Causal effects of landing parameters on runway occupancy time using causal machine learning models

Lim, Zhi Jun; Goh, Sim Kuan; Dhief, Imen; Alam, Sameer

2020

Lim, Z. J., Goh, S. K., Dhief, I., & Alam, S. (2020). Causal effects of landing parameters on runway occupancy time using causal machine learning models. Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI), 2713-2722. doi:10.1109/SSCI47803.2020.9308243

https://hdl.handle.net/10356/146537

https://doi.org/10.1109/SSCI47803.2020.9308243

Causal Effects of Landing Parameters on Runway Occupancy Time using Causal Machine Learning Models

Zhi Jun Lim*, Sim Kuan Goh†, Imen Dhief‡, Sameer Alam§
Air Traffic Management Research Institute, School of Mechanical and Aerospace Engineering
Nanyang Technological University, Singapore
Email: { *zhijun001 | †skgoh | ‡imen.dhief | §sameeralam }@ntu.edu.sg

***Abstract*—Limited runway capacity is a common problem faced by most airports worldwide. The two important factors that affect runway throughput are the wake-vortex separation and Runway Occupancy Time (ROT). Therefore, to improve runway throughput, Wake Turbulence Re-categorisation program (RECAT) was introduced to reduce the minimum separation distance required between successive aircraft on final approach. As a result, the constraining impact of ROT on runway throughput has now become significant. The objective of this paper is to identify data-driven intervention to reduce the ROT of landing aircraft. Specifically, we propose a data- driven approach to estimate the causal effect of landing parameters on ROT. We propose categorisation of each landing parameter into groups using Gaussian process models and employ Generalised Random Forest (GRF) to estimate the average treatment effect and the standard deviation of each landing parameters. Experimental results show that a few procedural changes to current landing procedure may reduce ROT. The results establish that slowing down the aircraft speed in the final approach phase leads to shorter ROT. In the final approach phase, ROTs of aircraft which are at least 10 knots slower than the average aircraft speed are on an average 2.63 seconds shorter. Furthermore, aircraft that are at least 10 knots faster than the average aircraft have on average 4 seconds longer ROTs. The second finding of this work is that flexible glide-slope angles should be introduced for the different aircraft types to achieve better ROT performance. Therefore, our findings also validate the industry need for Ground-Based Augmented System landing system which provides landing guidance with flexible glide-slopes.**

***Index Terms*—causal machine learning, generalized random forest, ROT**

I. Introduction

The continuous growth of air traffic worldwide solicits major airports to efficiently optimize their resources to accommodate the increase in traffic demands. Nevertheless, airports are usually constrained by the physical capacity of runways and taxiways [1], [2]. A cost-efficient alternative is to increase runway throughput by looking at its limiting factors, namely the wake-vortex separation and the Runway Occupancy Time (ROT) [3].

In this regard, the Wake Turbulence Re-categorization program (RECAT) [4] was introduced to increase runways capacity and efficiency. Re-categorisation aims to decrease the International Civil Aviation Organization (ICAO) wake-turbulence separation standard requirements between two aircraft based on an enhanced aircraft wake vortex categorization. Therefore, it permits to increase airport capacity by accommodating more arrivals and departures during peak traffic hours. The benefit resulting from deploying RECAT has been proved in several pieces of researches [5]–[7]. However, decreasing the wake vortex separation by implementing RECAT program emphasizes the constraining impact of ROT on the runway throughput [8]. Authors in [9] have likewise underscored the importance of reducing the runway occupancy times of landing aircraft. They have claimed that the efficiency gained from applying reduced separation between aircraft will not be fully achieved unless the runway occupancy times are optimized.

Due to its critical importance, the ROT-impacting factors have been widely addressed in the literature. For instance, several traditional machine learning techniques have been deployed for a data-driven prediction of ROT, capturing correlation within the data [10]–[12]. Nonetheless, the prediction and correlation results do not provide insight into how the ROT can be reduced.

Considering this lack of crucial information, the current work investigates the potential ways to reduce ROT by applying Causal Machine Learning (CML) techniques [13]. In fact, CML shifts the focus from data-driven prediction to data-driven decision as the causal/treatment effect reflects how a variable/treatment affects ROT. The primary goal of this work is to investigate how ROT is affected by the two following factors :

- the aircraft approach speed, and
- changing the 3° glide-slope.

The structure of this paper is organized as follows. First, section II presents a literature review of the main previous works relevant to our topic. Second, section III defines the main aspects related to the presented study and states the tackled problem. Third, section IV presents a data analysis that includes a thorough description of the data-set used in our model. Then, section V describes our proposed methodology. Experiments and results are presented in section VI. Finally, conclusions and perspectives are drawn in section VII.

II. Existing Work

The runway occupancy time is a key element for improving the efficiency of airport runway throughput. For instance, a study conducted with Singapore Changi airport (WSSS) flight data [14] has confirmed that runway capacity strongly depends on ROT. Thus, identifying the main factors influencing ROT has attracted the attention of many studies in the past, and researches have evolved over the years. Several tracks have been investigated involving various statistical analyses and different prediction models.

The first study on ROT dated back to 1978, Koenig et al. [15] investigated the main patterns that contributed to a higher ROT. The data used in their work is visually collected between 1972 and 1973 at US airports. Experiments demonstrated that ROT is highly correlated with pilot intend. In 1983, Weiss et al. [16] investigated the impact of two other parameters on ROT, namely the aircraft's wake vortex category and the runway surface condition (dry/wet). It was found that runway surface condition has an insignificant impact on the ROT.

A study developed by the National Aeronautics and Space Administration (NASA) [17] studied the ROT sensitivity to different operational factors. The ROT data was collected using Dynamic Runway Occupancy Measurement System (DROMS). It was concluded that the top five influencing factors are: Ice/flood runway surface condition, entrance/exit ground speed, number of exits, high-speed exit locations and spacing, and aircraft type. In 1999, Lee et al. [18] used DROMS to collect ROT measurements at Atlanta airport and to analyze factors influencing ROT. They confirmed that ROT is equally influenced by the aircraft weight and speed, but barely affected by the airlines and headwind/tail-wind conditions.

Recently, more accurate ROT measurements are becoming available via modern detection systems around airports such as the Airport Surface Detection Equipment - Model X (ASDE-X) [19], providing new opportunities to investigate and analyze ROT. For instance, Kumar et al. [20] investigated the correlation between ROT and the aircraft momentum and concluded that small aircraft could spend as much time on runways as large aircraft. This result has also been confirmed in [3]. The latter study investigated the correlation between ROT and wake-vortex category in order to evaluate their influence in the runway throughput capacity. Furthermore, they pointed out that ROT depends on the location of runway exits and the number of high-speed taxiways while denying the impact of meteorological conditions.

In [11], the authors used the ASDE-X data collected for 36 US airports to implement a data-driven approach aiming to understand the main factors driving ROT. The random forest algorithm was applied to determine the dominant factors influencing ROT. Consistent with previously discussed studies, it was found that the top factors impacting ROT are the runway exit location, the aircraft type, the airline, the aircraft's final approach speed, and the presence of the following aircraft.

The work in [21] introduced a real-time model that aims to identify and predict abnormal runway occupancy times for different aircraft types and weather conditions. Their findings are consistent with the ROT factors found in [11].

All the previously mentioned studies have restricted their research horizon into determining the main factors influencing the ROT. Most of these factors, such as aircraft type, airline, exit location and weather conditions, are fixed parameters that cannot be changed. We believe that ROT has a crucial impact in runway throughput, and thus runway efficiency. However, focusing on the static criteria that have impact on ROT might not have a valuable insight into how we can reduce ROT. To the best of our knowledge, no research has been directed to investigate the variability in ROT as a function of its deriving factors. Therefore, the current work studies the cause/effect relationship between ROT and its key factors. Our main objective is to investigate potential alternatives to

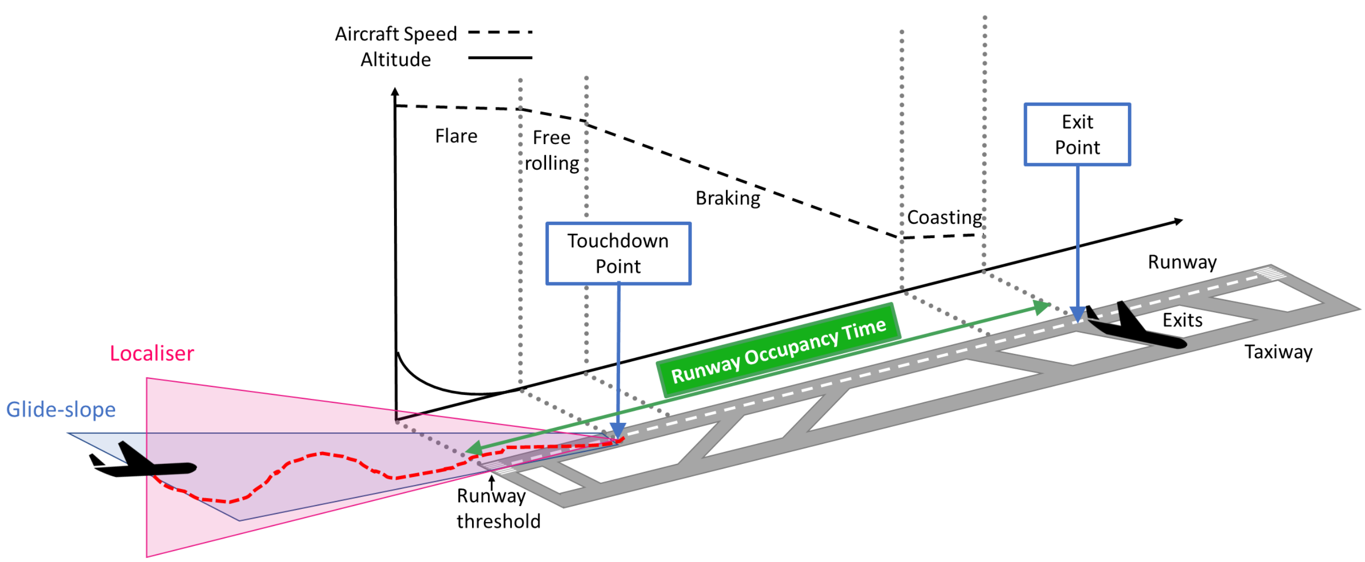


Fig. 1: An illustration of research problem context showing the Landing process with runway occupancy time and the two parameters considered in the research: aircraft speed and deviation from glide-slope

minimize ROT. To this end, only flexible ROT factors that can be tuned are considered.

The first factor that we are considering is the aircraft approach speed. In fact, most of the previously discussed work in the literature have proved the strong correlation between ROT and the aircraft speed. The second factor is the aircraft deviation from the Instrument Landing System (ILS) glide-slope. According to International Air Transport Association (IATA), when an aircraft lands in an unstable configuration, the aircraft may miss the touchdown zone and the runway center-line or touchdown too fast or too hard [22]. The consequences of such unstable landing may lead to structural damage due to 'bounced' landing or runway excursions [23]. Thus, it is important for the aircraft crew to follow a set of standard operating procedures (SOP) during landing to achieve a stabilized approach. One of the SOP recommended by IATA is that the aircraft should have minimal deviation from the ILS glide-slope during landing [24]. As the relationship between unstable approaches and ROT is not widely explored, we are interesting in investigating the causal effect of deviation from glide-slope (as a proxy for unstable approach) on ROT.

III. Research framework

A. Approach and landing procedures

Fig. 1 presents the concept diagram for estimating the causal effect of landing parameters on ROT. For an instrument approach landing, the aircraft should be aligned according to the Instrument Landing System (ILS) during the final approach/landing phase. The ILS uses the localizer antenna to provide horizontal guidance and the glide-slope antenna to provide vertical guidance [25]. The localizer ensures that the aircraft is aligned with the runway centerline while the glide-slope ensures that the aircraft is descending at an operationally acceptable angle which is commonly set at 3°. In addition to the localizer and the glide-slope instructions, the pilot should ensure that the aircraft final approach speed is appropriate while maintaining the aircraft track angle.

The aircraft landing phase can be divided into three phases [26]: the round-out (flare), the touchdown, and the after-landing roll.

The flare is a smooth transition phase between the final approach and the touchdown on the runway. Typically, the pitch of the aircraft slightly increases during this phase to ensure that the flight path gradually transitions to be parallel with the runway. Furthermore, The flight thrust must be adjusted simultaneously to decrease the rate of descent and airspeed.

If the approach and flare are carried out properly, the aircraft will be positioned to touchdown inside the touchdown target zone. The touchdown point is defined as the location where the main gears of the aircraft first establish contact with the runway. The aircraft then proceeds to roll-out by decelerating the aircraft to exit the runway. The roll-out process may consist of free-rolling, braking, and coasting.

The ROT is defined as the time interval between the time the aircraft crosses the threshold and the time the aircraft tail vacates the runway [27] as illustrated in Fig. 1.

B. Problem statement

Although the Instrument Landing System (ILS) has been widely used to aid navigation during landing, the causal relationship between ILS guided landing parameters and ROT remains unclear. An example of a landing parameter that we want to investigate is the effect of a deviation from the ILS glide-slope on ROT. More specifically, our objective is to investigate whether there are any significant increase (or decrease) in ROT between aircraft flying above (or below) the glide-slope and aircraft flying on the glide-slope. However, it is hardly possible to perform a large scale experimental study where we conduct randomized control trials on the landing flights in which we assign whether a flight flies above, below or on the glide-slope during landing. Instead, this research is designed as an observational study where we analyze the causal effect based on historical air traffic data. This causal study can provide insights on how the intervention or advisory on ILS navigation can improve ROT.

To derive the causal relationships, we adopt the Potential Outcome Framework, which is also known as the Rubin Causal Model [28]. This model addresses decision-making problems by quantifying the causal effects of treatment or intervention. The framework has enabled numerous applications ranging from quantifying the effectiveness of treatment in the medical settings [29] to policy evaluation (i.e., monetary fund programs on child poverty [30]).

In the potential outcome framework, treatment effect is defined as follows:

$$\gamma_i = Y_i(1) - Y_i(0)$$

where $Y_i(0)$ and $Y_i(0)$ denote the potential outcome of the individual receiving treatment or no treatment, given a collection of observed covariates. In practice, only one of the two potential outcomes can be observed in the data. This data missing problem is known as the fundamental problem of causal inference, where the counterfactual data is unobservable. Nonetheless, the average treatment effect and treatment effect heterogeneity can be identified, under certain assumptions (i.e., consistency and stable unit treatment value). The average treatment effect is the mean of a treatment strategy to the entire group, while treatment effect heterogeneity identifies treatment effect on subgroups (e.g., aircraft type, weather, etc.).

In this paper, we adopt the aforementioned framework and apply causal machine learning (CML) to estimate the ROT's causal effect. CML is a machine learning approach to perform causal inference, identifying cause and effect relationships between variables. Compared to classical machine learning that focuses on the data-driven prediction by capturing correlation in the data, CML prioritizes causal inference for data-driven decisions. Hence, CML addresses learning problems (i.e., ROT) concerning treatment strategy (e.g., navigation advisory to reduce ROT) or mapping of subgroups (e.g., aircraft type, weather, etc.) to treatment strategy, based on observational data (i.e., features extracted A-SMGCS and METAR data). Generalized Random Forest (GRF) [13] is utilized to derive the causal relationship, detailed in Section V, using the air traffic (A-SMGCS) and weather (METAR) data.

IV. Data analysis

The proposed data-driven causal analysis relies on the quality of the data. Therefore, this section is devoted to describing the data used in our model. It includes the data sources, the techniques applied to remove outliers, and the main characteristics of the data-set.

A. Data Source

To estimate the causal effect for landing parameters on Runway Occupancy Time, we need data on aircraft trajectories and weather information during the landing phase of the flights. Thus, for this study, we analyse the Advanced Surface Movement Guidance Control System (A-SMGCS) data and the Meteorological Aerodrome Reports (METAR) respectively.

1) A-SMGCS: To extract the runway occupancy time and the landing information of the flights, Advanced Surface Movement Guidance Control System (A-SMGCS) data from Singapore Changi Airport is analysed. A-SMGCS is an airport surface safety system that is deployed in the airport to monitor and provide guidance for movement of traffic at and in the vicinity of the airport. One benefit of using the A-SMGCS data is that the data has a 1-second update rate which ensures the processed data is of high fidelity as data interpolation is unnecessary.

2) METAR: In this paper, weather information of landing aircraft is extracted from historical Meteorological Aerodrome Reports courtesy of Iowa Environmental Mesonet from Iowa State University [31] which made them freely available online. The METAR of a specific aerodrome reports the meteorological conditions of the aerodrome and the report is updated and provided to the users at regular intervals which is typically every 30 minutes.

B. Data cleaning

For this study, we utilise one month (i.e., October 2017) of A-SMGCS data and focus on landing flights for runway 02L in Singapore Changi Airport. A total of 1610 medium category aircraft and 1635 heavy category aircraft are extracted to analyse the landing behaviour and the respective ROTs.

For each category of flight (i.e. medium and heavy only as Singapore Changi airport do not serve light aircraft), outliers, with respect to ROT, are removed by applying the standard deviation method. Thus, the observations that are 3 standard deviations away from the mean ROT are treated as outliers and are removed from the data. The remaining flight set after data cleaning includes 1598 medium category aircraft and 1624 heavy category aircraft.

C. Data exploration

The purpose of data exploration is to investigate the main characteristics of the data-set. There are three high-speed exits for Runway 02L in Singapore Changi Airport. For the considered data-set, the third high-speed exit (i.e. the high-speed exit furthest away from the runway threshold) is only utilised by the heavy category aircraft. Fig. 2 illustrates the speed profiles of the flights ans the speed profiles starts from the runway threshold and ends when the flight exited the runway.

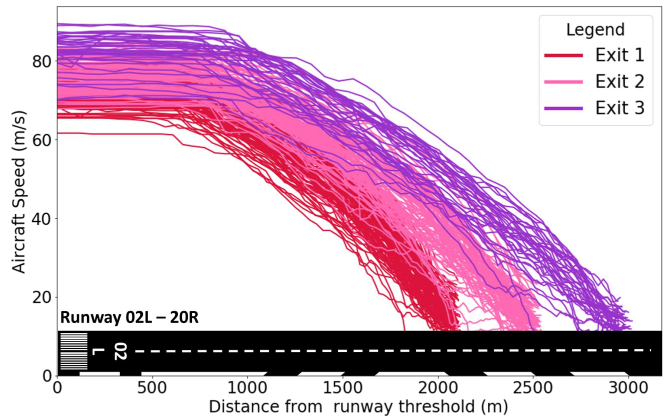


Fig. 2: Speed Profile of aircraft during landing after crossing runway threshold

TABLE I: Mean and Standard Deviation for Runway Occupancy Time (in seconds) for medium and heavy aircraft and the different runway exits

	Medium		**Heavy**	
	Mean	**S.d.**	**Mean**	**S.d.**
Overall	48.60	5.00	52.44	7.21
Exit 1	48.32	4.83	48.81	4.59
Exit 2	55.49	4.22	59.15	5.72
Exit 3	-	-	67.12	4.77

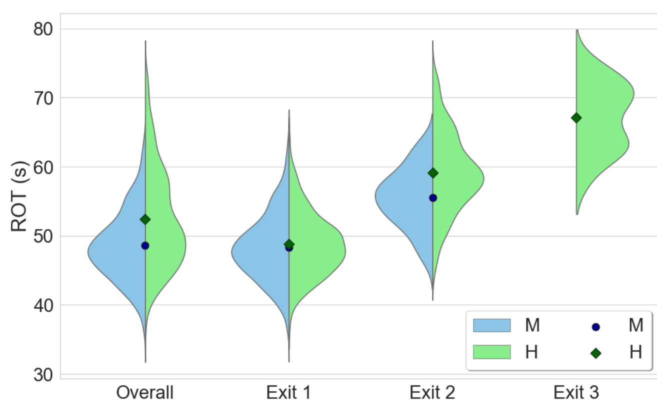


Fig. 3: Runway Occupancy Time distributions for medium and heavy aircraft and the different runway exits

Table I tabulates the overall means and standard deviations of ROT of the medium and heavy aircraft and their respective

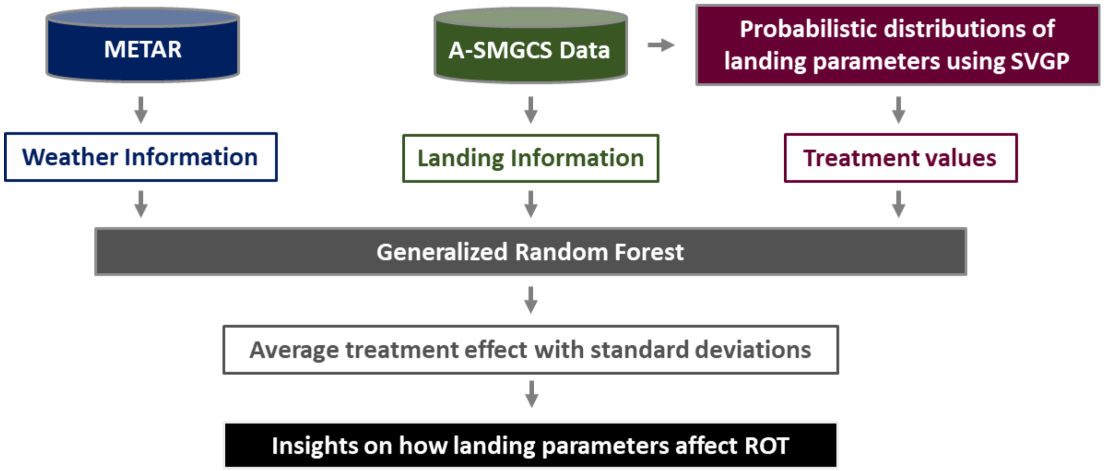


Fig. 4: Methodology framework for data-driven estimation of the causal effect of landing parameters on runway occupancy time using causal machine learning

mean and standard deviation for the different runway exits while Fig. 3 uses violin plots to illustrate the overall ROT distributions of the medium and heavy category aircraft, as well as their distributions for the different runway exits.

In general, the average ROT of heavy category aircraft (i.e., 52.44 seconds) is longer than that of the medium category aircraft (i.e., 48.6 seconds). Furthermore, there is also larger variation in the ROT of heavy category aircraft (i.e., 7.21 seconds) than the medium category aircraft (i.e., 5 seconds).

From Table I and Fig. 3, it is evident that flights that take the further runway exits have longer ROT on average. Such observation is in sync with our understanding as an aircraft have to travel a longer distance on the runway if it were to utilise the further runway exits and this results in a longer ROT.

For flights that exited the runway using exit 1 (i.e., the first exit that an aircraft landing on runway 02L can utilise), the ROT distributions of medium and heavy category aircraft are very similar to each other. The difference in the average ROT is almost negligible. However, for the flights that utilised runway exit 2, the ROT of heavy category aircraft is longer and has greater variation than that of the medium category aircraft.

V. Methodology

In this paper, we develop a data-driven framework that estimates the causal effect of landing parameters on runway occupancy time using Generalized Random Forest (GRF). The proposed methodology framework is illustrated in Fig. 4. Our framework includes processing of A-SMGCS and METAR data during the landing phase as well as causal effect analysis of ROT using GRF.

A. Generalized Random Forest

Generalized Random Forest (GRF) [13] is a recent method for non-parametric statistical estimation proposed by Athey et al. in 2018. GRF is a machine learning algorithm builds on top of the fundamental ideas behind Random Forest (RF) [32] such as ensemble learning, sub-sampling, recursive partitioning, and random split selection and modifies the original RF in numerous ways.

The first key differences between GRF and the original RF is that the prediction is not obtained by merely taking the average or majority votes over the trees. Rather, every observation in the forest is weighted by the frequency which it falls within the same leaf as the target observation. As discussed in [33], introducing adaptive weighting function improves the computing efficiency and the prediction accuracy of the algorithm.

The second difference is that forests in GRF are honest. In the original RF, all training data in the same sub-sample is used to build a tree. While building a tree, the same sub-sample that is used for choosing the best splits in the tree is also used to populate the leaves of that tree. However, in GRF, the training data in the sub-samples are randomly split into half where the first half is used for choosing the best split in the trees and the second half is used to populate the leaves of the trees. Some of the leaves might end up empty during the re-population of leaves when using the second half of the sub-sample, i.e. no example in the second half sub-sample falls within these leaves. As a result, these leaves are pruned away. A forest that is built with such trees is known as an honest forest and the trees within the forest are known as honest trees. The motivation of building honest trees is to reduce the bias in predictions. Besides, Athey et al. [13] proved that the treatment effect estimates are asymptotically normal for an honest forest. This implies that the treatment effect estimates follow a normal distribution as the number of sample size approach infinity. The advantage of such characteristics is that variance and confidence interval of the estimates can be determined which is useful to analyse the statistical significance of the treatment effect estimates.

The third difference between GRF and the classic RF is the splitting criterion. In the original RF, the feature and the value

TABLE II: Features considered for the analysis

Type	Name	Description
Landing aircraft information	Airline	First three letters of the callsign (e.g. SIA)
	Model	Aircraft Model (e.g. A320, B777)
	Terminal	The terminal where the gate assigned is located (e.g. Cargo Terminal, Passenger Terminal 1)
	Exit Number	Categorical feature which indicates the runway exit that the aircraft took to vacate Runway 02L
Following aircraft information	Wake Category	Medium or Heavy (e.g. 'M' or 'H')
	Distance	Distance (in m) of the following aircraft from the landing aircraft
	Speed	Following aircraft speed (in m/s)
Meteorological information	Headwind speed	Wind velocity (in knots) parallel to the runway
	Crosswind speed	Wind velocity (in knots) perpendicular to the runway
	Visibility	Visibility (in miles)
	Rain	Binary feature in which 1 indicates the presence of rain and 0 indicates otherwise

to split at each node in the tree are chosen to minimise the loss function which can be mean square error for regression problem or cross-entropy loss for classification problem. In other words, the splitting criterion finds the best split at each node of the tree to maximise the prediction accuracy. In contrast, the splitting criterion of GRF attempts to achieve two objectives simultaneously. The best split for each node is defined by the one that can maximise the heterogeneity of the treatment effect across the child nodes while maximising the prediction accuracy of the treatment effects. The modified splitting criterion addresses the fundamental problem of causal inference where treatment effect on the individual level is not observable.

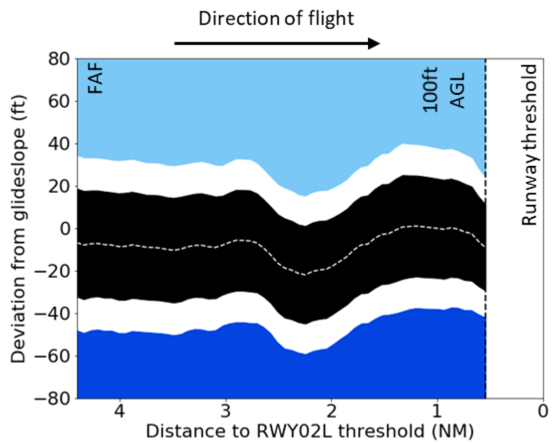
For the estimation of the average treatment effects, GRF algorithm does not naively take the average of the personalized treatment effects. Instead, it employs a doubly robust average treatment effect estimator which improves average treatment effect estimation efficiency and reduces variability which results in more accurate standard error estimates [34]. For this study, the augmented inverse-propensity weighting method [35] is employed as the doubly robust average treatment effect estimator.

B. Covariates

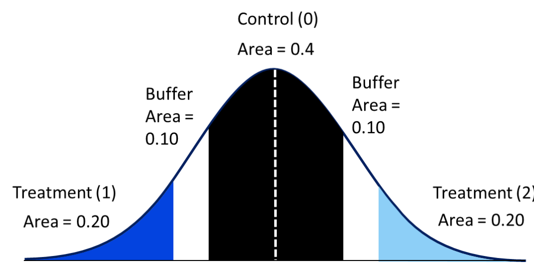
To study the causal effect of landing parameters on ROT, we need to account for the other factors that may also have an impact on ROT. These factors can be grouped into three big categories which are landing aircraft, following aircraft and the meteorological conditions. The set of features that are considered in this work are detailed in Table II.

C. Treatment values for landing parameters

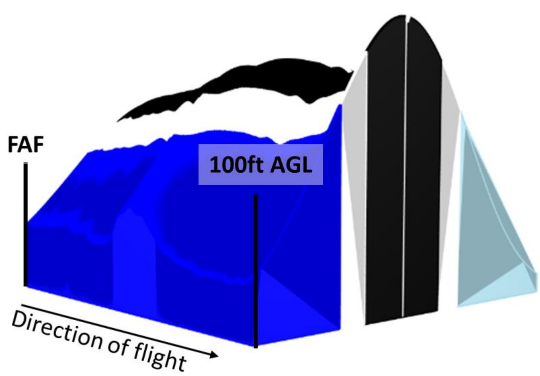
From the landing profile of each flight, we categorise the flights into either the treatment or the control group for each of the landing parameters. To do so, the probabilistic distributions of these landing parameters have to be computed. In a previous work done by Singh et al. [36], authors have designed a methodology which generates the probabilistic distributions for the landing parameters. The study adopted the Sparse Variation Gaussian Process (SVGP) to learn the probabilistic distribution of four parameters along the landing trajectories guided by ILS. The GP model was subsequently used for the detection of anomaly flights (i.e., unstable approach). The current paper extends the probabilistic modelling presented



(a) Top View



(b) Cross-sectional View



(c) 3D view

Fig. 5: Probabilistic distribution of deviation from glide-slope for medium category aircraft in three different views

in [36] for ROT analysis. Using the probabilistic bounds, we partition the flights' trajectories into the treatment or control group for each of the landing parameters.

The probability distribution of an example of a landing parameter (i.e., deviation from glide-slope) for medium category aircraft is illustrated in Fig. 5 with 3 different views.

Fig. 5a illustrates the probabilistic distribution from the top view in which the x-axis represents the distance to the runway 02L threshold in nautical miles (NM) and the y-axis represents the deviation from the glide-slope in feet. The direction of flight is from left to right and the probabilistic distribution starts from the Final Approach Fix (FAF) and ends when the flights are 0.542 NM away from the runway threshold (i.e., aircraft are approximately 100 feet above ground level (AGL)). Positive y-values (i.e. positive deviation from the glide-slope) means that the aircraft is above the glide-slope while negative y-values implies that the aircraft is below the glide-slope. The white dashed line depicts the line of average (i.e. mean) deviation from the glide-slope along the trajectory. The black region represents the control group segment (i.e. flights are flying not too far away from the glide-slope) while the dark blue region represents the treatment 1 segment (i.e. flights are flying well below the glide-slope) and the light blue region represents the treatment 2 segment (i.e. flights are flying well above the glide-slope). A flight has to be in the black region for at least 90% of the trajectory in order for it to be considered in the control group. The same condition applies for the dark blue region (i.e. treatment 1 group) and the light blue region (treatment 2 group). For each landing parameter, a flight can only be either in the control group, treatment 1 group, treatment 2 group, or no assigned group. A flight is not assigned to any group if 90% of its trajectory does not fall inside only one of the coloured regions.

An example of a cross-sectional view of the probabilistic distribution at an arbitrary distance away from the runway 02L threshold is shown in Fig. 5b. This view demonstrates clearly how the different regions are determined. The dark blue region is at the left tail of the distribution and has an area of 0.2, the black region is in the middle of the distribution and has an area of 0.4 and the light blue region is at the right tail of the distribution with an area of 0.2. The two white segments represent the buffer areas. These buffer areas ensure no overlapping between the control group and the treatment groups. Furthermore, they help to maintain the diversity of the landing parameter's value (i.e., deviation from glide-slope) between control and treatment groups.

The 3-dimensional view of the probabilistic distribution, which combines the top view and the cross-sectional view of the distribution, is shown in Fig. 5c.

VI. Experiment & Results

A. Experiment

With the partitions computed for both control and treatment groups, the number of flights belonging to each group is determined for the two landing parameters (i.e., deviation from glide-slope and aircraft speed). Table III records the number of flights in each of the groups for medium and heavy aircraft. As mentioned in Section III, the flight set includes 1598 medium flights and 1624 heavy flights. However, only a few flights, in both medium and heavy categories, are assigned to the control or treatment groups, as seen in Table III. One possible reason for such observation is that the aircraft are constantly adjusting their position and speed during the landing phase to ensure that their landing parameters are within the required range for landing. Thus, most flights do not have 90% of their trajectories fall inside only one of the coloured regions for each of the landing parameter distributions.

The probabilistic distributions of the two landing parameters (i.e., deviation from glide-slope and aircraft speed) for medium and heavy category aircraft are illustrated in Fig. 6. The black regions represent the control groups, the darker coloured regions represent the treatment 1 groups and the lighter coloured regions represent the treatment 2 groups.

TABLE III: Number of flights in each treatment and control group for both medium and heavy category aircraft

	Glide-slope			Speed		
Treatment	**1**	**Control**	**2**	**1**	**Control**	**2**
Description	Below	Baseline	Above	Slower	Baseline	Faster
Medium	186	265	199	74	70	43
Heavy	189	279	186	67	28	47

For the current experiments, we estimate the average treatment effects for each of the treatments by comparing them with their respective control groups. We also account for other factors that might affect ROT in the GRF model by including the covariates defined in Section V. Furthermore, when one of the landing parameters (e.g., deviation from glide-slope) is being analyzed as treatment, the other landing parameter (e.g., aircraft speed) is also accounted for as a covariate.

B. Results & Discussions

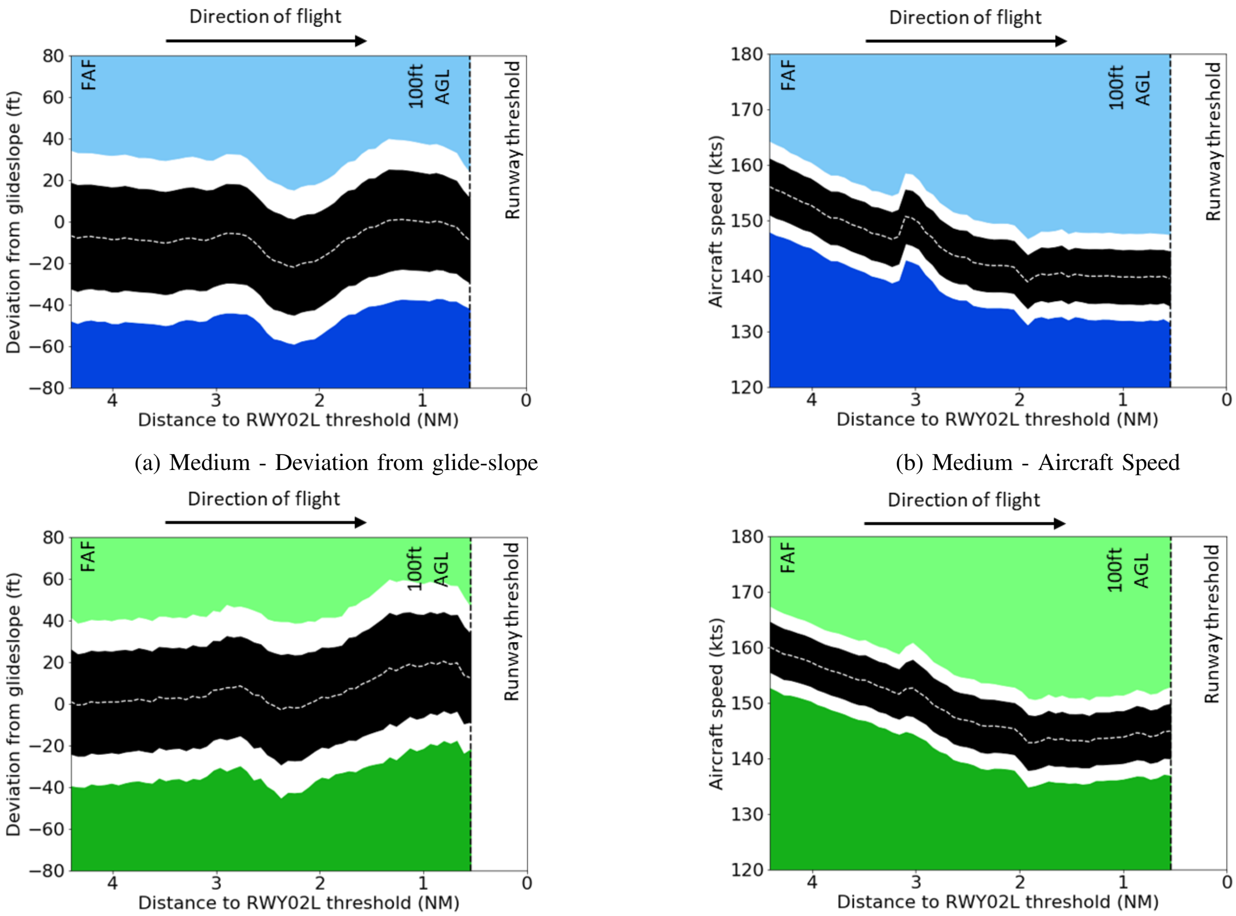
The average treatment effects along with the standard errors for each treatment groups are tabulated in Table IV.

TABLE IV: Average treatment effects of landing parameters on ROT for medium and heavy category aircraft

	Glide-slope		Speed	
Treatment	**1**	**2**	**1**	**2**
Description	Below	Above	Slower	Faster
Medium	-0.917 (0.451) **	0.554 (0.463)	-2.633 (0.890) ***	3.980 (0.751) ***
Heavy	-0.263 (0.463)	-1.110 (0.489) **	-0.444 (1.378)	0.785 (1.068)

(Standard errors in parentheses)
$^{***}p < 0.01$, $^{**}p < 0.05$

For the medium category aircraft, the GRF model estimates that, on average, the ROT for aircraft in treatment 1 group for the glide-slope parameter (i.e., flying below the glide-slope) is 0.92 seconds (p-value = $.021$) shorter than the aircraft in the control group (i.e. flying approximately on the glide-slope). However, there is no significant difference in ROT between the aircraft in treatment 2 and the control group. The results



(a) Medium - Deviation from glide-slope

(b) Medium - Aircraft Speed

(c) Heavy - Deviation from glide-slope

(d) Heavy - Aircraft Speed

Fig. 6: Probabilistic distributions of deviation from glide-slope and aircraft speed for medium and heavy category aircraft

also suggest that to improve the medium category aircraft ROT performance, the aircraft should have slower airspeed in the final approach phase. On average, the ROT of an aircraft with slower airspeed during the final approach phase (i.e., treatment 1) is 2.63 seconds (p-value = $.002$) shorter than the ROT of an aircraft in the control group. Moreover, the ROT of an aircraft with faster airspeed (i.e., treatment 2) is almost 4 seconds (p-value $< .001$) longer than an aircraft in the control group.

For the heavy category aircraft, the results show that the ROT of aircraft flying above the glide-slope during the final approach phase is, on average, 1.11 seconds (p-value = $.011$) shorter than those flying approximately on the glide-slope. There is also no significant difference in ROT between aircraft flying below and on the glide-slope. Unlike the medium category aircraft results, for heavy category aircraft, there is no significant treatment effect for the speed parameter. One possible reason is that there is a lack of enough data points for heavy category aircraft in the control group. As seen in Table III, there is only 28 heavy category aircraft in the control group for the speed parameter.

The results suggest the medium category aircraft to fly below the glide-slope and the heavy category aircraft to fly above the glide-slope during the final approach phase to achieve shorter ROT. However, this recommendation should not be confused with the standard procedure to intercept the ILS signal. In fact, the ILS signal is usually intercepted from below the glide-slope in order to avoid capturing false glide-slopes that exist at $6°$ and $9°$. In order to achieve efficient ROT performance, our proposed model advises two different instructions for the two aircraft categories. In fact, once intercepting with the $3°$ glide-slope, the medium category aircraft should fly around 40 feet below the glide-slope after passing the final approach fix (FAF) (as seen from Fig. 6a). In contrast, the heavy category aircraft should fly approximately 40 feet above the glide-slope after FAF (as seen from Fig. 6c). Even though the magnitude of the average treatment effect is small (i.e., 0.92 seconds and 1.11 seconds for medium and heavy categories, respectively), this value only indicates the reduction of ROT for a single aircraft on average. Nonetheless, the accumulated benefit can yield to a significant increase in the runway throughput efficiency. Fig. 7 illustrates the suggested procedural change for the two aircraft types along with the published Instrument Landing System landing procedure for Runway 02L of Singapore Changi Airport.

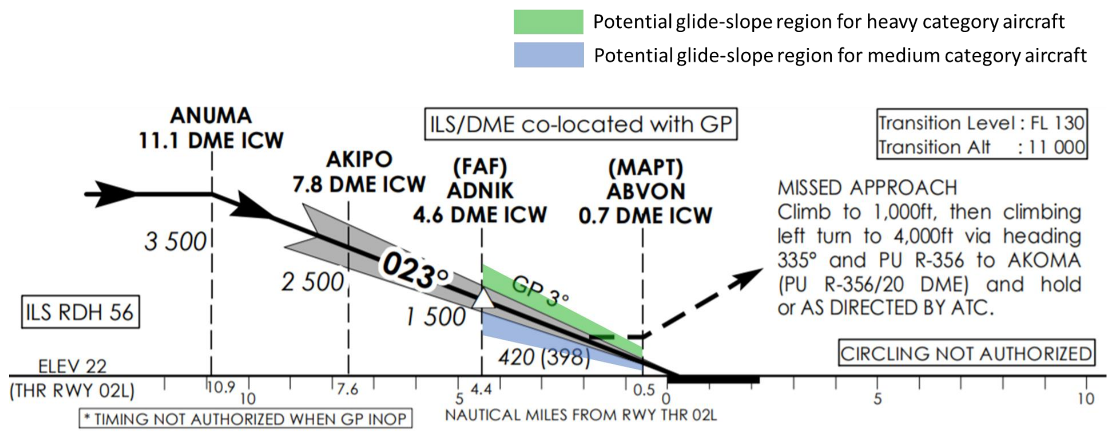


Fig. 7: Instrument approach chart extracted from aeronautical information publication (AIP), Singapore for Runway 02L of Singapore Changi Airport and the suggested procedural changes for the different aircraft types

The results also highlight the need for a flexible glide-slope system in which the glide-slope angle can change according to the arrival aircraft type. The current ILS system has a rigid glide-slope which is commonly set at 3°. The GBAS (Ground-Based Augmented System) Landing System is an alternative to ILS which provides navigation aid to landing aircraft with greater flexibility. The system is not yet widely adopted by airports and aircraft but GBAS is operational in at least 22 airports worldwide and GBAS capability is either standard or an option for most new commercial transport aircraft (eg., Boeing 737-NG, 747-8 and 787 and Airbus A320, A330/340, A350, and A380) [37]. Unlike the ILS where a pair of localiser and the glide-slope antennas are required to be placed at every runway that has instrument approach, the GBAS only requires one GBAS ground facility and a few GBAS reference receivers to serve multiple runways simultaneously. Besides that, GBAS Landing System (GLS) is able to provide navigation guidance for several approach glide angles and displaced threshold [38] while yielding smoother and more precise guidance outputs than ILS [39]. Thus, the results from our study validate the industry need to adopt GLS in the airports to provide flexible glide-slope guidance during the landing phase to improve ROT efficiency and runway throughput.

Besides adjusting the glide-slope angle, the results also suggest that lower airspeed in the final approach phase is more preferable to achieve efficient ROT for medium category aircraft. This observation is in line with the energy model of aircraft. Aircraft are travelling at a higher speed during the approach phase, they will have higher energy. This implies that the aircraft need to dissipate more energy during the landing phase in order to reach the optimal exit speed. Thus, the probability that aircraft with higher final approach speed misses the initial exit is greater and this forces the aircraft to take the exit further down the runway which increases the ROT of the aircraft. From this study, we can observe that the benefit brought by slower airspeed and the penalty of higher airspeed are quite substantial. Medium category aircraft with air-speeds that are around 10 knots slower than the average airspeed have -2.63 seconds reduction in ROT while aircraft that are faster than the average airspeed by around 10 knots will have around 4 increment in their ROT.

VII. Conclusion & Future Work

This paper presents a data-driven approach to estimate the causal effect of landing parameters on Runway Occupancy Time using Causal Machine Learning. The proposed approach can be divided into two main modules. First, the treatment groups are determined using the distributions of the aircraft landing parameters during the final approach phase. Subsequently, the Generalized Random Forest is applied to estimate the treatment effects on ROT while accounting for other potential factors. The proposed methodology relies only on historical air traffic data, in approach and landing phase, and METAR data, rather than conducting experimental studies, to investigate the causal/treatment effect of landing parameters on ROT. Thus, it prevents premature commitments to conduct a large-scale experiment.

Based on the analysis of A-SMGCS data of Runway 02L of Singapore Changi Airports, two conclusions can be drawn with respect to our research questions. First, slowing down the aircraft speed in the final approach phase, while maintaining the other factors constant, results in shorter ROT. Second, flexible glide-slope angles should be introduced for the different aircraft types to achieve better ROT performance. The second finding therefore support the sector's need to implement a landing system that can provide flexible glide-slopes such as the GLS.

The current work presents a preliminary study of implementing Causal Machine Learning for data-driven decisions in the Air Traffic Management context. In future works, we plan to enrich our data-set with more data from different months, different runways, and different airports in order to investigate the sensitivity of the estimated causal effect of landing parameters on Runway Occupancy Time. Furthermore, we plan to extend the proposed framework by investigating the causal effect of other decision variables on ROT, such as the following aircraft type, it's speed, and it's distance from the landing aircraft. This extension will aid air traffic controllers in terms of sequencing and managing the flow of landing the flights.

Acknowledgement

This research has been supported by the Civil Aviation Authority of Singapore (CAAS) and the Air Traffic Management Research Institute (ATMRI), Singapore.

References

[1] Bruno F Santos, Maarten MEC Wormer, Thomas AO Achola, and Richard Curran. Airline delay management problem with airport capacity constraints and priority decisions. *Journal of Air Transport Management*, 63:34–44, 2017.

[2] Paul Upham, Callum Thomas, David Gillingwater, and David Raper. Environmental capacity and airport operations: current issues and future prospects. *Journal of Air Transport Management*, 9(3):145–151, 2003.

[3] Tamas Kolos-Lakatos. *The influence of runway occupancy time and wake vortex separation requirements on runway throughput*. PhD thesis, Massachusetts Institute of Technology, 2013.

[4] Frederic Rooseleer and Vincent Treve. Recat-eu european wake turbulence categorisation and separation minima on approach and departure. *EUROCONTROL Headquarters, Brussels*, 2015.

[5] Bas v Meijden, Paul C Roling, and Richard Curran. Flexible runway use modeling using pairwise recat-eu separation minima. In *2018 Aviation Technology, Integration, and Operations Conference*, page 3840, 2018.

[6] Nastaran Coleman, Dave Knorr, and Almira Ramadani. Statistical model to estimate the benefit of wake turbulence re-categorization. In *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019), Vienna, Austria*, 2019.

[7] Navid Mirmohammadsadeghi, Junqi Hu, and Antonio Trani. Enhancements to the runway capacity simulation model using the asde-x data for estimating airports throughput under various wake separation systems. In *AIAA Aviation 2019 Forum*, page 3044, 2019.

[8] Junqi Hu, Navid Mirmohammadsadeghi, and Antonio Trani. Runway occupancy time constraint and runway throughput estimation under reduced arrival wake separation rules. In *AIAA Aviation 2019 Forum*, page 3046, 2019.

[9] Adib Kanafani. Operational procedures to increase runway capacity. *Journal of Transportation Engineering*, 109(3):414–424, 1983.

[10] Navid Mirmohammadsadeghi and Antonio Trani. Prediction of runway occupancy time and runway exit distance with feedforward neural networks. In *AIAA Scitech 2020 Forum*, page 1402, 2020.

[11] Nicolas P Meijers and R John Hansman. A data-driven approach to understanding runway occupancy time. In *AIAA Aviation 2019 Forum*, page 3045, 2019.

[12] Dario Martinez, Seddik Belkoura, Samuel Cristobal, Floris Herrema, and P Wachter. A boosted tree framework for runway occupancy and exit prediction. In *Eighth SESAR Innovation Days*, 2018.

[13] Susan Athey, Julie Tibshirani, Stefan Wager, et al. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019.

[14] Henry SL Fan. Effect of local operational constraints on runway capacity-a case study. *Journal of advanced transportation*, 26(2):169–184, 1992.

[15] Steven E Koenig. Analysis of runway occupancy times at major airports. Technical report, MITRE CORP MCLEAN VA METREK DIV, 1978.

[16] William E Weiss and JN Barrer. Analysis of runway occupancy time and separation data collected at la guardia, boston, and newark airports. Technical report, MITRE CORP MCLEAN VA, 1984.

[17] S.H. Goldthorpe. *Sensitivity of Runway Occupancy Time (ROT) to Various Rollout and Turnoff (ROTO) Factors. Volume 2; Complete Set of Plotted Data*. 1997.

[18] Derrick D Lee, Alexander Smith, Rick Cassell, and Bassam Abdul-Baki. Nasa low visibility landing and surface operations (lvlaso) runway occupancy time (rot) analysis. In *Gateway to the New Millennium. 18th Digital Avionics Systems Conference. Proceedings (Cat. No. 99CH37033)*, volume 1, pages 5–D. IEEE, 1999.

[19] Eli Perl. Review of airport surface movement radar technology. In *2006 IEEE Conference on Radar*, pages 4–pp. IEEE, 2006.

[20] Vivek Kumar, Lance Sherry, and Rafal Kicinger. Runway occupancy time extraction and analysis using surface track data. *George Mason University*, 2009.

[21] Herrema Floris Friso, Curran Richard, Hendrikus G Visser, Treve Vincent, and Desart Bruno. Predicting abnormal runway occupancy times and observing related precursors. *Journal of Aerospace Information Systems*, 15(1):10–21, 2018.

[22] International Air Transport Association (IATA). *Unstable approaches: risk mitigation policies, procedures and best practices*. International Air Transport Association, Montreal—Genevas, 2016 2nd Edition.

[23] National Transportation Safety Board. Aviation safety alert 077: Stabilized approaches lead to safe landings, 2019.

[24] International Air Transport Association (IATA). *Operational safety audit (IOSA) standards manual*. International Air Transport Association, Montreal—Genevas, 2016 11th Edition.

[25] *Aeronautical telecommunications: international standards and recommended practices*. ICAO, 1998.

[26] Federal Aviation Administration. *Airplane Flying Handbook, Chapter 8 Approaches and Landing*. Skyhorse Publishing Inc.

[27] Eurocontrol. *Enhancing Airside Capacity: The Complete Guide*. Eurocontrol, 2004.

[28] Jasjeet S Sekhon. The neyman-rubin model of causal inference and estimation via matching methods. *The Oxford handbook of political methodology*, 2:1–32, 2008.

[29] Donald B Rubin. Estimating causal effects from large data sets using propensity scores. *Annals of internal medicine*, 127(8_Part_2):757–763, 1997.

[30] Adel Daoud and Fredrik D. Johansson. Estimating treatment heterogeneity of international monetary fund programs on child poverty with generalized random forest. *SocArXiv*, 2019.

[31] Iowa Environmental Mesonet. Asos-awos-metar data download. https://mesonet.agron.iastate.edu/request/download.phtml. Accessed on 1 July 2020.

[32] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[33] Yi Lin and Yongho Jeon. Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association*, 101(474):578–590, 2006.

[34] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 01 2018.

[35] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

[36] Narendra Pratap Singh, Sim Kuan Goh, and Sameer Alam. Real-time unstable approach detection using sparse variational gaussian process. In *2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)*, pages 1–10. IEEE, 2020.

[37] Federal Aviation Administration. Satellite navigation — ground based augmentation system (gbas). https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/laas/, April 2018. Accessed on 1 July 2020.

[38] ICAO. Guide for ground based augmentation system implementation. https://www.icao.int/SAM/eDocuments/GBASGuide.pdf, May 2013. Accessed on 1 July 2020.

[39] Michael Felux, Thomas Dautermann, and Hayung Becker. Gbas landing system - precision approach guidance after ils. *Aircraft Engineering and Aerospace Technology*, 85, 07 2012.