33RD CONGRESS
OF THE INTERNATIONAL COUNCIL
OF THE AERONAUTICAL SCIENCES
STOCKHOLM, SWEDEN, 4-9 SEPTEMBER, 2022

ICAS 2022 SWEDEN

APPLYING MACHINE LEARNING TO TAXI-TIME PREDICTION AT TOKYO INTERNATIONAL AIRPORT

Furuto Kato[1] & Eri Itoh[1,2]

[1]The University of Tokyo, 7-3-1, Hongoh, Bunkyo-ku, Tokyo, 113-8656, Japan
[2]ENRI, 7-42-23 Jindaiji higashi-machi, Chohu, Tokyo 182-0012, Japan

Abstract

Reducing uncertainties in the estimation of aircraft departure taxi-time is one of the most important measures for maximizing the runway throughput while also saving fuel consumption and mitigating air traffic congestion in airport operations. A previous study applied queuing models to three major airports in the United States to predict taxi times of departure aircraft. However, the prediction error margins were within five minutes for just 60% to 70% of flights, which is not accurate enough to be able for application to real-world aircraft departure management. This study proposes a methodology for applying Machine Learning (ML) methods as possible alternatives for accurate prediction of departure aircraft taxi time. We used flight plans, radar data, and spot assignment charts of actual flights which departed from the case study airport, Tokyo International Airport, and developed a taxi-out model for departure aircraft traffic. For the taxi- time estimation, we applied five methods: Linear Regression, Elastic Net, Random Forest Regression, Gradient Boosting Regression, and XGBoost, and compared the prediction accuracy for each to determine the best ML method. Some past studies suggested that decision-tree-based methods such as random forest regression and gradient boosting regression are better suited for flight time prediction of aircraft cruising or approaching destination airports. We assumed that these decision-tree-based methods would also predict taxi time of aircraft on the ground more accurately than linear regression models.

Keywords: air traffic management, machine learning, taxi time, prediction

1. Introduction

Air transportation demand has been growing rapidly around the world. According to a survey published by the FAA [1], the air travel passenger levels are increasing by around 4% per year in the U.S. Although this growth fell far short of the forecast due to the COVID-19 pandemic in 2020, the demand is expected to recover by 2024 and continue on this trajectory [2]. Major global hub airports are already struggling to meet high demand, and departure delays are constantly occurring during congested hours [3].

Expanding the airport infrastructure such as by constructing new runways is one possible option to increase the airport capacity, but this costs huge amounts of money and takes significant time. For this reason, improving the airport operation efficiency has been suggested as a measure to deal with the congestion with limited infrastructure.

One of the ideas to improve the airport operation efficiency is integrated arrival and departure management, which achieves the optimal departure aircraft flow entering the departure runway in a unit of time [4][5]. One element of this is to apply departure metering, which assigns suitable hold times for departure aircraft at their gates. This idea is expected to reduce both departure delay and fuel consumption while taxiing. However, departure metering is only possible if taxi times for departure aircraft can be accurately predicted. Due to this requirement, many previous studies have tried to predict aircraft taxi times using various methodologies.

Sadeep et al. [8] applied queuing models to predict taxi time of departure aircraft from three major

US airports. The results showed that the prediction error of only 60% to 70% of flights had prediction error margins greater than five minutes.
Another research applied several machine learning algorithms to aircraft taxi time prediction at Charlotte Douglas International Airport ([6]). They analyzed number of data elements to identify the key factors affecting the taxi times and the unique operational characteristics of the airport. Although they compared five machine learning methods with carefully chosen features, random forest regression, which showed the best prediction performance among the methods they adopted, still failed to predict more than 30% of departures within an error margin of 5 minutes, although performance was slightly better than the aforementioned results of queuing models. These results suggest that modifying the prediction models and finding better feature sets are necessary to improve the taxi time prediction accuracy at large and busy airports.
Xinwei W et al. [7] compared multiple prediction models using real-world data for several international airports, and investigated the impact of various features on the prediction accuracy. They concluded that high accuracy can be achieved with a group of features that are generally important across all airports along with a small number of features specific to particular target airports. The generally important features included taxiing distance, total turns in the taxiing routes, and number of aircraft departing within a certain period of time. This paper also showed that weather condition features had little impact on taxi time, based on comparisons of 11 weather-related features.
Tokyo International Airport (RJTT), also known as Haneda Airport, is one of the busiest airports in the world. With rapidly growing air transportation demand, this airport for which improving the efficiency of airport operations to mitigate congestion and delay is extremely important. The previous studies cited above suggested that we need to identify key factors affecting taxi time at the particular airport we are interested in and also improve the methods used to achieve more accurate predictions. Considering these points, this study aims to develop a prediction method for RJTT that can achieve accurate taxi time prediction with prediction error margins small enough to be suitable for real-world applications in departure metering.
This paper is organized as follows. Section 2.explains the characteristics of RJTT. This section also covers the results of data analysis related to taxiing time at this airport. Section 3.presents the prediction methods and Section 4.covers the results. Finally, Section 5.contains our conclusions and possible topics for future study.

2. Data Analysis of Departure Traffic at Tokyo International Airport

2.1 Ground and Runway Operations at Tokyo International Airport

RJTT has four runways and three terminals, as shown in Figure 1. The runways include two parallel north-south runways (34L/16R, 34R/16L) and 2 northeast-southwest runways (04/22, 05/23). How these four runways are used is determined based on the wind directions (Figure 2). In the northerly wind conditions, runways 05 and 34R are used for southbound and northbound departures, respectively, and the parallel 34L and 34R runways are used for In arrivals. In southerly wind conditions, there are two patterns used. The new pattern, which is to be the main one for operations in the near future, uses three runways (16L/R, 22) for departures and two runways (16L/R) for arrivals. The data shows that around 70% of all yearly departures from RJTT use runways 05 and 34R. The data also shows that departures and arrivals to and from airports south of RJTT such as Osaka and Fukuoka are dominant (about 70% of the total flights), so runways 05 and 34L are busy in northerly wind conditions.
The previous research has confirmed that which departure runway is used is one of the most important factors for predicting taxi times of departure aircraft. For this reason, we decided to try taxi time prediction for departures from runway 05, which is the most used departure runway at RJTT.
One of the important characteristics to note is that taxiing time to runway 05 may differ greatly depending on the departure terminal used. Figure 1 shows the basic taxiing routes to runway 05 from Terminals 2 and 3. It is apparent from the figure that the taxiing distance from the farther terminal (Terminal 3) is much longer than from the closer terminal (Terminal 2). Also, every departure from Terminal 3 must cross runway 34L, on which more than 70% of arrivals land. This suggests that there might be different taxi time prediction key factors for departures from different terminals.

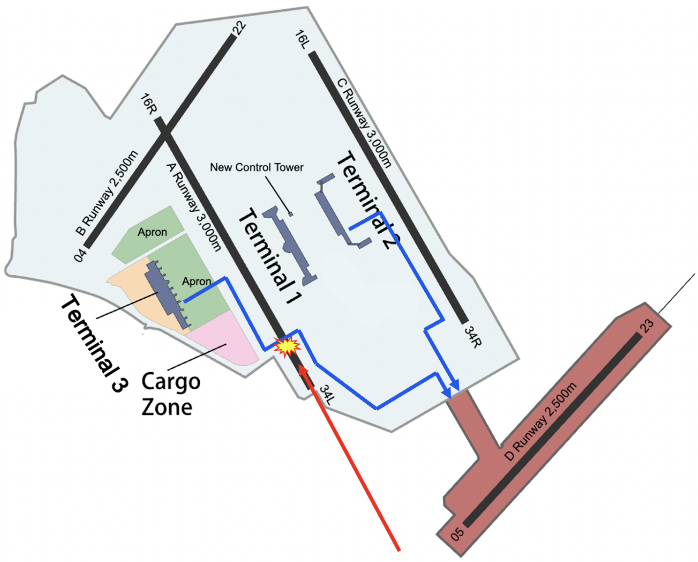
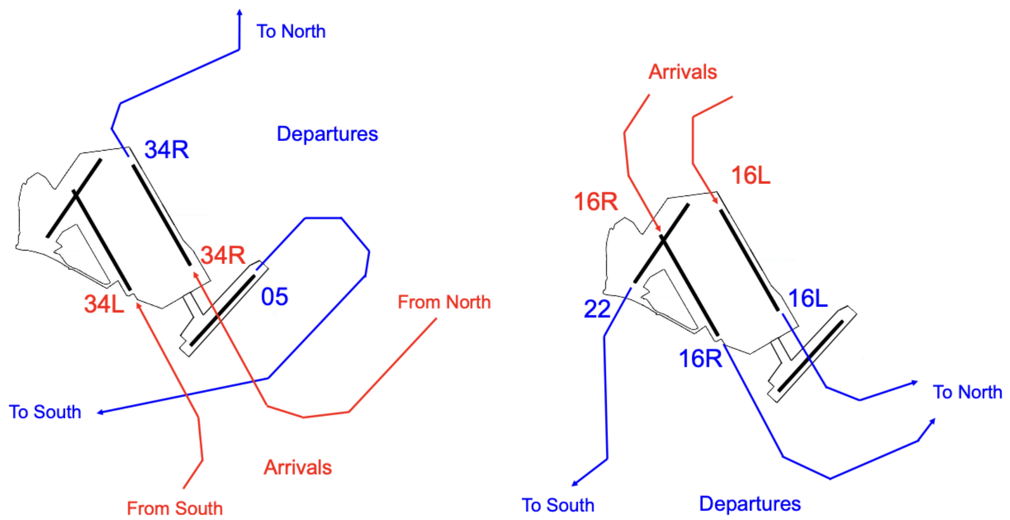


Figure 1 – Runway and terminal configurations of RJTT



(a) North wind operation (b) Southerly wind operation

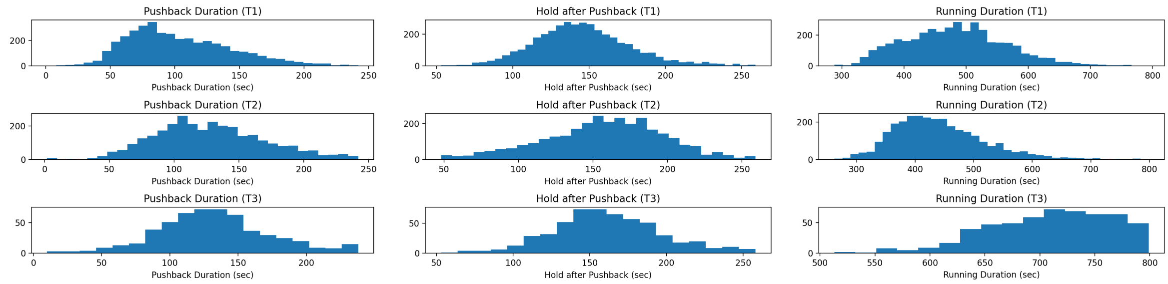
Figure 2 – The runway utilization patterns at RJTT

2.2 Taxiing Time Data Analysis

Taxiing time in this study is defined as the time period between the moment an aircraft starts pushing back from the gate and the moment it reaches the line around 0.5 NM in front of the runway hold line. Since our goal is to mitigate departure queues by applying the prediction results to departure metering, we needed to predict taxiing time without the effects of the queuing time. The line we used as the end of taxiing roughly represents the tail of the longest departure queue in front of runway 05 as observed in the data.

The taxiing operation defined above can be divided into three phases: pushback, taxi preparation, and running. Pushback is the phase in which an aircraft is pushed from the gate onto a taxiway by a towing truck. An aircraft holds its position for a few minutes after pushback, in order to complete taxi preparations both in the cockpit (starting the engines and setting up the computers) and on the ground (detaching from the towing truck). Afterward, the aircraft can finally start running to the departure runway with its own engines. We paid attention to this point, suspecting that key factors affecting the time of each phase might differ and therefore developing different prediction models for

each phase to account for this in hopes of improving the overall prediction accuracy.
In this study, we used flight plans, radar data, and spot assignment charts of actual flights which departed from RJTT during the 38-day period between September 2019 and February 2020. The data contained the location of each aircraft, as well as its ground speed, heading, and altitude recorded every second. Categorical data such as the type of aircraft, departure spot and runway, airline, and destination was used as well. We divided the time history data of each aircraft into the aforementioned three phases based on the ground speed (Figure 3).



(a) Pushback time (b) Taxi preparation time (c) Running time

Figure 3 – Histogram of the times for each taxiing phase (Terminal 1, 2, 3 from the top to the bottom)

3. Taxi Time Prediction for Departures at Tokyo International Airport

In the following two sections, we will explain our RJTT taxi time prediction experiment. Section 3. describes the features, methods, and performance metrics we used in this experiment. We conducted several experiments, and the results are explained and discussed in Section 4.

3.1 Factors for Taxi Time Prediction

Table 1 shows the factors we used for taxi time prediction in this study. These factors were chosen based on previous studies ([6] [7]). Airline and Destination are category factors consisting of some airlines and destinations in each category. Examples of Airline include "JAL," which referes to Japan Airlines only, and "Other_Japan," which refers to Japanese airlines except for JAL, ANA, and SKY. Some examples of Destination categories are "Japan", "Other_Asia", "Europe." Spot ID refers to the alphabetical IDs of close spot groups as shown in Figure 4. Numerical factors include those related to airport congestion levels such as the number of recent departures and arrivals, those related to the taxiing distance such as the direct distance, and those related to weather. The congestion factors had a relatively large impact in the previous study ([7]), and we assumed that they would have an impact on taxi time at RJTT to some extent as well. Weather data was obtained from the database provided by the the Japan Meteorological Agency. Many previous studies used more direct features such as taxi distance and pushback distance, but we did not because taxiing route may not yet be determined at the moment of prediction; before pushback starts.

Table 1 – Factors used for prediction

Type	Factor Name	Explanation
Categorical	Airline	Group of Airlines
	AircraftWeightCategory	Weight category of the aircraft
	Destination	Group of destinations
	SpotID	Group ID of departure spot
	DepTimeHour	Departure time (0 to 23)
Numerical	DirectDistance	Direct distance between spot and runway (km)
	Temperature	Celsius temperature at departure time
	Rainfall	Amount of rainfall (mm) at departure time
	WindVelocity	Wind velocity (m/s) at departure time
	CloseDepCount	Number of recent departures
	CloseArrCount34L	Number of recent arrivals to 34L

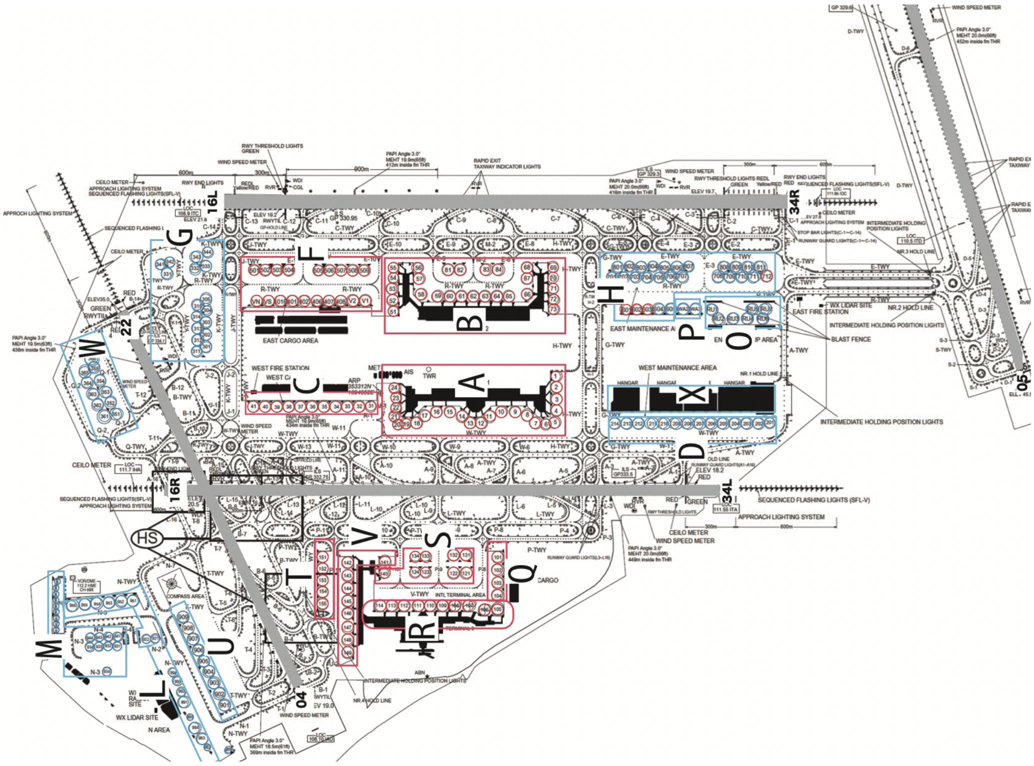


Figure 4 – Spot group ID [9]

3.2 Machine Learning Methods

In this study, we compared five machine learning methods: Linear Regression (LR), Elastic Net (EN) [10], Random Forest Regression (RFR) [11], Gradient Boosting Regression (GBR) [12], and Extreme Gradient Boosting, also known as XGBoost (XGB) [13]. The last three methods (RFR, GBR, XGB) are decision-tree-based methods. Various studies have shown that decision-tree-based methods performed better than linear regression methods (LR, EN) for predicting aircraft taxi time and flight time ([6] [7]). Therefore, we tested RFR and GBR, which have been frequently used in previous studies, and XGB, a relatively new method which has not been tested very much for taxi time prediction as of yet.

3.3 Performance Metrics

We evaluated the prediction results using Mean Absolute Error (MAE), Root Mean Squared Error ($RMSE$), R^2, and accuracy. These performance metrics are calculated using the following equations.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n}y_i\right)^2} \tag{3}$$

$$accuracy = \left(1 - \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right) \times 100\,(\%) \tag{4}$$

where n is the number of data samples, y_i is the actual value of sample i and $\hat{y}_i$ is the predicted value of sample i.

4. Prediction Results

4.1 Predicting the Entire Taxiing Duration

Figure 5 and Table 2 shows the prediction results for each prediction method. The best value of each metric is highlighted in bold. As shown in Table 2, decision-tree-based methods predicted taxi time more accurately than linear regression methods. This result corresponded to the aforementioned results from past studies, which showed that decision-tree-based methods were better suited for prediction of air traffic flow. XGB achieved the best performance in all of the four performance metrics, with GBR and RFR trailing close behind.

Table 3 shows lists of the top three most important factors for each method. One interesting point to note is that both of SpotID and DirectDistance have a relatively large importance on RFR and GBR, but only SpotID has a large impact on XGB. This may be explained by the fact that XGB is well-suited to handling factors which have strong correlation to each other. We included three weather-related factors, but none of these had a major impact on the taxi time, corroborating the findings of [7].

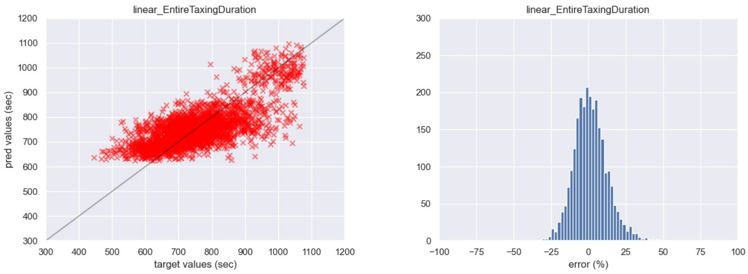
The MAE by XGB is less than a minute, which is much smaller than the margin of error in previous studies. Although this may be partly because we did not include runway queue in our prediction targets, this result still suggests that ML with decision-tree-based methods can predict taxi time with sufficient accuracy for departure metering applications.

Table 2 – Prediction results and metrics for the four methods

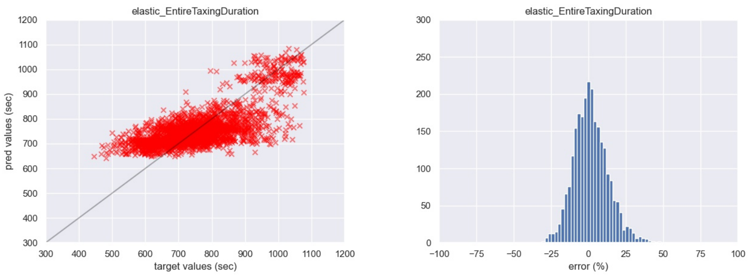
	LR	EN	RFR	GBR	XGB
$MAE(sec)$	63.74	66.21	58.00	57.44	**56.77**
$RMSE(sec)$	81.21	83.93	74.53	73.95	**73.29**
R^2	0.528	0.496	0.603	0.609	**0.616**
$accuracy(\%)$	91.37	90.98	92.19	92.26	**92.36**

Table 3 – Top three most important factors

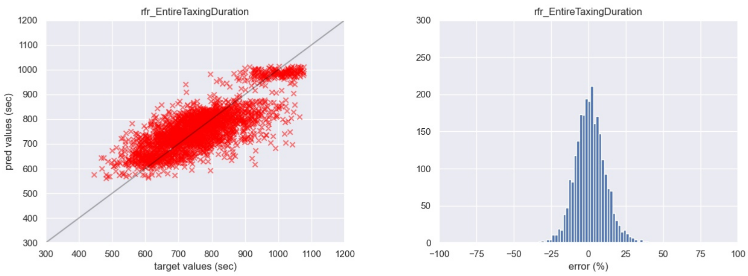
RFR		GBR		XGB	
SpotID	0.33	SpotID	0.49	SpotID	0.63
DirectDistance	0.32	DirectDistance	0.35	Destination	0.14
Destination	0.16	Airline	0.05	DirectDistance	0.09



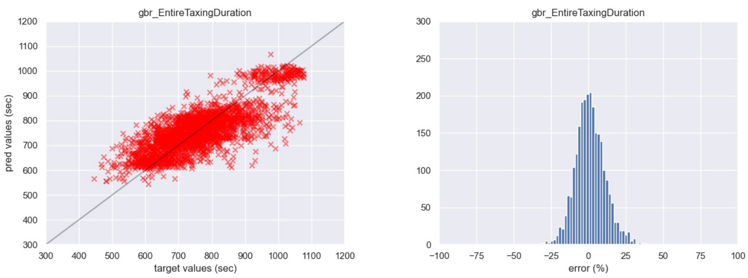
(a) LR



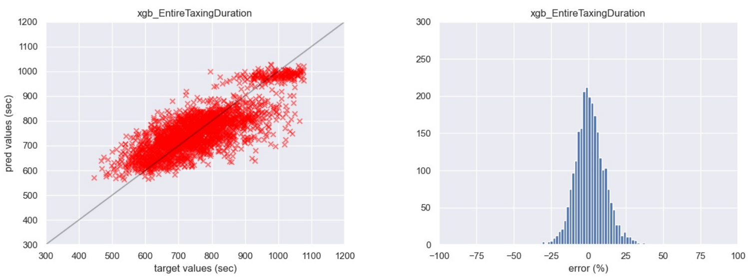
(b) EN



(c) RFR



(d) GBR



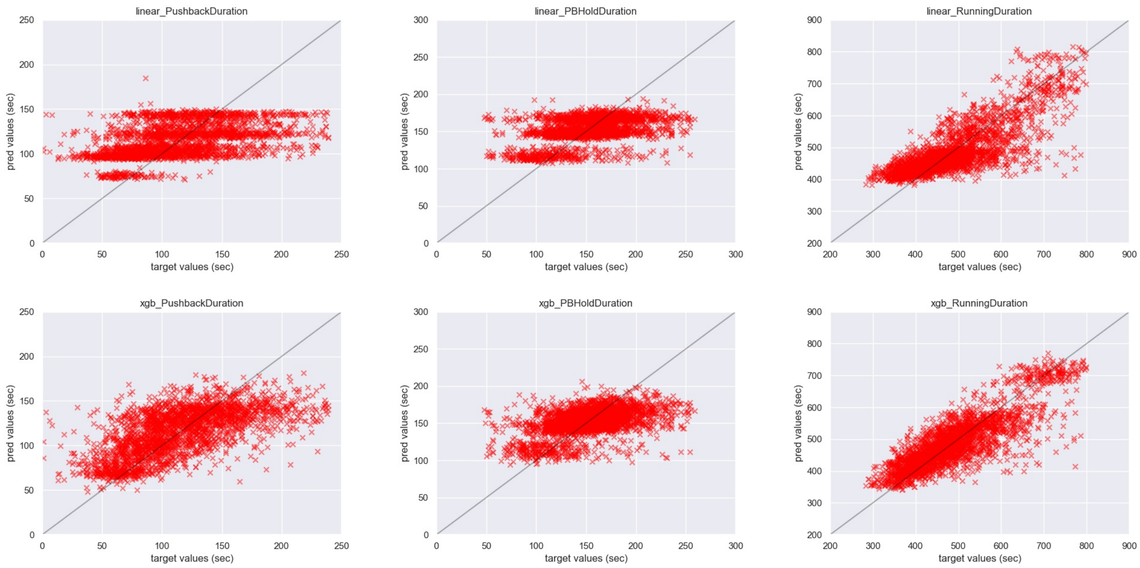
(e) XGB

Figure 5 – Left: Scatter plots of prediction results (actual value - predicted value), Right: histograms of prediction error margins (%)

4.2 Predicting Each Phase of Taxiing Operation and from Different Terminals with Different ML Models

As mentioned in Section 2,each taxiing phase seems to have different factors which are particularly important. In addition, departures from different terminals are impacted differently by these factors as well because of the different taxiing routes and runway crossings. Focusing on these points, we attempted to improve the prediction accuracy by applying different models to different taxiing phases and to aircraft from different terminals. We chose LR and XGB for this experiment because they achieved the best performance in the previous experiment in the linear-regression-based methods and decision-tree-based methods, respectively. First, all the flight data samples were divided into nine groups; three phases each for aircraft departing from three terminals. Next, we trained different regression models using each set of data. Finally, we predicted the testing data and integrated the results from the nine models.

Figure 6 and Table 4 show the prediction results for each taxiing phase. From Table 4 we can see that running time is predicted more accurately compared to the other two phases. This is reasonable because the time of pushback and taxi preparation largely depends on how smooth the ground handlers' operations proceed, which cannot be measured from aircraft movement data. In pushback and running preparation, XGB performed much better than LR. On the other hand, the difference in margin of error between XGB and LR was not so large for taxi preparation time. This indicates that there are almost no factors that have an impact on this time, and therefore the prediction results do not depend on the prediction methods used.



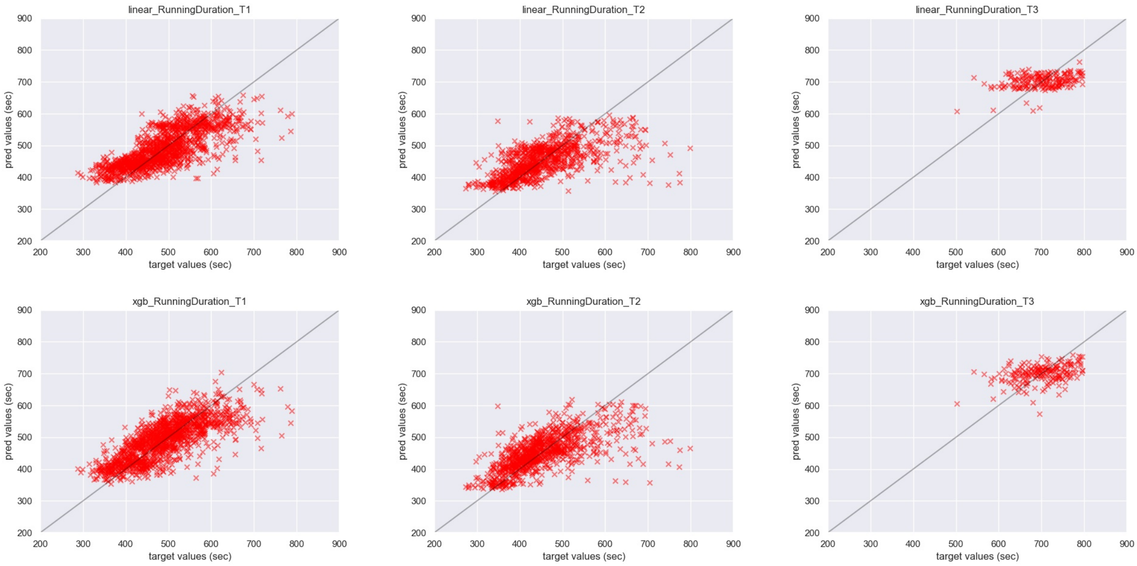
(a) Pushback (b) Taxi preparation (c) Running

Figure 6 – Scatter plots of prediction results for different taxiing phases (top: LR, bottom: XGB)

Table 4 – Prediction results and metrics for each phase

	Pushback		Taxi preparation		Running	
	LR	XGB	LR	XGB	LR	XGB
$MAE(sec)$	30.62	25.66	24.90	24.33	47.86	42.48
$RMSE(sec)$	38.50	33.19	32.63	31.77	62.81	56.83
R^2	0.191	0.245	0.186	0.228	0.626	0.694
$accuracy(\%)$	55.17	62.50	81.70	82.26	90.12	91.33

Figure 7 and Table 5 show the prediction results of departures from different terminals. The result for Terminal 3 has the largest deviation (lowest R^2) but the smallest $RMSE$. This large deviation can be explained by the existence of runway crossing. From Figure 7 we can clearly see that departures from Terminal 3 take a lot longer to reach the runway than those from the other terminals. This is interesting because we expected that the longer the average running time was, the larger the prediction error would be, but the result was the complete opposite. One possible explanation is that we have close arrival counts on 34L, which clearly has a major impact on departures from Terminal 3 but little impact on departures from the other 2 terminals.



(a) Terminal 1 (b) Terminal 2 (c) Terminal 3

Figure 7 – Scatter plots of running time prediction results for different terminals

Table 5 – Prediction results and metrics for different terminals

	T1		T2		T3	
	LR	XGB	LR	XGB	LR	XGB
$MAE(sec)$	42.58	41.61	45.88	44.24	43.26	41.53
$RMSE(sec)$	54.83	53.65	65.63	63.48	51.73	51.15
R^2	0.546	0.566	0.388	0.428	0.107	0.126
$accuracy(\%)$	91.20	91.46	90.11	90.51	93.67	93.94

Finally, Table 6 shows the comparison between the results for using only one model versus using divided models. We can see that in LR, divided models performed better in all the performance metrics. On the other hand, there were no improvements in the results of XGB. The result of a single XGB model was still better than the improved LR models.

Table 6 – The results for one model and divided models

	LR		XGB	
	1 model	divided models	1 model	divided models
$MAE(sec)$	63.74	60.88	**56.77**	58.40
$RMSE(sec)$	81.21	77.93	**73.29**	74.96
R^2	0.528	0.572	**0.616**	0.604
$accuracy(\%)$	91.37	91.82	**92.36**	92.18

5. Conclusions

In this research, we developed prediction models for aircraft taxiing time at Tokyo International Airport. We selected 11 factors based on previous studies. This factor sets contained airport-related, aircraft-related, and weather-related elements. For our prediction methods, we compared five different approaches, including a relatively new XGB along with frequently-used methods such as LR and

RFR. The results showed that decision-tree-based methods, especially XGB, achieved significantly improved performance compared to linear-regression-based methods in all of the metrics we used. In these experiments, airport-related factors such as departure spot and the distance between the spot and the runway had big impact on the taxi time. Weather-related features like temperature and rainfall amount were also included in the factor sets but none of them significantly affected taxi time. We also tried to improve the prediction accuracy by dividing the data samples by taxiing phases and departure terminals, then training different models for each set of data. This slightly improved the LR prediction accuracy but was not effective for improving XGB. Still, the mean absolute error of XGB was less than 1 minute, which we think is small enough for departure metering applications.
Our next project will be to develop prediction models for other runways at RJTT. Departure runways have been proven to have a major impact on taxi time, so we may need to develop completely different models for different departure runways. We will then apply the prediction models to departure metering and investigate the effects of departure metering on taxi delay and fuel consumption. We are currently developing an air traffic flow simulator using AirTOp software which will be used to simulate departure metering on this simulator to evaluate its effectiveness.

6. Acknowledgements

This research was supported by Grant-in-Aid for Scientific Research (20H04237). We would like to express our sincere appreciation for the data and technical support provided by the Civil Aviation Bureau of the Ministry of Land, Infrastructure, Transport and Tourism.

7. Copyright Statement

References

[1] FAA. Aerospace Forecast Fiscal Years 2020-2040, March 2020.
[2] IATA. "COVID-19 Outlook for air travel in the next 5 years," May 2020.
[3] Cynthia, B et al, Demand and capacity management in air transportation, EURO J Transp Logist, April 2012.
[4] Itoh, E.; and Mitici, M. Analyzing tactical control strategies for aircraft arrivals at an airport using a queuing model, Journal of Air Transport Management, 89, 101938, 2020.
[5] Itoh, E.; Mitici, M.; and Michael, S. Modeling Aircraft Departure at a Runway Using a Time-Varying Fluid Queue, Aerospace, February 2022.
[6] Hanbong Lee and Waqar Malik,"Taxi-out Time Prediction for Departures at Charlotte Airport Using Machine Learning Techniques", AIAA Aviation, June 2016.
[7] Xinwei W et al, "Aircraft Taxi Time Prediction: Feature Importance and their Implications", Transportation Research Part C Emerging Technologies, March 2021.
[8] Sadeep, B.; Hamsa, B.; Emily, J.; Tom, R. Impact of Off-Block Time Uncertainty on the Control of Airport Surface Operations, Transportation Science, April 2018.
[9] Ministry of Land, Infrastructure, Transport and Tourism. AIS Japan—Japan Aeronautical Information Service Center. 2021. Available online: `https://aisjapan.mlit.go.jp` (accessed on 24 May 2022).
[10] Zou Hui and Hastie Trevor, "Regulation and Variable Selection via the Elastic Net", Journal of the Royal Statistical Society, Series B, 67(2), pp. 301-320, 2005.
[11] Leo Breiman, "Random Forests", Machine Learning 45, 5-32 (2001)
[12] Llew M; Jonathan B; Peter B; Marcus F, "Boosting Algorithms as Gradient Descent"
[13] Tianqi, C.; Carlos, G. XGBoost: A Scalable Tree Boosting System, August 2016.