanda Airport, by assigning each arc in the graph to one of the three distances. The 'Distance', 'Distance0', 'Distance1', 'Distance2', $\log_{10}(\text{Distance})$, $\log_{10}(\text{Distance0})$, $\log_{10}(\text{Distance1})$ and $\log_{10}(\text{Distance2})$ values were all included in the analysis. The resulting regression model yielded an improved R_{Adj}^2 value of 0.604 (p -value $< 2.2\text{e-}16$).



Figure 3 Graph representing the airport ground layout for Stockholm-Arlanda Airport.

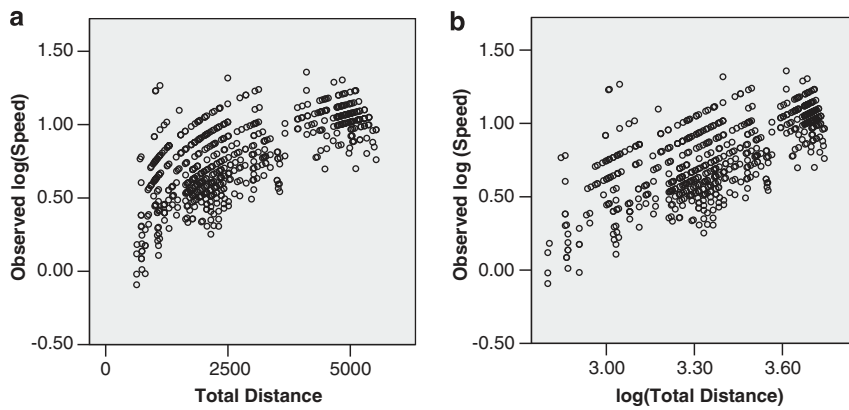


Figure 4 Scatterplots showing the logarithmic transformation: (a) Distance; (b) $\log_{10}(\text{Distance})$.

3.3.2. *Angle*. The total amount of turning that an aircraft had to achieve was another promising predictor of taxi speed, since aircraft obviously have to slow down to make turns. A factor was introduced to measure the total turning angle (in degrees), calculated as the total angular deviations between adjacent arcs on the shortest path for the aircraft. Again, the graph model of the airport layout was used for this, as shown in Figure 5. This turned out to be another major factor ($R_{\text{Adj}}^2 = 0.470$,

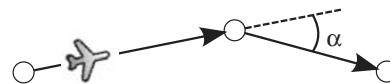


Figure 5 Measuring turning angle of aircraft on one node.

$p\text{-value} < 2.2e-16$) and the importance was improved further when $\log_{10}(\text{Angle})$ was considered ($R_{\text{Adj}}^2 = 0.482$, $p\text{-value} < 2.2e-16$).

3.3.3. *Departures versus arrivals.* As shown in Figure 2, the speed for departures can differ significantly from the speed for arrivals. In contrast to the factors that have been introduced so far, this information is nominal rather than being a continuous variable. A dummy variable called ARR was introduced, defined to be 1 for arrival aircraft and 0 for departure aircraft. The regression showed an R^2_{Adj} value of 0.380 for this single factor, demonstrating its importance (p -value $< 2.2e-16$).

3.3.4. *Amount of traffic.* Another important factor affecting the taxi speed of aircraft is the amount of traffic on the airport surface while the aircraft is taxiing. As a first attempt for an indicator of surface load, we divided the operational hours into three different categories. The indicator ‘Traffic_high’ was set to 1 for hours where more than 50 aircraft were moving and to 0 otherwise. ‘Traffic_medium’ was set to be 1 for hours with between 36 and 50 moving aircraft and 0 otherwise. Both indicators were set to zero for the last category representing low surface load (the same categorisation is visible in Figure 2). This approach with these variables resulted in an R^2_{Adj} value of only 0.007 and a p -value of 0.036.

A more advanced measure was introduced based on the paper by Idris *et al* (2002). The value N_i counts the number of other aircraft that are taxiing on the airport surface at the time that the particular aircraft i started to taxi, as shown in Equation (4), where the Iverson bracket denotes the value 1 if the condition in square brackets is satisfied and is 0 otherwise. The parameters t^i_{start} and t^i_{end} represent the time at which aircraft i starts and ends its taxi operation.

$$N_i = \sum_{j \in Aircraft \setminus \{i\}} [t^i_{start} \in (t^j_{start}, t^j_{end})] \quad (4)$$

The value Q_i was also adopted to count the number of other aircraft which cease taxiing during the time aircraft i is taxiing, as shown in Equation (5), again using the Iverson bracket.

$$Q_i = \sum_{j \in Aircraft \setminus \{i\}} [t^j_{end} \in (t^i_{start}, t^i_{end})] \quad (5)$$

Since the paper by Idris *et al* (2002) was restricted to taxi-out times, this approach was further developed to cope with separate departures and arrivals. Eight integer variables were used to allow consideration of the effects of the counts of arrivals and departures depending upon whether the current aircraft was an arrival or departure. These were named $N_{DEP, \#DEP}$, $N_{DEP, \#ARR}$, $N_{ARR, \#DEP}$, $N_{ARR, \#ARR}$, $Q_{DEP, \#DEP}$, $Q_{DEP, \#ARR}$, $Q_{ARR, \#DEP}$ and $Q_{ARR, \#ARR}$. In this notation, the N or Q indicated whether it was the count of already moving aircraft or of aircraft that ceased their movement. The first index for each value

represented the type of aircraft under consideration (*ARRival* or *DEP*arture). The second index indicated whether it was the count of arrivals or departures ($\#ARR$ or $\#DEP$) which was to be considered for counting, that is, for a departing aircraft, all of the variables with a first index *ARR* are treated as if they are 0 and for arriving aircraft all of the variables with the first index of *DEP* are treated as if they are 0.

A highly significant regression model considering only these eight factors led to an R^2_{Adj} value of 0.422 (p -value $< 2.2e-16$). Further investigation was performed to determine whether the model could be further improved by considering only aircraft destined for, or originating from, the same runway as the aircraft under consideration. In that case, the fit was worse ($R^2_{Adj} = 0.382$, p -value $< 2.2e-16$). One possible explanation for this is that often one runway is used for departures and another one for arrivals, in which case half of the factors have the same value as in the unrestricted case and the other half have the value 0, resulting in less information being considered by the model than in the unrestricted case.

3.3.5. *Less important factors.* A number of other elements were taken into consideration, for example whether the model could be improved by using the square of some of the values or by including some interaction terms but no improvement was found. Another approach was to consider the number of engines of the aircraft ($R^2_{Adj} = 0.007$, p -value = 0.039) or by using the wake vortex categorisation of the aircraft ($R^2_{Adj} = 0.032$, p -value = 4.4e-05). These results for the European airports that we studied fit the findings of Idris *et al* (2002) (for a North American airport), where a poor correlation was observed between taxi time and aircraft type, and the type determines both the number of engines and the wake vortex categorisation.

Further analysis studied the effect of the different runways and stand groups. Although nothing relevant was found for Stockholm-Arlanda Airport, some effects were found at Zurich Airport by analysing different operational modes (which runway(s) is/are being used for take-offs/landings). The details are reported later in the analysis of the whole model for Zurich Airport.

3.4. Multiple regression with several factors

This section presents multiple regression models for Stockholm-Arlanda Airport and Zurich Airport and ends with a consideration of the validity of the necessary assumptions to apply the regression. The discussion of the results and the applicability of the model can be found in Sections 4 and 5.

The goal of the multiple regression approach was to find the most important factors explaining the variability of the real data sets.

Extensive analysis was performed using different stepwise selection methods based on the factors described in Section 3.3 (depending on p -values, Akaike’s Information Criterion (AIC) and the Bayesian Information Criterion (BIC)). The authors decided to present models that are as practical as possible for use at airports (requiring less information) and that are easy to interpret. The following models fulfil this aim and are less than 2.2% away from the best models found (according to the R^2_{Adj} value).

3.4.1. Stockholm-Arlanda Airport. The final regression model for Stockholm-Arlanda Airport is given in Table 1. The first column indicates the variables, the second column the estimated unstandardised coefficients and the third column the corresponding estimated standard errors. The fourth column shows the estimated standardised coefficients for all non-dummy variables (ie the estimated coefficients if the variables were standardised so that their variance was 1). This measure can be used to analyse which factor has the largest positive or negative impact on $\log_{10}(Speed)$. In contrast to the unstandardised coefficients, they have no units and can therefore be compared directly. The last column shows the significance of each variable based on a t -test.

The model has a good R^2_{Adj} value of 0.863 (p -value $< 2.2e-16$). This means that around 86% of the variance of the $\log_{10}(Speed)$ values can be explained by the model. The fit of the prediction can be seen in Figure 6.

3.4.2. Zurich Airport. As indicated in Section 3.3.5, the current operational mode of the runways is a potentially significant factor at Zurich Airport. As long as no heavy winds occur, Zurich Airport operates strictly with three operational modes: before 7:00 runway 34 is used for arrivals and runways 32 and 34 for departures; during the day runways 14 and 16 are used for arrivals and 28 and 16 for departures; and after 21:00 only runway 28 is

used for arrivals and runways 32 and 34 are used for departures (see Figure 1(b)). We modelled the three operational modes using two dummy variables, $O_{Morning}$ to represent the morning period and $O_{Evening}$ to represent the evening period. Each variable was set to 1 during the corresponding period and 0 otherwise, so during the day period both variables were set to 0.

In contrast to Stockholm-Arlanda Airport, statistical analysis showed only small improvements by classifying the total distances into different components, so they were excluded from the final model. This was expected from the airport layout since it has fewer straight sub-paths without junctions.

The fit for Zurich Airport is given in Table 2, and shows an even better fit than for Stockholm-Arlanda Airport, with an R^2_{Adj} value of 0.878 (p -value $< 2.2e-16$). The scatterplot of the relationship between the observed values and the predicted values can be seen in Figure 7.

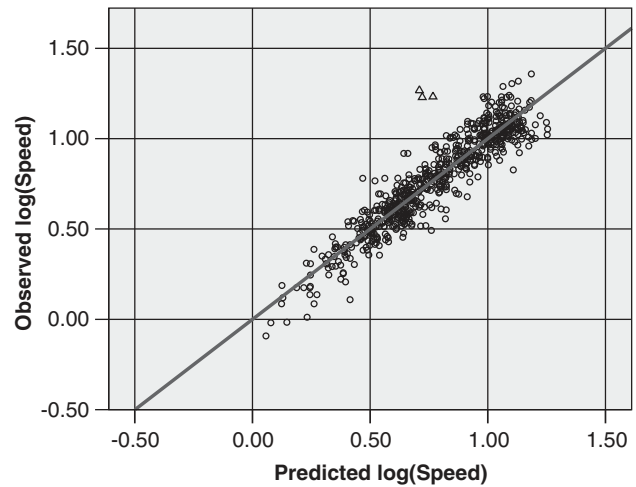


Figure 6 Scatterplot showing the linear fit of the regression model in Table 1 for Stockholm-Arlanda Airport.

Table 1 Coefficients for Stockholm-Arlanda Airport

	Coefficient $\hat{\beta}_i$	Std. dev.	Standardised coefficient	Sig.
(Constant)	-2.349	0.091		***
\log_{10} (Distance)	0.922	0.029	0.842	***
ARR	0.211	0.015		***
$N_{DEP, \#DEP}$	0.031	0.005	0.260	***
$N_{ARR, \#DEP}$	0.029	0.004	0.230	***
$N_{DEP, \#ARR}$	0.049	0.011	0.209	***
$N_{DEP, \#ARR}$	0.036	0.006	0.176	***
Distance2	-5e-05	8e-06	-0.188	***
$Q_{DEP, \#ARR}$	-0.034	0.004	-0.268	***
$Q_{ARR, \#ARR}$	-0.066	0.011	-0.279	***
$Q_{ARR, \#DEP}$	-0.052	0.006	-0.280	***
$Q_{DEP, \#DEP}$	-0.044	0.005	-0.397	***

Sig. indicates if the p -value is < 0.05 (*), < 0.01 (**) or < 0.001 (***).

Table 2 Coefficients for Zurich Airport

	Coefficient $\hat{\beta}_i$	Std. dev.	Standardised coefficient	Sig.
(Constant)	-2.601	0.250		***
$\log_{10}(\text{Distance})$	1.161	0.091	0.731	***
ARR	0.260	0.018		***
$N_{DEP, \#DEP}$	0.025	0.005	0.234	***
$N_{ARR, \#ARR}$	0.054	0.008	0.208	***
$N_{ARR, \#DEP}$	0.019	0.004	0.143	***
$N_{DEP, \#ARR}$	0.029	0.007	0.101	***
O_{Evening}	0.049	0.013		***
O_{Morning}	-0.075	0.019		***
$\log_{10}(\text{Angle})$	-0.143	0.039	-0.083	***
Distance	-7e-05	2e-05	-0.181	**
$Q_{DEP, \#ARR}$	-0.032	0.004	-0.208	***
$Q_{ARR, \#DEP}$	-0.067	0.006	-0.285	***
$Q_{ARR, \#ARR}$	-0.081	0.007	-0.318	***
$Q_{DEP, \#DEP}$	-0.046	0.004	-0.466	***

Sig. indicates if the p -value is <0.05 (*), <0.01 (**) or <0.001 (***).

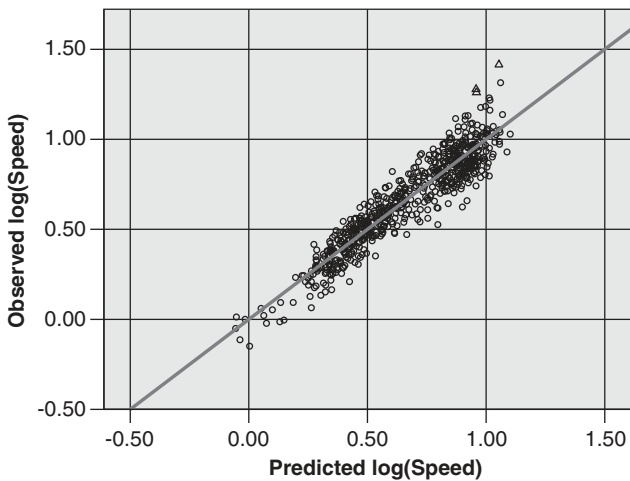


Figure 7 Scatterplot showing the linear fit of the regression model in Table 2 for Zurich Airport.

3.4.3. Validation of statistical assumptions. The estimated regression coefficients are unbiased if $E(\varepsilon_i) = 0$ for all $i = 1, \dots, n$. The residual plots in Figure 8 indicate that this assumption is approximately valid (with perhaps a slight lack of fit for small speeds). Hence, one can be confident that the estimated regression coefficients and resulting predictions are (almost) unbiased.

The standard errors for the estimated coefficients are valid if the following three assumptions hold: $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$ for all $i = 1, \dots, n$, and $Cov(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$. The residual plot in Figure 8(a) indicates that the constant variance assumption is approximately valid for Stockholm-Arlanda Airport. For Zurich Airport, there seems to be some increase in the variance with increasing predicted speeds. Owing to the time-dependent nature of the data, it is likely that there is some correlation in the statistical errors. The Durbin-Watson test indeed indicated

positive serial correlation for both airports. Generalized least squares models using autoregressive AR(1) and AR(2) models for the residuals were fitted to account for this correlation, and the results were compared with Tables 1 and 2. Estimates of the coefficients and standard errors at both airports are very consistent.

Finally, the p -values are valid if in addition to the assumptions above the statistical errors have a normal distribution. Moreover, even without the normality assumption they hold approximately if the sample size is sufficiently large, due to the central limit theorem. The Q-Q-plots in Figure 9 show that the residuals are approximately normally distributed. A discussion about the outliers (indicated with triangles) is presented in Section 4.2. Formal Shapiro-Wilk tests (Shapiro and Wilk, 1965) were also performed to test the normality assumption, where the outliers were excluded. These tests supported the findings from the figures and indicated no evidence for departure from normality (p -values 0.083 and 0.463 for Stockholm-Arlanda Airport and Zurich Airport, respectively). However, due to potential (small) violations of the assumptions of constant variance, the p -values for Zurich Airport might be slightly off.

The taxi distance appears on both sides of the multiple linear regression models, due to the decision to use speed as the dependent variable. However, since it seems clear that distance might influence speed but not the other way round, we assume that there are no endogeneity problems.

3.5. Cross-validation

A common way of testing how well a model performs in predicting new data is the so-called PRESS statistic, suggested by Allen (1971):

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2. \tag{6}$$

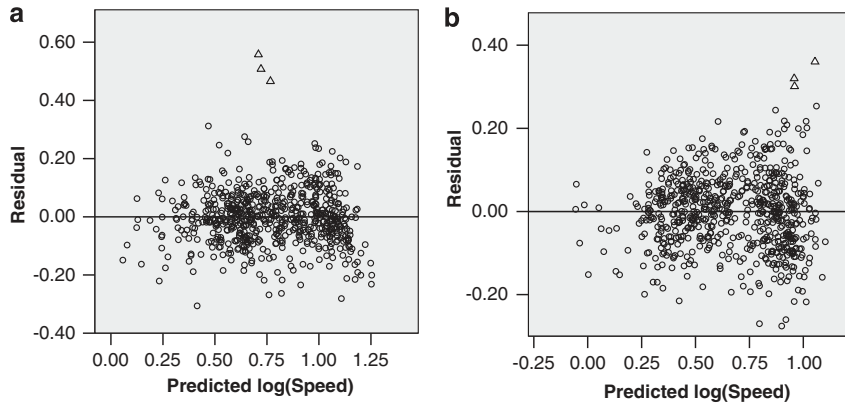


Figure 8 Residual plots showing the validation of the assumptions: (a) Stockholm-Arlanda Airport; (b) Zurich Airport.

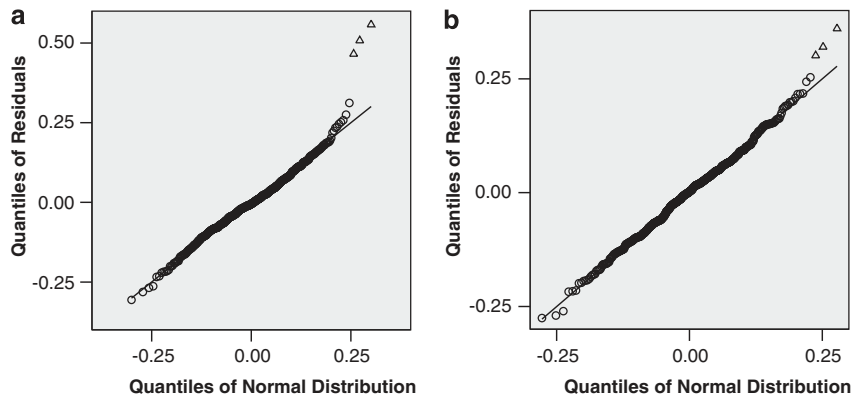


Figure 9 Normal Q-Q-plots showing the validation of the assumptions: (a) Stockholm-Arlanda Airport; (b) Zurich Airport.

It sums the squared differences between the observed variables y_i and the predicted variables $\hat{y}_{(i)}$ for each of the sample points i , where the prediction $\hat{y}_{(i)}$ only uses the data of the remaining observations. It can be categorised as a leave-one-out cross-validation. The PRESS statistic can be used to calculate an R^2 value for a prediction:

$$R_{\text{Pred}}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (7)$$

The R_{Pred}^2 value was 0.860 for Stockholm-Arlanda Airport and 0.875 for Zurich Airport. This means that, for similar settings at the airport (the same operational modes, similar weather conditions and so on), these models could explain around 86 and 87.5%, respectively, of the variability in predicting new observations due to the combination of the statistical analysis with the incorporation of the ground layout model.

3.6. Prediction accuracy

A second data set was made available for Zurich Airport after the model had been fitted to the existing data set. The second data set consisted of 5613 aircraft movements that occurred during one week's operation between the 27th of June and the 3rd of July 2011. Even though we used the same coefficients as reported in Table 2, and they were generated using the old data, the approach was still able to demonstrate a high R_{Adj}^2 value of 0.864 for the prediction. Keeping the same factors as in Table 2, but re-estimating the coefficients for the new data set, the R_{Adj}^2 could only be improved to 0.899. These results demonstrate that the model was not only able to fit historic data well but that it can also be used to make accurate taxi speed predictions, especially when keeping in mind that the two data sets were from periods which were almost 4 years apart.

4. Interpretation of the models

First of all, it can be seen from Tables 1 and 2 that the two fitted regression models are very similar and have the same

general structure, indicating the potential usage for other airports. All the factors in the tables are highly significant (p -value < 0.01).

4.1. Coefficient meanings

We now interpret some of the coefficients to gain insight into the effects of specific factors. The straightforward interpretation of this model could possibly encourage airport operators to use this approach to support their needs.

4.1.1. Distances. The most important factor for both airports was the logarithmic transformation of the total distance. In general, the average taxi speed was higher the further an aircraft had to taxi. This finding is new compared with the results from other research, where the focus was on airports with longer queues, which probably dominated the effect of the distance. Even with the assumption of using the shortest path for each aircraft, the results look promising and would probably look even better by utilising the actual distance rather than the shortest path.

4.1.2. Departures versus arrivals. Another important factor in the models for both airports was the differentiation between arriving and departing aircraft. Since departures often need to wait in a queue, their average speed is smaller in comparison with arriving aircraft, which are forced to clear the runway as soon as possible and taxi directly to the stands.

4.1.3. Angle. The logarithmic transformation of the total turning angle that an aircraft had to complete was observed to be a significant slowing factor at Zurich Airport. The inclusion of this factor significantly improved the accuracy of the prediction.

4.1.4. Amount of traffic. All of the different Q values were observed to have a negative effect upon the taxi speed. In general, more aircraft travelling around the airport means that each individual aircraft's speed is reduced. Factors which particularly slowed taxi speeds were $Q_{DEP, \#DEP}$ and $Q_{ARR, \#ARR}$, representing the number of aircraft which have the same target (runways or stands) but end their taxi operation first. The N variables were found to counteract some of the effect of the Q variables, together modelling those aircraft which both start to taxi earlier and which reach their destination earlier. Our results showed differences between the North American airport studied by Idris *et al* (2002) and the European airports considered in this research, since the number of

arrivals did not affect the taxi-out time in their study whereas there was a strong correlation in our analysis. This may be related to the airport layouts or the runway queue lengths.

4.1.5. Operational mode. In the case of Zurich Airport, the influence of the different operational runway modes was incorporated into the model. It can be observed that aircraft taxi faster in the evening than during the day, and faster during the day than in the morning. There is insufficient information at the moment to determine whether the effect is due to the different runway modes or whether other elements such as visibility or different aircraft mixes at different times of the day are affecting the taxi speeds.

4.2. Unexplained variability

Around 13% of the variability in taxi speeds cannot be explained by our models. Some potential explanations are listed below:

- The taxi behaviour can vary between different airlines and pilots. Additional data should allow us to analyse this in more detail in the future.
- In the case of Stockholm-Arlanda Airport the taxi time information was only to the minute rather than to the second, but the model uses continuous time for the speed predictions. The data of Zurich Airport had detailed times at the runway, but again the times at the stand/gates were only to the minute. This matching of continuous time to discrete values is unlikely to provide extremely accurate predictions.
- We assumed that aircraft travelled along the shortest path and that there were no unexpected changes. This assumption will be valid in general but can lead to occasional errors.

An analysis of the outliers at Stockholm-Arlanda Airport showed that the three worst fits (the three triangles in Figures 6, 8(a) and 9(a)) were for aircraft landing at runway 26 and taxiing to stand group F. The taxi times were extremely short: 1 min for one of the aircraft and 2 min for the other two. Given the minute granularity on the data, it is perhaps unsurprising that the estimations were least accurate for these aircraft. Similarly, the most extreme outliers at Zurich Airport (the three triangles in Figures 7, 8(b) and 9(b)) were also related to very short taxi times.

4.3. Related applications

The same approach was also used to estimate taxi times for London Heathrow Airport (LHR), which is one of the

busiest international airports in the world. A multiple linear regression approach was used to predict taxi times for Heathrow (Atkin *et al.*, 2011b), using a data set that covered one week's operations (9391 movements) for summer 2010. The dependent variable was $\log_{10}(\text{Speed})$ and $\log_{10}(\text{Distance})$ and the N and Q values were used as explanatory variables. For Heathrow, it was found to be better to have separate regression models for departures and arrivals, and to separate cases depending upon which runway the aircraft were starting from or landing at. The R_{Adj}^2 value was 0.929 for departing aircraft and 0.835 for arriving aircraft, totalling to 0.882. Experiments with leave-one-out cross-validation, as explained in Section 3.5, indicated that the R_{Pred}^2 values were at most 0.1% smaller than the R_{Adj}^2 values, leaving them very high.

5. Applicability of this research

The two main applications for this research are for total taxi time prediction and for use in a ground movement decision support system. We consider both of these in this section.

5.1. Improved total taxi time prediction

To the best of our knowledge, there is no existing taxi time prediction function to compare against for both departing and arriving aircraft, but we have the lookup table which is used for Zurich Airport. This considers only the sources and destinations and gives average taxi-in and taxi-out times. However, it has a granularity of 1 min and deliberately underestimates times. In order to eliminate the deliberate underestimates, we used linear regression to find a linear scaling that best fitted their table to the observed data. This resulted in an improved R_{Adj}^2 value of 0.180, with a scaling of $ax + b$, where a is 0.883 and b is 2.210. In contrast, the approach presented in this paper, when applied to taxi times (rather than $\log_{10}(\text{Speed})$) resulted in an R_{Adj}^2 value of 0.793, thus explaining the variability in taxi times at this airport to a much greater extent than the lookup table and indicating the benefits of the consideration of more factors. The function generated by our multiple linear regression is, therefore, more appropriate for predicting total taxi time.

The results were also compared with the results from the application of a reinforcement learning algorithm by Balakrishna *et al.* (2009) at other airports. They presented results for the ± 3 or ± 5 min prediction accuracy for the taxi-out times (see Table 3), measuring the percentage of departing aircraft with a time difference between the predicted time and the observed time which is smaller than the given threshold value. An average of 95.7% was found for Detroit International Airport (DTW) and an average of

Table 3 Comparison of prediction accuracy

	Within ± 3 min	Within ± 5 min
Stockholm-Arlanda Airport (simplified)	73.2%	88.4%
Zurich Airport (simplified)	82.6%	91.6%
Stockholm-Arlanda Airport	94.4%	98.9%
Zurich Airport	95.6%	99.4%
Stockholm-Arlanda Airport (full)	96.1%	99.2%
Zurich Airport (full)	96.8%	99.7%
Detroit International Airport	89.9%–97.1%	—
Tampa International Airport	89.9%–95.7%	—
John F. Kennedy International Airport	—	20.7%–100%
Boston Logan International Airport	—	65.6%

The first block shows the result for the two studied airports where the prediction model is simplified by not considering the airport layout (particularly by not considering the factors about the distances and the turning angles). The results for the best found models are indicated in the last three blocks.

93.8% for Tampa International Airport (TPA) for ± 3 min accuracy. The results for John F. Kennedy International Airport (JFK) were not very consistent and much less promising, showing ± 5 min prediction accuracy between 20.7 and 100% for different days and parts of the day. Additionally, Idris *et al.* (2002) predicted 65.6% of the taxi-out times at Boston Logan International Airport (BOS) within ± 5 min of the actual time. In contrast, our regression model found an average ± 3 min accuracy of 94.4% for Stockholm-Arlanda Airport and 95.6% for Zurich Airport, considering both departures and arrivals simultaneously.

Reported taxi times at Stockholm-Arlanda Airport were from 1 to 16 min for arrivals and 3 to 20 min for departures. The seven cases which were not predicted within ± 5 min accuracy were all departures with very long taxi times with the highest deviation of 7.40 min. Figure 10 shows the deviations of the estimated to the actual taxi times where the deviations are ordered. The rounded deviations are also shown (the step function), where the estimated taxi times are rounded to the nearest minute, to match the accuracy of the historic input data from Stockholm-Arlanda Airport, since many stakeholders are only interested to this level of accuracy. Taxi times at Zurich Airport ranged from 1 to 12 min for arrivals and 4 to 24 min for departures. Again, the four worst predictions were for aircraft with long taxi times and only one prediction was not within ± 6 min accuracy (but this has less than 8 min deviation).

The results labelled '(simplified)' in Table 3 also show the prediction accuracy of our approach for both Stockholm-Arlanda Airport and Zurich Airport without taking the actual graph layout of the airports into account.

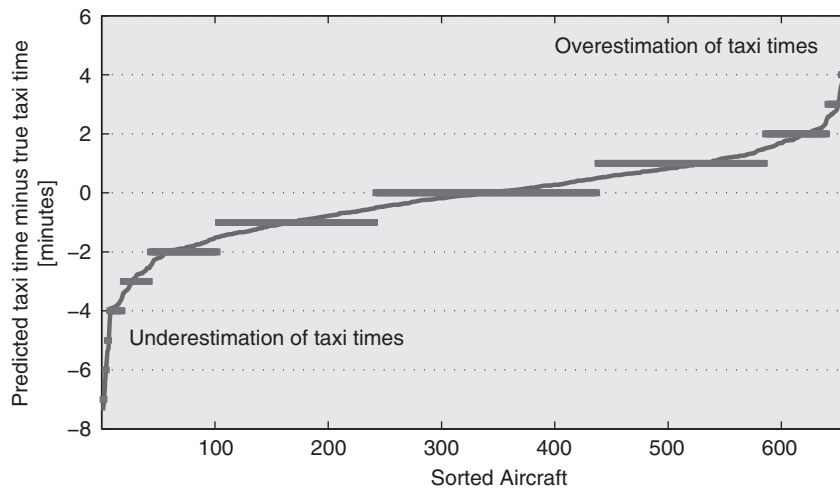


Figure 10 Taxi time prediction accuracy at Stockholm-Arlanda Airport.

A simplified regression analysis was performed without the different distance measures and the measures related to the turning angle. The big improvements when the layout is considered emphasise the need for layout-based factors for airports where queuing is not dominating the whole ground movement process.

In contrast, the results labelled '(full)' in Table 3 correspond to the model with the best R_{Adj}^2 value when considering all possible factors, rather than attempting to simplify the model. These indicate that the R_{Adj}^2 would increase by around 2.2%. However, the aim of this research was to provide a practical model that was easy to interpret and hence the focus was not entirely on getting the model with the best accuracy.

As discussed in the introduction of this paper, several other airport-related decision support systems as well as a wide variety of stakeholders at an airport (eg runway controllers, gate allocators, cleaning crews, de-icing crews, bus drivers, etc) will benefit from better taxi time predictions.

5.2. Use for ground movement decision support

As discussed at the beginning of this paper, algorithms that aim to optimise ground movement at airports need a model for predicting taxi times when there are no delays, since the interaction between aircraft would be explicitly considered by the model anyway. Such predicted uninterrupted taxi times can then be used to find a globally good solution by adding some delays or detours to aircraft where contention with other aircraft is indicated by the algorithm. The presented regression model allows such uninterrupted taxi time modelling by setting all N and Q values to 0.

Regression models work well within their range of observed data, but have to be handled with care for predictions at the boundaries and for extrapolations. Importantly, both data sets contain a number of observations with all N

and Q values equal to 0 (for three departures and nine arrivals at Stockholm-Arlanda Airport and six departures and for four arrivals at Zurich Airport) and these values are spread throughout the taxi speed range.

Once the regression approach has been implemented in a ground movement search methodology, it will be interesting to test the new system against the actual operations at the specific airport, and to fine tune the parameters to match the taxi times even more.

6. Conclusions

With the current emphasis upon improving the predictions for on-stand times and take-off times (Eurocontrol, 2010), an improved method for taxi time prediction is both important and timely. This paper analysed the variation in taxi speed and, consequently, the variations in taxi times, and considered not only departures but, for the first time, also arrivals. Data from Stockholm-Arlanda and Zurich Airport, both major European hub airports, was used for this research and the potential significant factors were identified and individually tested. Multiple linear regression was used to find a function which could more accurately predict the taxi times than existing methods. An emphasis was placed upon ensuring that the function was easy to interpret and simple to use for operators at airports and researchers. Key for the analysis was the incorporation of information about the surface layout, since, in contrast to other airports which have previously been studied, the runway queuing was not dominating the entire taxi time.

The average speed between the gate and runway (and between the runway and gate) was found to be highly correlated to the taxi distance, with higher speeds being expected for longer distances. Arrivals had higher taxi speeds than departures, due to departure queues at the runway, and the quantity of traffic at the airport was also

found to have a significant impact upon the average taxi speed, as identified by several variables in the resulting model. Finally, the total turning angle and the operating mode (which runways were in use) were also highly correlated to the average taxi speed.

Consideration of taxi time accuracy does not appear to have been sufficiently incorporated into the current state-of-the-art research in ground movement decision support systems at airports. Better predictions would, if nothing else, reduce the amount of slack which had to be allowed for taxi time inaccuracies, allowing tighter schedules to be created. Historic data is vital for model calibration, but such data usually include the effects of various inter-aircraft dependencies. When a decision support system takes care of the dependencies between the aircraft, predicted taxi speeds should not themselves include the effects of these dependencies. However, it is not usually obvious how to quantify and eliminate these effects. Among other uses, the approach which has been presented here could potentially be used for exactly such situations, allowing individual effects to be removed from consideration. The development of such a facility was the prime motivation for this research.

Since this work considers a combined statistical and ground movement model, which seems to accurately predict the effects of turns and congestion as well as total travel distances, we note here that these results can also feed into ground movement models, to improve the accuracy of the predictions for the effects of re-routing or delays. We plan to consider this in future research.

In addition to our current work on ground movement simulation, we aim to use this research to generate more realistic benchmark scenarios for ground movement. These will not only stimulate researchers to compare their ground movement algorithms but to do so with scenarios that are closer to reality. Further research should explore more sophisticated ways of fine-tuning the parameters to further increase the value of the approach for decision support systems for ground movement at airports, or other prediction approaches such as fuzzy rule-based systems or time series analysis.

Acknowledgements—This research was funded by EPSRC (The Engineering and Physical Sciences Research Council). The authors wish to thank the involved stakeholders, namely Swedavia AB and Flughafen Zürich AG for providing the real data sets from those airports, and NATS for providing the Heathrow data. We would also like to thank the anonymous reviewers who have helped to improve this paper.

References

- Allen DM (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics* **13**(3): 469–475.
- Atkin JAD, Burke EK, Greenwood JS and Reeson D (2007). Hybrid metaheuristics to aid runway scheduling at London Heathrow airport. *Transportation Science* **41**(1): 90–106.
- Atkin JAD, Burke EK and Greenwood JS (2010a). TSA allocation at London Heathrow: The relationship between slot compliance, throughput and equity. *Public Transport* **2**(3): 173–198.
- Atkin JAD, Burke EK and Ravizza S (2010b). 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, pp 131–138.
- Atkin JAD, Burke EK and Ravizza S (2011a). 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.
- Atkin JAD, Burke EK and Ravizza S (2011b). A statistical approach for taxi time estimation at London Heathrow Airport. In: *Proceedings of the 10th Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP)*, Nymburk, Czech Republic.
- Balakrishna P, Ganesan R and Sherry L (2009). Application of reinforcement learning algorithms for predicting taxi-out times. In: *Proceedings of the 8th ATM R&D Seminars*, Napa, USA.
- Balakrishnan H and Jung Y (2007). A framework for coordinated surface operations planning at Dallas-Fort Worth International Airport. In: *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Hilton Head, SC, USA.
- Burgain P, Feron E and Clarke J-P (2009). Collaborative virtual queue: Benefit analysis of a collaborative decision making concept applied to congested airport departure operations. *Air Traffic Control Quarterly* **17**(2): 195–222.
- Cormen TH, Leiserson CE, Rivest RL and Stein C (2001). *Introduction to Algorithms*. 2nd edn, MIT Press and McGraw-Hill: Cambridge, MA.
- Eurocontrol (2010). *Airport CDM implementation manual*. Technical Report, version 3.1, Eurocontrol. http://www.euro-cdm.org/library/cdm_implementation_manual.pdf.
- Gotteland J-B and Durand N (2003). Genetic algorithms applied to airport ground traffic optimization. *Proceedings of the Congress on Evolutionary Computation (CEC)*, Canberra, Australia pp 544–551, doi: 10.1109/CEC.2003.1299623.
- Idris HR, Clarke JP, Bhuva R and Kang L (2002). Queuing model for taxi-out time estimation. *Air Traffic Control Quarterly* **10**(1): 1–22.
- Lesire C (2010). Iterative planning of airport ground movements. In: *Proceedings of the 4th International Conference on Research in Air Transportation (ICRAT)*, Budapest, Hungary, pp 147–154.
- Montgomery DC, Peck EA and Vining GG (2001). *Introduction to Linear Regression Analysis*. 3rd edn, John Wiley & Sons, Inc: New York.
- Rappaport DB, Yu P, Griffin K and Daviau C (2009). Quantitative analysis of uncertainty in airport surface operations. In: *Proceedings of the 9th AIAA Aviation Technology, Integration, and Operations Conference*, Hilton Head, SC.
- Ravizza S and Atkin JAD (2011). *Exploration of the ordering for a sequential airport ground movement algorithm*. Technical Report, University of Nottingham. <http://eprints.nottingham.ac.uk/1543/>.
- Roling PC and Visser HG (2008). Optimal airport surface traffic planning using mixed-integer linear programming. *International Journal of Aerospace Engineering* **2008**(1): 1–11.
- SESAR (2006). *Milestone deliverable D1—Air transport framework: The current situation*. Technical Report, Eurocontrol, 3rd edn, http://www.eurocontrol.int/sesar/public/standard_page/documentation.html.

- Shapiro SS and Wilk MB (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**(3–4): 591–611.
- Simaiakis I and Balakrishnan H (2009). Queuing models of airport departure processes for emissions reduction. In: *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Chicago, IL.
- Smeltink JW, Soomer MJ, de Waal PR and van der Mei RD (2004). An optimisation model for airport taxi scheduling. In: *Proceedings of the INFORMS Annual Meeting*, Denver, CO.
- Tu Y, Ball MO and Jank WS (2008). Estimating flight departure delay distributions—A statistical approach with long-term trend and short-term pattern. *Journal of the American Statistical Association* **103**(481): 112–125.

*Received June 2011;
accepted August 2012 after two revisions*