Traffic & Transportation

Prediction of Taxi-in Time and Analysis of Influencing Factors for Arrival Flights at Airport with a Decentralised Terminal Layout

Xiaowei TANG[1], Mengfan YE[2], Shengrun ZHANG[3], Kurt FUELLHART[4]

Original Scientific Paper
Submitted: 28 Sep. 2023
Accepted: 11 Mar. 2024

[1] tangxiaowei@nuaa.edu.cn, Nanjing University of Aeronautics and Astronautics, College of Civil Aviation

[2] Mhymf27@163.com, Nanjing University of Aeronautics and Astronautics, College of Civil Aviation

[3] zhangshengrun_seg@126.com, Nanjing University of Aeronautics and Astronautics, College of Civil Aviation

[4] Kurt.Fuellhart@UGent.be, Ghent University, Department of Geography; Unison Consulting, Inc., Laguna Hills

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Accurately predicting taxi-in times for arrival flights is crucial for efficient ground handling resource allocation, impacting flight departure timeliness. This study investigates terminal layout characteristics, specifically decentralised layouts, to predict and analyse arrival flight taxi-in times. We develop a surface traffic flow calculation method considering arrival and departure flights, eliminating fixed thresholds. We introduce runway-crossing operations for decentralised airports, creating new prediction variables. We consider factors like runway, aircraft type, airline, taxi distance, and time periods. Gradient Boosting Regression Tree predicts taxi-in times, while Lasso analyses factor impact. Our approach yields highly accurate predictions for decentralised airports, with Surface traffic flow and Runway-crossing variables significantly influencing taxi-in times. This research informs airport managers in decentralised layouts, enabling tailored management strategies.

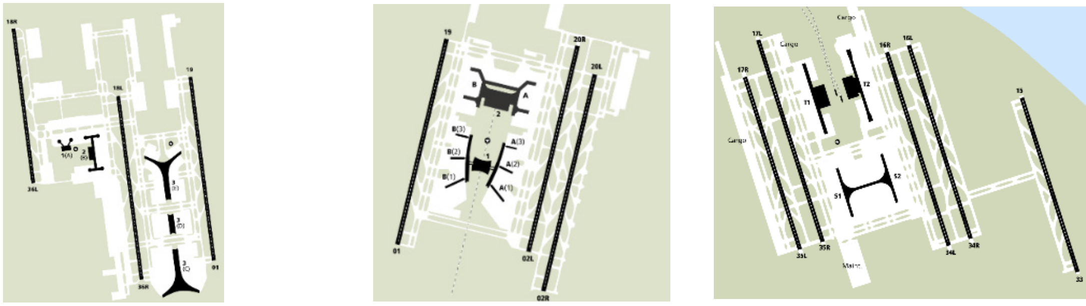
KEYWORDS

air transportation; taxi-in time prediction; terminal layout; airport surface movement; lasso; gbrt; airport management.

1. INTRODUCTION

Most researchers have primarily focused on predicting taxi-out times for departing flights [1–6] with relatively less attention given to predicting taxi-in times for arriving flights [7, 8]. Given that the arrival process sets the preceding of ground handling operations and the subsequent departures, the accuracy of taxi-in time predictions holds significance. It directly impacts the pre-allocation of ground handling resources, subsequently affecting the on-time performance of departure processes.

The arrival process of flights involves aircraft landing on the runway and taxiing through the taxiways until it reaches the stand. This process is particularly challenging for large airports with multiple runways due to numerous uncertainties. The variations in the configuration of multiple runways and the related terminal layout can structurally influence the taxi-in time prediction. This paper focuses on airports with independent parallel runways and a decentralised terminal layout in a Chinese context, for which Beijing Capital International Airport (PEK) is a typical illustration (*Figure 1a*). In this case, a single runway can simultaneously accommodate both arriving and departing flights on the condition that the runway spacing complies with requirements of independent parallel approaches and departures. A "runway crossing" issue occurs when arriving/departing flights require taxiways to bypass or conflict with an active runway. At PEK, both arriving and departing flights can confront the runway crossing issue in two directions, which is different from that of airports with a centralised terminal layout, e.g. the other two biggest airports in China (i.e. Guangzhou Baiyun International Airport (CAN)) and Shanghai Pudong International Airport (PVG) (*Figure 1b*). In the case of the latter, it is arriving flights that require taxiways to bypass or conflict with an active departure runway, i.e. unidirectional runway-crossing. Therefore, for airports with independent parallel runways and a decentralised terminal layout, larger runway-crossing occurrences and the complexity in the operation process increase the intricacy of predicting taxi-in times.



PEK *CAN* *PVG*

a) Decentralised layout *b) Centralised layouts*

Figure 1 – The layout of three airports

Existing research has shown that both intensity and complexity of surface traffic flow are crucial factors influencing the prediction of taxi-in times [1, 9–13]. Intensity indicators that significantly impact the current flight's taxi-in time include instantaneous traffic flow [1], the number of other flights on the same runway [10], the number of flights departing within the preceding 20 minutes [11] and the number arrival aircraft up to 1.3 minutes before the aircraft in question [12]. In these studies, the number of flights was measured by predefining a fixed time period. However, these fixed thresholds cannot be directly applied to other airports with different configuration layouts. Hence, the complexity perspective that considers the time-varying characteristics of surface traffic flow should be investigated [14]. The challenge is to examine the spatial-temporal characteristics of surface traffic flow by simultaneously modelling the interrelationships between arriving and departing flights. This paper introduces a novel approach to accurately represent the dynamic characteristics of surface traffic flow specifically for airports with independent parallel runways and explores its impact on the taxi-in time.

Selecting appropriate modelling techniques is crucial for identifying factors affecting taxi-in times for arriving flights and for accurate predictions. Existing literature has employed various modelling techniques to predict taxi-in times for flights, including multivariate linear regression [13], Lasso regression [15] in statistical regression models, as well as machine learning models such as Gradient Boosting Regression Trees (GBRT) [16]. Lasso regression, as one of the representative statistical regression models capable of variable selection, utilises its oracle property to extract significant factors from among multiple influencing factors [17]. Beyond identifying influencing factors, the accuracy of prediction models is also paramount. GBRT, one of the most representative decision tree prediction methods in machine learning, often offers higher predictive accuracy than statistical regression models [18], albeit at the expense of some interpretability [19]. Consequently, this paper employs Lasso regression for factor analysis and GBRT for prediction.

In summary, this paper focuses on predicting taxi-in times for arriving flights and analysing influencing factors for airports with a decentralised terminal layout. Emphasis is placed on two aspects: surface traffic flow and runway crossing operations. The empirical study is PEK. The remainder of the paper is structured as follows: In Section 3, we propose a method for calculating surface traffic flow and introduce features related to runway-crossing operations based on the decentralised terminal layout of PEK. Additionally, we construct other features for predicting taxi-in times. Section 4 analyses the correlation of various influencing factors and performs data pre-processing. Section 5 presents the results of and provides our interpretations of the outcomes, followed by the conclusion in chapter 6.

2. LITERATURE REVIEW

Taxi time consists of taxi-out and taxi-in time, and there are similarities in the influencing factors between taxi-out time for departing flights and taxi-in time for arriving flights [16]. To provide a clear overview of the latest advancements in taxi time prediction, *Table 1* outlines the research landscape and target airports for studies in the domain of taxi time prediction.

Table 1 – Overview of selected studies in taxi time prediction

Year	Authors	Predictors	Methodology	Study airports
2008[20]	Xu et al.	GDP holding time; carrier delay; departure demand ratio; swap aircraft rate	Piecewise linear model	34 U.S. airports
2010[12]	Jordan et al.	Surface traffic flow; type of airlines; taxi distance; weather	Sequential forward floating subset selection and Ordinary Least Square (OLS)	DFW
2010[1]	Balakrishna et al.	Surface traffic flow; time periods; average taxi time of adjacent periods	Reinforcement learning (RL)	TPA
2011[21]	Srivastava et al.	Taxi distance; queue position; mean of taxi-out time; arrival rate; weather	Uniform flow model and split flow model	JDK
2013[13]	Ravizza et al.	Surface traffic flow; taxi distance; turning angle	OLS	ARN and ZRH
2014[22]	Ravizza et al.	Surface traffic flow; taxi distance; turning angle	OLS, least median squared linear regression, support vector regression (SVR) and M5 model trees	ARN and ZRH
2016[23]	Lee and Malik	Time periods; runway used; aircraft type; taxi distance; terminal concourse; aircraft type; month; weather	OLS, SVR, k-nearest neighbors (KNN), random forest (RF), and neural networks (NN)	CLT
2017[10]	Feng and Meng	Surface traffic flow; runway used; taxi distance; average taxi time of adjacent periods	KNN and SVR	PEK
2018[11]	Herrema et al.	Surface traffic flow; time periods; aircraft type; airlines; air traffic control	NN, regression tree, RL, and multi-layer perceptron (MLP)	CDG
2018[14]	Yin et al.	Surface traffic flow; surface instantaneous flow indices; surface cumulative flow indices; aircraft queue length indices; and slot resource demand indices	OLS, SVM and RF	PVG
2018[2]	Diana	Runway used; total operations; total delays; percent of on-time gate departures; average taxi time; fleet mix	Ensemble machine learning, OLS, ridge, Lasso and elastic net	SEA
2021[16]	Wang et al.	Surface traffic flow; runway used; taxi distance; aircraft type; month; type of airlines; turning angle; speed of other aircraft that have recently taken off; weather	Multilayer perceptron, OLS, Polynomial regression, GBRT, RF	MAN, ZRH and HKG
2021[24]	Chen et al.	Surface traffic flow; aircraft type; type of airlines; taxi distance; speed of other aircraft that have recently taken off; stands	SVR, KNN, decision tree	PEK
2022[25]	Zhao et al.	Surface traffic flow; time periods; runway used; aircraft type; airlines; number of Hot Spot passes; cross taxi; whether to cross the runway; stands; weather	Extreme gradient boosting	CAN

Note: The column "Study Airports" is presented in the form of IATA codes.

Inspired by queuing models, Idris [26] linked surface traffic flow to taxi-out time for departures. He developed a method using pushback and take-off points. Jordan [12] applied statistical regression to determine a time window for counting flights on runways. Balakrishna calculated instantaneous surface traffic flow at multi-runway airports by considering flights on runways and taxiways [1]. Ravizza [13] extended Idris's approach to include arrivals and departures. Feng [10] and Herrema [11] used fixed 15- minute and 20-minute thresholds for surface traffic flow calculations.

Accounting for the interrelationships between arriving and departing flights is crucial in multi-runway airports in a collaborative optimisation circumstance [27]. For airports with independent parallel runways, as a single runway can simultaneously accommodate both arriving and departing flights, the number of departing flights can inevitably influence the taxi-in time of arrival flights. Moreover, once a flight has been landed and is taxiing, the priority for a departing flight to use runways is higher. When operation conflicts occur, the arrival flights have to wait, resulting in longer taxi-in time. To consider the impact of instantaneous arrivals and departures on the current flight's taxi-in time [16], this paper proposes a novel method to measure the intensity and complexity of the surface traffic flow for airports with parallel runways.

In addition to surface traffic flow, several factors related to airport operations have been considered in multiple taxi time prediction studies. These factors include runway used [23, 10, 2, 16, 25], aircraft type [2, 11, 16, 23–25], type of airlines [11, 12, 16, 24, 25], taxi distance [10, 12, 13, 16, 21, 23–25], and time periods [1, 11, 23, 25]. "Runway used" signifies the landing runway choice. "Aircraft type" categorises planes based on wingspan [4] or weight [23]. "Type of airlines" differentiates between domestic and foreign carriers to represent the language communication cost between foreign crew and local air traffic controllers. "Taxi distance" measures the path length between landing and stand position, a well-established taxi time predictor [9]. "Time periods" classify landing times into distinct segments. Balakrishna et al. [1] integrated time periods into a reinforcement learning model. However, complete validation was not possible due to machine learning's interpretability constraints. Lee et al. [23] discussed impact of time periods on taxi time at the Charlotte Douglas International Airport (CLT) and introduced them predicting departure taxi-out times.

Few studies have explored the influence of airport configuration on aircraft taxiing. However, based on the Total Airport Management (TAM) concept [28], it is essential to incorporate the entirety of the airport processes including airside, landside and terminal [29]. Pina [30] verified that accurately predicting taxing times for landing flights helps distribute ground handling resources and estimate the in-block time. Tang [31] examined terminal layout effects at PEK on apron taxiing, focusing solely on the apron and overlooking the entire movement area. Centralised layouts separate ground taxiing from runway operations, while decentralised layouts involve runway crossings or bypasses, leading to ground taxiing and runway conflicts. This paper introduces the "runway crossing" feature to assess the configuration's impact on ground taxiing.

Statistical regression models are commonly employed for aircraft taxi time prediction. Jordan [12] used OLS to predict taxi times at Dallas-Fort Worth International Airport (DFW), which has a centralised terminal layout. Ravizza [13] identified taxi distance as a key factor influencing taxi times using OLS and data from Stockholm Arlanda Airport (ARN) and Zurich Airport (ZRH). Lasso regression, known for variable selection, has gained attention [17]. Diana [2] compared regression and machine learning methods, highlighting OLS and Lasso's interpretability. Li [15] compared Lasso and OLS for PEK, where Lasso outperformed OLS notably at 3- and 5-minute error levels. Machine learning algorithms have also gained prominence with open-source libraries. Balakrishna [1] designed a reinforcement learning model for Tampa International Airport (TPA), achieving 81% accuracy within a 5-minute error. Ravizza [22] used SVR for ARN and ZRH, while Lee [23] compared OLS, SVR, RF and NN using CLT data, with RF showing superior accuracy among machine learning methods. Wang [16] evaluated OLS, RF and GBRT for multi-runway airports (MAN, ZRH, Hong Kong International Airport (HKG)), observing better machine learning performance than statistical regression.

In summary, taxi time prediction studies utilise OLS, Lasso regression and GBRT. OLS results can be less reliable when strong correlations exist among independent variables, as in models with intercorrelated factors like surface traffic flow and time periods. Lasso addresses this issue effectively [32]. GBRT, a decision tree method, is robust to multicollinearity and, as a Boosting technique, iteratively enhances performance, often surpassing RF and SVR [18]. Hence, Lasso regression and GBRT are suitable for taxi-in time prediction. This study applies Lasso for analysing factors, GBRT for making predictions and OLS as a baseline to check for multicollinearity among variables.

3. DATA AND VARIABLES

The data for this study are sourced from the Airport-Collaborative Decision Making (A-CDM) system of PEK, one of the busiest airports in China. The A-CDM system of the airport aggregates operational data from various stakeholders, including the airport, airlines, air traffic control (ATC) and ground services. *Figure 2* illustrates the runway configuration and the taxi paths across runways at PEK. The green rectangles denote corresponding runway identifiers, while the blue and orange arrows indicate the one-way taxi flow of aircraft crossing runways. PEK features three parallel runways, each with sufficient vertical spacing to facilitate independent parallel approaches and take-offs. As an airport with a decentralised terminal layout, Terminal 1 (T1) and Terminal 2 (T2) are situated between runways 18R/36L and 18L/36R, while Terminal 3 (T3) is located between runways 18L/36R and 01/19.

To mitigate the impact of COVID-19 control measures on data analysis, we collected operational data from PEK between 1 May 2019 and 31 December 2019, before the pandemic. This time frame covers the peak season of flight operations within a year. The collected data fields include flight number, arrival runway, stand, aircraft type category, airlines, actual landing time (ALDT), actual in block time (AIBT), actual off block time (AOBT), actual take-off time (ATOT) and others. Based on these data fields, we construct response and explanatory variables to predict taxi-in times accurately.

The taxi-in time for arriving flights is defined as the difference between the AIBT and the ALDT of the aircraft. It is calculated as follows:

$$T_{taxi\text{-}in\,time} = t_{AIBT} - t_{ALDT} \tag{1}$$

where T is the taxi time of arrival flights. After removing records with missing fields, a total of 105,907 valid sample data points were obtained. *Table 2* provides the minimum, median, maximum, mean, 25th percentile and the 75th percentile values for the overall sample. Considering the 25th and 75th percentile values described in *Table 2*, it can be observed that the distribution of taxi-in times for arriving flights at PEK is concentrated between 7 and 17 minutes.

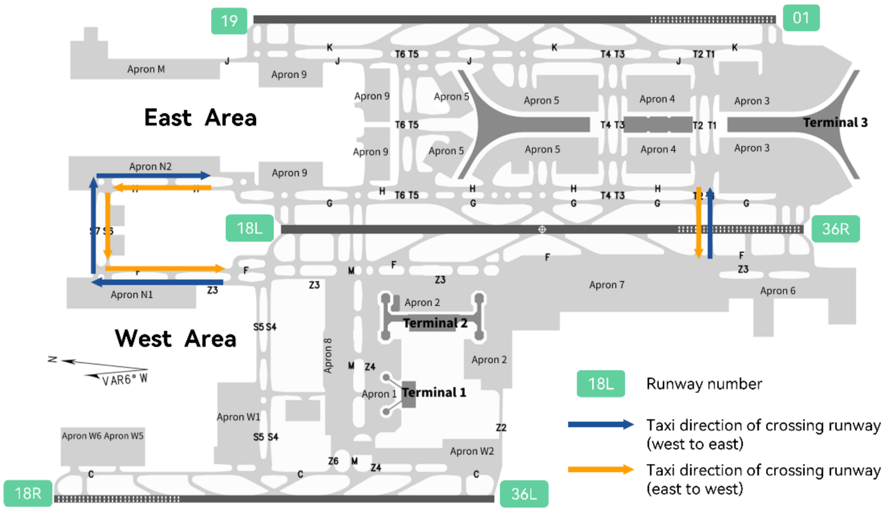


Figure 2 – Airport diagram of PEK

Table 2 – Statistical indicators of taxi-in time for flights at PEK

Parameter	Min	Med(Q2)	Max	Avg	25%(Q1)	75%(Q3)
Taxi time/min	1	11	59	12.35	7	17

3.1 Predictor variables

To accurately predict taxi times, it is necessary to extract features from various aspects. To achieve this goal, we have established 24 predictive variables, with a particular focus on "Surface traffic flow" and "Runway-crossing."

Table 3 – The predictor variables in the model

Variable	Type	Description
Surface traffic flow	Discrete	Number of arrival or departure flights that are moving on the runway, taxiway or apron during the time period between current flight landing and on block.
Runway-crossing	Categorical	Whether the flight crosses a runway. It is defined to be 1 for yes and 0 for no.
Time period	Categorical	Time period of a day. The clusters are [2:00–8:59], [1:00–1:59, 9:00–9:59], [0:00–0:59, 10:00–23:59] for three time periods, labelled as I, II, III respectively. Local time has been used.
Distance	Continuous	The distance of aircraft taxis from the landing point to the stand.
Runway used	Categorical	Runway number of arrival flights landing. Including 19, 01, 18L, 36R, 18R, 36L a total of six runways.
Aircraft type	Categorical	Aircraft type used on PEK arrival flights. Includes C (wing span is greater than or equal to 24 meters and less than 36 meters), D (wing span is greater than or equal to 36 meters and less than 52 meters), E (wing span is greater than or equal to 52 meters and less than 65 meters) and F (wing span is greater than or equal to 65 meters and less than 80 meters) aircraft types.
Type of airline	Categorical	Type of airlines. It is defined as 1 for domestic airlines and 0 for foreign airlines. This variable is designed to examine the difference between domestic and foreign airlines caused by language communication and familiarity level to local airports. For instance, a Chinese air traffic controller may need relatively longer response time when communicating with a foreign pilot in English.

3.2 Surface traffic flow

Designing a method to compute surface traffic flow for multi-runway airports presents challenges involving simultaneously considering arrival and departure flights and avoiding fixed threshold settings. To tackle these challenges, we introduce an approach for calculating surface traffic flow applicable to such airports. In this approach, we use the current flight's landing time and in-block time as the computation's start and endpoints. This calculation determines the count of various arrival and departure flights within this interval, obviating the need for fixed thresholds. Higher surface traffic flow values for the current flight indicate increased congestion during its surface movement.

According to *Equation 1*, we define the taxi time interval for the i-th arrival aircraft as $[t_{ALDT}(i), t_{AIBT}(i)]$, the taxi time interval for the j-th arriving aircraft as $[t_{ALDT}(j), t_{AIBT}(j)]$ and the taxi time interval for the k-th departing aircraft as $[t_{AOBT}(k), t_{ATOT}(k)]$. Here, t_{AOBT} represents the AOBT of the aircraft and t_{ATOT} represents the actual takeoff time.

Figure 3 illustrates four potential categories of arrival flights impacting the current flight, and likewise, four potential categories of departure flights affecting it, as displayed in their respective taxiing states. The calculation methods for these eight flight types are detailed in *Table 4*. *Figure 3* and *Table 4* illustrate the count of all flights j that meet the criteria for flight i and represents the count of all flights k satisfying the conditions for flight i.

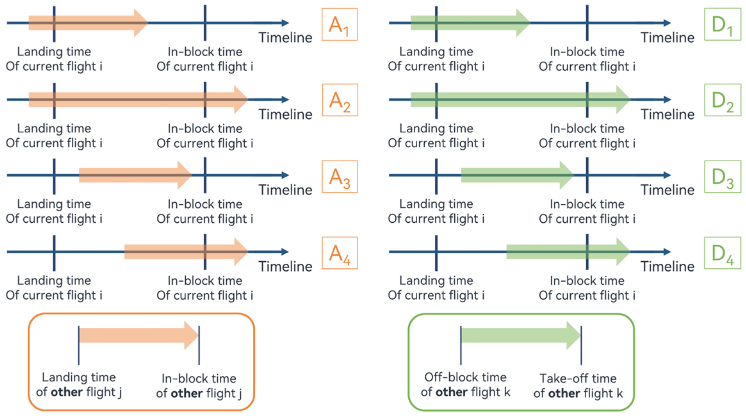


Figure 3 – Graphical description of the variables in surface traffic flow

Table 4 – Classification and description of surface traffic flow variable

Variable	Definition	Description
A_1	$A_1(i)=\sum count(j)$ $\begin{cases} count(j)=1, t_{ALDT}(j)<t_{ALDT}(i)<t_{AIBT}(j) \\ count(j)=0, otherwise. \end{cases}$	Number of other arrival aircraft landing before current aircraft landing, in block after current aircraft landing and before current aircraft in block.
A_2	$A_2(i)=\sum count(j)$ $\begin{cases} count(j)=1, t_{ALDT}(j)<t_{ALDT}(i)\cap t_{AIBT}(j)>t_{AIBT}(i) \\ count(j)=0, otherwise. \end{cases}$	Number of other arrival aircraft landing before current aircraft landing, in block after current aircraft in block.
A_3	$A_3(i)=\sum count(j)$ $\begin{cases} count(j)=1, t_{ALDT}(j)>t_{ALDT}(i)\cap t_{AIBT}(j)<t_{AIBT}(i) \\ count(j)=0, otherwise. \end{cases}$	Number of other arrival aircraft landing after current aircraft landing, in block before current aircraft in block.
A_4	$A_4(i)=\sum count(j)$ $\begin{cases} count(j)=1, t_{ALDT}(j)<t_{AIBT}(i)<t_{AIBT}(j) \\ count(j)=0, otherwise. \end{cases}$	Number of other arrival aircraft landing after current aircraft landing and before current aircraft in block, in block after current aircraft landing.
D_1	$D_1(i)=\sum count(k)$ $\begin{cases} count(k)=1, t_{AOBT}(k)<t_{ALDT}(i)<t_{ATOT}(k) \\ count(k)=0, otherwise. \end{cases}$	Number of other departure aircraft off block before current aircraft landing, take-off after current aircraft landing and before current aircraft in block.
D_2	$D_2(i)=\sum count(k)$ $\begin{cases} count(k)=1, t_{AOBT}(k)<t_{ALDT}(i)\cap t_{ATOT}(k)>t_{AIBT}(i) \\ count(k)=0, otherwise. \end{cases}$	Number of other departure aircraft off block before current aircraft landing, take-off after current aircraft in block.
D_3	$D_3(i)=\sum count(k)$ $\begin{cases} count(k)=1, t_{AOBT}(k)>t_{ALDT}(i)\cap t_{ATOT}(k)<t_{AIBT}(i) \\ count(k)=0, otherwise. \end{cases}$	Number of other departure aircraft off block after current aircraft landing, take-off before current aircraft in block.
D_4	$D_4(i)=\sum count(k)$ $\begin{cases} count(k)=1, t_{AOBT}(k)<t_{AIBT}(i)<t_{ATOT}(k) \\ count(k)=0, otherwise. \end{cases}$	Number of other departure aircraft off block after current aircraft landing and before current aircraft in block, take-off after current aircraft landing.

3.3 Runway crossing

Figure 2 illustrates PEK airport's three parallel runways, dividing it into two distinct areas. The East Area lies between Runway 01/19 and Runway 18L/36R, while the West Area extends between Runway 18L/36R and Runway 18R/36L. Arrival flights via Runway 18L/36R can access the East or West Area via separate taxi paths without runway crossings. However, arrivals from outer Runways 01/19 or 18L/36R may need to cross areas depending on stand assignments.

Two methods facilitate crossing Runway 18L/36R at PEK. The first involves using End-Around Taxiways (EAT): aircraft entering the West Area from the East use Taxiway S6. In contrast, those entering the East Area from the West use Taxiway S7, increasing taxi distance. The second method entails direct runway crossing, which may require waiting to avoid interfering with departing flights on Runway 18L/36R. Irrespective of the chosen method, flights crossing runways experience longer taxi times than others. Approximately 30% of PEK flights operate in the runway-crossing mode, their average taxi time exceeding twice that of non-crossing flights (*Table 5*). This data underscores the extended taxi times for crossing flights. In our model, we denote these as "Runway crossing" (C).

It is crucial to note that the "Runway-crossing" definition aligns with PEK's decentralised terminal layout and may not apply universally to multi-runway airports. For instance, airports like Heathrow Airport (LHR) and Nanjing Lukou International Airport (NKG) feature terminals predominantly positioned between two runways, requiring minimal or no active runway crossings during operations.

Table 5 – Percentage of runway-crossing flights in PEK and their average taxi-in time

Flights category	Proportion (%)	Average taxi-in time (min)
Runway-crossing flights	29	19.8
Non-runway-crossing flights	71	9.3

3.4 Time periods

Our model's time periods are denoted as variable TP. We collected landing times of all arrivals at PEK from 1 May to 31 December 2019. Each day was divided into 24 one-hour time periods. Based on the average number of arriving flights and the type of flights (i.e. passenger or cargo), a k-means clustering analysis was conducted to cluster the disaggregated time periods. The 24 one-hour time periods were categorised into the following segments: [2:00–8:59], [1:00–1:59, 9:00–9:59], and [0:00–0:59, 10:00–23:59], and labelled as I, II and III, respectively. TP-III experienced the highest average arrivals, while cargo flights dominated TP-I. Cargo flights at PEK tend to have more distant stands from runways than passenger flights, resulting in relatively longer taxi times in TP-I and TP-III. Boxplot of *Figure 4* illustrates taxi time variations across these periods.

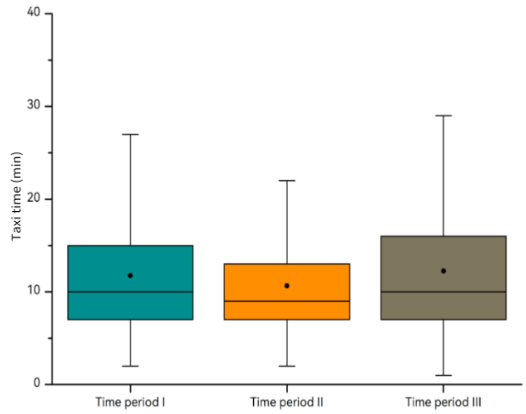


Figure 4 – Box plot of the taxi-in time for different time periods

4. METHODOLOGY

We first conducted a correlation analysis on various features. Based on the results of the correlation analysis, we constructed Linear Regression, Lasso and GBRT models separately, and subsequently compared and evaluated their predictive performance. Finally, we conducted pre-processing on the data, including one-hot encoding and standardisation.

4.1 Correlation analysis

Due to potential correlations among the variables in *Table 3*, we performed a correlation analysis on the variables in *Table 3*. Spearman's coefficient assessed correlations for continuous and categorical variables, Cramer's V for categorical variables and Eta squared for associations involving continuous, categorical and discrete variables. This paper has referred to Iversen's criterion, where they suggest that when the correlation coefficient falls between -0.3 to -0.7 or 0.3 to 0.7, it indicates a moderate correlation [33]. *Tables 6–8* display the results where the absolute coefficients greater than 0.3 are highlighted in bold.

Table 6 – Spearman coefficients for each pair of continuous variables and discrete variables

Variable	Distance	A1	A2	A3	A4	D1	D2	D3	D4
Distance	1	0.46	-0.50	0.64	0.55	0.55	-0.39	0.50	0.50
A1	0.46	1	-0.37	0.57	0.62	0.47	-0.28	0.37	0.42
A2	-0.50	-0.37	1	-0.58	-0.36	-0.48	0.46	-0.47	-0.45
A3	0.64	0.57	-0.58	1	0.61	0.62	-0.46	0.61	0.58
A4	0.55	0.62	-0.36	0.61	1	0.52	-0.30	0.42	0.52
D1	0.55	0.47	-0.48	0.62	0.52	1	-0.21	0.45	0.71
D2	-0.39	-0.28	0.46	-0.46	-0.30	-0.21	1	-0.48	-0.23
D3	0.50	0.37	-0.47	0.61	0.42	0.45	-0.48	1	0.46
D4	0.50	0.42	-0.45	0.58	0.52	0.71	-0.23	0.46	1

Table 7 — Cramer's V coefficients for each pair of categorical variables

Variable	Type of airlines	Runway-crossing	Aircraft type	Runway used	Time periods
Type of airlines	1	0.03	0.33	0.10	0.08
Runway-crossing	0.03	1	0.08	0.25	0.08
Aircraft type	0.33	0.08	1	0.11	0.06
Runway used	0.10	0.25	0.11	1	0.29
Time periods	0.08	0.08	0.06	0.29	1

Table 8 — Eta squared for each pair of continuous variables and categorical variables and for each pair of discrete variables and categorical variables

Variable	Distance	A1	A2	A3	A4	D1	D2	D3	D4
Type of airlines	<0.01	0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01
Runway-crossing	0.67	0.19	0.20	0.46	0.27	0.29	0.15	0.23	0.25
Aircraft type	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Runway used	0.01	<0.01	0.02	0.01	0.01	0.01	0.01	<0.01	0.01
Time periods	<0.01	<0.01	0.01	<0.01	0.02	0.07	0.07	<0.01	0.07

Tables 6–8 reveal correlations among variables. Notably, Distance correlates with Surface traffic flow and Runway-crossing. Longer taxi distances indicate higher surface traffic flow due to more conflicts during taxiing. Runway-crossing is affected by the detour via S6 and S7 taxiways, chosen by most flights crossing runways, increasing taxi distance.

Correlations also exist among the eight predictive variables within surface traffic flow. Multicollinearity challenges linear regression's effectiveness with these variables. Therefore, regularisation techniques and machine learning, like GBRT, are essential. Regularisation methods shrink coefficients of correlated variables, preserving significant ones. GBRT, an ensemble learning method based on decision trees, naturally handles high feature correlation by prioritising the most crucial features for branching. This makes it suitable for addressing multicollinearity and high dimensionality.

4.2 Linear regression

The objective of the linear regression is to minimise the sum of squared residuals of the response variable. The objective function is shown as follows:

$$D = \sum_{i=1}^{m} \left(y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2 \tag{2}$$

where m is the sample size, n is the number of predictor variables, y_i is the observed value of the response variable, and x_{ij} is the observed value of the predictor variable. B_0 represents the intercept of the model, and B_j represents the coefficient corresponding to the predictor variable. The commonly used linear regression model is the OLS model, which aims to minimise the objective function D.

4.3 Lasso

Lasso is built upon OLS to create an L1-regularised model. The Lasso model formulates an objective function as follows:

$$D = \sum_{i=1}^{m} \left(y_i - \beta_0 - \sum_{j=1}^{n} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{n} |\beta_j| \tag{3}$$

where D represents the objective function, and the aim of training the Lasso model is to minimise D. However, unlike OLS, the Lasso model introduces a penalty term $\lambda \sum_{j=1}^{n} |\beta_j|$ into the objective function. The parameter λ determines the extent of regularisation applied to the regression model. In the R software, the "glmnet" package is used to compute the Lasso model. The "cv.glmnet" function aids in selecting the most suitable λ value through n-fold cross-validation. As a regularisation method, Lasso produces reliable coefficients for each predictor variable in its results, even when there is strong correlation among predictor variables [32].

4.4 Gradient Boosting Regression Tree

GBRT, a common machine learning decision tree algorithm, constructs multiple regression trees collaboratively through iterative processes. Each iteration splits the dataset into two subsets, and feature selection minimises squared errors. GBRT's flexibility in handling variable types and superior predictive accuracy compared to traditional models are advantages. Yet, the interpretability of its results is reduced due to aggregating weighted tree sums. We implement the GBRT model using the "gbm" package in R.

4.5 Pre-processing

We employed One-Hot encoding to transform categorical variables into dummy variables and standardised the sample data. One-Hot encoding is a method that converts categorical variables into multiple dummy variables. After performing One-Hot encoding, a particular original variable is removed and replaced by n new dummy variables, where n depends on the number of distinct feature values in the original variable. Within the new set of dummy variables, only one dummy variable has a value of 1, while the rest have values of 0.

To standardise the units of different variables and to prevent certain variable coefficients from being excessively large or small, which could affect the interpretation and comparison of variable effects, a process of min-max normalisation was applied to the data where necessary.

5. RESULTS

5.1 Learning curves of models

Learning curves visualise changes in training and testing model scores with varying training set sizes for each algorithm. They help detect underfitting or overfitting caused by outliers during training and determine the required training samples for accurate predictions. R-squared and Root Mean Squared Error (RMSE) gauge model performance. R-squared explains variance, while RMSE measures prediction accuracy, sensitive to outliers. High R-squared and low RMSE denote well-fitted models. We plotted learning curves for training samples from 50 to 5000, focusing on statistical and GBRT models (*Figure 5*). We followed Viering's [34] criteria to assess curve stability: R-squared and RMSE differences between testing and training sets within 0.02, and no significant score changes with increasing samples.

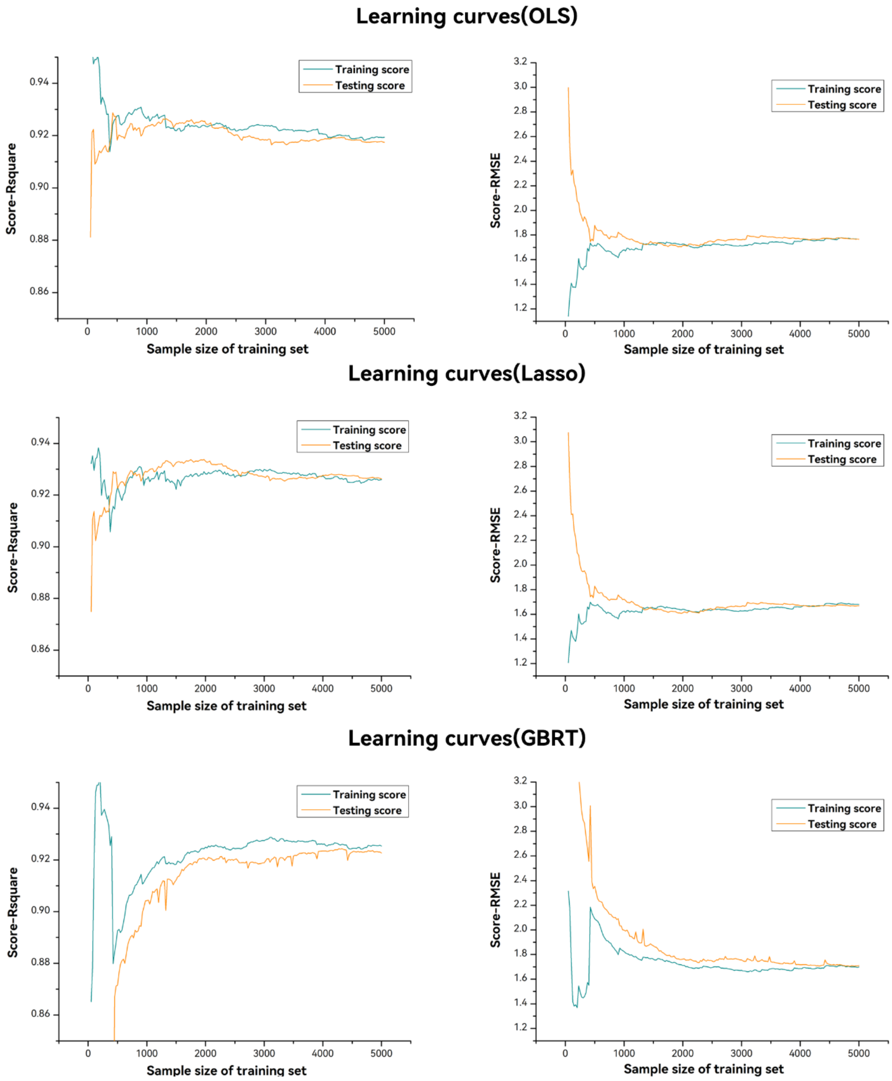


Figure 5 – Learning Curves of four models

GBRT converges slower on both training and testing sets than statistical models. For small datasets (<1000 samples), GBRT's convergence improvement is insignificant, raising the overfitting risk. But with larger training sets, all models converge. As the training set size grows, scores decrease for statistical models due to diminishing returns. Around 1500 samples, OLS and Lasso curves stabilise, while GBRT keeps converging. Around 3500 samples, GBRT surpasses others. These results from GBRT need more data to build a robust PEK data model. Choosing the right method based on sample size is vital for accuracy: smaller samples favour statistical methods, and larger samples favour GBRT. *Figure 5* shows no underfitting or overfitting issues with larger samples.

5.2 Evaluation of prediction results

R^2, RMSE, Mean Absolute Error (MAE), and Prediction Accuracy are utilised as the metrics to evaluate the predictive performance of different models. The Prediction accuracy is defined as the ratio of the number of model predictions within a certain set range of the actual taxi-in time to the total number of predicted samples. Existing research on departure taxi time prediction mainly focuses on prediction accuracy between ±3 minutes and ±5 minutes. However, considering that the average taxi time for arrival flights at the same airport is significantly lower than that for departure flights, we chose narrower ranges of ±1 minute, ±2 minutes and ±3 minutes as the standards for evaluation. The predictive performance of the three models is presented in *Table 9*.

Table 9 – Prediction performance metrics evaluation of four models

Performance measures		OLS	Lasso	GBRT
RMSE/min		1.690	1.617	1.568
MAE/min		1.177	1.124	1.088
R^2		0.908	0.932	0.938
Prediction accuracy	±1 min	48.8%	57.0%	58.6%
	±2 min	81.0%	85.5%	86.5%
	±3 min	90.5%	94.5%	95.5%

The GBRT model consistently exhibits the best predictive performance. The GBRT model achieves a prediction accuracy of 86.5% within a range of ±2 minutes and 95.5% within a range of ±3 minutes. Based on these results, the GBRT model is preferable for exit time prediction. However, predicting taxi time demands both accuracy and interpretable results for airport traffic management. Traditional stats lag GBRT in PEK data, yet GBRT lacks interpretability. Linear regression, notably Lasso, provides strong interpretability. Thus, we use Lasso results for variable interpretation.

Table 10 – Coefficients of the Lasso model for predicting taxi-in time (sorted in descending order of absolute values)

Variable	Coefficients	Coefficients of TP-I	Coefficients of TP-II	Coefficients of TP-III
A_3	27.859	35.352	53.705	26.289
D_3	24.440	35.277	51.778	24.829
A_4	8.091	5.707	10.662	8.281
D_1	7.749	6.046	9.076	8.762
A_2	-4.107	-5.951	-12.409	-3.572
D_4	3.896	5.707	8.687	4.193
Time periods-II	2.053	/	/	/
D	1.981	1.758	2.378	1.506
D_2	-1.748	-5.021	-12.761	-1.595
Runway-18R	1.221	0.817	0	1.223
Time periods-III	-1.166	/	/	/
Runway-crossing	1.016	0.679	2.590	0.975
Runway-18L	0.996	1.410	0.367	0.932
Runway-19	0.893	0.880	1.825	0.867
A_1	0.776	0.075	1.359	0.856
Aircraft type-D	-0.547	-0.549	-2.329	-0.070
Runway-01	-0.234	-0.389	-2.268	-0.169
Aircraft type-C	-0.220	-0.074	-0.397	-0.038
Runway-36R	-0.143	0.352	-0.580	-0.033
Type of airlines	-0.077	-0.415	-0.237	0.050
Aircraft type-E	0.001	0.065	0.072	0.148
Aircraft type-F	0	1.420	2.144	0
Runway-36L	0	0	0.085	0
Time periods-I	0	/	/	/

Compared to other features, the feature of surface traffic flow is the most important and impactful. The six features with the highest absolute coefficients are related to surface traffic flow: A_3, D_3, A_4, D_1, A_2 and D_4. Among the eight features related to surface traffic flow, A_3 and D_3 have the most significant impact on taxi-in time. This is because these two categories of flights directly occupy the taxiing resources of the current flight on the apron. However, for A_2 and D_2, the coefficients are negative. These two categories of flights land before the current flight and start taxiing after the current flight has landed and reached its stand. This indicates that these flights need to queue behind the current flight. Another possible reason is that these categories of flights are not on the same runway as the current flight, resulting in a smaller impact on the taxi-in time of the current flight.

Table 10 indicates that the runway-crossing operation mode significantly impacts taxi-in time, surpassing traditional variables such as aircraft type and type of airlines. The runway-crossing operation mode increases taxi-in time, aligning with our hypothesis since distance and time are typically correlated. However, we also observed that the coefficient of the runway-crossing is considerably different from those of A_3 and D_3. This discrepancy is due to the higher correlation between runway-crossing and Distance. Consequently, the coefficient of runway-crossing has been somewhat compressed in the Lasso model.

The impact of time periods is also relatively significant. The Lasso estimation shows that the taxi-in time in TP-II was longer than that in TP-I, while the taxi-in time in TP-III was statistically shorter than that in TP-I, which seems to be converse to the descriptive analysis in Fig. 4. After checking for other factors that have impacts on taxi-in time but might also correlate with the variable of time period, the real impact of the latter is revealed. This also indicates the explanatory capability of the Lasso model. In order to explore which factors statistically increase the taxi-in time in TP-II, we further run three other models for TP-I, TP-II and TP-III, respectively (*Table 10*). The results show that three variables, i.e. A3, D3 and aircraft type-F have larger impacts on taxi-in time in TP-II than the other two time periods. On the one hand, A3 and D3 represent the highest conflict level on the current flight from other flights on the surface. The larger the number of these two types of flights is, the longer the taxi-in time is. On the other hand, aircraft type-F have longer taxi-time than other types. Therefore, it can be seen that introducing controlling factors e.g. the surface traffic flow and aircraft type in the model can help identify the true impact of the exploring variables.

Taxi distance is generally considered one of the more influential variables in previous research, but in the Lasso model, its coefficient differs significantly from the coefficients of A_3 and D_3. This situation can be attributed to two main reasons. Firstly, the correlation between Distance and Surface traffic flow is relatively high. Yet, the Lasso model emphasises identifying the importance of Surface traffic flow compared to Distance, leading to the coefficient of Distance being shrunk. Secondly, the existing literature focuses on target airports such as ARN and HKG, which follow a centralised terminal layout that differs from the operational rules of PEK. This disparity caused a shift in the relative importance of taxi distance. Nonetheless, it remains undeniable that both Surface traffic flow and Distance are crucial variables in taxi time prediction studies.

The model results also indicate that the type of airlines has an impact on taxi-in time. Negative coefficients corresponding to the predictor variables suggest that flights operated by domestic airlines generally have shorter taxi-in times compared to those operated by foreign airlines. This phenomenon could be attributed to smoother communication between domestic airlines dispatchers and ATC towers, better familiarity with ground configurations, and fewer procedures to be arranged before reaching the stand. Both Type C and Type D aircraft have negative coefficients, while Type E aircraft have a coefficient greater than 0. This suggests that, holding other variables constant, taxi-in times for Type C and Type D flights are generally shorter than for Type E and Type F flights. This is because larger aircraft types have lower flexibility in ground operations. Unlike smaller aircraft types, they are subject to stricter gate assignment rules, and the number of available gates for their use is more limited. Additionally, larger aircraft types tend to have lower speeds during turns in ground operations than smaller aircraft.

It is worth noting that the choice of runway also plays a crucial role in influencing taxi-in time. Runways 19, 18L and 18R positively correlate with taxi-in time, while runways 01 and 36R negatively correlate with taxi-in time. Interestingly, the Lasso model deems runway 36L as an unrelated predictor variable. An analysis of PEK's layout indicates that when arrival flights use runways 19, 18L and 18R from the north, they perform a turn before entering the apron after leaving the runway. This increases the number of aircraft turns and the taxi distance. Conversely, they might taxi to the apron more swiftly when arrival flights approach from the south.

6. CONCLUSIONS

Existing studies on taxi-in-time prediction often lack sufficient attention to the characteristics of airport terminal layouts. In fact, compared to centralised terminal layouts, some flights at airports with decentralised terminal layouts require additional runway-crossing movements during actual operations. To address the gaps in existing methods for calculating surface traffic flow, this paper first proposes an improved multi-runway surface traffic flow calculation method based on a review of existing literature. Subsequently, an analysis is conducted for airports with decentralised terminal layouts, and new taxi-in-time prediction features are introduced. Given the potential high correlation among some influencing factors of taxi-in time, we employ Lasso regression to analyse these factors and use GBRT to predict the taxi-in time of arrival flights.

We chose PEK, one of the busiest airports in China, as our research subject. The results demonstrate that the GBRT model achieves a prediction accuracy exceeding 95% within a ±3-minute range. Despite Lasso being a statistical regression model with slightly lower prediction accuracy than GBRT, its ±3-minute prediction accuracy still surpasses 90%. The results of the Lasso regression model reveal that the feature of Surface traffic flow we designed is the most significant factor affecting flight taxi-in time. The introduced feature of Runway-crossing related to the decentralised layout also exhibits substantial coefficients, indicating a substantial impact of runway-crossing on flight taxi-in time. Finally, we discuss the importance of other features and provide insights for the airport operational management department.

These findings can assist airlines or ground handling departments in estimating aircraft in-block time with greater accuracy. Accurate prediction of taxi-in time for arriving flights can reduce the uncertainty in aircraft turnaround time. Additionally, it can aid airport authorities in allocating ground handling resources to aircraft at the right time, thereby enhancing resource utilisation efficiency. We have also validated that the airport's terminal layout is a crucial influencing factor for flight taxi-in time. Airport analysts should prioritise runway-crossing in their predictions of taxi-in time for airports with a decentralised terminal layout. In addition, as airports with other runway configurations (e.g. converging or crossing) or terminal layouts (e.g. centralised) become increasingly congested, the method proposed to measure the dynamic sufficient traffic flow can also apply for them by simultaneously integrating arrivals and departures.

ACKNOWLEDGEMENTS

This work was funded by the National Natural Science Foundation of China-Civil Aviation Administration of the China Joint Research Fund for Civil Aviation under grant number U2233208 and U2333204, the Civil Aviation Safety Capacity Building Project named "Optimisation of flight turnaround operations standards under the normal mechanism of pandemic prevention and control" ((2023) No. 155), and Nanjing University of Aeronautics and Astronautics (NUAA) 2023 Graduate Student Research and Practice Innovation Program Project (xcxjh20230736).

REFERENCES

[1] Balakrishna P, Ganesan R, Sherry L. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures. *Transportation Research Part C: Emerging Technologies*. 2010;18(6):950-962. DOI: 10.1016/j.trc.2010.03.003.

[2] Diana T. Can machines learn how to forecast taxi-out time? A comparison of predictive models applied to the case of Seattle/Tacoma International Airport. *Transportation Research Part E: Logistics and Transportation Review*. 2018;119:149-164. DOI: 10.1016/j.tre.2018.10.003.

[3] Lian G, et al. RETRACTED: A new dynamic pushback control method for reducing fuel-burn costs: Using predicted taxi-out time. *Chinese Journal of Aeronautics*. 2019;32(3):660-673. DOI: 10.1016/j.cja.2018.12.013.

[4] Tang X, et al. Taxi-in time prediction of arrival flight. *Journal of Beijing University of Aeronautics and Astronautics*. 2022. DOI: 10.13700/j.bh.1001-5965.2022.0625 [Accessed 26th Sept. 2023].

[5] Xia Z, Huang L. Prediction of departure flights' taxi-out time based on intelligent algorithm optimized BP. *Mathematical Problems in Engineering*. 2022;2022:1-12. DOI: 10.1155/2022/6254251.

[6] Park D, Kim J. Influential factors to aircraft taxi time in airport. *Journal of Air Transport Management*. 2023;106:102321. DOI: 10.1016/j.jairtraman.2022.102321.

[7] Andersson K, Carr F, Feron E, Hall WD. Analysis and modeling of ground operations at hub airports. *3rd Air Traffic Management Research and Development Seminar, 3-6 Jun. 2000, Napoli, Italy*. https://www.atmseminar.org/past-seminars/3rd-seminar/papers/ [Accessed 13th Jan. 2024].

[8] Andersson K. *Potential benefits of information sharing during the arrival process at hub airports*. Masters thesis.

Massachusetts Institute of Technology; 2000.

[9] Jiao Q, Li N. Taxi time prediction by using data driven approach: A new perspective. *SSRN Journal*. 2022. DOI: 10.2139/ssrn.4084964.

[10] Feng X, Meng J. Flight taxi-out time prediction based on KNN and SVR. *Journal of Southwest Jiaotong University.* 2017;52(5):1008-1014. DOI: 10.3969/j.issn.0258-2724.2017.05.023.

[11] Herrema F, et al. Taxi-out time prediction model at Charles de Gaulle airport. *Journal of Aerospace Information Systems.* 2018;15(3):120-130. DOI: 10/gc6dnh.

[12] Jordan R, Ishutkina MA, Reynolds T. AG statistical learning approach to the modeling of aircraft taxi time. *29th Digital Avionics Systems Conference,3-7 Oct. 2010, Salt Lake City, UT, USA*. IEEE; 2010. p. B.1-1-1.B.1-10. DOI: 10/fk68gp.

[13] Ravizza S, Atkin JAD, Maathuis MH, Burke EK. A combined statistical approach and ground movement model for improving taxi time estimations at airports. *Journal of the Operational Research Society.* 2013;64(9):1347-1360. DOI: 10/f467nd.

[14] Yin J, et al. Machine learning techniques for taxi-out time prediction with a macroscopic network topology. *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), 23-27 Sept. 2018, London, England, UK*. 2018. p. 1–8. DOI: 10.1109/DASC.2018.8569664.

[15] Li N, Jiao Q, Zhang L, Fan R. Taxi time prediction of departure aircraft. *Journal of Chongqing Jiaotong University (Natural Science)*. 2021;40(3):1-6.

[16] Wang X, et al. Aircraft taxi time prediction: Feature importance and their implications. *Transportation Research Part C: Emerging Technologies*. 2021;124:102892. DOI: 10/gq2mpk.

[17] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001;96(456):1348-1360. DOI: 10.1198/016214501753382273.

[18] Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*. New York, NY: Springer; 2009.

[19] Wu X, et al. Top 10 algorithms in data mining. *Knowl Inf Syst*. 2008;14(1):1-37. DOI: 10.1007/s10115-007-0114-2.

[20] Xu N, Sherry L, Laskey KB. Multifactor model for predicting delays at U.S. airports. *Transportation Research Record*. 2008;2052(1):62-71. DOI: 10.3141/2052-08.

[21] Srivastava A. Improving departure taxi time predictions using ASDE-X surveillance data. *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*, *16-20 Oct. 2011, Seattle, WA, USA*. IEEE; 2011. p. 2B5-1-2B5-14. DOI: 10.1109/DASC.2011.6095989.

[22] Ravizza S, et al. Aircraft taxi time prediction: Comparisons and insights. *Applied Soft Computing*. 2014;14:397–406. DOI: 10/gq2mpm.

[23] Lee H, Malik W, Jung YC. Taxi-out time prediction for departures at Charlotte airport using machine learning techniques. *16th AIAA Aviation Technology, Integration, and Operations Conference, 13-17 Jun. 2016, Washington, D.C.* 2016. DOI: 10/gq2mph.

[24] Chen Z, Tang X, Lin Y, Ren S. Prediction method and model of aircraft taxi-out time based on decision tree. *Journal of Wuhan University of Technology (Transportation Science & Engineering).* 2021;45(3):448–453.

[25] Zhao Z, et al. Prediction method of aircraft dynamic taxi time based on XGBoost. *Advances in Aeronautical Science and Engineering.* 2022;13(1):76–85. DOI: 10.16615/j.cnki.1674-8190.2022.01.08.

[26] Idris H, Clarke JP, Bhuva R, Kang L. Queuing model for taxi-out time estimation. *Air Traffic Control Quarterly*. 2002;10(1):1–22. DOI: 10/gq2mpn.

[27] Gilbo EP, Center V, Howard KW, Corp A. Collaborative optimization of airport arrival and departure traffic flow management strategies for CDM. *3rd Air Traffic Management Research and Development Seminar, 3-6 Jun. 2000, Napoli, Italy*. https://www.atmseminar.org/past-seminars/3rd-seminar/papers/ [Accessed 20th Jan. 2024].

[28] Günther Y, et al. Total airport management (operational concept & logical architecture) version 1.0. 2006.

[29] Yoo HS, et al. Benefit assessment of the integrated demand management concept for multiple New York metroplex airports. *AIAA Scitech 2020 Forum, 6-10 Jan. 2000, Orlando, FL, USA*. 2020. DOI: 10.2514/6.2020-1400.

[30] Pina P, Pablo JMD. Benefits obtained from the estimation and distribution of realistic taxi times. *6th Air Traffic Management Research and Development Seminar, 27-30 Jun. 2005, Baltimore, MD, USA*. https://www.atmseminar.org/past-seminars/6th-seminar/papers/ [Accessed 20th Jan. 2024].

[31] Tang X, Chen Z, Zhang S, Ding Y. Impact of apron spatial configuration on flight departure taxi time at busy airports. *Journal of Transportation Systems Engineering and Information Technology*. 2022;22(5):309–317. DOI: 10.16097/j.cnki.1009-6744.2022.05.032.

[32] Schreiber-Gregory D, Jackson HM. Regulation Techniques for multicollinearity: lasso, ridge, and elastic nets. *Proceedings of the SAS Conference Proceedings: Western Users of SAS Software, 5-7 Sept. 2018, Denver, CO, USA*. 2018. p. 8-11.

[33] Iversen GR, Gergen M. *Statistics: The conceptual approach*. New York, NY: Springer; 1997.

[34] Viering T, Loog M. The shape of learning curves: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022. http://arxiv.org/abs/2103.10948. [Accessed 26th Sept. 2023].

唐小卫 叶梦凡 张生润 Kurt Fuellhart

分散式航站区布局机场的进港航班滑入时间预测研究及影响因素分析

摘要

准确预测进港航班的滑行时间对于高效分配地面保障资源至关重要，且影响着飞机能否准时起飞。本研究调查了航站楼的布局特征，特别是机场的分散式布局，以预测和分析进港航班的滑行时间。我们制定了一种考虑进港和离港航班的场面流量计算方法，避免了固定阈值的设置。我们引入了适用于分散式航站区布局机场的跨跑道运行，创建了新的预测变量。我们考虑了诸如跑道配置、飞机类型、航空公司、滑行距离和运行时间等因素。梯度提升回归树用于预测滑行时间，而Lasso用于分析各因素的影响。我们的方法能够准确预测分散式机场中航班的滑行时间，场面流量和跨跑道变量在影响滑行时间方面起到了显著作用。这项研究为分散式布局的机场管理者提供了有效的理论支持，使他们能够制定针对性的管理策略。

关键词：

航空运输；滑入时间预测；航站楼布局；机场场面运行；Lasso；梯度提升回归树；机场管理